



HAL
open science

Mixing times of Markov chains: acceleration and cutoff in random environment

Bastien Dubail

► **To cite this version:**

Bastien Dubail. Mixing times of Markov chains: acceleration and cutoff in random environment. Probability [math.PR]. Ecole Normale Supérieure, 2023. English. NNT: . tel-04434707

HAL Id: tel-04434707

<https://inria.hal.science/tel-04434707>

Submitted on 2 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'École Normale Supérieure et l'Université d'Aix-Marseille

**Temps de mélange de chaînes de Markov : accélération
et cutoff en environnement aléatoire**

Soutenue par

Bastien Dubail

Le 13/10/2023

École doctorale n° 386

**Sciences Mathématiques de
Paris Centre**

Spécialité

Mathématiques

Composition du jury :

Alexandre PROUTIERE Professor, KTH Royal Institute of Technology	<i>Président</i>
Perla SOUSI Professor, University of Cambridge	<i>Rapporteuse</i>
Sourav CHATTERJEE Professor, Stanford University	<i>Rapporteur</i>
Anna BEN HAMOU Maître de Conférences, Sorbonne Université	<i>Examinatrice</i>
Justin SALEZ Professeur, Université Paris-Dauphine	<i>Examineur</i>
Charles BORDENAVE Directeur de recherche, Aix-Marseille Université	<i>Co-directeur de thèse</i>
Laurent MASSOULIE Directeur de recherche, INRIA	<i>Directeur de thèse</i>

Temps de mélange de chaînes de Markov : accélération et cutoff en environnement aléatoire

Bastien Dubail

*École Normale Supérieure, Département d'Informatique, INRIA
PSL Research University, Paris, France
Aix-Marseille Université, Marseille, France*

Thèse réalisée sous la supervision de Laurent Massoulié et Charles Bordenave

Résumé

Cette dissertation a pour objet l'étude de différentes questions reliées aux marches aléatoires et aux temps de mélange de chaînes de Markov.

Notre première contribution est l'introduction d'une nouvelle technique, appelée linéarisation, dans l'étude d'une marche aléatoire sur un groupe discret. Inspirée par des approches de linéarisation similaires utilisées pour l'étude des algèbres d'opérateurs et des matrices aléatoires, cette méthode se voit ici proposée un tout nouveau champ d'applications. Lorsqu'un groupe est donné par un ensemble fini de générateurs, nous montrons comment la linéarisation peut simplifier l'étude des marches aléatoires à support fini en les réduisant à des marches aux plus proches voisins, moyennant la considération d'un espace d'états élargi. Nous démontrons l'efficacité de cette technique en établissant des formules explicites pour des quantités asymptotiques telles que la vitesse et l'entropie de la marche, dans le cas des groupes libres et des produits libres de groupes finis.

Notre deuxième contribution porte sur un mécanisme d'accélération de dynamiques markoviennes, consistant à intercaler des étapes de dynamique déterministe entre les étapes aléatoires d'une chaîne de Markov. Lorsque la dynamique déterministe est elle-même choisie de façon aléatoire, il a été prouvé que le temps de mélange devient logarithmique en la taille de l'espace d'états avec grande probabilité, ce qui pour de nombreuses chaînes de Markov équivaut à une accélération exponentielle. Notre travail tente de "dérandomiser" ce résultat en étudiant une version multi-dimensionnelle de l'exemple historique pour lequel cette accélération a été étudiée pour la première fois. Dans cet exemple explicite, nous suggérons ainsi que l'accélération est l'effet du caractère chaotique de la dynamique déterministe appliquée.

La troisième partie de notre travail porte sur le phénomène de cutoff, qui est la convergence abrupte vers l'équilibre d'une chaîne de Markov. Nous étudions un phénomène de cutoff pour une chaîne en environnement aléatoire qui est un cadre propice à l'apparition du phénomène. Le modèle étudié consiste à superposer deux chaînes de Markov sur un même espace d'états, après avoir permuté les états de l'une de façon uniforme. Sous des hypothèses peu contraignantes, nous prouvons un phénomène de

cutoff au temps dit entropique. En particulier nous ne faisons aucune hypothèse de réversibilité ou d'irréversibilité, c'est-à-dire que nous autorisons la chaîne à revenir sur ses pas. Jusqu'à présent, l'une ou l'autre de ces hypothèses devait généralement être faite pour pouvoir établir le phénomène de cutoff. Nous proposons ainsi une stratégie de preuve unique faisant le pont entre ces deux cas opposés. L'argument repose en grande partie sur un nouveau résultat de concentration de la mesure pour le groupe symétrique, développé spécifiquement pour ce problème, et qui constitue ainsi une autre contribution de cette thèse.

Abstract

The aim of this dissertation is to explore various aspects concerning random walks and mixing times of Markov chains.

First, our research introduces a novel technique called "linearization" in the analysis of random walks on discrete groups. Drawing inspiration from similar linearization approaches in operator algebras and random matrices, this opens up a whole new field of applications of this method. When a group is given by a finite set of generators, we show how linearization can simplify the study of finitely supported random walks by reducing to nearest-neighbour walks, at the cost of considering a slightly more general state space. We demonstrate the effectiveness of this technique by providing explicit formulas for asymptotic quantities like the speed and entropy of the walk, in the case of free groups and free products of finite groups.

Our second contribution concerns a mechanism for accelerating Markov dynamics, consisting in inserting deterministic dynamics steps between the random steps of a Markov chain. When the deterministic dynamics is itself chosen randomly, it was proved that the mixing time becomes logarithmic in the size of the state space with high probability, which for many Markov chains implies an exponential speed-up. Our work attempts to "derandomize" this result by studying a multi-dimensional version of the historical example where such acceleration was first observed. In this explicit example, we thus suggest that the acceleration is the effect of the chaotic aspect of the applied deterministic dynamics.

The third part of our work deals with the cutoff phenomenon, which is the abrupt convergence of a Markov chain towards equilibrium. We study a cutoff phenomenon for a chain in a random environment, a setting known to favor the phenomenon's appearance. The model studied consists in superposing two Markov chains on the same state space, one being subject to a uniform permutation of the states. Under mild assumptions we prove such chains exhibit a cutoff phenomenon at the so-called entropic time. In particular we do not assume reversibility or non-backtrackingness, which can be thought of as the opposite of reversibility. Until now, one or other of these assumptions was generally made in order to establish the cutoff phenomenon.

We propose a unified proof strategy that bridges these opposing cases. This relies heavily on a novel measure concentration result for the symmetric group, developed explicitly for this problem, thereby making another contribution of this thesis.

Remerciements

Mes premiers remerciements vont à Charles et Laurent, pour m'avoir encadré pendant ma thèse et les stages qui l'ont précédée. Je les remercie pour leur soutien sans faille tout au long de cette odyssée, pour leur enseignement, qui va bien au delà du seul plan scientifique, ainsi que pour leur humour et leur permanente bienveillance. Je leur suis également très reconnaissant pour leurs conseils avisés en terme de stratégie et l'aide apportée durant la recherche de post-doc. Je sais que j'ai encore beaucoup à apprendre et que je peux compter sur eux.

Je remercie Perla Sousi, Anna Ben Hamou, Sourav Chatterjee, Justin Salez et Alexandre Proutière qui me font un véritable honneur en ayant accepté de faire partie de mon jury. Leurs travaux ont été et continuent d'être une immense source d'inspiration pour moi. Je remercie particulièrement Perla Sousi et Sourav Chatterjee d'avoir accepté la lourde tâche de rapporteurs et apporté leur expertise, en plein été, à la lecture de ce manuscrit. Je tiens à remercier également Alexandre Proutière d'avoir accepté de me prendre en tant que post-doctorant à KTH. L'aperçu que j'en ai eu me fait entrevoir de belles perspectives et je me réjouis déjà des défis à venir.

Il me paraît tout naturel d'exprimer ensuite ma reconnaissance envers toutes les personnes qui ont été directement auprès de moi durant ce périple. Je remercie mes camarades doctorants et post-docs, dont la chaleureuse compagnie aura été déterminante dans la préparation de cette thèse. Merci au «groupe le plus calé en bombardement» (de papiers), j'ai nommé Luca, Eric, Mathieu, Matthieu et Jakob, mes très chers compagnons de bureau, qui ne cessent de m'étonner de par leurs multiples talents. Merci pour tous les échanges scientifiques, les moments de franche rigolade, les discussions lors des moments de doute... autant d'ingrédients qui m'auront aidé à effectuer ma thèse dans les meilleures conditions. Merci aux doctorants Claire, Romain, Vianey, David, Lucas, Thomas, Amaury, Batiste, Antoine, avec qui j'ai aimé conversé de maths, de musique, de sport... autour d'une bière ou d'un repas à l'AGRAF, et mené des opérations de «team-building», séminaire ou groupe de lecture. Je remercie le grand homme de culture qu'est le docteur Shilov, présent à mes côtés depuis le début, avec qui on

arrête assez vite de compter les anecdotes mémorables tant il y en a. Спасибо, Илья! Merci à Bertille pour ses fréquentes visites depuis les étages supérieurs, apportant toujours du réconfort aux moments opportuns. Merci à Adam, Sarah, Joffrey, Juan, Yanis, Yenni, Anaïs, Antoine que je n'ai pas manqué de retrouver à chacune de mes visites à Marseille. Que de bons moments j'ai passé avec eux à l'I2M, au cours Ju ou dans les Calanques. Merci à Ludovic et Simon, mes aînés de thèse que j'ai eu la chance de cotoyer et qui ont été pour moi des modèles à suivre.

Je suis extrêmement reconnaissant envers toute l'équipe des Argonautes, dont le dynamisme et la bonne humeur auront apporté une énergie et une motivation sans pareil pour la réalisation de cette thèse. Un merci tout particulier à Hélène B., Argonaute jusqu'à cet été, qui aura contribué de façon essentielle à faire de l'équipe ce qu'elle est aujourd'hui, par son travail, son dévouement et son inébranlable enthousiasme. Merci à notre nouvelle recrue Marina ainsi qu'à Hélène M. de nous prêter main forte aussi souvent que nécessaire. J'adresse ma gratitude également à Lise-Marie, pour sa disponibilité, son sourire et son accueil à chacune de mes venues au DI ENS. Merci à Evelyne et Marie-Christine, toujours aux petits soins lors de mes visites à Marseille. J'adresse plus généralement des remerciements à toutes les petites mains et services grâce auxquels nous pouvons venir à bout de bien des tracasseries administratives. Merci à Marion, Alexandre, Christine et Johanna d'égayer encore un peu plus la vie du troisième étage par leurs sourires. Merci à l'AGOS, instigatrice de nombreuses soirées mémorables. Merci aux impresarios de la science, Lucile et toute l'équipe de la Comm'.

Ces années studieuses se sont aussi accompagnées de quelques moments de répit salvateurs. Je tiens ainsi à remercier quelques uns de mes amis qui sans le savoir peut-être ont joué un rôle dans cette aventure. Merci à Emmanuel et Nicolas, éternels compagnons de la «Stratosteam», ainsi qu'à Alexis, mon fidèle colloc de Lyon, pour les moments partagés pendant toutes ces années, notamment celles du covid. Merci à Cédric de m'avoir initié à l'art du slap. Merci au Jeune Kayak Parisien, véritable petite famille qu'il m'est triste de devoir quitter aujourd'hui. Merci à Mathias et Fred de m'avoir enseigné l'art de vivre jkliste et fait découvrir les magnifiques rivières du Morvan. Merci à Lara, Colas, Paul, Camille et Yannick pour tous ces moments mémorables vécus ces deux dernières années.

Enfin, je remercie ma famille. Mes parents bien sûr, pour leur soutien indéfectible durant toutes mes études, mais aussi le triumvirat formé par Jérôme, Pauline et Hélène, qui y sont sans doute un peu pour quelque chose dans l'existence aujourd'hui de ce manuscrit.

Contents

Résumé détaillé	13
Marches aléatoires sur les groupes	13
Accélération de chaînes de Markov	16
Concentration sur le groupe symétrique	20
Le phénomène de cutoff	21
1 Introduction	25
1.1 Random walks on groups	25
1.2 Acceleration of Markov chains	28
1.3 Concentration on the symmetric group	31
1.4 The cutoff phenomenon	33
2 Linearization of random walks on groups	37
2.1 Introduction	37
2.1.1 The linearization trick and coloured random walks on groups	38
2.1.2 Main results	40
2.2 A linearization trick	47
2.2.1 Linearization in the non-reversible case	48
2.2.2 Linearization in the reversible case	51
2.3 Coloured random walks on free products of groups	53
2.3.1 The harmonic measure on the boundary	54
2.3.2 Markovian structure of the harmonic measure	55
2.3.3 Solving the traffic equations	57
2.4 Computing drift and entropy	60
2.4.1 Convergence in direction for inhomogeneous products of matrices	61
3 Accelerating abelian random walks with hyperbolic dynamics	67
3.1 Introduction	67
3.1.1 Main results	67

3.1.2	Related works	70
3.1.3	Discussion	71
3.2	A lower bound for all n	74
3.3	An upper bound for all n	74
3.3.1	Fourier analysis on the torus	75
3.3.2	Expansiveness	76
3.3.3	Symbolic Representations	78
3.3.4	Proof of Theorem 3.1	79
3.4	Irreducibility and aperiodicity	86
3.5	An upper bound for almost all n	88
3.6	Markov partitions	90
3.6.1	Hyperbolic dynamics	91
3.6.2	Markov Partitions	92
4	Concentration for low-degree functions on the symmetric group	97
4.1	Introduction	97
4.1.1	Main results	97
4.1.2	Related works	100
4.2	Control of the smoothness parameters	101
4.3	Proof of the concentration inequality	103
4.3.1	Induction with the method of exchangeable pairs	103
4.3.2	Almost sure bound on F	108
4.3.3	Tensor representation	109
5	Cutoff for mixtures of permuted Markov chains	113
5.1	Models and main results	113
5.1.1	Cutoff for mixtures of permuted Markov Chains	113
5.1.2	A reversible model	116
5.1.3	Related works	118
5.1.4	Proof outline	120
5.2	First moment argument, quasi-trees	124
5.2.1	Quenched vs annealed probability, first moment argument	124
5.2.2	The two-lift chain and half-integer time steps	125
5.2.3	The graph point of view	126
5.2.4	Quasi-trees	128
5.2.5	The covering quasi-tree	130
5.2.6	Random setting	132
5.2.7	Coupling and sequential generation	133

5.3	The entropic method: main arguments	135
5.3.1	Nice trajectories	135
5.3.2	Concentration of drift and entropy	137
5.3.3	Concentration of nice paths	140
5.4	Counter-example to uniform entropic time	144
5.5	Quasi-tree I: Escape probabilities	146
5.5.1	Escape probabilities	147
5.5.2	Escape probabilities: reversible model	151
5.5.3	Escape probabilities: general model	154
5.5.4	General case: $o(n)$ transient or close to transient states	162
5.6	Quasi-tree II: Markovian regeneration structure	164
5.6.1	Markov renewal processes	164
5.6.2	Regeneration structure	175
5.6.3	Mixing of the regeneration chain	182
5.7	Quasi-tree III: concentration of drift and entropy	186
5.7.1	Typical paths in quasi-trees	186
5.7.2	Concentration of the drift	189
5.7.3	Concentration of the entropy	192
5.8	First steps towards nice paths	199
5.8.1	The case of typical starting states	199
5.8.2	Bootstrapping annealed bounds with parallel chains	200
5.8.3	From $o(1)$ to $o(1/n)$ bounds	203
5.8.4	Nice paths from arbitrary states	204
5.9	Nice paths	207
5.9.1	Forward neighbourhood	207
5.9.2	Nice paths: definition	211
5.9.3	Nice paths are typical	213
5.9.4	Concentration of nice paths	216
5.9.5	Expectation of nice paths	221

Résumé détaillé

Ce mémoire est la synthèse du travail réalisé lors de ma thèse. Il est basé sur trois articles qui explorent différents aspects de la théorie des chaînes de Markov : [30], réalisé avec Charles Bordenave, [65] réalisé avec Laurent Massoulié, et un troisième qui n'a pas encore été publié. Les chaînes de Markov sont des processus stochastiques qui n'ont pas de mémoire du passé. En tant que tel, ils sont le type de processus stochastique le plus simple avec une dépendance non triviale dans le temps et ont toujours été un objet central de la théorie des probabilités. Il existe de nombreuses références excellentes sur le sujet : les monographies de Durrett [66] et Norris [142] sont certainement un bon point de départ, avec de nombreux exemples et applications. Les livres de Levin, Peres et Wilmer [117] sur les temps de mélange, Aldous et Fill [6] et Lyons et Peres [128] sur les chaînes réversibles et sujets connexes sont d'autres pierres angulaires de la littérature, qui ont été largement utilisés lors de la réalisation de cette thèse. Nous considérons exclusivement des chaînes de Markov à temps discret avec un espace d'états discret. D'un point de vue général, notre étude des chaînes de Markov est motivée par la question d'identifier les mécanismes sous-jacents qui expliquent le comportement spécifique d'une chaîne. En particulier, nous essayons de comprendre les liens existant entre une chaîne de Markov et la géométrie de son espace d'états. Nos travaux tombent tous dans ce cadre général mais abordent des questions distinctes. Pour cette raison, nous essaierons ici de donner un aperçu concis de nos contributions et de donner des références générales. Chaque chapitre se verra plus tard consacré sa propre introduction ainsi qu'une revue de la littérature plus spécifiquement liée aux travaux présentés.

Marches aléatoires sur les groupes

Le terme de marche aléatoire fait généralement référence à un type particulier de chaînes de Markov que l'on aime visualiser comme évoluant sur une structure géométrique comme un graphe. Dans cette catégorie, les marches aléatoires sur les groupes sont des chaînes de Markov qui présentent beaucoup de symétrie et qui apparaissent naturelle-

ment dans divers problèmes théoriques et appliqués. La marche aléatoire simple sur une droite, peut-être la plus canonique des chaînes de Markov, est une marche aléatoire sur le groupe additif \mathbb{Z} . Les marches aléatoires sur les groupes sont la généralisation directe des sommes de variables aléatoires iid à des espaces plus généraux. Plus précisément, étant donné un groupe G équipé d'une mesure de probabilité p , considérons la chaîne de Markov $X = (X_t)_{t \geq 0}$ sur G définie par

$$\forall g, h \in G, \forall t \geq 0 : \quad \mathbb{P}[X_{t+1} = h \mid X_t = g] = p(g^{-1}h).$$

De manière équivalente, X_t peut être écrit comme le produit $X_t = Z_0 Z_1 \dots Z_t$ d'"increments" indépendants sur G tous distribués avec la loi p , exceptée potentiellement Z_0 . Une telle chaîne est appelée une marche aléatoire de convolution sur G , souvent abrégée simplement en marche aléatoire sur le groupe G .

Notre premier travail concerne plus spécifiquement le cas où le groupe G est dénombrable et donné par un ensemble fini de générateurs S . Définissons le support de la marche aléatoire comme le support de la mesure p , c'est-à-dire l'ensemble des éléments qui ont une masse non nulle sous p . On dit que la marche aléatoire est de support fini, ou qu'elle a une portée finie, si la mesure p est de support fini. C'est une marche aléatoire aux plus proches voisins si p est supportée par l'ensemble des générateurs S .

La première contribution de cette thèse est motivée par l'observation générale que l'étude d'une marche aléatoire est souvent considérablement plus simple si elle est supposée être aux plus proches voisins. L'analyse du comportement asymptotique des marches aléatoires sur les groupes donne un bon exemple de ce que peuvent être les difficultés supplémentaires. Une grande partie de cette théorie remonte aux travaux de Furstenberg [76], Avez [12], Guivarc'h [82], Derriennic [54], Kaimanovich et Vershik [102]. L'une des caractéristiques essentielles des marches aléatoires de convolution est la transitivité : par le biais de la multiplication dans le groupe, l'état de départ de la marche aléatoire peut toujours être translaté vers l'élément neutre. Par conséquent, la loi de la marche est essentiellement indépendante du point de départ. A partir de cette observation, des théorèmes ergodiques tels que le théorème sous-additif de Kingman établissent l'existence des limites presque sûres

$$v := \lim_{t \rightarrow \infty} \frac{|X_t|}{t}, \quad h := \lim_{t \rightarrow \infty} -\frac{\log \mathbb{P}[X'_t = X_t \mid X]}{t}.$$

sous condition que $\sum_{g \in G} p(g) |g| < \infty$ et où X' dénote une copie indépendante de la chaîne. Étant donné $g \in G$, $|g|$ dénote ici le nombre minimal d'éléments nécessaires pour écrire g comme un produit de générateurs. Cela peut être interprété comme une distance entre g et l'élément d'identité. La quantité v est pour cette raison appelée la vitesse, ou dérive, de la marche aléatoire. La limite h est l'entropie asymptotique, ou

entropie d'Avez, nommée d'après Avez qui l'a introduite dans [12]. L'entropie asymptotique contient une grande partie de l'information sur le comportement asymptotique de la marche. Par exemple, Kaimanovich et Vershik [102] et Derriennic [54] ont prouvé que la σ -algèbre de queue de la marche aléatoire est non triviale si et seulement si $h > 0$.

Considérons maintenant le cas spécifique où $G = \mathbb{F}_d$ est le groupe libre sur d générateurs. L'ensemble de générateurs considéré est celui des d générateurs canoniques et de leurs inverses. Ce groupe peut être représenté par un graphe, son graphe de Cayley, qui n'est rien d'autre qu'un arbre régulier de degré $2d$. Une marche aléatoire générale peut faire des sauts arbitrairement grands dans cet arbre, une marche à support fini fait des sauts bornés, tandis qu'enfin une marche aux plus proches voisins se déplace uniquement le long des arêtes. Cette dernière propriété, couplée à la structure de l'arbre, est extrêmement utile : pour atteindre un sommet y dans l'arbre à partir d'un sommet x , la marche doit passer par tous les sommets intermédiaires. Cela ouvre la voie à des arguments de récurrence, à partir desquels il est possible d'obtenir des formules explicites pour la vitesse et l'entropie de la marche [114]. Dans le cas général d'un support fini, de tels arguments échouent, et pour cette raison, aucune formule explicite ne semblait réalisable.

Dans notre travail, nous introduisons une nouvelle technique de "linéarisation" pour étudier les marches aléatoires sur les groupes qui permet de se ramener au cas de marches aux plus proches voisins. Le nom "linéarisation" vient de l' "astuce de linéarisation" utilisée notamment en algèbres d'opérateurs, de laquelle notre travail s'inspire [89, 83]. Le prix à payer pour ne considérer que le cas de marches aux plus proches voisins est que la chaîne de Markov résultante n'est pas une marche aléatoire sur un groupe mais une chaîne de Markov un peu plus générale que nous appelons une marche aléatoire colorée. En partant d'une marche aléatoire à support fini sur un groupe G , la technique de linéarisation amène à étudier une chaîne de Markov sur $G \times [r]$ pour un certain entier $r \geq 1$. Cette nouvelle chaîne hérite d'une certaine forme de transitivité dans la coordonnée du groupe, ce qui lui assure suffisamment de structure pour appliquer essentiellement les mêmes arguments que pour les marches aléatoires classiques sur les groupes. Plus précisément, une marche aléatoire colorée $Y = (Y_t)_{t \geq 0}$ sur $G \times [r]$ est définie par un ensemble de matrices $(p_g)_{g \in G}$ de taille $r \times r$ à coefficients positifs telles que $\sum_{g \in G} p_g$ est stochastique et pour tous $g, h \in G, u, v \in [r]$,

$$\mathbb{P}[Y_{t+1} = (h, v) : | : Y_t = (g, u)] = p_{g^{-1}h}(u, v).$$

Le cas $r = 1$ correspond à une marche aléatoire classique sur un groupe. Nous appelons ces chaînes des marches aléatoires colorées car nous interprétons la deuxième coordonnée comme une couleur qui change à mesure que la marche évolue sur le groupe

G. Notre contribution en ce qui concerne les chaînes de Markov colorées est double. D'abord, nous donnons des procédures explicites de linéarisation pour étudier les marches aléatoires à support fini au moyen de marches colorées aux plus proches voisins. Le principe de base est très simple : les grands sauts effectués par les marches à support fini sont divisés en plusieurs étapes individuelles. L'ajout d'une coordonnée de couleur permet ensuite de déterminer à quelle étape du grand saut se trouve la chaîne. Deuxièmement, nous donnons un exemple illustrant comment les marches aléatoires colorées peuvent être étudiées de manière similaire aux marches classiques. Cet exemple consiste en l'analyse des propriétés asymptotiques de la marche à support fini au moyen de la théorie ergodique. Nous obtenons ainsi des formules explicites pour la vitesse et l'entropie des marches à support fini sur les groupes libres et le produit libre de groupes finis, qui n'étaient disponibles jusqu'à présent que dans le cas de marches aux plus proches voisins.

Accélération de chaînes de Markov

La théorie des temps de mélange trouve son origine dans le résultat fondamental selon lequel les chaînes de Markov, sous certaines hypothèses facilement réalisées, convergent en loi vers un état d'équilibre. Un exemple populaire est donné par les mélanges de cartes : il semble intuitif que plus un paquet de cartes est battu, mieux il sera mélangé. Mais combien de battages sont réellement nécessaires ? Les temps de mélange fournissent des réponses rigoureuses à cette question, dans un contexte beaucoup plus large. Cette question est également d'importance pratique, au cœur des méthodes dites de Monte Carlo par chaînes de Markov (MCMC).

Le cadre formel est le suivant : soit $X = (X_t)_{t \geq 0}$ une chaîne de Markov sur un espace d'états fini ou dénombrable. Supposons que X est récurrente positive, irréductible et apériodique. X_t converge alors en distribution vers l'unique mesure invariante de X . L'étude des temps de mélange vise à préciser la convergence en quantifiant le temps nécessaire pour atteindre la distribution stationnaire. Pour quantifier la convergence, une mesure doit être choisie sur l'ensemble des mesures de probabilité. La distance la plus couramment considérée est la distance en variation totale : étant donné deux mesures de probabilité μ, ν sur un ensemble dénombrable S , on définit leur distance en variation totale par

$$\|\mu - \nu\|_{\text{TV}} := \sup_{A \subseteq S} |\mu(A) - \nu(A)|.$$

La convergence peut ainsi être quantifiée par le temps nécessaire pour que la distance passe en dessous d'un certain seuil. Étant donné P le noyau de transition de la chaîne,

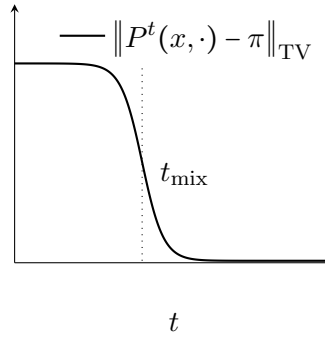


Figure 1: Le phénomène de cutoff.

$\varepsilon \in (0, 1)$ et un état initial $x \in S$, on définit le temps de mélange $t_{\text{mix}}(x, \varepsilon)$ par

$$t_{\text{mix}}(x, \varepsilon) := \inf \{t \geq 0 : \|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq \varepsilon\}.$$

On considère souvent l'état de départ le plus défavorable qui maximise le temps de mélange, ce qui conduit à considérer

$$t_{\text{mix}}(\varepsilon) := \max_{x \in S} t_{\text{mix}}(x, \varepsilon).$$

Lorsqu'il est enseigné dans un cours d'introduction sur les chaînes de Markov, le théorème de convergence est souvent démontré pour les chaînes finies en établissant une convergence exponentielle vers l'équilibre : supposant que S est fini, il existe $\alpha \in [0, 1)$ et $C > 0$ tels que

$$\max_{x \in S} \|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq C\alpha^t.$$

De cette inégalité, il peut être tentant de penser que la convergence est toujours exponentiellement rapide et que la notion de temps de mélange n'est en fin de compte pas si décisive. Ce serait une grave erreur. Alors que la convergence est en effet exponentielle asymptotiquement, elle semble souvent loin d'être exponentielle en pratique lorsque l'on se restreint à un intervalle de temps fini. Dans certains cas, la distance peut même diminuer comme une sorte de fonction seuil, comme celle illustrée dans la figure 1 : c'est le phénomène de cutoff, dont nous discuterons plus tard. Dans bien des cas, la chaîne de Markov étudiée dépend d'un paramètre qui mesure la taille ou la dimension de l'espace d'états. De telles chaînes de Markov sont par exemple la marche aléatoire simple sur un graphe à n sommets, sur l'hypercube n -dimensionnel, le mélange d'un paquet de n cartes, etc. À cet égard, il est crucial de comprendre comment les temps de mélange évoluent avec ce paramètre n . La question porte ainsi principalement sur l'obtention de développements asymptotiques des temps de mélange lorsque $n \rightarrow \infty$. Nous renvoyons à [117] et [136] pour des références sur les temps de mélange des chaînes de Markov finies.

Nous passons maintenant plus précisément au sujet de notre deuxième travail, qui porte sur l'accélération des chaînes de Markov. Ce sujet est motivé par des questions très pratiques dans le domaine des méthodes de Monte Carlo par chaînes de Markov (MCMC) [148, 147]. Les méthodes MCMC sont une classe de schémas d'échantillonnage basé sur la convergence des chaînes de Markov. Dans de nombreux cas pratiques, la distribution de probabilité cible est trop complexe sur le plan computationnel pour être réalisée directement. Ce problème se pose particulièrement dans le contexte de la simulation de modèles probabilistes en physique statistique [131] ou en statistiques bayésiennes [139] : dans ces contextes, les états consistent généralement en des configurations de haute dimension modélisées par des distributions de Gibbs : une configuration est dans l'état ψ avec une probabilité définie à une constante de normalisation près par

$$\mathbb{P}[\psi] \propto e^{-H(\psi)}$$

où $H(\psi)$ est l'énergie de la configuration. La probabilité d'une configuration peut donc être calculée à une constante de normalisation de près, la fonction de partition, tandis que l'estimation de celle-ci nécessite un nombre important de calculs, bien trop grand pour être réalisé numériquement. Les méthodes MCMC tentent de contourner cette difficulté en exploitant la convergence des chaînes de Markov vers une mesure stationnaire : afin d'échantillonner selon une distribution de probabilité cible π , l'idée est d'échantillonner plutôt une chaîne de Markov ayant π comme mesure invariante et d'attendre que la chaîne converge vers son équilibre. Dans le cas des distributions de Gibbs, la nature locale des phénomènes généralement impliqués permet de concevoir des chaînes qui contournent le problème du calcul de la constante de normalisation : c'est l'échantillonneur de Gibbs, également appelé dynamique de Glauber dans le cas des systèmes de spins. Même dans des contextes plus généraux, trouver une chaîne de Markov avec une mesure invariante prescrite est en théorie résolu par l'algorithme de Metropolis-Hastings [134, 86], (voir aussi [55] et [117, Chapitre 3]). Cependant, la convergence pour de nombreux cas d'intérêt est souvent trop lente pour être utilisée en pratique. Il faut donc concevoir de nouvelles chaînes spécifiquement adaptées au problème à résoudre. Une autre option consiste à essayer d'améliorer les dynamiques déjà disponibles : c'est l'accélération des chaînes de Markov. Nous renvoyons à l'article [58] de Diaconis pour une introduction au sujet. De nombreuses méthodes d'accélération consistent à donner une certaine inertie à la chaîne, ce qui signifie en pratique considérer des chaînes non réversibles sur des espaces d'états augmentés, voir [47, 60, 7, 140] et les références qui y sont citées. Ces mécanismes d'accélération sont commun à ceux proposés pour les algorithmes d'optimisation et de passage de messages [146].

Dans notre travail, nous nous concentrons sur un mécanisme d'accélération quelque

peu étonnant, qui consiste à insérer des étapes d'une dynamique déterministe entre celles de la chaîne de Markov. L'accélération résultante est spectaculaire : pour une chaîne de Markov générale, le temps de mélange peut être réduit jusqu'à devenir logarithmique en la taille de l'espace d'états en prenant pour la dynamique déterministe une application bijective choisie... au hasard. Ce résultat est dû à Chatterjee et Diaconis [45], mais l'histoire ne commence pas là. L'accélération par une application déterministe a été identifiée pour la première fois par Chung, Diaconis et Graham [49] dont la motivation était d'étudier le temps de mélange d'un processus apparaissant dans la génération de nombres pseudo-aléatoires. Leur modèle consiste en la chaîne de Markov X sur $\mathbb{Z}/n\mathbb{Z}$ définie par l'équation de récurrence : $X_{t+1} = aX_t + B_t$, avec (B_t) une suite d'incrémentes iid uniformément distribués sur $\{-1, 0, 1\}$. Pour $a = 2$, ils ont prouvé qu'après $t = O(\log n \log \log n)$ étapes, la distribution de X_t est proche de l'uniforme, ce qui montre déjà une accélération considérable par rapport à la marche aléatoire simple : cette dernière, obtenue avec $a = 1$, est connue pour mélanger en environ n^2 étapes. Ce n'est pas l'accélération promise en $O(\log n)$ étapes, mais Chung, Diaconis et Graham prouvent aussi que pour "la plupart des suites de n ", le temps de mélange est en fait vraiment de l'ordre de $\log n$.

Le phénomène d'accélération de chaînes de Markov par des fonctions déterministes a été récemment remis au goût du jour. Dans [45], Chatterjee et Diaconis considèrent le modèle suivant : soit P une matrice stochastique sur $[n]$ et S la matrice d'une permutation de n éléments uniformément distribués sur le groupe symétrique. Sous des hypothèses peu contraignantes, les auteurs prouvent que le temps de mélange de la chaîne de matrice de transition PS est $O(\log n)$ avec grande probabilité. Cependant, ils notent également que des exemples explicites sont particulièrement difficiles à trouver. Dans notre travail, nous étudions un analogue multidimensionnel du processus de Chung-Diaconis-Graham : nous considérons une chaîne X sur $(\mathbb{Z}/n\mathbb{Z})^d$, $d \geq 1$ définie par l'équation de récurrence :

$$X_0 := 0 \quad X_t = AX_{t-1} + B_t \pmod{n} \quad \forall t \geq 1$$

Ici $A \in \text{GL}_d(\mathbb{Z})$ est une matrice hyperbolique inversible, par quoi nous entendons que A est inversible, son inverse a des coefficients entiers et A n'a pas de valeur propre de module 1, tandis que les variables B_t sont des variables aléatoires iid sur \mathbb{Z}^d distribuées selon une loi μ . De tels cas multidimensionnels ont été précédemment étudiés par Ascì [8, 9], Hildebrand [100] et Klyachko [106] qui ont prouvé une accélération en $O((\log n)^2)$ étapes sous de faibles hypothèses sur μ . Notre contribution est d'établir des limites similaires au cas unidimensionnel : nous obtenons un mélange en $O(\log n \log \log n)$ étapes et même $O(\log n)$ étapes pour "la plupart des suites de n ". De plus, nous sug-

gérons une piste sur la manière de construire des applications explicites à l'origine d'une accélération : notre preuve est basée sur les propriétés de la dynamique déterministe $x \mapsto A^\top x$ sur le tore continu $(\mathbb{R}/\mathbb{Z})^d$, qui est connue pour présenter un comportement chaotique. Nous suggérons donc que l'accélération est dans ce cas à expliquer par le comportement pseudo-aléatoire de la dynamique déterministe, et que d'autres exemples pourraient potentiellement être obtenus de cette manière.

Concentration sur le groupe symétrique

L'une des lois fondamentales de la théorie des probabilités, la loi des grands nombres, affirme que la moyenne empirique d'une famille de variables aléatoires iid converge vers son espérance. Il s'agit d'une instance du phénomène de concentration de mesure : des fonctions qui dépendent de manière régulière d'un grand nombre de variables aléatoires indépendantes ont tendance à se comporter de manière déterministe et à concentrer autour de leur moyenne. La concentration de la mesure est devenue un aspect central de la théorie des probabilités moderne, nous renvoyons aux monographies de Ledoux [111] et Boucheron, Lugosi et Massart [34] pour une introduction au sujet. Souvent, le terme de concentration de la mesure se réfère de manière plus précise à l'utilisation d'inégalités de concentration, qui quantifient à quel point une fonction donnée est proche de sa moyenne. Le phénomène de concentration de la mesure a également une forte teinte géométrique : dans le cas de variables aléatoires indépendantes, au lieu d'interpréter la concentration comme la propriété de nombreuses variables aléatoires interagissant entre elles, on peut la voir comme la propriété d'une mesure (la mesure produit) sur un espace de grande dimension (un espace produit). Le phénomène de concentration de la mesure va donc bien au-delà du monde des variables aléatoires indépendantes. Notre travail sur le phénomène de cutoff, présenté dans la section suivante, nous a amenés à considérer un tel problème de concentration de mesure pour la mesure uniforme sur le groupe symétrique. Les phénomènes de concentration pour les groupes et plus généralement pour les variables aléatoires faiblement dépendantes ont été étudiés par Chatterjee dans sa thèse de doctorat [41, 43, 42], qui a développé des techniques pour obtenir des inégalités de concentration basées sur la méthode de Stein. Parmi ces travaux, Chatterjee a établi une inégalité de type Bernstein pour la mesure uniforme sur le groupe symétrique qui peut être énoncée comme suit : soit $A \in M_n(\mathbb{R}_+)$ une matrice à coefficients positifs, $\sigma \in S_n$ une permutation uniforme et considérons la variable aléatoire

$$Z = \sum_i A_{i\sigma(i)}.$$

Pour tout $t \geq 0$

$$\mathbb{P}[|Z - \mathbb{E}[Z]| \geq t] \leq 2 \exp\left(\frac{-t^2}{2\|A\|(t + 2\mathbb{E}[X])}\right),$$

où $\|A\| = \max_{i,j \in [n]} A_{i,j}$.

La variable aléatoire Z peut être vue comme provenant d'une fonction linéaire sur le groupe symétrique, en ce sens qu'elle est une combinaison linéaire d'indicatrices $S_{ij} := \mathbb{1}_{\sigma(i)=j}$, qui sont les coefficients de la matrice de permutation S . Dans notre travail, nous proposons une version d'ordre supérieur de ce résultat, c'est-à-dire une inégalité de concentration pour des variables aléatoires qui s'écrivent comme des polynômes réguliers et de faible degré en les entrées de S . Si notre point de départ a été la représentation des polynômes de degré supérieur comme des applications linéaires sur des produits tensoriels d'espaces vectoriels (ce qui constitue une autre astuce de linéarisation) le résultat final ne nécessite pas l'introduction d'une matrice A . Au lieu de cela, nous considérons directement des fonctions ϕ sur l'espace des matrices réelles $M_n(\mathbb{R})$ de taille $n \times n$. Comme les fonctions doivent être appliquées à des matrices de permutation, certaines contraintes peuvent être imposées, comme le fait que le degré est au plus un en chaque entrée. Il est cependant crucial que les coefficients du polynôme soient positifs. Pour une telle fonction polynomiale ϕ et S la matrice d'une permutation uniforme, le théorème 4.1 établit une inégalité de concentration qui prend la forme d'une inégalité de Bernstein similaire à l'inégalité unidimensionnelle ci-dessus. Le paramètre de régularité qui joue le rôle de $\|A\|$ ayant une expression plus complexe, nous ne tenterons pas de donner une déclaration plus précise ici. Ce paramètre de régularité fait intervenir le degré ainsi que la norme ℓ^∞ du gradient de ϕ , mais nécessite également la considération de fonctions auxiliaires dont les degrés s'échelonnent jusqu'à celui de ϕ . Il s'agit d'une conséquence de la preuve, basée sur une récurrence sur le degré. S'inspirant de la méthode de Chatterjee, la preuve consiste in fine à prouver une borne supérieure sur la fonction génératrice des moments. En utilisant une représentation tensorielle linéaire de la fonction polynomiale, la méthode de Stein des paires échangeables aboutit à une formule de récurrence sur la fonction génératrice des moments à partir de laquelle une borne supérieure est obtenue par récurrence.

Le phénomène de cutoff

Découvert par Diaconis, Shahshahani et Aldous [61, 4, 5], le phénomène de cutoff est une caractéristique étonnante observée dans un grand nombre de chaînes de Markov, qui consiste en une convergence abrupte vers l'équilibre. La définition formelle est la suivante. Comme d'habitude lorsque l'on considère le temps de mélange d'une chaîne de Markov, on considère en réalité une famille de chaînes de Markov indexée par un

entier n . La chaîne présente le phénomène de cutoff si pour tout $\varepsilon \in (0, 1)$, son temps de mélange satisfait

$$\frac{t_{\text{mix}}(\varepsilon)}{t_{\text{mix}}(1 - \varepsilon)} \xrightarrow{n \rightarrow \infty} 1.$$

La chaîne de Markov la plus simple à présenter le phénomène de cutoff est peut-être la marche aléatoire simple sur l'hypercube n -dimensionnel $\{0, 1\}^n$, qui a été démontrée avoir un cutoff au temps $\frac{1}{4}n \log n$ pour sa version continue par Aldous. Des exemples historiques incluent le célèbre processus des transpositions aléatoires dont le cutoff au temps $\frac{1}{2}n \log n$ a été prouvé par Diaconis et Shahshahani [61], ou le mélange américain (riffle shuffle), modélisé par Gilbert-Shannon et Reeds. Aldous [4] a d'abord établi le cutoff au temps $\frac{3}{2}n \log n$, après quoi une analyse plus approfondie a été menée par Bayer et Diaconis [15]. De leur travail découle le résultat quelque peu informel mais immensément populaire selon lequel un jeu de 52 cartes doit être mélangé exactement 7 fois avec le battage américain pour être bien mélangé. On pense que le phénomène de cutoff est une caractéristique très courante des chaînes de Markov et aujourd'hui de nombreux exemples soutiennent cette croyance. Pour n'en nommer que quelques-uns, le cutoff a été établi pour de nombreuses marches aléatoires sur des groupes finis, pour lesquelles nous renvoyons à [57, 150, 48], les chaînes de naissance et de mort [63], le modèle de Bernoulli-Laplace [62, 70, 3], certaines dynamiques de Glauber pour des systèmes de spin, en particulier pour le modèle d'Ising à haute température [116, 64, 123, 124], certains systèmes de particules en interaction unidimensionnels, pour lesquels nous renvoyons à l'article de Lacoïn [108] et à la thèse de doctorat de Schmid [151].

Cependant, toutes les chaînes de Markov ne présentent pas de cutoff : c'est le cas par exemple des chaînes de faible dimension que sont les marches aléatoires simples sur une grille ou un tore (cela peut être prouvé grâce à la condition de produit donnée après). La question centrale des recherches sur le phénomène de cutoff est ainsi celle de déterminer une caractérisation simple du cutoff. Idéalement, une bonne caractérisation devrait être assez simple pour éviter le calcul précis du temps de mélange, ne nécessitant de connaître que l'ordre du terme dominant. À ce jour, nous ne disposons pas de telle caractérisation mais d'importants progrès ont été réalisés. Une condition nécessaire connue pour les chaînes réversibles est la condition de produit. Une chaîne de matrice de transition P et mesure invariante π est réversible si elle satisfait l'équation

$$\pi(x)P(x, y) = \pi(y)P(y, x)$$

pour tous les états x, y . Soit λ_* la valeur propre de P de valeur absolue maximale mais distincte de la valeur propre triviale 1. On définit le *trou spectral absolu* comme $\gamma := 1 - \lambda_*$ et le temps de relaxation $t_{\text{rel}} := \gamma^{-1}$. La condition de produit est la propriété

que pour tout $\varepsilon \in (0, 1)$,

$$\frac{t_{\text{mix}}(\varepsilon)}{t_{\text{rel}}} \xrightarrow{n \rightarrow \infty} \infty.$$

La condition de produit a d'abord été conjecturée être une équivalence, mais il existe des contre-exemples, comme [117][Exemple 18.7]. Une autre caractérisation de cutoff a été établie par Basu, Hermon et Peres dans [14] pour les chaînes réversibles en termes de temps d'atteintes, mais elle nécessite le calcul précis du temps de mélange.

La compréhension du phénomène a fait d'importants progrès lorsque l'attention s'est portée vers le cas de chaînes génériques. Ce changement a été initié par le travail fondateur de Lubetzky et Sly [122] établissant le cutoff pour les marches aléatoires sur des graphes réguliers aléatoires. Comme ces chaînes n'ont pas de probabilités de transition fixes et connues à l'avance, la stratégie de preuve ne peut pas s'appuyer sur une connaissance explicite de la matrice de transition, comme c'était le cas pour les travaux précédemment cités. Au lieu de cela, le résultat s'appuie sur un phénomène de concentration entropique. Une heuristique peut être décrite comme suit : idéalement l'objectif est de prouver l'existence d'une constante $h > 0$ telle que $-\log P^t(X_0, X_t)$ se concentre autour de th , pour tous les temps t jusqu'au temps de mélange. Si cela est vrai, cette propriété implique que X_t est essentiellement uniformément distribuée sur un ensemble de taille e^{th} . À condition que la mesure invariante de X_t soit suffisamment dispersée dans l'espace des états, comme c'est le cas par exemple pour la mesure uniforme, cela implique immédiatement que $\log n/h$ étapes sont nécessaires, mais aussi suffisantes pour le mélange. Depuis les travaux de Lubetzky et Sly, la méthode entropique a donné lieu à une longue lignée de papiers établissant le cutoff pour des chaînes génériques au temps entropique $\log n/h$. À partir de ces travaux, que nous présentons dans le Chapitre 5, une percée récente a été faite par Salez [149] qui a donné une condition suffisante de cutoff s'exprimant en termes de concentration entropique. Similaire à la condition de produit ci-dessus, elle peut en théorie être vérifiée sans la connaissance précise du temps de mélange. De cela, Salez déduit le cutoff pour une large famille de chaînes satisfaisant une hypothèse de courbure non négative.

La dernière de nos contributions présentées dans cette thèse s'inspire du papier de Hermon, Sly et Sousi [92] dont l'objet d'étude est la marche aléatoire simple sur les graphes auxquels sont rajoutés des arêtes aléatoires correspondant à un appariement uniforme des sommets. Sous des hypothèses faibles sur la suite de graphes, la marche aléatoire simple présente un cutoff au temps entropique. Notre travail dans le domaine du cutoff est une tentative de généraliser ce modèle à des chaînes plus générales. En partant de deux matrices stochastiques déterministes P_1, P_2 sur $[n]$, nous étudions les propriétés de mélange d'une combinaison aléatoire de P_1 et P_2 : en notant S la matrice d'une permutation uniforme σ et $p \in (0, 1)$, nous analysons le mélange de la chaîne de

matrice de transition

$$\mathcal{P} = pP_1 + (1-p)SP_2S^{-1}.$$

En réalité, nous étudions un modèle légèrement plus général où nous permettons à la probabilité p de dépendre de l'état actuel x de la chaîne et de son image $\sigma(x)$. Cette généralisation est nécessaire pour inclure le cas de marches aléatoires sur une superposition de graphes. En particulier, en prenant pour P_1, P_2 les matrices de la marche aléatoire simple sur un graphe et sur une famille d'arêtes respectivement, un bon choix de p redonne le modèle de Hermon, Sly et Sousi. Nous prouvons un phénomène de cutoff au temps entropique dans le sens suivant : nous établissons que dans notre modèle, il existe une mesure π telle que pour tous $\varepsilon \in (0, 1)$, $t = \log n/h + a_n\sqrt{\log n}$, pour tout point de départ $x \in [n]$, la chaîne est avec une grande probabilité à une distance au moins $1 - \varepsilon$ en variation totale de la mesure π si $a_n \rightarrow -\infty$, et est sinon ε -proche de π si $a_n \rightarrow +\infty$ et $t = O(\log n)$. Nous conjecturons que la mesure limite π est elle-même ε -proche d'une vraie mesure invariante, ce qui permettrait d'en déduire des estimations de temps de mélange correspondantes. Malheureusement, au moment où cette thèse est écrite, cette conjecture fait encore l'objet de recherches actives. Le problème principal vient du fait que la propriété de mélange ne vaut que pour un point de départ typique : l'état initial de la chaîne doit être fixé avant de générer l'environnement ou peut être choisi uniformément au hasard, mais ne peut pas être choisi arbitrairement après que l'environnement a été généré.

Chapter 1

Introduction

This dissertation is the synthesis of the work achieved during my thesis. It is based on three papers that explore different aspects of the theory of Markov chains: [30], done jointly with Charles Bordenave, [65] done jointly with Laurent Massoulié, and a third one that has not been published yet. Markov chains are stochastic processes that have no memory of the past. As such, they are the simplest kind of stochastic processes with a non-trivial dependence in time and for this reason have always been a central object in probability theory. There exists many excellent references on the subject: the monographs of Durrett [66] and Norris [142] are certainly a good place to start, with many examples and applications. The books by Levin, Peres and Wilmer [117] on mixing times, Aldous and Fill [6] and Lyons and Peres [128] on reversible chains and related topics are other cornerstones of the litterature, which have been used extensively during the realization of this thesis. We consider exclusively discrete-time Markov chains with discrete state space. From a general perspective, our study of Markov chains is motivated by the question of identifying the underlying mechanisms that account for the specific behaviour of a chain. In particular, we are especially interested in understanding the relation between the chain and the geometry of its state space. Our works all fall in that general setting but tackle distinct questions. For this reason, we will try to give here a concise overview of our contributions and give general references. Then each chapter will be dedicated a proper introduction and review of the litterature more specifically related with our works.

1.1 Random walks on groups

The term random walks generally refers to a particular kind of Markov chains that one likes to visualize as evolving on a geometric structure like a graph. Within this category, random walks on groups are Markov chains that exhibit a lot of symmetry

and arise naturally in various theoretical and applied problems. The simple random walk on the line, perhaps the most canonical of Markov chains, is a random walk on the additive group \mathbb{Z} . Random walks on groups are the direct generalization of sums of iid random variables to more general spaces. Specifically, given a group G equipped with a probability measure p , consider the Markov chain $X = (X_t)_{t \geq 0}$ on G defined by

$$\forall g, h \in G, \forall t \geq 0: \quad \mathbb{P}[X_{t+1} = h \mid X_t = g] = p(g^{-1}h).$$

Equivalently, X_t can be written as the product $X_t = Z_0 Z_1 \dots Z_t$ of independent "increments" on G all distributed with law p , except potentially Z_0 . Such a chain is called a convolution random walk on G , often abbreviated as simply random walk on the group G .

Our first work is concerned more specifically with the case where the group G is countable and given by a finite set of generators S . Define the support of the random walk as the support of the measure p , that is the set of elements that have positive mass under p . The random walk is said to be finitely supported, or to have finite range, if the measure p is finitely supported. It is a nearest-neighbour random walk if p is supported on the set of generators S .

The first contribution of this thesis is motivated by the general observation that the study of a random walk is often considerably simpler if it is supposed to be nearest-neighbour. The analysis of the asymptotic behaviour of random walks on groups gives good example of what the difficulties can be. A large part of this theory can be traced back to the work of Furstenberg [76], Avez [12], Guivarc'h [82], Derriennic [54], Kaimanovich and Vershik [102]. One of the essential features of convolution random walks is transitivity: by means of group multiplication, the starting state of the random walk can always be translated to the identity element. Consequently, the law of the walk is essentially independent of the starting point. From this observation, ergodic theorems such as Kingman's subadditive theorems establish the existence of the almost sure limits

$$v := \lim_{t \rightarrow \infty} \frac{|X_t|}{t}, \quad h := \lim_{t \rightarrow \infty} -\frac{\log \mathbb{P}[X'_t = X_t] X}{t}.$$

provided $\sum_{g \in G} p(g) |g| < \infty$ and where X' denotes an independent copy of the chain. Given $g \in G$, $|g|$ denotes here the minimal number of elements that are needed to write g as a product of generators. This can be interpreted as a distance between g and the identity element. The quantity v is for this reason called the speed, or drift, of the random walk. The limit h is the asymptotic entropy, or Avez entropy, named after Avez who introduced it in [12]. The asymptotic entropy contains much of the information about the asymptotic behaviour of the walk. For instance Kaimanovich and Vershik

[102] and Derriennic [54] proved that the tail σ -algebra of the random walk is non-trivial if and only $h > 0$.

Consider now the specific case where $G = \mathbb{F}_d$ is the free group on d generators. The generating set considered is that of the d canonical generators and their inverses. This group can be represented by a graph (its Cayley graph), which is nothing but a $2d$ -regular tree. A general random walk can make arbitrarily large jumps in this tree, a walk with finite support makes bounded jumps, while finally a nearest-neighbour moves along the edges only. This last property coupled with the tree structure is extremely useful: to reach a vertex y in the tree from a vertex x , the walk has to go through all vertices in between. This paves the way for recurrence arguments, from which it is possible to get explicit formulas for the drift and entropy of the walk [114]. In the finite support case, such arguments fail, and for this reason no explicit formulas seemed achievable.

In our work we introduce a novel technique to study random walks on groups which allows to reduce to the nearest-neighbour case. The name "linearization" comes from the so-called "linearization trick" used notably in operator algebras, where our works finds its inspiration [89, 83]. The price to pay for considering only the nearest-neighbour case is that the resulting Markov chain is not a random walk on a group but a slightly more general Markov chain which we call a colored random walk. Basically, starting from a random walk with finite support on a group G , we will be left to study a Markov chain on $G \times [r]$, for some integer $r \geq 1$. The new Markov chains inherits some transitivity property in the group coordinate from the starting walk, so it keeps sufficient structure to apply the arguments for classical random walks on groups. Namely, a colored random walk $Y = (Y_t)_{t \geq 0}$ on $G \times [r]$ is defined by a set of $r \times r$ matrices $(p_g)_{g \in G}$ with non-negative entries such that $\sum_{g \in G} p_g$ is stochastic and for all $g, h \in G, u, v \in [r]$,

$$\mathbb{P}[Y_{t+1} = (h, v) \mid Y_t = (g, u)] = p_{g^{-1}h}(u, v).$$

The case $r = 1$ corresponds to a classical random walk on a group. We call such chains colored random walks because we interpret the second coordinate as a color that changes as the walk evolves on the group G . Our contribution in regard to colored Markov chains are twofold. First we give explicit procedures to study finitely supported random walks by means of nearest-neighbour colored walks. The basic principle is very simple: large jumps made by finitely supported walks are splitted into several individual steps. The addition of a color coordinate then allows to determine what stage of a large jump the chain is at. Secondly we give an example of how colored random walks can be studied in essentially the same manner as classical walks. This example consists in the

analysis of asymptotic properties of finitely supported walk by means of ergodic theory. From this we obtain explicit formulas for the drift and entropy of finitely supported walks on free groups and free product of finite groups, which were so far only available in the nearest-neighbour case.

1.2 Acceleration of Markov chains

The theory of mixing times originates from the fundamental result that Markov chains, under mild assumptions, converge in distribution to an equilibrium state. A popular example of this is given by card shuffles: it seems intuitive that the more shuffles are performed, the better mixed the deck will be. But how many shuffles are really needed? Mixing times provide rigorous answers this question, in a much broader context. This question is also of practical importance, at the core of the so-called Markov Chain Monte Carlo (MCMC) methods.

The formal setting is as follows: let $X = (X_t)_{t \geq 0}$ be a Markov chain on a finite or countable state space. Assume X is positive recurrent, irreducible and aperiodic. Then X_t converges in distribution towards the unique invariant measure of X . The study of mixing times aims to make the convergence more precise by quantifying the time needed to reach the stationary distribution. To quantify the convergence, a metric needs to be chosen on the set of probability measures. The metric most commonly considered is the total variation distance: given two probability measures μ, ν on a countable set S , define their total variation distance as

$$\|\mu - \nu\|_{\text{TV}} := \sup_{A \subseteq S} |\mu(A) - \nu(A)|.$$

Convergence can thus be quantified by the time necessary for the distance to pass below a given threshold. Letting P denote the transition kernel of the chain, $\varepsilon \in (0, 1)$ and a starting state $x \in S$, define the mixing time $t_{\text{mix}}(x, \varepsilon)$ as

$$t_{\text{mix}}(x, \varepsilon) := \inf\{t \geq 0 : \|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq \varepsilon\}.$$

One often considers the worst possible starting state which maximizes the mixing time, leading to the consideration of

$$t_{\text{mix}}(\varepsilon) := \max_{x \in S} t_{\text{mix}}(x, \varepsilon).$$

When taught in an introductory course about chains, the convergence theorem is often proved for finite chains by establishing uniform ergodicity, which asserts exponential convergence to equilibrium: assuming S is finite there exists $\alpha \in [0, 1)$ and $C > 0$ such that

$$\max_{x \in S} \|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq C\alpha^t.$$

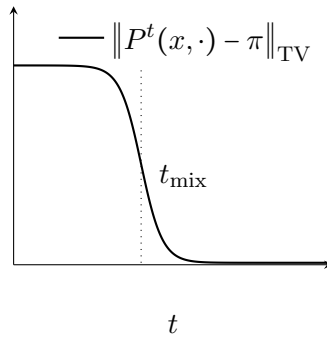


Figure 1.1: The cutoff phenomenon.

From this equation, it may be tempting to think that convergence is always exponentially fast and the notion of mixing times not so decisive after all. This would be a fatal mistake. While the convergence is indeed exponential asymptotically, it often looks far from being exponential in practice when one restricts to a finite interval of time. In some cases, the distance can even decrease as some kind of a step function, as in Figure 1.1: this is the cutoff phenomenon, discussed later. In most cases of interest, the Markov chain depends on a parameter that measures the size or dimension of the state space. Such Markov chains are for instance the simple random walk on graphs with n vertices, on the n -dimensional hypercube, shuffles of a deck of n cards, etc. In that regard, it is crucial to understand how mixing times scale with this parameter n . All in all, the theory is thus mainly about obtaining asymptotic developments of the mixing times in the limit $n \rightarrow \infty$. We refer to [117] and [136] for references on mixing times of finite Markov chains.

We now move on more closely to the topic of our second work, which is about the acceleration of Markov chains. This subject finds its motivations in very practical matters that are Markov Chain Monte Carlo (MCMC) methods [148, 147]. MCMC methods consist in the design of sampling schemes to sample from specific target distributions. Most often than not, the probability distribution of interest is computationally too complex to be performed directly. This issue arises particularly in the context of simulating probabilistic models in statistical physics [131] or in Bayesian statistics [139]: in these settings states generally consist of high-dimensional configurations modeled with Gibbs distributions: a configuration is in state ψ with probability defined up to a normalizing constant by

$$\mathbb{P}[\psi] \propto e^{-H(\psi)}$$

where $H(\psi)$ is the energy of the configuration. The probability of a configuration can thus be well estimated up to the normalizing constant, the partition function, which requires an enormous number of computations, often too large to be realized

numerically. MCMC methods address this difficulty by leveraging the convergence of Markov chains towards a stationary measure: in order to sample from a target probability distribution π , the idea is to sample instead from a Markov chain with π as invariant measure and wait for the chain to converge to its equilibrium. In the case of Gibbs distributions, the local nature of the phenomena generally involved allow to design chains that bypass the issue of computing the normalizing constant: this is the Gibbs sampler, also called Glauber dynamics in the case of spin systems. Even in more general settings finding a Markov chain with prescribed invariant measure is in theory solved by the Metropolis-Hastings algorithm [134, 86], (see also [55] and [117][Chapter 3]). However the convergence for many cases of interest is often too slow to be used in practice. One therefore needs to design new chains specifically tailored to the problem at hand. Another option consists in trying to improve already available dynamics: this is the acceleration of Markov chains. We refer to the survey [58] by Diaconis for an introduction to the subject. Numerous works consider speed-ups obtained by giving some inertia to the chain, which means in practice considering non-reversible chains on augmented (lifted) state spaces. See [47, 60, 7, 140] and the references therein. We note that similar acceleration mechanisms have been considered for optimization and message-passing algorithms [146].

In our work, we focus on a somewhat peculiar acceleration mechanism, that consists in inserting steps of a deterministic dynamics between those of the Markov chain. The resulting speed-up is spectacular: for general Markov chains, the mixing time can be brought down to logarithmic order in the size of the state space by taking for the deterministic dynamics a bijective map chosen ... at random. This result is due to Chatterjee and Diaconis [45], however this is not the origin of this story. The acceleration by a deterministic map was first identified by Chung, Diaconis and Graham [49] whose motivation was to study the mixing time of a process arising in the generation of pseudo-random numbers. Their model consists in the Markov chain X on $\mathbb{Z}/n\mathbb{Z}$ defined by the recursion: $X_{t+1} = aX_t + B_t$, with (B_t) a sequence of iid increments uniformly distributed on $\{-1, 0, 1\}$. For $a = 2$ they proved that after $t = O(\log n \log \log n)$ steps the distribution of X_t is close to uniform, which already shows a dramatic speed-up over the simple random walk: the latter is obtained with $a = 1$ and is known to mix in around n^2 steps. This is not the promised acceleration to $O(\log n)$ steps, however Chung, Diaconis and Graham also prove that for "most sequences of n ", the mixing time is in fact really of order $\log n$.

Recently, attention has been brought back to the potential speed-up obtained by applying deterministic functions to Markov chains. In [45] Chatterjee and Diaconis consider the following setting: let P be a stochastic matrix on $[n]$ and S the matrix

of a permutation of n elements uniformly distributed on the symmetric group. Under mild assumptions, they prove that the mixing time of the chain with transition matrix PS is $O(\log n)$ with high probability. However they also note that explicit examples are particularly challenging to find. In our research, we study a multi-dimensional analog of the Chung-Diaconis-Graham process: we consider a chain X on $(\mathbb{Z}/n\mathbb{Z})^d$, $d \geq 1$ defined by the recursion:

$$X_0 := 0 \quad X_t = AX_{t-1} + B_t \pmod{n} \quad \forall t \geq 1$$

Here $A \in \text{GL}_d(\mathbb{Z})$ is an invertible hyperbolic matrix, by which we mean that A is invertible, its inverse has integer entries and A has no eigenvalue of modulus 1, whereas B_t are iid random variables on \mathbb{Z}^d with some distribution μ . Such multi-dimensional cases have been previously studied by Ascì [8, 9], Hildebrand [100] and Klyachko [106] who proved acceleration to $O((\log n)^2)$ under mild assumptions on μ . Our contribution is to establish bounds that match the one-dimensional case: we obtain mixing in $O(\log n \log \log n)$ and $O(\log n)$ steps for "most sequences of n ". Furthermore, we suggest a lead as to how explicit maps causing speed-up may be constructed: our proof is based on the properties of the deterministic dynamical $x \mapsto A^\top x$ on the continuous torus $(\mathbb{R}/\mathbb{Z})^d$, which is known to exhibit chaotic behaviour. We thus suggest that the speed-up is in this case to be explained by the pseudo-random behaviour of the deterministic dynamics, and that other examples may potentially be obtained this way.

1.3 Concentration on the symmetric group

One of the fundamental laws of probability, the law of a large numbers, asserts that the empirical mean of a family of iid random random variables converges to its expectation. This is an instance of the concentration of measure phenomenon: functions that depend smoothly on a large numbers of independent random variables tend to behave deterministically and concentrate around their mean. Concentration of measure has become a central aspect of modern probability, we refer to the monographs of Ledoux [111] and Boucheron, Lugosi and Massart [34] for an introduction to the subject. Often, concentration of measure refers more closely to the use of concentration inequalities, which quantify how close a given function is with respect to its mean. It also comes with a geometric flavour: in the case of independent random variables, instead of interpreting concentration as the property of many random variables interacting with each other, it can be seen as the property of one measure (the product measure) on a given high-dimensional space (a product space). The concentration of measure thus goes beyond the world of independent random variables. Our work on the cutoff phenomenon,

discussed in the next section, led us to consider such a problem of concentration of measure for the uniform measure on the symmetric group. Concentration phenomena for groups and more generally for weakly dependent random variables were studied by Chatterjee in his PhD. thesis [41, 43, 42], who developed techniques to obtain concentration inequalities based on Stein's method. Among those works, Chatterjee established a Bernstein-like inequality for the uniform measure on the symmetric group that can be stated as follows: let $A \in M_n(\mathbb{R}_+)$, $\sigma \in S_n$ is a uniform permutation and consider the random variable

$$Z = \sum_i A_{i\sigma(i)}$$

Then for all $t \geq 0$

$$\mathbb{P}[|Z - \mathbb{E}[Z]| \geq t] \leq 2 \exp\left(\frac{-t^2}{2\|A\|(t + 2\mathbb{E}[X])}\right),$$

where $\|A\| = \max_{i,j \in [n]} A_{i,j}$. The random variable Z can be seen as arising from a linear function on the symmetric group, in the sense that it is a linear combination of indicators $S_{ij} := \mathbb{1}_{\sigma(i)=j}$, which are entries of the permutation matrix S . In our work, we propose a higher-order version of this result, that is a concentration inequality for random variables writing as smooth and low-degree polynomials in the entries of S . Our starting point was the representation of higher degree polynomials as linear maps on tensor vector spaces (which is another linearization trick) but the final result is not stated in terms of a matrix A . Instead we consider directly functions ϕ on the space of $n \times n$ real matrices $M_n(\mathbb{R})$. Since functions are to be applied to permutation matrices, some constraints can be imposed, such as the fact that the degree is at most one in each entry. It is crucial however that coefficients are non-negative. For such a polynomial function ϕ and S the matrix of a uniform permutation, Theorem 4.1 establishes a concentration inequality that takes the form of a Bernstein inequality similar to the one-dimensional inequality above. The smoothness parameter that plays the role of $\|A\|$ having a more involved expression we will not try to give a more precise statement here. This smoothness parameter involves the degree as well as the ℓ^∞ norm of the gradient of ϕ , but also requires the consideration of auxiliary functions with degrees up to that of ϕ . This is a consequence of the proof, which is based on an induction on the degree. Following Chatterjee's arguments, the proof consists eventually in proving an upper bound on the moment generating function. Using a linear tensor representation of the polynomial function, Stein's method of exchangeable pairs yields a recursion formula on the moment generating function from which an upper bound is obtained by induction.

1.4 The cutoff phenomenon

Discovered by Diaconis, Shahshahani and Aldous [61, 4, 5], the cutoff phenomenon is an intriguing feature observed in a large number of Markov chains, that consists in the abrupt convergence to equilibrium. The formal definition is the following. As usual when considering mixing time of a Markov chain one actually considers a family of Markov chains indexed by an integer n . The chain exhibits cutoff if for all $\varepsilon \in (0, 1)$, its mixing time satisfies

$$\frac{t_{\text{mix}}(\varepsilon)}{t_{\text{mix}}(1-\varepsilon)} \xrightarrow{n \rightarrow \infty} 1.$$

The simplest Markov chain to exhibit cutoff is perhaps the simple random walk on the n -dimensional hypercube $\{0, 1\}^n$, shown to have cutoff at time $\frac{1}{4}n \log n$ for its continuous version by Aldous. Historical examples include the famous random transposition process proved to exhibit cutoff at time $\frac{1}{2}n \log n$ by Diaconis and Shahshahani [61], and the riffle shuffle, modeled by Gilbert-Shannon and Reeds. Aldous [4] first established the cutoff at time $\frac{3}{2}n \log n$, after which a more thorough analysis was conducted by Bayer and Diaconis [15]. From their work comes the immensely popular statement that to mix a deck of 52 cards, 6 shuffles is not enough but 7 shuffles is more than satisfactory. The cutoff phenomenon is believed to be a very common feature of Markov chains and today numerous examples support this opinion. To name but a few, cutoff has been established for many random walks on finite groups, for which we refer to [57, 150, 48], birth and death chains [63], the Bernoulli-Laplace model [62, 70, 3], Glauber dynamics of spin systems, in particular for the Ising model at high temperature [116, 64, 123, 124], one-dimensional interacting particle systems, for which we refer to the survey of Lacoïn [108] and Schmid's PhD dissertation [151].

However not all Markov chains exhibit cutoff: examples are for instance low-dimensional chains such as simple random walks on grids and torii (this can be shown thanks to the product condition given after). Most of the research on the cutoff phenomenon has thus been fueled by the question of determining a simple characterization of cutoff. Ideally, a good characterization should be simple enough to avoid the sharp computation of the mixing time, so the prefactor in the leading term needs not to be known exactly. To this date, no such characterization is known but important progress has been made. A known necessary condition for reversible chains is the product condition. A chain with transition matrix P and invariant measure π is reversible if it satisfies the detailed balance equation

$$\pi(x)P(x, y) = \pi(y)P(y, x)$$

for all states x, y . Let λ_* be the eigenvalue of P with maximal absolute value but which is distinct from the trivial eigenvalue 1. Define the *absolute spectral gap* as $\gamma := 1 - \lambda_*$

and the relaxation time $t_{\text{rel}} := \gamma^{-1}$. The product condition is the requirement that for all $\varepsilon \in (0, 1)$,

$$\frac{t_{\text{mix}}(\varepsilon)}{t_{\text{rel}}} \xrightarrow{n \rightarrow \infty} \infty.$$

The product condition was at first conjectured to be an equivalence, but counter examples exist, e.g. [117][Example 18.7]. Another characterization of cutoff was established by Basu, Hermon and Peres in [14] for reversible chains in terms of hitting time, unfortunately this requires the sharp computation of the mixing time.

The understanding of the phenomenon improved drastically when attention shifted to the case of generic chains. This shift was initiated by the seminal work of Lubetzky and Sly [122] who established cutoff for random walks on random regular graphs. As such chains do not have fixed transition probabilities, the proof strategy cannot rely on an explicit knowledge of the transition matrix, as was the case for the previously cited works. Instead, the result is based on an *entropic concentration* phenomenon. A heuristic is as follows: ideally the goal is to prove the existence of a constant $h > 0$ such that $-\log P^t(X_0, X_t)$ concentrates around th , for all t up to the mixing time. If true, this statement implies that X_t is essentially uniformly distributed on a set of size e^{th} . Provided the invariant measure of X_t is sufficiently spread-out on the state space, as is the case for instance for the uniform measure, this immediately implies that $\log n/h$ steps are necessary, but also sufficient for mixing. Since the work of Lubetzky and Sly, a whole series of papers have used the entropic method to prove cutoff for generic chains at entropic time $\log n/h$. From these works, which we review in Chapter 5, a recent breakthrough was made Salez by [149] who gave a sufficient "product-like" condition for cutoff expressed in terms of entropic concentration which in theory can be checked without the sharp knowledge of the mixing time. From this, Salez deduces the cutoff for a large family of chains satisfying a non-negative curvature assumption.

The last of our contributions presented in this thesis is inspired by the paper of Hermon, Sly and Sousi [92] who considered the simple random walk on graphs to which is added a uniform random matching of the vertices. Under mild assumptions on the graph sequence, the simple random walk exhibits cutoff at entropic time. Our work in the area of cutoff is an attempt to generalize this model to more general graphs and in fact more general chains. Starting from two deterministic stochastic matrices P_1, P_2 on $[n]$, we investigate the mixing properties of a randomized mixture of P_1 and P_2 : letting S be the matrix of a uniform permutation σ and $p \in (0, 1)$, we analyze the mixing of the chain with transition matrix

$$\mathcal{P} = pP_1 + (1 - p)SP_2S^{-1}.$$

In fact, we study a slightly more general model where we allow the probability p

to depend on the current state x of the chain and its image $\sigma(x)$. This relaxation is necessary for the model to include the case of the superposition of graphs. In particular, taking for P_1, P_2 the matrices of the simple random walk on a graph and on a family of edges respectively, we can recover the model of Hermon, Sly and Sousi. We prove a cutoff phenomenon at entropic time in the following sense: we establish that in our model, there exists a measure π such that for all $\varepsilon \in (0, 1)$, $t = \log n/h + a_n\sqrt{\log n}$, for any starting point $x \in [n]$, the chain is with high probability at distance at least $1 - \varepsilon$ in total variation to the measure π if $a_n \rightarrow -\infty$, and is otherwise ε -close to π if $a_n \rightarrow +\infty$ and $t = O(\log n)$. We conjecture that the limiting measure π is itself ε -close to a genuine invariant measure, which would allow to deduce proper mixing time estimates. Unfortunately, at the time this thesis is written, this conjecture is still being actively pursued. The main issue arises from the fact that the mixing property holds only for a typical starting point: the initial state of the chain needs to be fixed before generating the environment or can be chosen uniformly at random, but cannot be chosen arbitrarily after the environment has been generated.

Chapter 2

Linearization of random walks on groups

This chapter is adapted from the paper [30], done jointly with Charles Bordenave.

2.1 Introduction

Let \mathcal{A} be a complex unital algebra and consider a non-commutative polynomial $\mathcal{P} = \mathcal{P}(x_1, \dots, x_n)$ in the variables $x_1, \dots, x_n \in \mathcal{A}$. In many cases, a detailed study of the relevant properties of \mathcal{P} is possible only when the degree of \mathcal{P} is small, typically when \mathcal{P} is of degree one, in which case \mathcal{P} is just a linear combination of the x_i 's. The linearization trick precisely consists in constructing another polynomial $\tilde{\mathcal{P}}$ of degree one, which can be related to the relevant properties of \mathcal{P} . It thus can be used to make computations that were not a priori possible for \mathcal{P} . The price to pay is an enlargement of the algebra: writing \mathcal{P} as a finite sum of monomials

$$\mathcal{P} = \sum_{i_1, \dots, i_k} \alpha_{i_1 \dots i_k} x_{i_1} \cdots x_{i_k},$$

with $\alpha_{i_1 \dots i_k} \in \mathbb{C}$, $\tilde{\mathcal{P}}$ is generally constructed as

$$\tilde{\mathcal{P}} = \sum_i \tilde{\alpha}_i \otimes x_i$$

where $\tilde{\alpha}_i \in M_r(\mathbb{C})$ are complex matrices. Therefore $\tilde{\mathcal{P}}$ is no longer in the algebra \mathcal{A} but in the algebra $M_r(\mathbb{C}) \otimes \mathcal{A}$ for some integer r that depends on \mathcal{P} (usually through its degree).

Under various names, the linearization trick has been used in several domains such as electrical engineering, random matrices, operators algebras, automata theory, etc, we refer to the survey [89]. A famous application of the linearization trick in operator

algebra, due to Haagerup and Thorbjørnsen [83], is the computation of the spectrum of a non-commutative polynomial. If $a \in \mathcal{A}$, the spectrum of a is defined as $\sigma(a) := \{z \in \mathbb{C}, zI - a \text{ is not invertible}\}$. Given a non-commutative polynomial $\mathcal{P}(x_1, \dots, x_n) \in \mathcal{A}$, it is possible to construct a linear $\tilde{\mathcal{P}} \in M_r(\mathbb{C}) \otimes \mathcal{A}$ for some integer r , such that the resolvent operator $(zI - \mathcal{P})$ is invertible in \mathcal{A} if and only if $(\Lambda(z) \otimes I_{\mathcal{A}} - \tilde{\mathcal{P}})$ is invertible in $M_r(\mathbb{C}) \otimes \mathcal{A}$, where $\Lambda(z)$ is the $r \times r$ matrix with only zero elements except the $(1, 1)$ entry which is z . Moreover, $(zI - \mathcal{P})^{-1}$ is then precisely the $(1, 1)$ entry of $(\Lambda(z) \otimes I_{\mathcal{A}} - \tilde{\mathcal{P}})^{-1}$ (seen as a $r \times r$ matrix of elements in \mathcal{A}). Such a construction can be found for instance in the monograph [135, Chapter 10].

As illustrated in this last example, we note that the relevant properties of \mathcal{P} (here its spectrum and its resolvent) dictates the linearization procedure. In this chapter, we introduce a new linearization of the Markov kernel of a random walk with finite range on a group. This linearization is a new Markov kernel on an enlarged state space and it corresponds to a nearest-neighbour random walk on the group. Many properties of the original random walk can be read on the new random walk. We illustrate our method on free groups and free product of finite groups. Notably, we give a new description of the harmonic measure for finite range random walks on these groups and obtain formulas for the drift and entropy as a byproduct.

2.1.1 The linearization trick and coloured random walks on groups

Let G be a finitely generated group with identity element e and we fix a finite set of generators S . We assume that S is symmetric, meaning that $g \in S$ implies $g^{-1} \in S$. Let $p = (p_g)_{g \in G}$ be a probability measure on G . Consider the Markov chain $(X_n)_{n \geq 0}$ on G with transition probabilities

$$\mathbb{P}[X_n = h \mid X_{n-1} = g] = p_{g^{-1}h},$$

for all $g, h \in G$. Such a Markov chain is called a right convolution random walk on G . Left convolution random walk can be considered as well but are essentially the same object, so the choice of right convolution random walks is merely a convention of ours.

The random walk is said to be finitely supported, or to have finite range, if the measure p is finitely supported. It is a nearest-neighbour random walk if p is supported on the set of generators S .

As for any Markov chain, the transition kernel of a right convolutional random walk can be seen as an operator acting on square integrable functions: $f = f(x) \mapsto (\mathcal{P}f)(x) = \sum_y P(x, y)f(y)$. In the present case, it can be written in terms of the right multiplication operators $\rho(g)$, $g \in G$, defined as: for all $f \in \ell^2(G)$,

$$\rho(g) \cdot f : x \mapsto f(xg).$$

Letting $\mathcal{P}(g, h) = p_{g^{-1}h}$ it is then possible to write

$$\mathcal{P} = \sum_{g \in G} p_g \rho(g). \quad (2.1)$$

This sum is finite if and only if the random walk is finitely supported. Furthermore, ρ is a group morphism so $\rho(gh) = \rho(g)\rho(h)$ for all $g, h \in G$ (ρ is the right regular representation). Writing each element g in the support of the walk as a product of elements of S , it is thus possible to write \mathcal{P} as a non-commutative polynomial in the operators $\rho(g), g \in S$. In other words, \mathcal{P} is an element of the left group algebra which is generated by the $\rho(g)$'s, $g \in G$. The polynomial \mathcal{P} is of degree 1 if and only if the random walk is nearest neighbour. With this point of view, the linearization trick potentially allows to study any finitely supported random walk by considering instead a “nearest-neighbour operator”

$$\tilde{\mathcal{P}} = \sum_{g \in G} \tilde{p}_g \otimes \rho(g) \quad (2.2)$$

for some well-chosen matrices $\tilde{p}_g \in M_r(\mathbb{C})$ such that $\sum_g \tilde{p}_g$ is a stochastic matrix, that is, the transition kernel of a Markov chain on $[r]$ where for $n \geq 1$ integer, we set $[n] = \{1, \dots, n\}$.

Then $\tilde{\mathcal{P}}$ is the transition kernel of a Markov chain $(\tilde{X}_n)_n$ on the state space $G \times [r]$, which at each step moves from a pair (g, u) to a pair (h, v) with probability $\tilde{p}_{g^{-1}h}(u, v)$. Equivalently, if one interprets $[r]$ as a set of colours, \tilde{X}_n can at each step be multiplied by $g \in G$ while the colour is updated from u to v with probability $p_g(u, v)$. Such a Markov chain will be called a coloured random walk and can thus be defined similarly to a classical convolution random walk. We note that such random walks already appeared under different names: random walks with internal degrees of freedom [107, 104, 105], covering Markov operators [103], matrix-valued random walks [51]. We note also that a similar idea of using coloured random walks to study finite range random walks on groups has already been considered by Jean Mairesse (personal communication).

Definition 2.1. Let $p = (p_g)_{g \in G}$ be a family of matrices with non-negative coefficients of $M_r(\mathbb{R})$ such that

$$P = \sum_{g \in G} p_g \quad (2.3)$$

is a stochastic matrix. A coloured random walk is a Markov chain $(Y_n)_n$ on the state space $G \times [r]$, with transition probabilities

$$\mathbb{P}[Y_n = (h, v) \mid Y_{n-1} = (g, u)] = p_{g^{-1}h}(u, v).$$

The support of the random walk is the set of elements $g \in G$ such that $p_g \neq 0$. The coloured walk is finitely supported if the support is finite, nearest-neighbour if the support is included in S .

By looking only at the colour coordinate, (Y_n) induces a Markov chain on the set of colours $[r]$ whose transition matrix is exactly the matrix P in (2.3).

Irreducibility assumptions. For standard convolution random walks, it is generally assumed that the support generates the whole group, which then makes the random walk irreducible. For coloured random walks, we will make the same assumption but suppose as well that the matrix P defines an irreducible Markov chain on the colours. Note this assumption does not necessarily make the coloured walk irreducible, for there may be unreachable pairs (g, u) .

Definition 2.2. A coloured walk is quasi-irreducible if the marginal of its support on G generates the whole group and the matrix P is irreducible.

Reversibility. An important property satisfied by some Markov chains is reversibility. Recall that, if X is a countable set, ν a measure on X and Q is a Markov chain, then Q is said to be reversible with respect to ν if for all $x, y \in X$, $\nu(x)Q(x, y) = \nu(y)Q(y, x)$. In other words, Q seen as an operator on $\ell^2(X, \nu)$ is self-adjoint. In our setting, \mathcal{P} defined in (2.1) is said to be reversible if it is reversible for the counting measure on G . This is equivalent to the condition: for all $g \in G$, $p_g = p_{g^{-1}}$. Similarly, consider a coloured random walk $\tilde{\mathcal{P}}$ of the form (2.2). Assume that P defined by (2.3) has invariant measure π on $[r]$. Then $\tilde{\mathcal{P}}$ is reversible if it is reversible for the product of π and the counting measure on G . This is equivalent to the condition, for all $g \in G$, $u, v \in [r]$,

$$\pi(u)p_g(u, v) = \pi(v)p_{g^{-1}}(v, u).$$

2.1.2 Main results

Linearizing random walks. The first main contribution of our work is to formalize a way to apply the linearization trick to random walks on groups.

Definition 2.3. Let G be a group generated by a finite set S , with identity element e . Let $(X_n)_{n \geq 0}$ be a random walk with kernel \mathcal{P} as in (2.1) and with finite support generating G . Let $(Y_n)_{n \geq 0}$ on $G \times [r]$ be a quasi-irreducible, nearest-neighbour coloured random walk with kernel $\tilde{\mathcal{P}}$ as in (2.2). We say that $(Y_n)_{n \geq 0}$ linearizes $(X_n)_{n \geq 0}$ (or $\tilde{\mathcal{P}}$ linearizes \mathcal{P}) if the following two property holds: (i) if $Y_0 = (e, 1)$ there exists a sequence of stopping times $(\tau_n)_{n \geq 0}$ with $\tau_0 = 0$ such that $(Y_{\tau_n})_{n \geq 0}$ is a realization of the random walk $(X_n)_{n \geq 0}$ with initial condition e , and (ii) these stopping times are a renewal process, that is the variables $\tau_{n+1} - \tau_n$ are iid with finite mean.

To be precise, in the above definition, when saying that Y_{τ_n} is a convolution random walk, we of course identify Y_{τ_n} with its G coordinate and forget about the colour, which in this case is constant equal to 1.

We remark also that due to the invariance of transition probabilities by translation by elements of G one can always translate the starting point to $e \in G$, so there is no loss of generality in supposing that the walks are started at $e \in G$.

Theorem 2.1. *Let G be a group generated by a finite set S , with identity element e . Consider a random walk with kernel \mathcal{P} as in (2.1) with finite support generating G . Then there exists $r \geq 1$ and a coloured random walk on $G \times [r]$ with kernel $\tilde{\mathcal{P}}$ as in (2.2) which linearizes \mathcal{P} . Moreover, if \mathcal{P} is reversible, then $\tilde{\mathcal{P}}$ can be chosen to be also reversible.*

Theorem 2.1 has been stated in a non-constructive manner for ease of notation. The proof of Theorem 2.1 in Section 2.2 will exhibit two simple linearization constructions which have a vector $p = (p_g)_{g \in G}$ as input and gives as output the integer r and the family of matrices $(\tilde{p}_g)_{g \in S}$. There is one construction in the general case and one construction which preserves reversibility. We refer to Remark 2.4 for the number of colours r needed in both constructions. We note also that the previous spectral linearization tricks reviewed in [89] did not preserve the Markov property and could not be used to prove Theorem 2.1.

There are possible extensions of the Markovian linearization trick. It is possible with the same proof techniques to linearize a coloured random walk on $G \times [r]$ with finite range and obtain a coloured nearest-neighbour coloured random walk on $G \times [r']$ with $r' \geq r$. In another direction, it is possible to linearize random walks on G which do not have a finite range provided that we allow a countable set of colours, see Remark 2.2 below. Finally, in this work we focus on groups only but our first linearization construction applies to random walks on monoids as well.

Explicit description of the harmonic measure The whole point of the linearization trick is that the complexity added by passing to matrices is expected to be much smaller than the benefit of having nearest-neighbour random walks. The second part of this chapter provides such an example where results extend naturally to the matrix setting, thus illustrating the use of the linearization trick. Namely, we extend the explicit description of the harmonic measure for nearest-neighbour convolution walks on groups with an underlying tree structure, such as free groups [102, 114] and free products of finite groups [129, 130, 78, 77]. Following [129], these two cases can be considered at once.

Write \mathbb{F}_d for the free group on d generators a_1, \dots, a_d and let G_1, \dots, G_m be finite groups. Consider the group $G = \mathbb{F}_d * G_1 * \dots * G_m$, with the set of generators

$$S := \bigcup_{i=1}^d \{a_i, a_i^{-1}\} \cup \left(\bigsqcup_{j=1}^m S_j \right). \quad (2.4)$$

In the above expression, for all $j = 1, \dots, m$, $S_j := G_j \setminus \{e\}$ is the set of elements of G_j distinct from the identity element, and \bigsqcup denotes a disjoint union. Such groups were considered in [129, 84] under the name *plain groups*.

Introducing the notation from [129],

$$\forall g \in S, \quad \text{Next}(g) := \begin{cases} S \setminus \{g^{-1}\} & \text{if } g \in \mathbb{F}_d \\ S \setminus S_i & \text{if } g \in S_i \end{cases}, \quad (2.5)$$

we see that every element $g \in G$ writes uniquely as a word $g = g_1 \dots g_n$ with $n = |g|$ and $g_{i+1} \in \text{Next}(g_i)$ for all $i = 1, \dots, n-1$.

Consider an irreducible convolution random walk $(X_n)_{n \geq 0}$ on G . If we set aside the trivial cases where G is isomorphic to \mathbb{Z} or $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$, G is a non-amenable group which implies that $(X_n)_{n \geq 0}$ is transient (see for instance [159, Chapter 1]). More precisely, arguments going back to Furstenberg [76] show that X_n converges a.s. to an infinite word $X_\infty = (\xi_i)_{i \geq 1}$ with the property that for all $i \geq 1$, $\xi_i \in S$ and $\xi_{i+1} \in \text{Next}(\xi_i)$. The law p^∞ of X_∞ is called the harmonic measure and it provides much information on the asymptotic properties of X_n .

In the case $(X_n)_{n \geq 0}$ is nearest-neighbour, the harmonic measure can be described very explicitly. In particular it is Markovian: writing $X_\infty^{(k)}$ for the truncature of X_∞ to its k first letters, the sequence $(X_\infty^{(k)})_{k \geq 1}$ is a Markov chain. The transition probabilities are as follows: given $g \in S$, let $\mu_g := p^\infty[\xi_1 = g] = \mathbb{P}[X_\infty^{(1)} = g]$. Then for all $g_1, \dots, g_k \in S$, satisfying $g_{i+1} \in \text{Next}(g_i)$ for all i ,

$$p^\infty[(\xi_i)_{i=1}^k = (g_i)_{i=1}^k] = \mathbb{P}[X_\infty^{(k)} = g_1 \dots g_k] = \mu_{g_1} \frac{\mu_{g_2}}{\sum_{h_2 \in \text{Next}(g_1)} \mu_{h_2}} \dots \frac{\mu_{g_k}}{\sum_{h_k \in \text{Next}(g_{k-1})} \mu_{h_k}}. \quad (2.6)$$

This equation determines the harmonic measure on cylinders and thus determines it completely.

On the other hand as soon as one supposes the random walk is only finitely supported, then this description completely falls apart and much less is known about the harmonic measure. A second contribution of our work is to show how these results can actually be extended even for general finitely supported random walks. Extending well-known techniques for nearest-neighbour random walks to the coloured setting, the linearization trick then allows to transfer results back to the case of finitely supported convolution walks.

Extensions of harmonic measures and of more general concepts (namely, the Poisson boundary) to the case of coloured random walks have already been considered in [103]. Here it takes the form of a measure p_u^∞ on infinite words in the group, $u \in [r]$ being the starting colour. Notice that it does not take colours other than through the starting colour. Yet colours need to be incorporated to recover the Markovian decomposition, which is achieved by considering a matrix version of the parameters μ_g , satisfying

$$p_u^\infty(\xi_1 = g) = \sum_{v \in [r]} \mu_g(u, v), \quad \forall u \in [r].$$

Supposing now that X_n is nearest-neighbour, the family of parameters $(\mu_g)_{g \in S}$ is uniquely characterized by a family of matrix relations (2.7), generalizing the so-called traffic equations of Mairesse [129], which arrive as a consequence of stationarity:

$$x_g = p_g \Delta(x)_g + \sum_{\substack{h, h' \in S \\ hh' = g}} p_h x_{h'} + \sum_{h \in \text{Next}(g)} p_{h^{-1}} x_h \Delta(x)_h^{-1} x_g, \quad \forall g \in S, \quad (2.7)$$

where $\Delta(x)_g$ is the diagonal matrix with entries, for $u \in [r]$,

$$\Delta(x)_g(u, u) := \sum_{h \in \text{Next}(g)} \sum_{v \in [r]} x_h(u, v). \quad (2.8)$$

In the sequel, $\Delta(\mu)_g$ will be simply written Δ_g .

These matrices are also used to extend Equation (2.6) to the coloured setting. The global description of the harmonic measure can be summed up in the following theorem. We write $\mathbb{1}$ for the vector of \mathbb{R}^r with all coordinates equal to 1 and $\mathbb{1}_u$ for the indicator at $u \in [r]$, that is the vector with zero coordinates except the u entry equal to 1.

Theorem 2.2. *For all starting colour $u \in [r]$ and all cylinder $\xi_1 \cdots \xi_n$,*

$$\mathbb{P}_u \left[X_\infty^{(n)} = \xi_1 \cdots \xi_n, u_n = v \right] = \mathbb{1}_u^\top \mu_{\xi_1} \Delta_{\xi_1}^{-1} \mu_{\xi_2} \cdots \Delta_{\xi_{n-1}}^{-1} \mu_{\xi_n} \mathbb{1}_v, \quad (2.9)$$

and

$$p_u^\infty(\xi_1 \cdots \xi_n) = \mathbb{1}_u^\top \mu_{\xi_1} \Delta_{\xi_1}^{-1} \mu_{\xi_2} \cdots \Delta_{\xi_{n-1}}^{-1} \mu_{\xi_n} \mathbb{1}. \quad (2.10)$$

Moreover, the family $\mu = (\mu_g)_{g \in S}$ is the unique family of matrices with non-negative entries which sum to a stochastic matrix and which are solutions of Equations (2.7).

One motivation for this work was the derivation for finitely supported walks of explicit formulas for various invariant quantities of interest, mainly entropy and drift. By 'explicit', we mean expressing drift and entropy in terms of the unique solutions of some finite dimensional fixed point equations. These can be obtained easily from the harmonic measure.

Suppose $\mathbb{E}[|X_1|] < \infty$, which is in particular true if the walk is finitely supported. The length with respect to the generating set S of $x = (g, u) \in G \times [r]$ is defined as

$$|x| := |g| = \min\{k, g = a_1 \cdots a_k, a_i \in S\}. \quad (2.11)$$

On the other hand, recall the definition of the entropy:

$$H(X_n) := - \sum_g \mathbb{P}[X_n = g] \log \mathbb{P}[X_n = g].$$

By sub-additivity arguments, eg. Kingman's theorem, one can prove the existence of the following limits:

$$\gamma = \lim_{n \rightarrow \infty} \frac{|X_n|}{n} \quad \text{a.s. and in } L^1, \quad (2.12)$$

$$h = \lim_{n \rightarrow \infty} - \frac{\log \mathcal{P}^n((e, u), X_n)}{n} \quad \text{a.s. and in } L^1. \quad (2.13)$$

The quantity h , introduced by Avez in [12] and called Avez entropy, or entropy rate, or asymptotic entropy, of the random walk, can in fact be interpreted as the drift for the Green pseudo-metric, see [24]. In this chapter it will often simply be referred to as the entropy of the random walk. We refer the reader to Kaimanovich [102] for a general reference on the topic in the case of convolution random walks, whereas extended notions of entropy and drift to the setting $G \times [r]$ were considered in [104, 105].

In the specific case of a linearizing random walk, the initial random walk can be recovered along a sequence of stopping times. Hence the law of large numbers immediately yields the following Abramov formulas. Such formulas are well known, see [85, 72].

Corollary 2.1. *Let $(X_n)_{n \geq 0}$ be a finitely supported random walk on G and $(Y_n)_{n \geq 0}$ a coloured random walk that linearizes $(X_n)_{n \geq 0}$ in the sense of Definition 2.3. The drift $\tilde{\gamma}$ and entropy \tilde{h} of $(Y_n)_{n \geq 0}$ can be related to the drift γ and entropy h of $(X_n)_{n \geq 0}$ by: $\gamma = \mathbb{E}[\tau_1] \tilde{\gamma}$ and $h = \mathbb{E}[\tau_1] \tilde{h}$.*

We notice also that the expected time $\mathbb{E}[\tau_1]$ has a simple expression in the two linearization constructions given in Section 2.2, see Remark 2.1 and Remark 2.3.

On the other hand, the following results provide formulas for the drift and entropy coloured random walks. They are direct extensions of known results in the colourless case (corresponding to $r = 1$) and the proof will be omitted. The usual scalar product is denoted by $\langle \cdot, \cdot \rangle$.

Theorem 2.3. *Let $(Y_n)_{n \geq 0}$ be a nearest-neighbour quasi-irreducible coloured random walk on $G \times [r]$ defined by a family of matrices $(p_g)_{g \in S}$. Let π be the unique invariant*

probability measure of the stochastic matrix $P = \sum_g p_g$. The drift and entropy of $(Y_n)_{n \geq 0}$ are respectively given by

$$\begin{aligned} \gamma &= \sum_{g \in S} \left\langle \mathbb{1}, \pi p_g \left(-\mu_{g^{-1}} + \sum_{h \in \text{Next}(g)} \mu_h \right) \right\rangle. \\ &= \sum_{g \in S} \sum_{u, v, w \in [r]} \pi(u) p_g(u, v) \left(-\mu_{g^{-1}}(v, w) + \sum_{h \in \text{Next}(g)} \mu_h(v, w) \right). \end{aligned}$$

and

$$h = - \sum_{\substack{g \in G \\ u, v \in [r]}} \pi(u) p_g(u, v) \int \log \left(\frac{dg^{-1} p_u^\infty}{dp_v^\infty} \right) dp_v^\infty. \quad (2.14)$$

For nearest-neighbour random walks, the previous integral formula can yield explicit expression for the entropy. However the non-commutativity of $M_r(\mathbb{C})$ creates new difficulties. For coloured random walks, adapting the computation naturally leads to the problem of determining the limit of infinite products of random matrices. This in turn leads to a formula of the entropy in terms of the parameters $(\mu_g)_{g \in S}$ and the law of a family indexed by $[r]$ of random probability measures on \mathbb{R}^r whose law is uniquely characterized by some invariance equation.

To state our result more precisely, we first introduce the hitting probabilities defined as follows. Given $g \in G, u \in [r]$, let $\tau_{(g, u)} := \inf\{n : X_n = (g, u)\}$ be the hitting time of a pair (g, u) , so that $\tau_g := \min_u \tau_{(g, u)}$ is the hitting time of g . For $g \in S$, set

$$q_g(u, v) := \mathbb{P}[\tau_g < \infty \text{ and } \tau_g = \tau_{(g, v)}]. \quad (2.15)$$

As we will check, the matrices q_g and μ_g satisfy a simple relation (2.24) and the hitting probabilities $(q_g)_{g \in S}$ are characterized by a quadratic equation (2.28) which can be solved quite explicitly for the free group. In the statement below, \mathcal{P}^+ denotes the open simplex of probability measures on $[r]$:

$$\mathcal{P}^+ := \left\{ x \in \mathbb{R}^r : \sum_{i \in [r]} x_i = 1 \text{ and } x_i > 0, \forall i \in [r] \right\}. \quad (2.16)$$

Theorem 2.4. *Suppose the matrices $(q_g)_{g \in S}$ satisfy (2.37) defined in Section 2.4. Then the entropy of the coloured random walk is given by*

$$\begin{aligned} h &= - \sum_{g \in S} \sum_{u, v, w \in [r]} \pi(u) p_g(u, v) \left[\mu_{g^{-1}}(v, w) \int \log \left(\frac{\langle \mathbb{1}_v, z \rangle}{\langle q_{g^{-1}}(u, \cdot), z \rangle} \right) d\nu_w(z) \right. \\ &\quad + \sum_{h \in \text{Next}(g)} \mu_h(v, w) \int \log \left(\frac{\langle q_g(v, \cdot), q_h z \rangle}{\langle \mathbb{1}_u, q_h z \rangle} \right) d\nu_w(z) \\ &\quad \left. + \sum_{h \in S: gh \in S} \mu_h(v, w) \int \log \left(\frac{\langle q_{gh}(v, \cdot), z \rangle}{\langle q_h(u, \cdot), z \rangle} \right) d\nu_w(z) \right], \end{aligned} \quad (2.17)$$

where $(\nu_u)_{u \in [r]}$ is the unique family of probability measures on \mathcal{P}^+ satisfying

$$\int f(z) d\nu_u(z) = \sum_{\substack{g \in S \\ v \in [r]}} \int f(q_g z) \mu_g(u, v) d\nu_v(z), \quad \forall u \in [r],$$

for all bounded measurable functions f on \mathcal{P}^+ .

The technical condition (2.37) is described in Section 2.4. Let us simply note that this condition is automatically satisfied for coloured random walks arising as a linearization of a finite range random walk on G (this is the content of Proposition 2.4 below). Applying Theorem 2.4 to a nearest-neighbour walk for which $r = 1, \pi = 1, \nu = \delta_1$, we get back formula (22) in [129]:

$$h = - \sum_{g \in S} p_g \left(\mu_{g^{-1}} \log \frac{1}{q_{g^{-1}}} + \sum_{h \in \text{Next}(g)} \mu_h \log q_g + \sum_{h \in S: gh \in S} \mu_h \log \frac{q_{gh}}{q_h} \right).$$

As alluded to above, in Theorem 2.4, the measure ν_g on \mathcal{P}_+ will arise as the convergence in direction toward a random rank one projector of the product of matrices $q_{g_1} q_{g_2} \cdots q_{g_n}$ where g_1, g_2, \dots in S are the successive letters of X_∞ with law p^∞ conditioned to start with $g_1 = g$.

For finite range random walks on free products of groups, fine qualitative results on drift and entropy were already available such as [109, 112] and even on general hyperbolic groups, see [113, 79, 132, 80] and references therein. As pointed in [114, 79], the computations for drift and entropy for nearest-neighbour random walk on the free group have been known for around 50 years but it was however unexpected that these explicit formulas could be extended to finite range random walks. To our knowledge, Theorem 2.3 and Theorem 2.4 provide the first formulas for finite range random walks on free products of group, see 2.1 for an illustration. It would be interesting to try to generalize this kind of formulas in other settings such as amalgamated free products of finite groups or virtually free groups.

In this work, we have chosen to focus through the harmonic measure on the drift and entropy but there are other numerical invariants associated to a random walk such as the spectral radius of \mathcal{P} or the index of exponential decay of the return probability: $\lim_{n \rightarrow \infty} \mathcal{P}^{2n}(e, e)^{1/(2n)}$ (they coincide for reversible walks). Our linearization technique could also be used to study these last two quantities but this is less novel since these quantities can be read on the resolvent operator and previous linearization techniques allow to compute resolvent of operators of the form (2.1), see [89] for such linearization and [115, 67] for examples of computation of resolvent operators of the form (2.2).

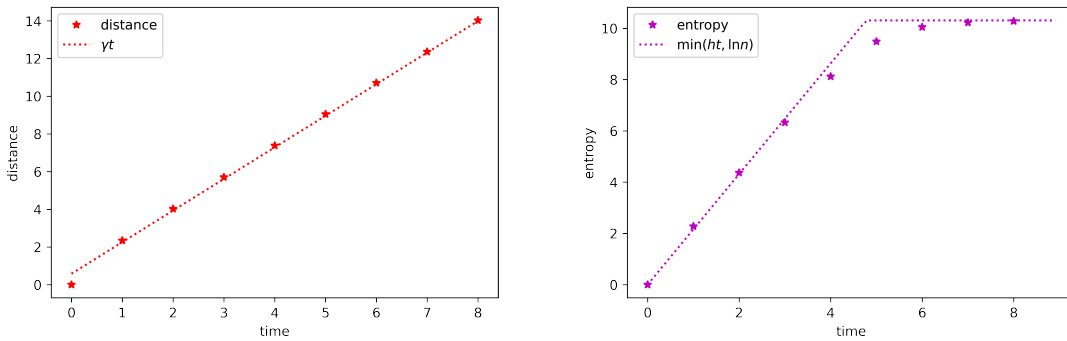


Figure 2.1: Numerical simulation of a finite range random walk $(X_t)_{t \geq 0}$ on $\mathbb{F}_2 = \langle a, b \rangle$ with 12 generators of length 1, 2 and 3, and with random transition probabilities. The linearization procedure gives $r = 16$. On the left, an average of $|X_t|$ over $n = 20000$ trajectories compared to the line γt . On the right, an average of $H(X_t)$ over n trajectories compared to the line ht . For computational efficiency, the entropy $H(X_t)$ is approximated with the entropy of a random walk on $[n]$ where $\lambda(g), g \in \{a, b\}$, are replaced by an independent uniform permutation matrix of size n (in other words, we consider a random action of \mathbb{F}_2 on a set of size n). This approximation is valid as long as $ht \leq \ln n$, see [31].

Organization of the chapter

In Section 2.2 we propose a general construction of linearized coloured walks, with some variations, that lead to Theorem 2.1. In Section 2.3 we define the harmonic measure and prove Theorem 2.2. Section 2.4 contains the arguments leading to Theorem 2.4.

2.2 A linearization trick: proof of Theorem 2.1

Let $(X_n)_{n \geq 0}$ be a finitely supported random walk on G . In this section we construct a nearest-neighbour coloured random walk $(Y_n)_{n \geq 0}$ from which $(X_n)_{n \geq 0}$ can be easily recovered. A natural strategy is to try to split transitions into nearest-neighbour steps, so that when projected on G , Y_n becomes a delayed version of X_n . The two constructions that follow are simple refinements of this idea. The first one improves on the naive approach in order to reduce the dimension of the matrices, whereas the second approach adapts the idea to preserve reversibility.

From now on, let a_1, \dots, a_d be the elements of the generating set S , and $a_0 := e$. Let K denote the support of the probability measure p and ℓ_K be the maximal length of an element of K . For each $g \in K$ choose arbitrarily a representative $s(g)$, which can always be assumed to be of length $|g|$. We will generally identify g with its representative and

write simply $g = a_{i_1} \cdots a_{i_k}$ to refer to the fact that $s(g) = a_{i_1} \cdots a_{i_k}$, for some $i_1, \dots, i_k \in [d]$. Similarly, set $p_{i_1 \dots i_k} := p_g$, with in particular $p_0 := p_e$ and $p_i := p_{a_i}$, whether $a_i \in K$ or not. Since we look for a nearest-neighbour coloured random walk, the coloured random walk $(Y_n)_{n \geq 0}$ is defined by the $r \times r$ matrices $\tilde{p}_i, i \in [0 : d]$.

General Idea Suppose X_n can jump directly from e to the element $g = a_{i_1} \cdots a_{i_k} \in K$. The basic idea consists in splitting this one transition into k successive steps: from e to a_{i_1} , then from a_{i_1} to a_{i_2} , etc. The question arising next is the following: once the walk is at a_{i_1} how can it tell whether this step was part of the move towards g or the first step of another transition? This problem is addressed by the introduction of colours. With this extra parameter, it becomes possible to distinguish between transitions that use the same elementary steps. A specific colour, written 1, will therefore serve as a reference colour: it is the colour telling the walk that it is not in the middle of transition. By opposition, all other colours are only transitional. In particular, the initial random walk is recovered as the trace of the coloured walk along colour 1.

The previous ideas are common to all the linearization procedures that we propose. The latter differ in the way colours are introduced and by the transition probability one assigns to elementary steps.

2.2.1 Linearization in the non-reversible case

The naive strategy sketched at the beginning of this section consists in imposing deterministic steps. Taking again the example of a transition $g = a_{i_1} \cdots a_{i_k} \in K$, we put a probability p_g on the transition from e to a_{i_1} and then probability 1 to all other transitions. This is possible if one allows sufficiently many colours, namely $|g| - 1$ colours, exclusive to every generator. Thus the coloured walk Y_n goes from $(e, 1)$ to $(a_{i_1}, 2)$ with probability p_g and then move deterministically from $(a_{i_1}, 2)$ to $(a_{i_2}, 3)$, etc. to eventually arrive at $(g, 1)$. The set of colours $\{2, \dots, |g|\}$ needs to be exclusive to g .

It is possible to improve on this idea, noticing that elements of K with common prefixes can share the same colours.

Recall that a representative of minimal length has been fixed for each element of K . These representatives are called the representatives of K . Given $g, h \in K, g = a_{i_1} \cdots a_{i_k}, h = a_{j_1} \cdots a_{j_{k'}}$, the prefix of g and h is defined as $g \wedge h = a_{i_1} \cdots a_{i_m}$ where $m := \max\{n \geq 0 : i_1 = j_1, \dots, i_n = j_n\}$. h is a strict prefix of g if $g \wedge h = h$ and $|g| > |h|$. For $k \in [\ell_K]$ and $i_1, \dots, i_k \in [d]$, let

$$[i_1 \cdots i_k] := \{g \in K, g \wedge a_{i_1} \cdots a_{i_k} = a_{i_1} \cdots a_{i_k}, |g| \geq k + 1\}$$

and

$$p_{[i_1 \cdots i_k]} := \sum_{g \in [i_1 \cdots i_k]} p_g$$

be the cumulative mass under p of all words in K for which $a_{i_1} \cdots a_{i_k}$ is a strict prefix.

Then for $k \in [\ell_K]$, we set

$$q_{i_1 \cdots i_k} = \frac{p_{[i_1 \cdots i_k]}}{p_{[i_1 \cdots i_{k-1}]}}$$

where the denominator is to be taken as equal to one in the case $k = 1$.

The essential idea is the following: we associate a colour $u_{i_1 \cdots i_k}$ to each strict prefix $[i_1 \cdots i_k]$. Suppose now the random walk starts at $(e, 1)$ and goes to a_{i_1} . With probability p_{i_1} it stays at colour 1 and with probability $q_{i_1} = p_{[i_1]}$ it takes colours u_{i_1} . Suppose this case occurs and the walk continues to $a_{i_1} a_{i_2}$. Then it can come back to colour 1 with probability $p_{i_1 i_2}/p_{[i_1]}$ or continue to colour $u_{i_1 i_2}$ with probability $q_{i_1 i_2}$. With our choice of transition probabilities, the overall probability to have moved from $(e, 1)$ to respectively $(a_{i_1} a_{i_2}, 1)$ and $(a_{i_1} a_{i_2}, u_{i_1 i_2})$ is thus respectively $p_{i_1 i_2}$ and $p_{[i_1 i_2]}$. Iterating this procedure yields the whole generalization, which is formalized in what follows.

Here we will make use of the operator point of view in order to write all the matrices defining $(Y_n)_{n \geq 0}$ at once. As $M_r(\mathbb{C}) \otimes \mathcal{A}$ is isomorphic to $M_r(\mathcal{A})$ for any unital algebra \mathcal{A} , the transition kernel \tilde{P} can be written as one matrix whose coefficients are operators on $\ell^2(G)$. In this case, the matrix has coefficients which are linear combinations of the multiplication operators. The matrices $p_i, i \in [0, d]$ can easily be deduced: to obtain p_i it suffices to replace $\rho(a_i)$ by 1 and the $\rho(a_j), j \neq i$ by 0.

For all $k \in [\ell_K]$, let

$$r_k := \text{Card} \{ (i_1, \dots, i_k) \in S^k, [i_1 \cdots i_k] \neq \emptyset \}$$

be the number of strict prefixes of elements of K that have length k . Then, we define the following matrices: $C(k)$ is the $r_{k-1} \times 1$ column matrix

$$C(k) := \left(\sum_{j \in [d]} \frac{p_{i_1 \cdots i_{k-1} j}}{p_{[i_1 \cdots i_{k-1}]}} \rho(a_j) \right)_{i_1 \cdots i_{k-1}},$$

which is indexed by strict prefixes of K of length $k-1$. Given a representative $i_1 \cdots i_{k-1}$, define the row matrix

$$L(i_1, \dots, i_{k-1}) := (q_{i_1 \cdots i_{k-1} j} \rho(a_j))_j$$

indexed by all j such that $i_1 \cdots i_{k-1} j$ is the strict prefix of a representative of K . Then use these row matrices to form the $r_{k-1} \times r_k$ diagonal block matrix $D(k)$, whose $i_1 \cdots i_{k-1}$ diagonal entry is $L(i_1, \dots, i_{k-1})$:

$$D(k) = \begin{pmatrix} \ddots & & & \\ & L(i_1, \dots, i_{k-1}) & & \\ & & \ddots & \\ & & & \ddots \end{pmatrix}.$$

Finally combine all the preceding matrices to construct:

$$\tilde{P} := \begin{pmatrix} \sum_{i \in [0:d]} p_i \rho(a_i) & D(1) & 0 & & 0 \\ C(2) & 0 & D(2) & & \\ C(3) & 0 & 0 & D(3) & \\ \vdots & & & \ddots & \\ C(\ell_K - 1) & & & & D(\ell_K - 1) \\ C(\ell_K) & & & & 0 \end{pmatrix} \quad (2.18)$$

Example 2.1. The construction will certainly be clearer on a concrete example. Consider a random walk on the group G given by the presentation $\langle a, b \mid ab = ba \rangle$, which is in fact isomorphic to \mathbb{Z}^2 . Suppose the random walk has support

$$K = \{e, a, b, a^2, a^2b\}.$$

To avoid large matrices, we forget about the inverses a^{-1}, b^{-1} but they should not be discarded in a general case.

Because of the relation $ab = ba$, there can be several ways to write an element of G as words in a and b , for instance $a^2b = aba$. Therefore we fix a representative for each element of the support. In the present case, group elements can be written uniquely as $a^k b^l$ so it is natural to choose these words as representatives.

Applying the preceding construction, one eventually obtains the following operator matrix:

$$\begin{pmatrix} p_e \rho(e) + p_a \rho(a) + p_b \rho(b) & q_a \rho(a) & 0 \\ \frac{p_{a^2}}{p_{[a]}} \rho(a) & 0 & q_{a^2} \rho(a) \\ \frac{p_{a^2b}}{p_{[a^2]}} \rho(b) & 0 & 0 \end{pmatrix} = \begin{pmatrix} p_e \rho(e) + p_a \rho(a) + p_b \rho(b) & (p_{a^2} + p_{a^2b}) \rho(a) & 0 \\ \frac{p_{a^2}}{p_{a^2} + p_{a^2b}} \rho(a) & 0 & \frac{p_{a^2b}}{p_{a^2} + p_{a^2b}} \rho(a) \\ \rho(b) & 0 & 0 \end{pmatrix}.$$

Proof of Theorem 2.1 (non-reversible case). On row $i_1 \dots i_k$, the sum of entries of the matrix $P = \sum \tilde{P}_i$ is

$$\sum_{j \in [d]} \frac{p_{i_1 \dots i_{k-1} j}}{P_{[i_1 \dots i_{k-1}]}} + \sum_{j \in [d]} q_{i_1 \dots i_{k-1} j} = \frac{1}{P_{[i_1 \dots i_{k-1}]}} \sum_{j \in [d]} (p_{i_1 \dots i_{k-1} j} + p_{[i_1 \dots i_{k-1}] j}) = 1.$$

Thus P is stochastic and \tilde{P} defines indeed a coloured random walk $(Y_n)_{n \geq 0}$.

Suppose now $Y_n = (g_n, u_n)$ is started at colour 1. Then define $(\tau_n)_{n \geq 0}$ as the successive return times at the first colour: $\tau_0 := 0$ and for $n \geq 1$

$$\tau_n := \inf\{m > \tau_{n-1}, u_m = 1\}$$

By Markov's property, the random variables $\tau_n - \tau_{n-1}$ are iid with the same law as τ_1 .

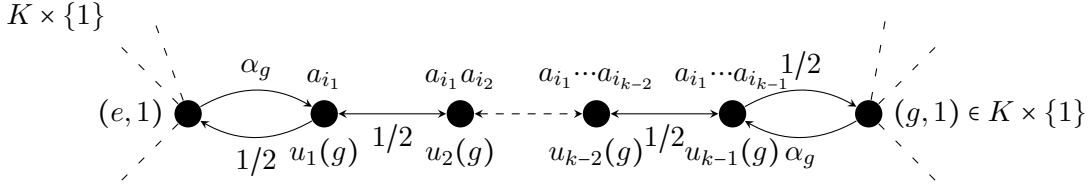


Figure 2.2: The linearizing reversible random walk

For all $g = i_1 \cdots i_k$ in K , the probability that $Y_{\tau_1} = g$ is

$$\begin{aligned} \mathbb{P}[Y_{\tau_1} = g] &= q_{i_1} q_{i_1 i_2} \cdots q_{i_1 \cdots i_{k-1}} \frac{p_{i_1 \cdots i_k}}{p_{[i_1 \cdots i_{k-1}]}} \\ &= p_{[i_1]} \frac{p_{[i_1 i_2]}}{p_{[i_1]}} \cdots \frac{p_{[i_1 \cdots i_{k-1}]}}{p_{[i_1 \cdots i_{k-2}]}} \frac{p_{i_1 \cdots i_k}}{p_{[i_1 \cdots i_{k-1}]}} \\ &= p_{i_1 \cdots i_k} = p_g. \end{aligned}$$

By Markov's property and the fact that the increments are iid one easily deduce that $Y_{\tau_n} \stackrel{(d)}{=} X_n$ for all $n \geq 0$. Finally, the irreducibility of X_n implies the quasi-irreducibility of Y_n . \square

Remark 2.1. The expectation of the hitting time τ_1 is very simple to compute: Y_1 can go through each $g \in K$ with probability p_g , in which case it needs $|g|$ steps to get back to the colour 1. Hence the expectation of τ_1 is just the average length of elements in K :

$$\mathbb{E}[\tau_1] = p_e + \sum_{g \in K} p_g |g| = p_e + \mathbb{E}[|X_1|].$$

Remark 2.2. If \mathcal{P} is not of finite range, then the construction produces a countable set of colours. The construction could nevertheless be useful if the expected time $\mathbb{E}[\tau_1] = p_e + \mathbb{E}[|X_1|]$ is finite.

2.2.2 Linearization in the reversible case

The previous constructions prevent the coloured random walk to be reversible. To correct this, we again assign new colours to every element of K with the following variation. We assume that for all $g \in K$, the representative of $g = a_{i_1} \cdots a_{i_n}$ is chosen such that the representative of g^{-1} is $a_{i_n}^{-1} \cdots a_{i_1}^{-1}$. We start with the neutral colour 1 and, for each pair (g, g^{-1}) with $|g| = |g^{-1}| \geq 2$, we add $|g| - 1$ new colours $u_1(g), \dots, u_{|g|-1}(g)$ and we set $u_k(g^{-1}) = u_{|g|-k}(g)$ for all $1 \leq k \leq |g| - 1$. Suppose g is written as $g = a_{i_1} \cdots a_{i_k}$. For all $h \in G$, the transition probability to go from $(h, 1)$ to $(ha_{i_1}, u_1(g))$ is set to some value α_g to be determined, such that $\alpha_g = \alpha_{g^{-1}}$ and $\sum_{g \in S} \alpha_g = 1$. All other transition probabilities on the segment joining $(x, 1)$ to $(xg, 1)$ are set to $1/2$, see Figure 2.2.

Thanks to the constraint $\sum_{g \in G} \alpha_g = 1$, $\tilde{\mathcal{P}}$ is indeed a Markov transition kernel. However, the coloured walk can now make steps that would not have been possible in the naive approach. For instance, it can go from $(e, 1)$ to some (g, u) and then come back to $(e, 1)$, even if $p_e = 0$ for the initial walk. Therefore, the stopping times cannot be taken directly as the return times at colour 1. Instead define, if $Y_n = (g_n, u_n)$ has transition kernel $\tilde{\mathcal{P}}$,

$$\tau_n := \begin{cases} \inf\{m \geq \tau_{n-1} : u_m = 1, g_m \neq g_{\tau_{n-1}}\} & \text{if } g_{\tau_{n-1}+1} \neq g_{\tau_{n-1}} \\ \tau_{n-1} + 1 & \text{if } g_{\tau_{n-1}+1} = g_{\tau_{n-1}}. \end{cases}$$

Theorem 2.1 in the reversible case is now contained in the following lemma.

Lemma 2.1. *Suppose X_n is a finitely supported random walk on G defined by a probability vector p and started at e . Setting for each $g \in K$*

$$\alpha_g := \begin{cases} \frac{(1-p_e)|g|p_g}{\sum_{h \in K} |h|p_h} & \text{if } g \neq e \\ p_e & \text{if } g = e \end{cases} \quad (2.19)$$

the preceding construction yields an operator $\tilde{\mathcal{P}}$ which defines a reversible coloured random walk $(Y_n)_{n \geq 0}$ satisfying $Y_{\tau_n} \stackrel{(d)}{=} X_n$ for all $n \geq 0$ if started at $(e, 1)$.

Proof. As already pointed, the construction defines a coloured random walk as long as $\sum_{g \in S} \alpha_g = 1$, which is true for α_g defined as in (2.19). Furthermore such a coloured walk is reversible: for all $(g, u) \in G \times [r]$, letting

$$\mu(g, u) = \begin{cases} 1 & \text{if } u = 1 \\ 2\alpha_h & \text{if } u = u_h(k) \text{ for some } k < |h|. \end{cases}$$

defines a reversible measure for the random walk, as it can be checked directly.

Consider now the coloured walk $(Y_n)_{n \geq 0}$ defined by (2.19), started at $Y_0 = (e, 1)$. Suppose first that $p_e = 0$, so that τ_1 is the hitting time of the set $(K \setminus \{e\}) \times \{1\}$.

At step 1 the coloured walk necessarily enters the segment joining $(e, 1)$ to $(h, 1)$ for some $h \in K$. If the walk escapes this segment at h , then $Y_{\tau_1} = h$. Otherwise, it necessarily goes back to the starting state $(e, 1)$. Now on each of these segments, the coloured walk is simply performing a simple random walk so the escape probabilities are given by the standard gambler's ruin problem. Namely the simple random walk on $[0 : n]$ reaches n before it gets to 0 with probability k/n when started at $k \in [0 : n]$. Therefore by Markov's property

$$\mathbb{P}_{(e,1)} [Y_{\tau_1} = g] = \alpha_g / |g| + \sum_{h \in K} \alpha_h (1 - 1/|h|) \mathbb{P}_{(e,1)} [Y_{\tau_1} = g].$$

We deduce that

$$\mathbb{P}_{(e,1)} [Y_{\tau_1} = g] = \frac{\alpha_g / |g|}{\sum_{h \in K} \alpha_h / |h|} = p_g.$$

For the general case $p_e \neq 0$, $Y_{\tau_1} = (e, 1)$ if and only if $Y_1 = (e, 1)$ which occurs with probability $\alpha_e = p_e$. For $g \neq 0$, consider the random walk conditioned to move at every step, written $(Y'_n)_{n \geq 0}$. If $q(x, y)$ is the transition probability for Y_n between two states x and y , then the transition probability for Y'_n is 0 if $y = x$ and $q(x, y)/(1 - q(x, x))$ otherwise. The previous argument apply to this walk, so

$$\mathbb{P}_{(e,1)} \left[Y'_{\tau'_1} = g \right] = \frac{\alpha_g / ((1 - p_e) |g|)}{\sum_{h \in K \setminus \{e\}} \alpha_h / ((1 - p_e) |h|)} = \frac{p_g}{1 - p_e},$$

with τ'_1 is the obvious extension of τ_1 to Y'_n . Coming back to Y_{τ_1} : to reach $g \neq e$ it is necessary that $Y_1 \neq e$, which occurs with probability $(1 - p_e)$. Conditional on that event it no longer matters whether the random walk comes back at e and possibly stays there, so one can reason with Y'_n instead. Hence

$$\mathbb{P}_{(e,1)} [Y_{\tau_1} = g] = (1 - p_e) \mathbb{P}_{(e,1)} \left[Y'_{\tau'_1} = g \right] = p_g.$$

The conclusion follows. \square

Remark 2.3. The expected time $\mathbb{E}[\tau_1]$, although not as simple as in the non-reversible case, can nonetheless be computed quite easily using for instance the electric network paradigm. Using [128][Prop 2.20], we found in the case $p_e = 0$

$$\mathbb{E}[\tau_1] = \sum_{g \in K} p_g |g|^2 = \mathbb{E}|X_1|^2. \quad (2.20)$$

Remark 2.4. In the naive approach, we have that the total number of colours is

$$r = 1 + \sum_{g \in K} (|g| - 1).$$

In the reversible construction, there is a factor $1/2$ on front of the sum (because we use the same colours for g and g^{-1}). In the improved construction of Subsection 2.2.1, this is an upper bound, the actual value is $r = 1 + \text{Card}(K') - \text{Card}(K)$ where K' is the set of words which are a suffix of some element in K (in the chosen representatives). In a concrete application, it is often possible to design linearization procedures which are more economical in terms of number of colours.

2.3 Coloured random walks on free products of groups

In this section we define and study the harmonic measure for coloured random walks on such groups, proving Theorem 2.2. Explicit descriptions of the harmonic measure on free groups were obtained in [102, 54, 114]. Similar approaches can be carried out for free products of finite groups, of monoids, of finite alphabets [78, 77, 130].

The main ingredient in the study of random walks on groups is the invariance by translation of the transition probabilities. This G -invariance of transition probabilities still holds for coloured random walks, allowing many arguments to extend to this setting. Indeed, once a good way of taking colours into account is found, one can essentially replace scalar parameters (such as p_g, μ_g , etc.) by matrices. The argumentation for convolution random walks can then be extended in a straightforward way as long as the computations are consistent with matrix products, which is the case most of the time. For this reason, we will often sketch or omit the proofs when they are too similar to already known arguments.

The case of free groups and free products of finite groups is very similar in nature and they can be handled together as done in [129]. Let G_1, \dots, G_m be finite groups and consider the group $G = \mathbb{F}_d * G_1 * \dots * G_m$, with the set of generators (2.4). Recall the definition of the map Next in (2.5). Every element $g \in G$ writes uniquely as a word $g = g_1 \dots g_n$ with $n = |g|$ and $g_i \in S, g_{i+1} \in \text{Next}(g_i)$ for all i . Such words will be called reduced.

2.3.1 The harmonic measure on the boundary

Consider a coloured random walk $(X_n)_{n \geq 0}$ on $G \times [r]$ defined by a family (p_g) of matrices. We assume that the walk is quasi-irreducible and nearest-neighbour. We assume furthermore that G is not isomorphic to \mathbb{Z} or $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$. This implies the coloured walk is transient, as shown by Proposition 2.1 below.

Define the boundary ∂G as

$$\partial G := \{\xi_0 \xi_1 \dots \xi_n \dots, \forall i \geq 0, \xi_i \in G, \xi_{i+1} \in \text{Next}(\xi_i)\}.$$

The multiplication action by G on itself can be extended to ∂G : given $g \in G, \xi = \xi_0 \xi_1 \dots \in \partial G$, define

$$g\xi := \begin{cases} g\xi_0 \xi_1 \dots & \text{if } \xi_0 \in \text{Next}(g) \\ (g\xi_0)\xi_1 \dots & \text{if } g\xi_0 \in S \\ \xi_1 \xi_2 \dots & \text{if } g = \xi_1^{-1}. \end{cases} \quad (2.21)$$

The boundary ∂G , which can be seen as a subset of $G^{\mathbb{N}}$, is equipped with the product topology and the corresponding σ -algebra. Given a measure ν on ∂G , let $g \cdot \nu$ be the image measure of ν under the multiplication by g , that is the measure defined by the fact that

$$\int f(\xi) d(g \cdot \nu)(\xi) := \int f(g\xi) d\nu(\xi),$$

for all bounded measurable function f on ∂G .

Definition 2.4. Consider a family $\nu = (\nu_u)_{u \in [r]}$ of probability measures on ∂G indexed by colours. ν is said to be stationary if for all $u \in [r]$,

$$\nu_u = \sum_{g \in G, v \in [r]} p_g(u, v) g \cdot \nu_v. \quad (2.22)$$

From now on, all measures considered on ∂G will be implicitly indexed by colours. The following result extends Theorem 1.12 in [114] to coloured random walks. The proof is exactly the same and will thus be omitted.

Proposition 2.1. *There exists a random variable $X_\infty \in \partial G$ such that X_n converges a.s. to X_∞ . The law of X_∞ is called the harmonic measure and is the unique stationary measure on the boundary ∂G . It will be denoted $(p_u^\infty)_{u \in [r]}$ where the index u is to be interpreted as the starting colour of $(X_n)_{n \geq 0}$.*

2.3.2 Markovian structure of the harmonic measure

A measure on ∂G is uniquely determined by its mass on cylinders. Given a measure ν , we will write $\nu(\xi_1 \cdots \xi_n)$ for the mass of the cylinder containing all infinite words which start with the prefix $\xi_1 \cdots \xi_n$.

In the colourless case, the tree structure of the group G implies the harmonic measure is Markovian. It can be computed entirely from the solutions of a set of equations derived from the stationarity of the harmonic measure with the following interpretation.

For all $k \geq 0$, let $X_\infty^{(k)}$ be the restriction of X_∞ to the first k -th letters. Thus for all $k \geq 1$ the mass under the harmonic measure of cylinders of size k is given by the law of $X_\infty^{(k)}$. On the other hand $(X_\infty^{(k)})_{k \geq 1}$ is a non-backtracking walk which one can interpret as the loop-erased random walk formed from $(X_n)_{n \geq 0}$. In the standard colourless setting, the tree structure of the group makes the loop-erased random walk a Markov chain whose transition probabilities can be computed easily.

In the coloured setting, one can expect to have similar properties but the loop-erased random walk of the process is no longer a Markov chain so this is not the right process to consider. Instead let $\tau_g := \inf\{n \geq 0, X_n = (g, \cdot)\}$ be the hitting time of g by the random walk X_n and u_k the colour at time $\tau_{X_\infty^{(k)}}$. Given $g \in S$, set

$$\mu_g(u, v) := \mathbb{P}_{(e, u)}[X_\infty^{(1)} = g, X_{\tau_g} = (g, v)]. \quad (2.23)$$

In words, $\mu_g(u, v)$ is the probability that the random walk, starting from (e, u) , visits g , with colour v for the first time, and later escapes at infinity in direction g . The process considered which is the equivalent of the loop erased random walk in the coloured setting is $(X_\infty^{(k)}, u_k)_{k \geq 1}$, which is a coloured Markov chain with increment distribution μ .

Proof of Theorem 2.2. Most of the proof is similar to the case of colourless walks, so we will only emphasize what is different. Details can be found in [114, 129, 130].

Recall Definition (2.8) and that $\Delta_g := \Delta(\mu)_g$. Given the definition of the matrices μ_g , Equation (2.9) is an easy consequence of Markov's property and the G -invariance of the transition probabilities of the random walk.

Let us now prove now that the matrices μ_g are characterized by Equation (2.7). Consider any family $(\nu_g)_{g \in S}$ of non-negative matrices, solutions of (2.7) and such that $\sum_{g \in S} \nu_g$ is a stochastic matrix. Then for all $u \in [r]$ define the measure ν_u^∞ on ∂G by setting for each cylinder $\xi_1 \cdots \xi_n$,

$$\nu_u^\infty = \mathbb{1}_u \nu_{\xi_1} \Delta(\nu)_{\xi_1}^{-1} \nu_{\xi_2} \cdots \Delta(\nu)_{\xi_{n-1}}^{-1} \nu_{\xi_n} \mathbb{1}$$

The fact that $\sum_{g \in S} \nu_g$ is a stochastic matrix proves that ν_u^∞ is indeed a probability measure on ∂G .

Mimicking the proof of the case of colourless walks, one can deduce that ν^∞ is stationary, which by uniqueness of stationary measures yields that $\nu_u^\infty = p_u^\infty$ for all $u \in [r]$.

At this stage we only proved that for all $g \in S, u \in [r]$, $p_u^\infty(g) = \sum_{v \in [r]} \mu_g(u, v) = \sum_{v \in [r]} \nu_g(u, v)$. To deduce $\mu_g = \nu_g$, we observe that the previous equality shows $\Delta_g = \Delta(\nu)_g$ for all $g \in S$. Consequently the matrices μ_g and ν_g can be seen as the fixed points of a quadratic map $f : M = (M_g)_{g \in S} \mapsto (f(M)_g)_{g \in S}$ from the set of non-negative matrices to itself, where

$$f(M)_g := A_g + \sum_{\substack{h, h' \in S \\ hh' = g}} p_h M_{h'} + \sum_{h \in \text{Next}(g)} p_{h^{-1}} M_h B_g M_g,$$

for all $g \in S$, and A_g, B_g are non-negative matrices that do not depend on M .

By the fact that f is quadratic, one can now use monotonicity arguments as in the proof of [129, Lem. 4.7] to deduce that $\mu_g = \nu_g$. \square

Hitting probabilities. Let us discuss here another way to compute the harmonic measure through hitting probabilities. It is an easy extension to the coloured case of equations that can be found for classical nearest-neighbour random walks in [114, 129, 130] but here the equations become matrix equations.

Recall Definition (2.15). The motivation for introducing these new matrices is that they can be easily related to the matrices μ_g by

$$\mu_g(u, v) = q_g(u, v) \Delta_g(v, v), \quad \forall g \in S. \quad (2.24)$$

This gives yet another way to write equations (2.9) and (2.10):

$$\mathbb{P}_u \left[X_\infty^{(k)} = \xi_1 \cdots \xi_k, u_k = v \right] = (q_{\xi_1} \cdots q_{\xi_{k-1}} \mu_{\xi_k})(u, v), \quad (2.25)$$

$$p_u^\infty(\xi_1 \cdots \xi_k) = \sum_v (q_{\xi_1} \cdots q_{\xi_{k-1}} \mu_{\xi_k})(u, v). \quad (2.26)$$

Furthermore, using (2.24), the traffic equation (2.23) can be rewritten as

$$\mu_g = p_g \Delta_g + \sum_{\substack{h, h' \in S \\ hh' = g}} p_h \mu_{h'} + \sum_{h \in \text{Next}(g)} p_{h^{-1}} q_h \mu_g, \quad \forall g \in S. \quad (2.27)$$

Observe that $y_g := \Delta_g \mathbb{1} = \mu_g \mathbb{1}$ and $y_g \in \mathbb{R}^r$ is equal to the diagonal of the diagonal matrix Δ_g . By construction, we have that $\sum_g y_g = \mathbb{1}$. If we evaluate the matrix equation (2.27) on the vector $\mathbb{1}$, we obtain a linear equation for $y = (y_g)_{g \in S}$, seen as a vector with coordinates in \mathbb{R}^r , of the form $Ty = 0$ where T is matrix on $S \times S$ with matrix-valued coefficients in $M_r(\mathbb{R})$. Finally, once, Δ_g is known, Equation (2.27) becomes linear in μ seen as a vector of matrices. Therefore the matrices μ_g can be completely recovered from the q_g .

On the other hand, the latter satisfy a simpler quadratic equation: conditioning on X_1 and applying Markov property, one obtains

$$x_g = p_g + \sum_{\substack{h, h' \in S \\ hh' = g}} p_h x_{h'} + \sum_{h \in \text{Next}(g)} p_{h^{-1}} x_h x_g. \quad (2.28)$$

Arguing as in the proof of Theorem 2.2, one can prove this equation characterizes the q_g as follows:

Lemma 2.2. *The family $q = (q_g)_{g \in S}$ is the unique solution to Equation (2.28) among family of matrices $(m_g)_{g \in S}$ such that $\mu_g(u, v) \leq m_g(u, v) \leq 1$ for all $u, v \in [r]$.*

2.3.3 Solving the traffic equations

We now propose methods to solve the equations of the previous sections. For a general free product of groups, we show how solutions can be computed numerically. When $G = \mathbb{F}_d * \mathbb{Z}/2\mathbb{Z} * \cdots * \mathbb{Z}/2\mathbb{Z}$, these equations can be solved explicitly in the case of classical nearest-neighbour random walks. For coloured random walks, complications arise due to matrix equations, so we can give explicit solutions up to the issue of determining a matrix square root. Thanks to the linearization trick, this problem is thus the only remaining obstacle preventing from obtaining an explicit description for finite range random walks.

General case. We have seen in the previous section that the harmonic measure is fully described by the family of matrices $\mu = (\mu_g)_{g \in S}$ on $[r]$ defined by (2.23). By Theorem 2.2, these matrices are uniquely characterized by the traffic equation (2.7) so it is possible to evaluate numerically μ by iterating the map defining the traffic equations.

The hitting probabilities $q = (q_g)_{g \in S}$ satisfy a simpler quadratic equation (2.28) than the traffic equation of μ . As explained in the proof of Lemma 2.2, it is possible to evaluate numerically from above and from below q by iterating the map f defined in (2.28) such that $q = f(q)$.

Case of $G = \mathbb{F}_d * \mathbb{Z}/2\mathbb{Z} * \cdots * \mathbb{Z}/2\mathbb{Z}$. In some special cases, one can find additional relations allowing to write the matrix μ_g as an explicit function of the matrices q_g . This case occurs for instance when $G = \mathbb{F}_d * \mathbb{Z}/2\mathbb{Z} * \cdots * \mathbb{Z}/2\mathbb{Z}$. Given $g \in S$, let d_g be the diagonal matrix with diagonal entries

$$d_g(u, u) := \sum_w q_g(u, w).$$

In words, $d_g(u, u)$ is the probability to ever reach g , starting from the pair (e, u) .

Proposition 2.2. *When $G = \mathbb{F}_d * \mathbb{Z}/2\mathbb{Z} * \cdots * \mathbb{Z}/2\mathbb{Z}$, for all $g \in S$ the matrix $(I - q_{g^{-1}}q_g)$ is invertible and*

$$\mu_g(u, v) = q_g(u, v) \sum_{w \in [r]} (I - q_{g^{-1}}q_g)^{-1}(I - d_{g^{-1}})(v, w) \quad \forall g \in S, u, v \in [r]. \quad (2.29)$$

Proof. Suppose $G = \mathbb{F}_d * \mathbb{Z}/2\mathbb{Z} * \cdots * \mathbb{Z}/2\mathbb{Z}$. Then for all $g \in S, u \in [r]$,

$$\mu_g(u, v) = q_g(u, v) \left(1 - \sum_{w \in [r]} q_{g^{-1}}(v, w) + \sum_{w, z \in [r]} q_{g^{-1}}(v, w) \mu_g(w, z) \right).$$

In particular $q_g(u, v) = 0$ implies $\mu_g(u, v) = 0$. Otherwise, rewrites this equation as

$$\frac{\mu_g(u, v)}{q_g(u, v)} = 1 - d_{g^{-1}}(v, v) + \sum_{w, z \in [r]} q_{g^{-1}}(v, w) q_g(w, z) \frac{\mu_g(w, z)}{q_g(w, z)}.$$

The right-hand being independent of u , so is the left-hand side. Consequently let $x_v := \mu_g(u, v)/q_g(u, v)$. Then one can again rewrite the above equation as the matrix linear equation

$$x = (I - d_{g^{-1}})\mathbb{1} + q_{g^{-1}}q_g x.$$

Provided $I - q_{g^{-1}}q_g$ is invertible, solving this equation yields the desired expression for $\mu_g(u, v)$.

Therefore we are left to prove that $I - q_{g^{-1}}q_g$ is invertible. This can be justified by the fact that $\sum_{v \in [r]} \sum_{n \geq 0} (q_{g^{-1}}q_g)^n(u, v)$ is the average number of times the walk goes

to g^{-1} and comes back to e when starting at colour u . By transience of the walk, this number must be finite. Hence the sum $\sum_{n \geq 0} (q_{g^{-1}} q_g)^n$ is convergent and the inverse of this matrix is precisely $I - q_{g^{-1}} q_g$. \square

We now further study the matrix equation (2.15) satisfied by the hitting probabilities $(q_g)_{g \in S}$. Supposing $G = \mathbb{F}_d * \mathbb{Z}/2\mathbb{Z} * \dots * \mathbb{Z}/2\mathbb{Z}$, for all $g \in S$, Equation (2.28) becomes:

$$q_g = p_g + \sum_{h \neq g^{-1}} p_{h^{-1}} q_h q_g. \quad (2.30)$$

In the colourless case, it is possible to reduce Equation (2.30) to a scalar equation, see for example [114]. We extend this computation to the coloured case. We define the matrix in $M_r(\mathbb{R})$,

$$z = I - \sum_{g \in S} p_{g^{-1}} q_g. \quad (2.31)$$

We now express q_g as a function of $(z, p_g, p_{g^{-1}})$ and find a closed equation satisfied by z . For simplicity, we assume that for all $g \in G$, the matrix p_g is invertible. If this is not the case, a similar argument holds but one should be careful with pseudo-inverses.

We define the matrices in $M_{2r}(\mathbb{R})$:

$$P_g = \begin{pmatrix} 0 & p_g \\ p_{g^{-1}} & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} z & 0 \\ 0 & z \end{pmatrix} \quad \text{and} \quad Q_g = \begin{pmatrix} 0 & q_g \\ q_{g^{-1}} & 0 \end{pmatrix}.$$

From our assumption, P_g is invertible and

$$P_g^{-1} = \begin{pmatrix} 0 & (p_{g^{-1}})^{-1} \\ p_g^{-1} & 0 \end{pmatrix}.$$

Applying (2.30) to g and g^{-1} , we find that, for all $g \in S$, $z q_g = p_g - p_g q_{g^{-1}} q_g$ and thus

$$Z Q_g = P_g - P_g Q_g^2.$$

We set $Z_g = P_g^{-1} Z$. The above equation rewrites

$$(Z_g + Q_g) Q_g = I. \quad (2.32)$$

In particular $Z_g = Q_g^{-1} - Q_g$ so Z_g and Q_g commute. We may thus solve the quadratic equation $Q_g^2 + Z_g Q_g - I = 0$ with Q_g as unknown as in the scalar case. Completing the square yields

$$(Q_g + Z_g/2)^2 = I + Z_g^2/4.$$

Therefore, for some proper choice of the matrix square root function, we get

$$Q_g = \frac{1}{2} \left(\sqrt[4]{4I + (P_g^{-1} Z)^2} - P_g^{-1} Z \right), \quad (2.33)$$

where $\sqrt{}$ is a notation to stress that the choice of the square root is unknown. First, as Z_g and Q_g commute, the eigenvalues completely determine the square root. Also, since Q_g is a block antidiagonal matrix, for every of its eigenvalue λ , $-\lambda$ must also be an eigenvalue with the same multiplicity (algebraic and geometric). In particular, we are left with at most 2^r choices for the square root to pick in (2.33) (a choice of sign for each eigenvalue pair $(\lambda, -\lambda)$).

There is one useful property to further determine the square root. Consider

$$R_g = Q_g P_g^{-1} = \begin{pmatrix} q_g p_g^{-1} & 0 \\ 0 & q_{g^{-1}} p_{g^{-1}}^{-1} \end{pmatrix}.$$

We observe from (2.30) that $q_g p_g^{-1} = (I - \sum_{h \neq g^{-1}} p_{h^{-1}} q_h)^{-1}$. The matrix $\sum_{h \neq g^{-1}} p_{h^{-1}} q_h$ is sub-stochastic: it has non-negative entries and the sum over each row is less or equal than one (from a starting colour, it corresponds to the probability that the coloured walk killed when visiting g comes back to e after some time). In particular, the matrix $\sum_{h \neq g^{-1}} p_{h^{-1}} q_h$ has spectral radius less than one. Thus all eigenvalues of $q_g p_g^{-1}$ and R_g have positive real part. Using (2.31), the same argument shows that z and Z have all their eigenvalues with positive real parts.

Finally, from Equation (2.31), we have $Z = I - \sum_g P_g Q_g = I - \sum_g P_g R_g P_g$. We deduce that

$$Z = I - \frac{1}{2} \sum_{g \in S} \left(P_g \sqrt{4I + (P_g^{-1} Z)^2} - Z \right). \quad (2.34)$$

Up to this issue of square root, we thus have found a fixed point equation satisfied by z (in Equation (2.34)) and expressed q_g as a function of $(z, p_g, p_{g^{-1}})$ (in Equation (2.33)). If $r = 1$, we can retrieve a known formula for colourless random walks. We note that Equation (2.34) should be compared to Proposition 3.1 in Lehner [115] where a related formula is derived in a self-adjoint case (it can be checked that $2z$ is the inverse of the diagonal term of the Green function $(I - \tilde{\mathcal{P}})^{-1}(e, e)$ where $(I - \tilde{\mathcal{P}})^{-1}$ is seen as an infinite matrix indexed by $G \times G$ with coefficients in $M_r(\mathbb{C})$).

2.4 Computing drift and entropy

For classical convolution random walks, the general formulas for the rate of escape and entropy given in Theorem 2.3 go back to the work of Furstenberg [76] and are well known. The proofs extend easily to coloured random walks, so we will not give further details. We refer to Ledrappier [114] and the influential paper by Kaimanovich and Vershik [102] for the case of free groups; to the works of Mairesse [129], Mairesse and Matheus [130], Gilch [78, 77] for more general free products. See also the paper of Nagnibeda and Woess [138].

We now turn to the derivation of (2.4) from the integral formula (2.14). The latter shows that the computation of the entropy ultimately comes down to the computation of some Radon-Nikodym derivatives. In the colourless case, this computation goes basically as follows. Consider a cylinder $\xi_1 \cdots \xi_n$. Compute $g \cdot p^\infty(\xi_1 \cdots \xi_n)$ distinguishing the different cases occurring:

- if $g = \xi_1$ then $g \cdot p^\infty(\xi_1 \cdots \xi_n) = p^\infty(\xi_2 \cdots \xi_n)$
- if $\xi_1 \in G_i, g \neq \xi_1$ and there exists $h \in G_i$ such that $gh = \xi_1$, then $g \cdot p^\infty(\xi_1 \cdots \xi_n) = p^\infty(h\xi_2 \cdots \xi_n)$
- otherwise, $g \cdot p^\infty(\xi_1 \cdots \xi_n) = p^\infty(g^{-1}\xi_1 \cdots \xi_n)$.

As one can check, the last case occurs if and only if $\xi_1 \in \text{Next}(g^{-1})$. Expand now the expressions above into products using from (2.26) that $p^\infty(\xi_1 \cdots \xi_n) = q_{\xi_1} \cdots q_{\xi_{n-1}} \mu_{\xi_n}$ to get

$$\frac{g \cdot p^\infty(\xi_1 \cdots \xi_n)}{p^\infty(\xi_1 \cdots \xi_n)} = \begin{cases} 1/q_g & \text{if } \xi_1 = g, \\ q_{g^{-1}} & \text{if } \xi_1 \in \text{Next}(g^{-1}), \\ q_h/q_{\xi_1} = q_{g^{-1}\xi_1}/q_{\xi_1} & \text{if } h \in S \text{ and } gh = \xi_1. \end{cases}$$

One can then easily express the integral (2.14) in terms of the scalars p_g, q_g and μ_g and obtain the formulas given in [114, 130].

For coloured random walks, the same computations can be made with matrices except that the cancellation between products no longer takes place. Instead, one has to deal with an infinite product of random matrices. For example, as one might guess from (2.21)-(2.26), it is true that

$$\frac{dg p_v^\infty}{dp_u^\infty}(\xi) = \lim_{n \rightarrow \infty} \frac{\mathbb{1}_v^\top q_{g^{-1}} q_{\xi_1} \cdots q_{\xi_{n-1}} \mu_{\xi_n} \mathbb{1}}{\mathbb{1}_u^\top q_{\xi_1} \cdots q_{\xi_{n-1}} \mu_{\xi_n} \mathbb{1}} \quad \text{a.s.}, \quad (2.35)$$

if $\xi_1 \in \text{Next}(g^{-1})$. Convergence as above is the object of the next subsection.

2.4.1 Convergence in direction for inhomogeneous products of matrices

In this final subsection, we prove Theorem 2.4. The computation of the Radon-Nikodym derivatives in the integral formula of the entropy requires to investigate infinite products of random, non-negative matrices (that is, with non-negative entries). We use first use results for non-negative, deterministic matrices [152] to justify limits of infinite matrix products. This yields formulas like (2.35) and a first expression for the entropy (2.39). Then we use results for products of random matrices [35, 40] to describe the law of the limits obtained, which yields eventually Theorem 2.4.

Convergence of the Radon-Nikodym derivatives. Instead of the general results in [152], it turns out that we can directly use results from [53, 109] which already involved matrix products to study finite range random walks. It does not seem however that the methods used in these references can be interpreted as an application of the linearization trick.

Let X be a finite set and Y a subset of $X \times X$ such that for all $x \in X$, the set $\{y \in X, (x, y) \in Y\}$ is non empty. Let $\Sigma := X^{\mathbb{Z}}$ be the space of doubly infinite sequences $(\xi_n)_{n \in \mathbb{Z}}$ with values in X , such that $(\xi_n, \xi_{n+1}) \in Y$ for all $n \in \mathbb{Z}$. Let $\sigma : (\xi_n)_{n \in \mathbb{Z}} \mapsto (\xi_{n+1})_{n \in \mathbb{Z}}$ denote the standard shift on Σ and given a function $f : \Sigma \rightarrow \mathbb{C}$ and $n \in \mathbb{N}$, write

$$S_n f := f + f \circ \sigma + \dots + f \circ \sigma^{n-1}.$$

Recall the definition of \mathcal{P}^+ in (2.16). We will use the following formulation of a result of [109] on convergence of matrix products, but note that a similar result was already used in the earlier reference [53].

Proposition 2.3 ([109, Prop 5.2]). *Let $(M_x)_{x \in X}$ be a family of $r \times r$ matrices with non-negative entries. Assume there exists integers $m \geq 0, k \geq 1$ and a function $B : X^k \rightarrow 2^{[r]} \setminus \{\emptyset\}$ (the set of non-empty subsets of $[r]$), such that for every $n \geq m$ and every family $x_1, \dots, x_{n+k} \in X$ with $(x_i, x_{i+1}) \in Y$ for all i ,*

$$(M_{x_1} M_{x_2} \cdots M_{x_{n+k}})_{u,v} > 0 \Leftrightarrow v \in B(x_{n+1}, \dots, x_{n+k}) \quad \forall u, v \in [r] \quad (2.36)$$

Then there exist constants $C > 0$ and $0 < \alpha < 1$, maps $\varphi, \gamma : \Sigma \rightarrow \mathbb{R}$ and $V, W : \Sigma \rightarrow \mathcal{P}^+$ such that for all $\xi \in \Sigma$,

$$\left\| e^{-S_n \varphi(\xi)} M_{\xi_1} M_{\xi_2} \cdots M_{\xi_n} - \gamma(\sigma_n \xi) V(\xi) W(\sigma^n \xi)^\top \right\| \leq C \alpha^n, \quad (2.37)$$

where $V = V(\xi_1, \xi_2, \dots)$ depends only on the “forward coordinates“, while on the other hand $W = W(\xi_0, \xi_{-1}, \dots)$ depends only on the “backward coordinates“. Furthermore,

$$M_{\xi_1} V(\sigma \xi) = e^{\varphi(\xi)} V(\xi). \quad (2.38)$$

Proposition 2.3 states shows under condition (2.36), the product $M_{\xi_1} M_{\xi_2} \cdots M_{\xi_n}$ tends up to renormalization factor to a rank one matrix whose range is spanned by the vector $V(\xi)$. This result is well known for powers of a matrix with positive entries, in which case the vector V is nothing but the Perron-Frobenius eigenvector. Proposition 2.3 is thus a generalization of the Perron Frobenius theory to inhomogeneous products of non-negative matrices with possibly zero columns (condition (2.36)). We refer to [109] for additional properties of the functions φ, γ with the vectors V, W which will not be needed here.

To apply Proposition 2.3 in context of linearized random walks, take $X = S$ and $(x, y) \in Y$ if and only if $y \in \text{Next}(g)$.

Corollary 2.2. *Suppose the matrices $(q_g)_{g \in S}$ satisfy the hypothesis of Proposition 2.3 and let $V : \partial G \rightarrow \mathcal{P}^+$ be the corresponding map. Then for all $u, v \in [r]$, $\xi \in \partial G$,*

$$\frac{dgp_v^\infty}{dp_u^\infty}(\xi) = \begin{cases} \frac{\langle \mathbb{1}_v, V(\sigma\xi) \rangle}{\langle q_g(u, \cdot), V(\sigma\xi) \rangle} & \text{if } \xi_1 = g, \\ \frac{\langle q_{g^{-1}}(v, \cdot), V(\xi) \rangle}{\langle \mathbb{1}_u, V(\xi) \rangle} & \text{if } \xi_1 \in \text{Next}(g^{-1}), \\ \frac{\langle q_h(v, \cdot), V(\sigma\xi) \rangle}{\langle q_{gh}(u, \cdot), V(\sigma\xi) \rangle} & \text{if } h \in S \text{ and } gh = \xi_1. \end{cases} \quad (2.39)$$

Proof. From (2.21)-(2.26), for any cylinder $\xi_1 \cdots \xi_n$

$$g \cdot p_v^\infty(\xi_1 \cdots \xi_n) = \begin{cases} \mathbb{1}_v^\top q_{\xi_2} \cdots q_{\xi_{n-1}} \mu_{\xi_n} \mathbb{1} & \text{if } \xi_1 = g, \\ \mathbb{1}_v^\top q_{g^{-1}} q_{\xi_1} \cdots q_{\xi_{n-1}} \mu_{\xi_n} \mathbb{1} & \text{if } \xi_1 \in \text{Next}(g^{-1}), \\ \mathbb{1}_v^\top q_h q_{\xi_2} \cdots \mu_{\xi_n} \mathbb{1} & \text{if } h \in S \text{ and } gh = \xi_1. \end{cases}$$

Let $\xi \in \partial G$. Combining the previous computation with Proposition 2.3, there exist $\lambda_n > 0$, uniformly lower bounded, and $\alpha < 1$ such that if $\xi_1 = g$,

$$\begin{aligned} \frac{g \cdot p_v^\infty(\xi_1 \cdots \xi_n)}{p_u^\infty(\xi_1 \cdots \xi_n)} &= \frac{\mathbb{1}_v^\top (q_{\xi_2} \cdots q_{\xi_{n-1}}) (\mu_{\xi_n} \mathbb{1})}{q_g(u, \cdot)^\top (q_{\xi_2} \cdots q_{\xi_{n-1}}) (\mu_{\xi_n} \mathbb{1})} \\ &= \frac{\mathbb{1}_v^\top V(\sigma\xi) \lambda_n + O(\alpha^n)}{q_g(u, \cdot)^\top V(\sigma\xi) \lambda_n + O(\alpha^n)} \\ &\xrightarrow{n \rightarrow \infty} \frac{\langle \mathbb{1}_v, V(\sigma\xi) \rangle}{\langle q_g(u, \cdot), V(\sigma\xi) \rangle}. \end{aligned}$$

The other cases are treated similarly.

On the other hand

$$\frac{g \cdot p_v^\infty(\xi_1 \cdots \xi_n)}{p_u^\infty(\xi_1 \cdots \xi_n)} = \int_{\partial G} \frac{dgp_v^\infty}{p_u^\infty} \frac{\mathbb{1}_{\xi_1 \cdots \xi_n}}{p_u^\infty(\xi_1 \cdots \xi_n)} dp_u^\infty$$

but as $n \rightarrow \infty$ the measure $\frac{\mathbb{1}_{\xi_1 \cdots \xi_n}}{p_u^\infty(\xi_1 \cdots \xi_n)} dp_u^\infty$ converges to the Dirac mass at ξ : for all integer k and $\epsilon > 0$, the mass of all k -cylinders except the cylinder $\xi_1 \cdots \xi_k$ is bounded by ϵ for n large enough. Thus the above limits give indeed the Radon-Nikodym derivative. \square

The next proposition asserts that for the coloured random walks that we are mainly interested in, Corollary 2.2 applies.

Proposition 2.4. *Suppose the coloured random walk $(Y_n)_{n \geq 0}$ is a linearized random walk obtained from a finite range random walk $(X_n)_{n \geq 0}$ via one of the procedures presented in Section 2.2. If $(X_n)_{n \geq 0}$ is irreducible, then the matrices $(q_g)_{g \in S}$, satisfy the hypothesis of Proposition 2.3.*

Proof. Extend the definition (2.15) of matrices q_g to the whole group and notice that if $g \in G$ writes as a reduced word $g = g_1 \cdots g_n$, then $q_g = q_{g_1} \cdots q_{g_n}$. We recall that given the set of generators considered (2.4) representatives are uniquely defined. Thus proving (2.36) comes down to prove that for $|g|$ large enough, $q_g(u, v) > 0$ if and only if v is in some subset of $[r]$ that depends only on the last letters of g . This statement regarding the coloured random walk can be transferred to a statement about the initial walk as follows.

Consider the constructions of the linearized random walks in Section 2.2: both are made so that if $X_0 = e$ and $Y_0 = (e, 1)$, one can couple $(Y_n)_{n \geq 0}$ and $(X_n)_{n \geq 0}$ to have $X_n = Y_{\tau_n}$ where τ_n is the n -th return time at colour 1. Furthermore, for all colour $v \in [r]$, there exists (at least one) $h \in K$ with representative $h = h_1 \cdots h_k h_{k+1} \cdots h_l$ such that $\mathbb{P}_{(e,1)}[Y_k = (h_1 \cdots h_k, v)] > 0$. The main difference between the proposed linearization procedures is the number of such h that one can associate to the colour v , but this has no consequence for what follows.

Let $K(v)$ be the set of such elements h . For each $h = h_1 \cdots h_k h_{k+1} \cdots h_l \in K(v)$, set $p(h) := h_1 \cdots h_k$ and $s(h) := h_{k+1} \cdots h_l$. Then consider the following events: given $a, g \in G$, $h = h_1 \cdots h_l \in K$, if there exists $k \leq l$ such that $ah_1 \cdots h_k = g$, let

$$A_{a,h}(g) := \{\exists n \geq 0 \mid X_n = a, X_{n+1} = ah\}$$

and $A_{a,h}(g) = \emptyset$ otherwise, and let $B_{a,h}(g)$ the event that $A_{a,h}(g)$ occurs before all other events $A_{b,h'}(g)$, $(b, h') \neq (a, h)$.

Suppose now Y_0 starts with colour u . By G -invariance of transition probabilities we can suppose that it is at (e, u) . Then the starting colour imposes the walk to transition towards $(s(h), 1)$ for some $h \in K(u)$. Once at state $(s(h), 1)$, Y_n will visit g if and only if X_n , started at $s(h)$, makes a transition that goes through g , ie if and only if the event $A_{a,h'}(g)$ occurs for some $a \in G, h' \in K$. Thanks to the irreducibility of X_n it is clear that such transitions occur, but the difficulty is to make sure that Y_n can arrive for the first time at g with the prescribed color v . This will occur with the events $B_{a,h'}(g)$ for $h' \in K(v)$ and $a = gp(h')^{-1}$. Given $x \in G, h \in K$, let $m_g(x, h) := \mathbb{P}_x[B_{gp(h)^{-1}, h}(g)]$. What precedes thus sums up in the lower bound

$$q_g(u, v) \geq \mathbb{P}_{(e,u)}[\tau_{(s(h),1)} < \infty] m_g(s(h), h') \quad \forall h \in K(u), h' \in K(v)$$

where $\mathbb{P}_{(e,u)}[\tau_{(s(h),1)} < \infty] > 0$ is the probability that Y_n visits $(s(h), 1)$. In particular, it now suffices to prove that for all $x, g \in G, h \in K$, the positivity of $m_g(x, h)$ depends only on h and the last letters of g , provided $|g|$ is large enough. This can be shown in the same way as Lemma 5.3 in [109] where a similar result is proved.

Let L be the largest length of elements in the support of X_n . By G -invariance of transition probabilities $m_g(x, h) = m_{x^{-1}g}(e, h)$ so one can in particular suppose $x = e$.

From the irreducibility of X_n , every element of the generating set S can be reached by the random walk in a bounded number of steps l , because S is finite. Therefore for $|g| > 2L + l$, any path of positive probability that goes from e to $gp(h)^{-1}$ without passing through g and then moves from $gp(h)^{-1}$ to $gp(h)^{-1}h = gs(h)$ can be extended with an initial loop between e and any $s \in S$. From this we deduce that $m_g(e, h) > 0$ if and only if $m_g(s, h) = m_{s^{-1}g}(e, h) > 0$. Choosing s as the first letter of g , we deduce that the positivity of $m_g(a, b)$ does not depend on it. Iterating this argument we obtain that it only depends on h and the $2L + l$ last letters of g . \square

Law of Radon-Nikodym derivatives. The previous result is purely deterministic and does not take into account that the process $(X_\infty^{(n)}, u_n)_{n \geq 0}$ is a Markov chain. Our goal in this paragraph is to determine the law of the probability vector V appearing in Corollary 2.2. The monograph [35] (see also [40]) state results for products of iid random matrices which, as we shall see, apply to Markovian products as well.

Remploying the general framework of Proposition 2.3, consider a Markov chain $(Z_n)_{n \geq 0}$ on the finite state space X with transition matrix Q and $Y := \{(x, y) \in X \times X, Q(x, y) > 0\}$. Let $(M_x)_{x \in X}$ be a family of $r \times r$ non-negative matrices, such that condition (2.36) holds.

Consider the sequence defined by $Y_n := M_{Z_n}, n \geq 1$. By Proposition 2.3, a.s. the product $Y_1 \cdots Y_n$ is non-zero and converges in direction to a rank one matrix spanned by a random vector $V \in \mathcal{P}^+$. If one multiplies the product on the left by another matrix M_x , one still has a products of matrices M_x , so one can expect the law of V to satisfy some invariance property.

For all non-negative matrix $A \in M_r(\mathbb{R})$ and all $z \in \mathcal{P}^+$, we define Az to be the normalized image of z (identified with a vector in \mathbb{R}^r) that makes it a probability vector, provided $Az \neq 0$. The latter case will not be an issue: condition (2.36) implies in particular that matrices M_x have no zero row, and vectors of \mathcal{P}^+ have no zero coordinate so $M_x z$ is well defined for all $x \in X, z \in \mathcal{P}^+$.

Since Z_n is a Markov chain, the law of V may depend on the starting state of Z_n . Hence it is natural to think of it as a coloured measure (colours being here the states $x \in X$).

Definition 2.5. A family $\nu = (\nu_x)_{x \in X}$ of probability measure on \mathcal{P}^+ is called a coloured measure. Let $Q * \nu$ denote the coloured measure defined by

$$\int f(z) d(Q * \nu)_x(z) = \sum_y \int f(M_y z) Q(x, y) d\nu_y(z)$$

for all bounded measurable function f on \mathcal{P}^+ and $x \in X$. The coloured measure ν is said to be invariant if $Q * \nu = \nu$.

Lemma 2.3. *Let ν_x be the law of V when Z_n is started at x , the coloured measure $(\nu_x)_{x \in X}$ is the unique coloured measure which is invariant with respect to Q .*

Proof. We use Markov property together with (2.38) to obtain the invariance of the law of V .

The proof of uniqueness is similar to the proof for the harmonic measure in Proposition 2.1. Consider another invariant coloured measure $(\rho_x)_{x \in X}$. By invariance, for all bounded measurable function f on \mathcal{P}^+ the sequence

$$M_n := \int f(Y_1 \cdots Y_n w) d\rho_{Z_n}(w), \quad n \geq 0$$

is a bounded martingale with respect to the filtration $\mathcal{F}_n := \sigma(Z_0, Z_1, \dots, Z_n)$, so M_n converges a.s. and in L^1 . This being true for all bounded measurable function, this implies the measures $Y_1 \cdots Y_n * \rho_{Z_n}$ converges weakly to a measure. Now, because the product $Y_1 \cdots Y_n$ converges in direction to V , the limit is necessarily the Dirac mass at V .

On the other hand, the martingale property with the L^1 convergence gives that for all bounded measurable functions f on \mathcal{P}^+ , for all $x \in X$,

$$\mathbb{E}_x[f(V)] = \mathbb{E}_x \left[\int f \delta_V \right] = \int f d\rho_x.$$

Therefore ρ_x must be the law of V when $Z_0 = x$, that is $\rho_x = \nu_x$, and is consequently unique. \square

We are finally ready to prove Theorem 2.4.

Proof of Theorem 2.4. We apply Lemma 2.3 in the context of coloured random walks, with $X = S \times [r]$ and $Q((g, u)(h, v)) = \mu_h(u, v)$. Due to the form of the transition probabilities the law of V is only indexed by colours. It remains to use Corollary 2.2 in conjunction with the integral formula of the entropy given by Theorem 2.3. \square

Chapter 3

Accelerating abelian random walks with hyperbolic dynamics

This chapter is adapted from the paper [65], done jointly with Laurent Massoulié.

3.1 Introduction

In [49], Chung Diaconis and Graham investigated the behaviour of a Markov chain on $\mathbb{Z}/n\mathbb{Z}$ defined by $X_{t+1} = aX_t + B_t$, with (B_t) a sequence of iid increments distributed on $\{-1, 0, 1\}$. They proved for $a = 2$ that after $t = O(\log n \log \log n)$ steps the distribution of X_t is close to uniform, thus showing a dramatic speed-up over the simple random walk, which needs $\Omega(n^2)$ steps. Recently, attention has been brought back to the potential speed-up obtained by applying deterministic functions to Markov chains [45, 87, 88, 18]. In this work, we study an analog of the Chung-Diaconis-Graham process in the multi-dimensional case. This case has been previously studied [8, 100, 9, 106] however the bounds proved in these papers did not match those of dimension one, or only in specific cases. This work improves these results and aims to provide a new look on the speed-up caused by the doubling map.

3.1.1 Main results

Let $d \geq 2$ be an integer, μ a finitely supported probability measure on \mathbb{Z}^d . We write $\text{GL}_d(\mathbb{Z})$ for the set of invertible matrices with integer entries and integer-valued inverse. A matrix is hyperbolic if it has no eigenvalue of modulus 1. Given $n \geq 1$ and $A \in \text{GL}_d(\mathbb{Z})$ an invertible hyperbolic matrix, consider the discrete-time random walk $(X_t)_{t \geq 0}$ on $(\mathbb{Z}/n\mathbb{Z})^d$ defined by

$$X_0 := 0 \quad X_t = AX_{t-1} + B_t \pmod n \quad \forall t \geq 1 \quad (3.1)$$

where B_t are iid random variables on \mathbb{Z}^d with distribution μ . Such random walks have been first considered by Chung, Diaconis and Graham in [49] in the case $d = 1$, $A = 2$.

A induces a bijection on the finite set $(\mathbb{Z}/n\mathbb{Z})^d$. It is easy to see from this that the uniform measure U on $(\mathbb{Z}/n\mathbb{Z})^d$ is invariant for X_t , so X_t will converge in law to U provided it is irreducible and aperiodic. Convergence to stationarity of finite Markov chains is classically measured by total variation distance: for any pair of measures p, q on a finite set X , one sets

$$\|p - q\|_{\text{TV}} := \max_{A \subseteq X} |p(A) - q(A)| = \frac{1}{2} \sum_{x \in X} |p(x) - q(x)|.$$

The mixing time is then defined as the time needed for a Markov chain to get at distance at most ϵ to its stationary distribution, for a fixed parameter ϵ .

Let $\text{supp } \mu$ denote the support of μ and consider H the smallest A -invariant subgroup of \mathbb{Z}^d that contains $\text{supp } \mu - \text{supp } \mu := \{x - y : x, y \in \text{supp } \mu\}$.

As a subgroup of \mathbb{Z}^d , H is itself isomorphic to \mathbb{Z}^k for some integer $k \leq d$; see for instance [110, Thm 7.8]. More precisely, there exists a basis of \mathbb{Z}^d , that is a family $(u_i)_{i=1}^d$ that generates \mathbb{Z}^d as a group, and positive integers $(a_i)_{i=1}^k$, such that H is the subgroup generated by $(a_i u_i)_{i=1}^k$. The integer $\text{rank } H := k$ is called the rank of H and is equal to the dimension of the \mathbb{Q} -subvector space of \mathbb{Q}^d spanned by H . Our first result gives an upper bound on the mixing time of the same order as the one given by Chung, Diaconis and Graham in [49]. It applies to any dimension $d \geq 2$ and improve the results of [9, 100].

Theorem 3.1. *Let $d \geq 2$ be an integer, μ a probability measure on \mathbb{Z}^d and $A \in \text{GL}_d(\mathbb{Z})$. Let $(B_t)_{t \geq 1}$ be iid random variables with law μ and consider the random walk $(X_t)_{t \geq 0}$ on $(\mathbb{Z}/n\mathbb{Z})^d$ defined by (3.1).*

Let $P^t(0, \cdot)$ denote the distribution of X_t and U the uniform measure on $(\mathbb{Z}/n\mathbb{Z})^d$. Let H be the smallest A -invariant subgroup of \mathbb{Z}^d that contains $\text{supp } \mu - \text{supp } \mu$. Suppose H is of rank d , generated by $(a_i u_i)_{i=1}^d$ for some basis $(u_i)_{i=1}^d$ of \mathbb{Z}^d and positive integers $(a_i)_{i=1}^d$. If A is hyperbolic, there exists a constant $C > 0$ such that, for all n coprime with all the a_i , if $t > C \log n \log \log n$ then $\|P^t(0, \cdot) - U\|_{\text{TV}} \rightarrow 0$ as $n \rightarrow \infty$.

Remark 3.1. The hyperbolicity assumption on A is essential. Theorem 3.11 of [9] shows that if A has an eigenvalue which is a root of unity, $\Omega(n^2)$ steps are necessary. The proof of Theorem 3.1 will use extensively the properties of the deterministic dynamical system $x \mapsto A^\top x$ on the continuous torus $\mathbb{R}^d/\mathbb{Z}^d$. Such dynamical systems are commonly referred to as hyperbolic toral automorphisms and are typical examples of dynamical systems exhibiting chaotic properties. Thus Theorem 3.1 and its proof suggest that the speed-up observed is the consequence of specific features of hyperbolic dynamical systems.

Remark 3.2. On the other hand, the invertibility assumption on A seems superfluous. We conjecture it suffices to have $\det A \neq 0$ and restrict to n coprime with $\det A$, as it is already the case for dimension 1 [49, 94, 95]. To prove Theorem 3.1 we make use of Markov partitions, which apply essentially to invertible maps. Many arguments in this work, even those about Markov partitions, only use "forward trajectories" and could thus be extended to the non-invertible case. The main issue is thus to show existence of Markov partitions for non invertible maps.

Remark 3.3. As one can guess, the condition given on H and n is necessary to have irreducibility and aperiodicity of the random walk. This will be proved subsequently in Proposition 3.6. Thus Theorem 3.1 shows that for all n , $O(\log n \log \log n)$ steps are sufficient to reach stationarity as long as convergence holds.

Remark 3.4. Let us remark also that there is no loss of generality in supposing the walk is started at 0. Indeed, should the random walk start at $x \in (\mathbb{Z}/n\mathbb{Z})^d$, we can always write $X_t = A^t x + \tilde{X}_t$ with \tilde{X}_t a random walk started at 0. Since $A^t x$ is deterministic, it does not affect total variation distance.

Our second result establishes a lower bound on the mixing time, under a condition of finite entropy. Given any measure p on a discrete set X , we recall the entropy of p is defined as

$$H(p) := - \sum_{x \in X} p_x \log p_x.$$

Theorem 3.2. *Let μ be a measure on \mathbb{Z}^d with finite entropy. Let $(X_t)_{t \geq 0}$ be as in (3.1). Then*

$$\|P^t(0, \cdot) - U\|_{\text{TV}} \geq 1 - \frac{tH(\mu) + \log 2}{\log n}. \quad (3.2)$$

The upper bound of Theorem 3.1 is valid for all n , as long as convergence to uniformity holds. As it was already the case for the Chung-Diaconis-Graham process in dimension one, it turns out that for almost all n tending to infinity only $O(\log n)$ steps suffice to reach stationarity. A property \mathcal{P} is satisfied by almost all integer if the proportion of integers smaller than n that satisfy \mathcal{P} tends to 1 as $n \rightarrow \infty$.

Theorem 3.3. *Consider the setting of Theorem 3.1. There exist a constant C such that for almost all n , if $t > C \log n$ then $\|P^t(0, \cdot) - U\|_{\text{TV}} \rightarrow 0$ as $n \rightarrow \infty$.*

Affine random walks defined by (3.1) allow in particular to study random walks defined by recursions of higher order. For instance the so-called Fibonacci random walk on $\mathbb{Z}/n\mathbb{Z}$, $X_{t+1} = X_t + X_{t-1} + B_{t+1}$, studied in [45], can naturally be written as an affine random walk, using the matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, which is hyperbolic. Theorems 3.1 and 3.3 thus imply that when convergence holds, $O(\log n \log \log n)$ or $O(\log n)$ steps

are sufficient to reach stationarity. This applies in particular to the case where B_t is uniform in $\{-1, 0, 1\}$ and improves the result of [45].

Remark 3.5. From Theorem 3.3, one may inquire if the upper bound of $O(\log n)$ is valid for all n . We conjecture it is not. In dimension $d = 1$, $A = 2$ and B_t uniform in $\{0, \pm 1\}$, [49] gives specific values of n_k , including $n_k = 2^k - 1$, for which $O(\log n_k \log \log n_k)$ is in fact the right order of magnitude. This phenomenon occurs because for such n_k , the map induced by $x \mapsto 2x$ on $\mathbb{Z}/n\mathbb{Z}$ has order $O(\log n_k)$. Thus we expect that for certain distributions of B_t , $O(\log n \log \log n)$ is indeed the right mixing time whenever the permutation induced by the matrix A has order $O(\log n)$. This could be an interesting question for future research.

3.1.2 Related works

On the Chung-Diaconis-Graham process, ie the case $d = 1$, great effort has been made to improve the bounds on the mixing rates, see [95, 96, 97, 98, 141]. The latest paper is [68], where Eberhard and Varjú eventually established the cutoff phenomenon at entropic time $\log n/h$ for almost all n and some entropic constant h .

There has also been interest for generalizations of this process: Hildebrand considered the case $X_t = a_t X_{t-1} + B_t$ where a new random multiplier a_t is applied at each step [94, 99], while Ascì [8, 9] and Hildebrand, McCollum [100] considered d -dimensional generalizations, obtaining upper bounds of order $O((\log n)^2)$. Ascì [9] and recently Klyachko [106] managed to prove order $O(\log n \log \log n)$ for specific cases in dimension 2. Our results improve on these works, giving a $O(\log n \log \log n)$ upper bound for any hyperbolic matrix $A \in \text{GL}_d(\mathbb{Z})$. One crucial argument in the one-dimensional case was the use of the binary decomposition in order to identify the dynamical system $x \mapsto 2x$ on \mathbb{R}/\mathbb{Z} with a shift, which a priori had no direct extension to the multi-dimensional case, thus accounting for the difference in the order of magnitude. Klyachko [106] uses a similar technique, β -ary expansions, to get the $\log n \log \log n$ bound. This work uses the theory of Markov partitions which provides a general framework where dynamical system can be identified to shifts. Both kind of expansions are related. Another related result is [59], where Diaconis and Graham investigated the case of an affine random walk on the hypercube. With our notations this would amount to taking $n = 2$ and $d \rightarrow \infty$, which is not a case covered here.

On the other hand, the effect of applying deterministic bijections on the mixing times of Markov chains has been recently reconsidered from a general point of view. In [32], Bordenave, Qiu and Zhang prove an upper bound on the second eigenvalue of the product of a bistochastic matrix with a uniform permutation. In [45] Diaconis and Chatterjee consider a generic Markov chain on n states, and give a geometric

condition on deterministic bijections which yields a mixing time of order $O(\log n)$. This condition is made to increase the expansion of the Markov chain to $\Omega(1)$, a property known to imply mixing in $O(\log n)$. Moreover, it holds with large probability when the bijection is chosen uniformly at random, so almost every bijection actually yields fast mixing. Those results were subsequently improved by Ben Hamou and Peres in [18], who established the cutoff phenomenon at entropic time in the case the bijection is picked uniformly at random.

Although speed-up occurs with almost every bijection, it was noticed however in [45] that the doubling map $x \mapsto 2x$ does not fall in that category, so expansion is not the reason accounting for the acceleration. Thus a second goal of our work is to suggest an alternative explanation as to why the doubling map improves the mixing rate. As explained in Remark 3.1, at the heart of our proof is the behaviour of the hyperbolic dynamical system $x \mapsto A^\top x$. Hence the speed-up established here may be specific to such maps and of a different nature than the acceleration proved for almost all bijection.

Finally, let us mention two other examples of explicit deterministic bijections implying speed-ups that have been studied. On the cycle $\mathbb{Z}/p\mathbb{Z}$ with p prime, He [87] proves an almost linear mixing time in p when the bijection considered is a rational function. This bound has been improved in [88] in the specific case of the inverse function $f : x \mapsto 1/x$ if $x \neq 0$, $f(0) := 0$, where He, Pham and Xu shows that $O(\log p)$ steps suffice to reach stationarity.

3.1.3 Discussion

In the two following paragraphs we discuss motivations and related models.

Random matrices: One can inquire about applying random matrices instead of deterministic ones, ie consider random walks of the form $X_t = A_t X_{t-1} + B_t$, with for instance $(A_t)_{t \geq 1}$ iid random matrices. For $d = 1$, such cases were considered in [94, 99].

The simplest randomness we can add is by taking $A_t = A^{\epsilon_t}$ where ϵ_t are iid Bernoulli random variables of parameter $1/2$. This may slow down the convergence to stationarity, for instance with B_t corresponding to increments of the simple random walk. With constant A , acceleration comes from applying successive powers A^t to the random increments B_k . Now with $A_t = A^{\epsilon_t}$ the exponent of A is to be replaced with the t -th step a random walk on \mathbb{Z} . Because of the diffusive behaviour of the random walk on \mathbb{Z} , it is of order at most $O(\sqrt{t})$. As a consequence, the mixing time becomes $O((\log n)^2)$. This can be proven with the same argumentation as in [99], which treats the case $d = 1$.

Another case which can be considered is when A_t are iid with distribution ν , such that the subgroup generated by $\text{supp } \nu$ is $\text{SL}_d(\mathbb{Z})$. In this case, it is possible to prove

acceleration to $(O(\log n))$, without restriction on n , while removing the constraint that the matrices have no eigenvalues of modulus 1. The idea is to consider the evolution of the pair (X_t, S_t) with $S_t := A_1 \cdots A_t$. Since the A_t are iid, the process (X_t, S_t) is a convolution random walk, ie a product of iid increments, on the semi-direct product $(\mathbb{Z}/n\mathbb{Z})^d \rtimes \mathrm{SL}_d(\mathbb{Z}/n\mathbb{Z})$, whose group operation is defined by $(x, A)(y, B) = (x + Ay, AB)$. This random walk can then be obtained as the reduction modulo n of a convolution random walk on the infinite group $\mathbb{Z}^d \rtimes \mathrm{SL}_d(\mathbb{Z})$. Such groups have been extensively studied for their expansion properties, as they give rise to deterministic constructions of families of expander graphs, including quotients $\mathrm{SL}_d(\mathbb{Z}/n\mathbb{Z})$. We refer to the monography of Tao [157] for an introduction to this topic.

Since expansion is related to the fast-mixing of simple random walks, the same tools can be used to prove fast mixing of affine random walks. Incidentally, such results have already been used in [88] to prove speed-up for the inverse mapping on $\mathbb{Z}/p\mathbb{Z}$. Let us give a few examples.

Let S be a finite symmetric set of generators of $\mathbb{Z}^d \rtimes \mathrm{SL}_d(\mathbb{Z})$ and consider the Cayley graph $G_n = (V_n, E_n)$ where $V_n = (\mathbb{Z}/n\mathbb{Z})^2$ and $(x, y) \in E_n$ is an edge if and only if there exists $s = (z, A) \in S$ such that $y = sx := Ax + z$. Consider the simple random walk, which at each step is multiplied by a uniform element of S . To ensure aperiodicity we can always impose laziness, ie allow the walk to stay put with positive probability. Theorem 2.4.1 in [157] proves that the graphs G_n are expanders, which implies that the (lazy) simple random walk mixes in $O(\log n)$ steps. In particular this applies to the random walk on $(\mathbb{Z}/n\mathbb{Z})^2$ which has holding probability $1/2$ and otherwise moves from (x, y) to $(x \pm 1, y)$, $(x, y \pm 1)$, $(x \pm y, y)$ or $(x, y \pm x)$ with probability $1/16$.

Let us finally mention the result of [118], which shows that the expansion of a simple random walk on a semi-direct product $(\mathbb{Z}/p\mathbb{Z})^d \rtimes \mathrm{SL}_d(\mathbb{Z}/p\mathbb{Z})$, p prime, and that of the projection to $\mathrm{SL}_d(\mathbb{Z}/p\mathbb{Z})$, are of the same order. Hence in this case the mixing properties of an affine random walk are essentially determined by the product of random matrices.

Interacting random walks: There has been another motivation to generalizing the result of [49] to the high-dimensional case. Suppose one has two Markov chains (X_t) , (Y_t) on the same finite state space. Would it be possible to make them "interact" in such a way as to accelerate the convergence to stationarity? A natural example of interaction is group multiplication. Thus one could investigate Markov chains $Z_t = (X_t, Y_t)$ on $G \times G$, with G a finite group, defined for instance such that, with some probability we add to Z_t a random increment, and with remaining probability we make its two coordinates interact, in the sense that either $X_{t+1} := Y_t X_t$ or $Y_{t+1} := Y_t X_t$ with some

probability. Such models were proposed in [45]: given a finite set S , P a transition matrix on S and $f : S^d \rightarrow S^d$ a function such that $f(\cdot, x_2, \dots, x_d)$ is a bijection for all x_2, \dots, x_d , the authors inquire about the random walk that moves deterministically from a state (X_1, \dots, X_d) to the state $(X_2, \dots, X_{d-1}, f(X_1, \dots, X_d))$ and then takes a step of P in the last coordinate. The latter model allows in particular to study Markov chains defined by recursion of higher order, like $X_{t+1} = X_t X_{t-1} B_{t+1}$ with (B_t) a sequence of i.i.d. random variables on the group. However at this stage there is no general result on the mixing rates of such Markov chains.

A case at hand is that of Abelian groups. When the group is Abelian, bijections such as $(x, y) \mapsto (y, x + y)$ are group homomorphisms. Moreover every finite Abelian group G is isomorphic to a product of cyclic groups $G \simeq \prod_i \mathbb{Z}/k_i\mathbb{Z}$, and one can ensure uniqueness of the k_i by imposing that k_i divides k_{i+1} for all i ; see [110, Thm 7.7]. Then $G^d \simeq \prod_i (\mathbb{Z}/k_i\mathbb{Z})^d$ and the previous homomorphisms are those obtained by taking a matrix $A \in GL_d(\mathbb{Z})$ and applying A to each i -th coordinate. Thus the projection on an affine random walk on the i -th coordinate yields an affine random walk on the torus $(\mathbb{Z}/k_i\mathbb{Z})^d$, to which we can apply Theorem 3.1.

What about non-linear interactions? The proof technique initiated in [49] which has been central to all works on affine random walks of $(\mathbb{Z}/n\mathbb{Z})^d$ relies heavily on linearity. The question therefore remains open. Let us remark though that when the bijection takes the form of random multiplications between coordinates in G^d , the model looks very similar to the so-called product replacement algorithm. The latter is designed to construct uniform sets of generators of a finite group. Given a k -uplet of generators of a group G , the algorithm consists at each step, to select two generators at random and multiply one by the other. Bounds of order $\log|G|$ on the mixing time are available when G is Abelian. In the general case this is conjectured to be the correct order, as it would be the consequence of Kazhdan's (T) property for the automorphism groups of free groups; see [125, 144]. As far as we know this is still an open question.

Organization of the chapter

We start by proving the lower bound in Section 3.2. Section 3.3, which forms the crux of the chapter, contains the essential arguments used in the proof of Theorem 3.1, including results from hyperbolic dynamics. It is followed by Section 3.4, where we establish that the necessity of conditions in 3.1 for irreducibility and aperiodicity. Then in Section 3.5 we prove Theorem 3.3. Finally, Section 3.6 provides the definitions and proofs of the results about hyperbolic dynamics that are needed in the other sections.

3.2 A lower bound for all n : proof of Theorem 3.2

Entropic arguments to get lower bounds on mixing times have become standard. The following inequality, due to Fannes and Audenaert [10], was already used in [88]. An entropic argument was also used in [68] to lower bound the mixing time and prove the cutoff phenomenon.

Proposition 3.1 (Fannes and Audenaert [10]). *Let μ, ν be two probability measures defined on a finite set of size $N \geq 1$. Let $\epsilon := \|\mu - \nu\|_{\text{TV}}$ denote their total variation distance. Then*

$$|H(\mu) - H(\nu)| \leq \epsilon \log(N - 1) + H(\epsilon) \quad (3.3)$$

where $H(\epsilon) = -\epsilon \log \epsilon - (1 - \epsilon) \log(1 - \epsilon)$ is the entropy of a Bernoulli random variable of parameter ϵ .

Proof of Theorem 3.2. We can bound the entropy of a Bernoulli random variable by $\log 2$, hence Proposition 3.1 shows

$$\|P^t(0, \cdot) - U\|_{\text{TV}} \geq 1 - \frac{H(X_t) + \log 2}{\log n}.$$

On the other hand the entropy of X_t grows at most linearly as $tH(\mu)$. Indeed we can construct X_t as $X_t = Y_t \bmod n$, where Y_t is an affine random walk on \mathbb{Z}^d defined by $Y_0 := 0, Y_t := AY_{t-1} + B_t$ for all $t \geq 1$. Using the well known fact that entropy is non-increasing under deterministic transforms yields $H(X_t) \leq H(Y_t)$. Using conditional entropies, we can then notice $(H(Y_t))_{t \geq 0}$ forms a subadditive sequence, as

$$\begin{aligned} H(Y_{t+s}) &\leq H(Y_{t+s}, Y_t) = H(Y_t) + H(Y_{t+s} | Y_t) \\ &= H(Y_t) + H(Y_s) \end{aligned}$$

To get the second line, notice that one can write $Y_{t+s} = A^s Y_t + Z_s$ with Z_s a random variable independent from Y_t and distributed like Y_s , so that $H(Y_{t+s} | Y_t) = H(Y_s)$. As a consequence of subadditivity, $H(Y_t) \leq tH(Y_1) = tH(\mu)$, which proves the result. \square

3.3 An upper bound for all n : proof of Theorem 3.1

Main idea of the proof. The basic idea is to use Fourier analysis: by the classical upper bound lemma of Diaconis and Shahshahani (see (3.4) below), the ℓ^2 -distance between the law $P^t := P^t(0, \cdot)$ of the random walk and the uniform measure can be expressed in terms of $\widehat{P^t}$, the discrete Fourier transform of the law. However, by the linearity of the dynamics, the process X_t is a sum of independent random variables, so its Fourier transform factorizes. The factors of $\widehat{P^t}$ are the Fourier transform $\hat{\mu}$ of

μ composed with successive powers of A^\top , the transpose of the matrix A . The goal will thus be to get the best upper bound on $|\hat{\mu}((A^\top)^s \rho)|$ for all $\rho \in (\mathbb{Z}/n\mathbb{Z})^d$, $\rho \neq 0$. Considering a function f , independent of n , defined on the continuous torus $\mathbb{R}^d/\mathbb{Z}^d$ such that $f(\rho/n) = |\hat{\mu}(\rho)|$ for all $\rho \in \mathbb{T}_n$, we will then identify a compact subset W^c of $\mathbb{R}^d/\mathbb{Z}^d$ on which f is bounded by a constant $\gamma < 1$. Thus we can bound $\widehat{P^t}(\rho)$ by γ^l , where l is the number of steps $s \leq t$ for which $(A^\top)^s \rho \in W^c$. For that matter, we will use the theory of hyperbolic toral automorphisms to code such orbits by bi-infinite sequences. By doing so, the map $x \mapsto A^\top x$ gets identified with a shift, which allows to bound the number of times the orbits cross W^c .

3.3.1 Fourier analysis on the torus

Let $n \geq 1$ be an integer. From now on we write $\mathbb{T}_n := (\mathbb{Z}/n\mathbb{Z})^d$. Let $(X_t)_{t \geq 0}$ be defined as in (3.1), and for all $t \geq 0$ write P^t for the law of X_t .

Let $\langle \cdot, \cdot \rangle$ be the usual inner product on \mathbb{R}^d . Given $\rho \in \mathbb{T}_n$ and a function f on G , the Fourier transform of f at ρ is

$$\hat{f}(\rho) := \sum_{x \in \mathbb{T}_n} f(x) e^{2i\pi \langle x, \rho \rangle / n}.$$

In this expression x, ρ are implicitly identified with their representative in $[0, n-1]^d$. Due to the 2π -periodicity of the complex exponential, the result is independent of the choice of representatives. Such identifications will be done frequently in the sequel without further explicit mention.

Bounds on mixing times can be obtained through Fourier transforms thanks to the upper bound lemma of Diaconis and Shahshahani [56]. Recall U denotes the uniform measure on \mathbb{T}_n . In the present setup it yields that for any measure ν on \mathbb{T}_n :

$$4 \|\nu - U\|_{\text{TV}} \leq \sum_{g \neq 0} |\hat{\nu}(g)|^2. \quad (3.4)$$

This bound will be applied to P^t . The linearity of the problem yields a nice factorization of the Fourier transform, as shown by the following lemma, initially due to Ascì [8]:

Lemma 3.1. *For all $t \geq 1$, for all $\rho \in \mathbb{T}_n$,*

$$\widehat{P^t}(\rho) = \prod_{j=0}^{t-1} \hat{\mu}((A^\top)^j \rho). \quad (3.5)$$

Proof. Applying the definition of the Fourier transform, for all $t \geq 1$,

$$\begin{aligned}\widehat{P^t}(\rho) &= \sum_{\substack{x \in \mathbb{T}_n \\ y \in \text{supp } \mu}} P^{t-1}(x) \mu(y) e^{2i\pi \langle Ax+y, \rho \rangle / n} \\ &= \sum_{x \in \mathbb{T}_n} P^{t-1}(x) e^{2i\pi \langle x, A^\top \rho \rangle / n} \sum_{y \in \text{supp } \mu} \mu(y) e^{2i\pi \langle y, \rho \rangle / n} \\ &= \widehat{P^{t-1}}(A^\top \rho) \widehat{\mu}(\rho)\end{aligned}$$

which gives the result by induction, since $\widehat{P^0} = 1$. \square

3.3.2 Expansiveness

Let $\mathbb{T}^d := \mathbb{R}^d / \mathbb{Z}^d$. A induces a bijective map on \mathbb{T}^d , which is identified with A .

The factorization of the Fourier transform provided by Lemma 3.1 motivates the study of the deterministic dynamical system $x \mapsto A^\top x$ on \mathbb{T}^d . In the next section, we will see that by partitioning \mathbb{T}^d into a finite number of "rectangles", we can represent any point by the sequence of rectangles its trajectory passes through. A key ingredient to make this representation work is the expansiveness property: the orbits of two distinct points must separate from each other at exponential rate. This in turn will ensure that provided rectangles of the partition are small enough, two distinct points cannot have their trajectories going through the exact same set of rectangles, hence representations are well-defined. Note that from Lemma 3.1 the results of the two next sections will be applied later on with A^\top .

Let us introduce the quotient metric on \mathbb{T}^d . Let p be the natural projection of \mathbb{R}^d onto \mathbb{T}^d . Given a norm $\|\cdot\|$ on \mathbb{R}^d , the quotient metric d , defined as

$$d(x, y) = \inf \{ \|\hat{x} - \hat{y}\|, \hat{x} \in p^{-1}(x), \hat{y} \in p^{-1}(y) \}$$

makes \mathbb{T}^d a compact metric space.

Since A has no eigenvalue of modulus one, by factorizing the characteristic polynomial of A into a product of irreducible factors and regrouping them, one can write it as the product of two real polynomials P_u, P_s , the eigenvalues of which all have modulus strictly above or below 1 respectively. Then P_u, P_s being coprime, it is a standard result in linear algebra that $\ker P_u P_s(A)$ is the direct sum of $\ker P_u(A)$ and $\ker P_s(A)$. However by Cayley Hamilton theorem $\ker P_u P_s(A) = \mathbb{R}^d$. Thus we obtain the following decomposition which is the very heart of hyperbolic dynamics.

Definition 3.1. \mathbb{R}^d can be decomposed as the direct sum of two subspaces E_s and E_u , invariant by A such that the restriction of A to each of these subspaces has eigenvalues of modulus $|\lambda| < 1$ and $|\lambda| > 1$ respectively. E_s , resp. E_u is called the stable subspace, resp. unstable subspace.

Lemma 3.2. *There exists a norm $\|\cdot\|$ on \mathbb{R}^d such that for all $x \in \mathbb{R}^d$ decomposing as $x = x_s + x_u$ with $x_s \in E_s$ and $x_u \in E_u$, $\|x\| = \max(\|x_s\|, \|x_u\|)$ and $\|A|_{E_s}\| < 1$, $\|A^{-1}|_{E_u}\| < 1$, where the norms considered are the operator norms induced by $\|\cdot\|$.*

Proof. Start from any norm $\|\cdot\|$ on \mathbb{R}^d . Since all the eigenvalues of $A|_{E_s}$ have modulus strictly smaller than 1, there exists an integer $l \geq 1$ such that $\|A|_{E_s}^l\| < 1$. This can be proved with the Jordan's normal form of the matrix. Similarly there exists $l' \geq 1$ such that $\|A|_{E_u}^{-l'}\| < 1$. Taking the maximum of l, l' we can suppose $l = l'$. Setting

$$\|x\|' := \max \left\{ \sum_{k=0}^{l-1} \|A|_{E_s}^k x_s\|, \sum_{k=0}^{l-1} \|A|_{E_u}^{-k} x_u\| \right\}$$

for all $x = x_s + x_u$ yields an adapted norm $\|\cdot\|'$. \square

A norm satisfying the properties of the lemma will be called adapted to A . From now on, we consider such an adapted norm $\|\cdot\|$ and the associated quotient metric d .

Finally let

$$\lambda := \max(\|A|_{E_s}\|, \|A^{-1}|_{E_u}\|) < 1 \quad (3.6)$$

Proposition 3.2. *Let $\epsilon > 0$ small enough so that p is injective on $B(0, \epsilon \|A\|)$. Consider $x, y \in \mathbb{T}^d$ and define*

$$K_+ := \sup\{k \geq 0 \mid \forall l \leq k, d(A^l x, A^l y) < \epsilon\}.$$

where by convention the supremum is taken as $-\infty$ if the corresponding set is empty, that is if $d(x, y) \geq \epsilon$. Otherwise let v be the unique representative of $x - y$ in $B(0, \epsilon)$ and decompose it as $v = v_s + v_u$ with $v_s \in E_s$ and $v_u \in E_u$. Then

$$\|v_u\| < \lambda^{K_+} \epsilon \quad (3.7)$$

In particular if $d(A^k x, A^k y) < \epsilon$ for all $k \geq 0$, then $y \in W_\epsilon^s(x)$.

The previous proposition is only concerned with future trajectories. Applying with A^{-1} in place of A , it yields a dual statement on the past trajectories. Taken simultaneously with the forward statement, this yields a "two-sided" result for toral automorphisms. The last statement is what is traditionally referred to as the expansiveness property of toral automorphisms.

Corollary 3.1. *Let $\epsilon > 0$ so that p is injective on $B(0, \epsilon \max(\|A\|, \|A^{-1}\|))$. Consider*

$$K := \sup\{k \geq 0 \mid \forall l \in [-k, k] d(A^l x, A^l y) < \epsilon\}.$$

Then

$$d(x, y) < \epsilon \lambda^K. \quad (3.8)$$

In particular if $d(A^k x, A^k y) < \epsilon$ for all $k \in \mathbb{Z}$ then $x = y$.

The smallest parameter ϵ satisfying Equation (3.8) is called the expansiveness constant of A and will be denoted ϵ_c .

Proof of Proposition 3.2. Assume without loss of generality that $d(x, y) < \epsilon$ to have $K_+ \geq 0$. Then for $k = 1$, $\|A^k v\| < \epsilon'$ so $A^k v$ is the unique representative in $B(0, \epsilon')$ of $A^k(x - y) \in \mathbb{T}^d$. Thus if $d(A^k x, A^k y) < \epsilon$ we deduce that $A^k v \in B(0, \epsilon)$. Iterating this argument, we get by induction that $A^k v \in B(0, \epsilon)$ for all $0 \leq k \leq K_+$.

Now since the norm is adapted to A , $\|v\| = \max(\|v_s\|, \|v_u\|)$ and thus $A^{K_+} v_u$ belongs to $B(0, \epsilon)$. Then

$$\|v_u\| = \|A^{-K_+} A^{K_+} v_u\| < \|A^{-K_+}|_{E_u}\| \epsilon = \lambda^{K_+} \epsilon.$$

□

3.3.3 Symbolic Representations

The main ingredient in [49] that allows to get a mixing time of $O(\log n \log \log n)$ is the use of the binary decomposition to study the orbits of elements of the form ρ/n under the doubling map $x \mapsto 2x$. The theory of symbolic representations and Markov partitions provides a generalization of the binary decomposition to the multidimensional setting and beyond. By making the dynamical system $x \mapsto Ax$ essentially conjugate to a shift, this will permit the same kind of analysis as for the doubling map.

The basic idea is as follows: given a finite partition $\mathcal{P} = \{P_1, \dots, P_m\}$ of \mathbb{T}^d , one may try to code an element $x \in \mathbb{T}^d$ by the subsets of \mathcal{P} containing the orbit of x . For instance if for all $k \in \mathbb{Z}$, $A^k x \in P_{x_k}$, then one may try to identify x with the sequence $(x_k)_{k \in \mathbb{Z}}$.

There is a main obstacle to this: without restriction on \mathcal{P} , two distinct elements may be coded by the same sequence. In the case of invertible hyperbolic automorphisms on the torus, it is possible to design well-suited families, called Markov partitions, to avoid this issue.

A Markov partition is not properly a partition: its sets, called rectangles, are closed sets with disjoint interiors but may have intersecting boundaries. Rectangles have additional properties but those are not necessary for the proofs of Theorems 3.1 and 3.3, so the definition is postponed to Section 3.6. The diameter of a Markov partition is the maximal diameter of its rectangles.

Given a Markov partition $\mathcal{R} = \{R_1, \dots, R_m\}$, define the adjacency matrix $\mathcal{A} = \mathcal{A}(\mathcal{R})$ as the $m \times m$ matrix with entries

$$\mathcal{A}_{ij} = \begin{cases} 1 & \text{if } \text{int}(R_i) \cap A^{-1}(\text{int}(R_j)) \neq \emptyset \\ 0 & \text{otherwise.} \end{cases} \quad (3.9)$$

Given an $m \times m$ adjacency matrix \mathcal{A} , define

$$\Omega_{\mathcal{A}} := \{\omega \in [m]^{\mathbb{Z}} : \forall k \in \mathbb{Z}, \mathcal{A}_{\omega_k, \omega_{k+1}} = 1\}. \quad (3.10)$$

Let $\theta : (\omega_k)_{k \in \mathbb{Z}} \mapsto (\omega_{k+1})_{k \in \mathbb{Z}}$ be the shift operator on $\Omega_{\mathcal{A}}$. $[m]$ denotes the set of integers from 1 to m and $\Omega_{\mathcal{A}}$ is given the product topology.

The following results are [37][Thm. 3.18, Thm. 3.12] applied to toral automorphisms, and will be proved in Section 3.6.

Proposition 3.3. *For any hyperbolic matrix $A \in \mathrm{GL}_d(\mathbb{Z})$, there exist Markov partitions of arbitrarily small diameter.*

Proposition 3.4. *Let \mathcal{R} be a Markov partition and $\mathcal{A} = \mathcal{A}(\mathcal{R})$ be the associated adjacency matrix. Provided the diameter of \mathcal{R} is small enough, for all $\omega \in \Omega_{\mathcal{A}}$, the intersection $\bigcap_{k \in \mathbb{Z}} A^{-k}(R_{\omega_k})$ is reduced to a single point, denoted $\pi(\omega)$. The map $\pi : \Omega_{\mathcal{A}} \rightarrow \mathbb{T}^d$ is continuous, finite-to-one, surjective and satisfies $\pi \circ \theta = A \circ \pi$.*

Remark 3.6. In addition to being finite-to-one, the set of points which admit several coding sequences is actually negligible, in the sense that for all ergodic θ -invariant measure with support $\Omega_{\mathcal{A}}$, the set of points which have two images or more under π has null measure; see [21][Thm 12.4].

A pair $(\Omega_{\mathcal{A}}, \pi)$, as given in Proposition 3.4, or simply the map π , will be called a symbolic representation of the dynamical system $x \mapsto Ax$. Similarly for $\xi \in \mathbb{T}^d$, a sequence $\omega \in \pi^{-1}(\xi)$ will be called a symbolic representation of ξ . $\pi^{-1}(\xi)$ denotes here the inverse image of the singleton $\{\xi\}$. We omit braces when considering inverse image of singletons.

We will finally need the two following results about Markov partitions which form Lemma 1 and Corollary 11 of [36]. They are proved in Section 3.6.

Proposition 3.5. (i) *If x belongs to a rectangle $R \in \mathcal{R}$, then there exists $\omega \in \Omega$ such that $x = \pi(\omega)$ and $\omega_0 = R$.*

(ii) *If $\pi(\omega)$ is a periodic point, then ω is periodic.*

3.3.4 Proof of Theorem 3.1

Let $\rho \in \mathbb{T}_n$ and $t \geq 1$. From the formula of the Fourier transform, multiply Equation (3.5) by its conjugate to obtain

$$\left| \widehat{P^t}(\rho) \right|^2 = \prod_{j=0}^{t-1} \left| \widehat{\mu}((A^T)^j \rho) \right|^2 \quad (3.11)$$

$$= \prod_{j=0}^{t-1} \left(\sum_{x, y \in \mathrm{supp} \mu} \mu(x) \mu(y) e^{2i\pi(A^j(x-y), \rho)/n} \right). \quad (3.12)$$

For $k \geq 0$ and $\xi \in \mathbb{T}^d$, define

$$f_k(\xi) := \sum_{x, y \in \text{supp } \mu} \mu(x)\mu(y)e^{2i\pi\langle A^k(x-y), \xi \rangle}$$

so that $|\widehat{P^t}(\rho)|^2 = \prod_{k=0}^{t-1} f_k(\rho/n)$. For all $k, l \geq 0$, $0 \leq f_l \leq 1$ and $f_k \circ (A^\top)^l = f_{k+l}$ hence for $k \in [0, d] = \{0, 1, \dots, d\}$,

$$\begin{aligned} |\widehat{P^{t+d}}(\rho)|^2 &= \prod_{l=0}^{k-1} f_l(\rho/n) \prod_{j=k}^{t+d-1} f_j(\rho/n) \\ &\leq \prod_{j=0}^{t+d-k-1} f_{j+k}(\rho/n) \\ &\leq \prod_{j=0}^{t+d-k-1} f_k((A^\top)^j \rho/n) \\ &\leq \prod_{j=0}^{t-1} f_k((A^\top)^j \rho/n). \end{aligned}$$

In particular, if we set

$$f := \min_{k \in [d]} f_k \tag{3.13}$$

then we have for all $t \geq 0$,

$$|\widehat{P^{t+d}}(\rho)|^2 \leq \prod_{j=0}^{t-1} f((A^\top)^j \rho/n). \tag{3.14}$$

Furthermore, the f_k are continuous and thus so is f .

Consider now

$$H := \langle \bigcup_{k \geq 0} A^k(\text{supp } \mu - \text{supp } \mu) \rangle \tag{3.15}$$

the smallest subgroup of \mathbb{Z}^d that contains $\text{supp } \mu - \text{supp } \mu$ and is invariant by A . By Cayley-Hamilton's theorem, we can in fact restrict the powers of A to have exponent less than d .

Let u_1, \dots, u_d be generators of H , although at this stage one does not require H to have rank d . They can be chosen of the form $u_i = A^{k_i}(x_i - y_i)$ with $k_i \leq d$ and $x_i, y_i \in \text{supp } \mu$. Then let

$$\alpha := \min_i \bigcup \{ \mu(x_i), \mu(y_i) \} > 0$$

and

$$W = \{ \xi \in \mathbb{T}^d : \forall h \in H, \langle h, \xi \rangle \in \mathbb{Z} \}.$$

W is a closed set of \mathbb{T}^d and is invariant by A^\top since H is invariant by A . If $\xi \notin W$, then one can find $h \in H$ such that $\langle h, \xi \rangle \notin \mathbb{Z}$. However h being a linear combination

with integer coefficients of the $A^{k_i}(x_i - y_i)$, we can directly assume that $h = A^{k_i}(x_i - y_i)$. Then we deduce that

$$\begin{aligned} f(\xi) &\leq 1 - 2\mu(x_i)\mu(y_i) + 2\mu(x_i)\mu(y_i) \cos 2\pi \langle A^{k_i}(x_i - y_i), \xi \rangle \\ &\leq 1 - 2\alpha^2 |1 - \cos 2\pi \langle A^{k_i}(x_i - y_i), \xi \rangle| \\ &< 1. \end{aligned} \tag{3.16}$$

By continuity of f we get for all $\eta > 0$

$$\sup_{\xi: d(\xi, W) \geq \eta} f(\xi) \leq \gamma \tag{3.17}$$

for some $\gamma = \gamma(\eta) < 1$. Thus our goal is now to prove that the points on the orbit of ρ/n spend a considerable amount of time away from the “bad set” W .

For this, we will use the symbolic representation of the orbits with an explicit description of the set W . So far Fourier transforms were expressed in the canonical basis but this choice is not well-suited for expressing the elements of H . On the other hand by considering a basis adapted to H , H becomes generated by multiple of the basis vectors the set W takes a very simple form. We use the following well known result (see [110, Chap. III Thm. 7.8]) which can be deduced from the Smith normal form of a matrix. We recall a basis of a subgroup $H \subseteq \mathbb{Z}^d$ is a minimal generating family of H . The terminology comes from the fact that subgroups of \mathbb{Z}^d are free modules over the ring \mathbb{Z} .

Lemma 3.3. *Let H be a subgroup of \mathbb{Z}^d . Then there exists a basis $(u_i)_{i=1}^d$ of \mathbb{Z}^d and positive integers a_1, \dots, a_k , such that a_i divides a_{i+1} for all $i = 1, \dots, k-1$ and $a_1 u_1, \dots, a_k u_k$, forms a basis of H . The family $(a_i)_{i=1}^k$ is uniquely determined by the previous conditions.*

Change of basis By the previous lemma, there exists a basis $(u_i)_{i \in [d]}$ of \mathbb{Z}^d and a family of positive integers $(a_i)_{i \in [d]}$ such that the subgroup H has basis $(a_i u_i)_{i=1}^d$. Let Q be the matrix formed from the vectors u_i , expressed in the canonical basis, so that $Qe_i = u_i$ for all $i \in [d]$.

We can now express everything in the basis $(u_i)_{i \in [d]}$: this comes down to replace the map A by $Q^{-1}AQ$ which has the same desired properties as A : it is in $\text{GL}_d(\mathbb{Z})$ and hyperbolic, while representation ρ has to be pulled-back by Q , ie it has to be replaced by $\rho \circ Q$. Notice now that the upper bound lemma (3.4) involves summing over all non-trivial representations, which form a set left invariant by the pull-back operation, so we can simply forget about this pull-back.

Hence, we can now suppose that all vectors expressions are given in a basis adapted to H , so H is generated by $(a_i e_i)_{i \in [d]}$. Then one easily obtains from the definition of W that

$$W = \left\{ \left(\frac{k_i}{a_i} \right)_{i=1}^d, k_i \in [0, a_i - 1] \forall i \in [d] \right\}. \quad (3.18)$$

Lemma 3.4. *There exists constants $c_1, c_2 > 0$ such that the following holds. Let $\rho \in \mathbb{T}_n \setminus \{0\}$ and y be either a point of W or $y = \rho'/n$ with $\rho' \in \mathbb{T}_n \setminus \{0\}$ distinct from ρ . Then either $d(\rho/n, y) \geq \epsilon$ or, letting v be the unique representative of $\rho/n - y$ in $B(0, \epsilon)$,*

$$\|v_u\| \geq \frac{c_1}{n^{c_2}}. \quad (3.19)$$

where $v = v_s + v_u$ is the decomposition of v into stable and unstable component.

Proof. Consider the case of $y = w \in W$. Then there exist integers $(k_i)_{i=1}^d$ such that $w = \left(\frac{k_i}{a_i} \right)_{i=1}^d$ while $\rho = (\rho_i)_{i \in [d]} \in \mathbb{T}_n \setminus \{0\}$. Suppose that $d(\rho/n, w) < \epsilon$. Assuming that n is coprime with the a_i , it is impossible to have $\rho/n = w$ and furthermore any non-zero coordinate has absolute value

$$|(\rho/n - w)_i| = \left| \frac{\rho_i a_i - k_i n}{a_i n} \right| \geq \frac{1}{a_i n}.$$

By equivalence of norms we deduce there exists a constant $c_1 = c_1(A, a_i) \in (0, 1)$ such that

$$\|v\| = \|\rho/n - w\| \geq \frac{c_1}{n}.$$

Since the norm is adapted, this implies that $\|v_u\| \geq c_1/n$, in which case the claimed result holds, or $\|v_s\| \geq c_1/n$. In that case, one still has $\|v_s\| \leq \epsilon$. Thus for $k > \log n / \log(\lambda^{-1})$ and n large enough,

$$\|A^k v_s\| \leq \lambda^k \|v_s\| < \frac{c_1}{n}.$$

Now if $\|A^l v\| < \epsilon$ for all $0 \leq l \leq k$, then $A^k v$ is the representative in $B(0, \epsilon)$ of $A^k \rho/n - A^k w$. Since the sets $\{\rho'/n, \rho' \in \mathbb{T}_n \setminus \{0\}\}$ and W are left invariant by A , $A^k v$ must also satisfy:

$$\|A^k v\| \geq \frac{c_1}{n}.$$

From the choice of k we deduce that $\|A^k v_u\| \geq c_1/n$. Then $\|A^k v_u\| \leq \|A\|^k \|v_u\|$ yields

$$\|v_u\| \geq \frac{c_1}{\|A\|^k n} = \frac{c_1}{n^{1+\log\|A\|/\log\lambda^{-1}}}.$$

Finally if $A^l v$ leaves the ball $B(0, \epsilon)$ for some $0 \leq l \leq k$, it can only expand in the unstable direction so necessarily $\|A^l v_u\| \geq \epsilon \geq c_2/n$ for n large enough and the same conclusion holds.

The case $y = \rho'/n$ is proved similarly. \square

Let us now turn to symbolic representations. Consider a Markov partition \mathcal{R} associated to A^\top , of diameter δ determined later on, as given by Proposition 3.3. Let $m := |\mathcal{R}|$ be the number of rectangles and (Ω, π) the symbolic representation given by the Markov partition (Proposition 3.4). Consider the set \mathcal{R}_0 of rectangles which contain a point of W and set $\mathcal{R}_1 := \mathcal{R} \setminus \mathcal{R}_0$. We define

$$\delta_0 := \min \left(\epsilon_c, \frac{\min_{x \neq y \in W} d(x, y)}{1 + \|A^\top\|} \right),$$

which is positive by finiteness of W . From now on, we identify $[m]$ with the set of rectangles \mathcal{R} . In particular, given a sequence $\omega \in \Omega$, we will write $\omega_k \in \mathcal{R}_0$ to indicate that $R_{\omega_k} \in \mathcal{R}_0$. Similarly, we may use the words letters and rectangles interchangeably.

Lemma 3.5. *Suppose the diameter of the partition satisfies $\delta < \delta_0$. Then*

(i) *for all $i \in \mathcal{R}_0$, there exists a unique $j \in \mathcal{R}_0$ such that $\mathcal{A}(i, j) = 1$;*

(ii)

$$W = \{ \pi(\omega) : \forall k \in \mathbb{Z}, \omega_k \in \mathcal{R}_0 \}.$$

Proof. We start proving (ii).

One inclusion in (ii) is easy. For all $w \in W$ by surjectivity of π there exists $\tau \in \Omega$ such that $w = \pi(\tau)$. Then by definition of π , $(A^\top)^k w \in R_{\tau_k}$ for all $k \in \mathbb{Z}$. However W is invariant by A^\top , so by definition of \mathcal{R}_0 $\tau_k \in \mathcal{R}_0$. This holds for every sequence $\tau \in \pi^{-1}(w)$.

Conversely, suppose $\tau \in \Omega$ is such that $\tau_k \in \mathcal{R}_0$ for all $k \geq 0$ and let $x = \pi(\tau)$. Then for all $k \geq 0$ there exists $w_k \in R_{\tau_k} \cap W$ at distance at most δ from $(A^\top)^k x$. The sequence $(w_k)_{k \geq 0}$ is a $\delta(1 + \|A^\top\|)$ -pseudo orbit, in the sense

$$\begin{aligned} \forall k \geq 0 \quad d(w_{k+1}, A^\top w_k) &\leq d(w_{k+1}, (A^\top)^{k+1} x) + d(A^\top (A^\top)^k x, A^\top w_k) \\ &\leq (1 + \|A^\top\|) \delta. \end{aligned}$$

Therefore if $\delta(1 + \|A^\top\|) < \min_{x \neq y \in W} d(x, y)$, the sequence $(w_k)_{k \in \mathbb{Z}}$ is necessary the orbit of a point $w \in W$. Then we obtain $d((A^\top)^k x, (A^\top)^k w) \leq \delta$ for all $k \in \mathbb{Z}$. If $\delta \leq \epsilon_c$, Proposition 3.2 implies $x = w$. Hence $x \in W$.

We now prove (i). Consider $i \in \mathcal{R}_0$. By definition R_i must contain a point $w \in W$. From Proposition 3.5 (i), w admits a symbolic representation τ such that $\tau_0 = i$. Let $j := \tau_1$. Then $\mathcal{A}(i, j) = 1$ and R_j contains $A^\top w \in W$ so $j \in \mathcal{R}_0$. This proves existence.

Suppose now $\mathcal{A}_{i, j'} = 1$ with $j' \in \mathcal{R}_0$. By the existence results one can find for all $l \geq 2$ $\tau^{(l)} \in \Omega$ such that $\tau_0^{(l)} = i, \tau_1^{(l)} = j'$ and $\tau_k^{(l)} \in \mathcal{R}_0$ for all $k \leq l$. By

compactness of Ω , $(\tau^{(l)})_l$ admits a subsequence converging towards $\tau' \in \Omega$, which satisfies $\tau'_0 = i, \tau'_1 = j'$ and $\tau'_k \in \mathcal{R}_0$ for all $k \geq 0$. By (ii), $\pi(\tau') =: w'$ is a point of W which is also in the rectangle R_i . However notice that $\delta < \delta_0 < \min_{x \neq y \in W} d(x, y)$ so every rectangle of the Markov partition contains at most one point of W and necessarily $w' = w$. We thus proved that τ, τ' , or any other sequence in $\mathcal{R}_0^{\mathbb{Z}}$ starting with i , represent the same element w . A last observation to make is that by Proposition 3.5 (ii), τ, τ' are periodic sequences. Thus τ, τ' are the concatenation infinitely many times of a same block of finite size, say B for τ and B' for τ' . Substituting any block B by a block B' and vice versa, one obtains other representations of w . Since π is finite-to-one, there can only be finitely many such representations, which implies $B = B'$ and in particular $j = j'$. Hence uniqueness. \square

Given an integer $k \geq 1$, $\xi \in \mathbb{T}^d$ and a symbolic representation $\omega = (\omega_i)_{i \geq 0} \in \pi^{-1}(\xi)$ of ξ . we decompose ω into blocks of k letters: $B_i(\omega) := \omega_{(i-1)k} \cdots \omega_{ik-1}$, which we call k -blocks. We set $B_i(\xi) := \{B_i(\omega), \omega \in \pi^{-1}(\xi)\}$ to be the set of all blocks that are the i -th block of some symbolic representation of ξ . When considering $\xi = \rho/n$, we may simply write $B_i(\rho)$ instead of $B_i(\rho/n)$.

Lemma 3.6. *Let \mathcal{R} be a Markov partition of diameter $\delta < \delta_0$, and some constant c . For all $n \geq 1$ let*

$$k = 1 + \lceil \frac{c_2}{\log \lambda^{-1}} \log n \rceil. \quad (3.20)$$

and decompose symbolic representations into blocks of k letters. For large enough n :

- (i) *For all $\rho \in \mathbb{T}_n \setminus \{0\}$, every block of $B_1(\rho)$ contains at least one letter in \mathcal{R}_1 .*
- (ii) *The sets of blocks $(B_1(\rho)_{\rho \in \mathbb{T}_n})$ are all disjoint.*
- (iii) *The family $(B_i(\rho)_{\rho \in \mathbb{T}_n})$ is independent of $i \in \mathbb{Z}$.*

Proof. (i) Consider any symbolic representation ω of $\rho \neq 0$. If $\omega_0, \dots, \omega_k \in \mathcal{R}_0$, from the proof of Lemma 3.5 (i), there exists $\omega' \in \Omega$ with $\omega'_l = \omega_l$ for all $0 \leq l \leq k$ and $\omega'_l \in \mathcal{R}_0$ for all $l \in \mathbb{Z}$. Lemma 3.5 (ii) tells us that $\pi(\omega') =: w$ is a point of W . w is at distance at most δ from ρ/n so we can consider the unique representative v of $\rho/n - w$ in $B(0, \epsilon_c)$. Since ω, ω' coincide up to rank k , (3.7) yields $\|v_u\| \leq \epsilon_c \lambda^k$, which contradicts (3.19) for $k \geq c_2 / \log \lambda^{-1} \log n$ and n large enough. Hence for k as in (3.20) $\omega_l \in \mathcal{R}_1$ for some $0 \leq l \leq k-1$.

- (ii) The proof is the same as (i), since if $\rho, \rho' \in \mathbb{T}_n$ are distinct, $\rho/n - \rho'/n$ has unstable component lower bounded by $\min(\epsilon_c, c_1 n^{-1-c_2})$ by Lemma 3.4.

- (iii) The matrix A^\top is a bijection on $\mathbb{T}_n \setminus \{0\}$, so the shift by k letters on Ω induces a permutation on the blocks.

□

The rest of the proof is now similar to the original argument of Chung, Diaconis and Graham [49]. Let k be as in (3.20).

By definition, rectangles of \mathcal{R}_1 are closed sets which contain no point of W , and \mathcal{R}_1, W are both finite sets, thus $\eta := \min_{R \in \mathcal{R}_1} d(R, W)$ must be positive. By equation (3.16) and the continuity of the function f (3.13), there exists $\gamma = \gamma(\eta, \alpha) \in (0, 1)$ such that f is bounded by γ on the compact set $\{\xi \in \mathbb{T}_d, d(x, W) \geq \eta\}$. In particular $f(\rho/n) \leq \gamma$ if ρ/n is contained in a rectangle of \mathcal{R}_1 .

Combining this observation with Lemma 3.6 and equation (3.14), we deduce that for all $\rho \in \mathbb{T}_n \setminus \{0\}$, for all $r \geq 0$

$$\begin{aligned} \left| \widehat{Prk'+d}(\rho) \right|^2 &\leq \gamma^{g(\rho/n)} \\ &\leq \prod_{i=1}^r \gamma^{g(B_i(\rho/n))} \end{aligned}$$

where $g(\rho/n)$ is the number of letters in \mathcal{R}_1 appearing in the first rk letters of any symbolic representation of ρ/n , and $g(B_i(\rho/n))$ is the maximal number of letters in \mathcal{R}_1 appearing in the i -th k' -block of some symbolic representation of ρ/n .

As in [49], we then make use of the following interchange lemma in order to regroup similar blocks when summing over $\rho \neq 0$: for all $a \leq a', b \leq b'$,

$$\gamma^{a+b'} + \gamma^{a'+b} \leq \gamma^{a+b} + \gamma^{a'+b'} \quad (3.21)$$

Using point (iii) of Lemma 3.6, we deduce the bound

$$\sum_{\rho \neq 0} \left| \widehat{Prk'+d}(\rho) \right|^2 \leq \sum_{\rho \neq 0} \gamma^{rg(B_1(\rho/n))}$$

Then, using point (i) and (ii) of Lemma 3.6, this sum can be bounded by the sum over all blocks of length k with at least one letter in \mathcal{R}_1 . Let $m_0 := |\mathcal{R}_0|$, $m_1 = |\mathcal{R}_1| = m - m_0$. Once the positions of the rectangles in \mathcal{R}_1 have been determined, there are m_1^j choices of such rectangles. On the other hand, Lemma 3.5 (i) shows that the sequences containing only rectangles in \mathcal{R}_0 are completely determined by their first letter. Thus the total

number of blocks with j letters in \mathcal{R}_1 is upper bounded by $\binom{k}{j}(m_1 m_0)^j$. Hence

$$\begin{aligned} \sum_{\rho \neq 0} \left| \widehat{P^{rk+d}}(\rho) \right|^2 &\leq \sum_{\rho \neq 0} \gamma^{rg(B_1(\rho/n))} \\ &\leq \sum_{j=1}^k \binom{k}{j} (m_1 m_0)^j \gamma^{rj} \\ &= \left((1 + m_0 m_1 \gamma^r)^k - 1 \right) \\ &\leq \left(e^{k m_0 m_1 \gamma^r} - 1 \right). \end{aligned}$$

Finally use the upper bound lemma (3.4) to get

$$\|P^{rk+d}(0, \cdot) - U\|_{\text{TV}} \leq \frac{1}{4} \left(e^{m_0 m_1 k \gamma^r} - 1 \right). \quad (3.22)$$

Thus for all $s > 0$ and $r \geq \frac{\log(m_0 m_1 k)}{\log(\gamma^{-1})} + s$ we obtain

$$\|P^{rk'+d}(0, \cdot) - U\|_{\text{TV}} \leq \frac{1}{4} \left(e^{\gamma^s} - 1 \right).$$

Since $k = O(\log n)$, this yields the result.

3.4 Necessary condition for irreducibility and aperiodicity

In this section we prove that the condition on the subgroup H given in Theorems 3.1 and 3.3 is necessary to ensure irreducibility and aperiodicity of the random walk. Proposition 3.6 and its proof should be reminiscent of the following result for convolution random walks on groups.

For a random walk on a finite group Γ defined as the product of iid random increments of law μ , irreducibility is equivalent to the fact that the subgroup generated by $\text{supp } \mu$ is the whole group Γ . Furthermore if it is irreducible, it is aperiodic if and only if $\text{supp } \mu$ is not contained in the coset of a proper normal subgroup, which is also equivalent to (see for example [137, p.97])

$$\bigcup_{k \geq 0} (\text{supp } \mu)^{-k} (\text{supp } \mu)^k = \Gamma,$$

where for all subset $K \subseteq \Gamma$ and $l \geq 0$, $K^l := \{x_1 \cdots x_k, \forall i x_i \in K\}$, $K^{-1} := \{x^{-1}, x \in K\}$ and $K^{-l} := (K^{-1})^l$. In what follows, additive notation is used for group operation.

Proposition 3.6. *Let H be the smallest A -invariant subgroup of \mathbb{Z}^d that contains $\text{supp } \mu - \text{supp } \mu$. Suppose it is generated by $(a_i u_i)_{i=1}^l$ for some basis $(u_i)_{i=1}^d$ of \mathbb{Z}^d and integers $(a_i)_{i=1}^l$. If the random walk X_t is irreducible and aperiodic then necessarily $l = d$ and n is coprime with all the a_i .*

Proof. The subgroup H is described by (3.15). Let P^t denote the law of X_t at time t . Consider the subgroup $N := H \bmod n$ of \mathbb{T}_n defined as the image of H under the natural projection. We claim the following:

Claim 1.

$$N = \bigcup_{k \geq 0} \text{supp } P^k - \text{supp } P^k$$

Claim 2. *If X_t is irreducible and aperiodic then $N = \mathbb{T}_n$.*

From the second claim and Lemma 3.3 we easily deduce the result. Consider a basis $(u_i)_{i=1}^d$ of \mathbb{Z}^d and integers a_1, \dots, a_l such that $(a_i u_i)_{i=1}^l$ forms a basis of H . Then N consists of the projections on \mathbb{T}_n of all linear combinations of the $a_i u_i$ with coefficients in $[0, n-1]$. In particular N has at most n^l points, which proves the necessity of $l = d$.

Suppose now $l = d$. The basis $(u_i)_{i=1}^d$ actually yields an isomorphism between N and the subgroup of \mathbb{T}_n generated by elements $(0, \dots, 0, a_i, 0, \dots, 0)$ where a_i is in position i . As a_i generates $\mathbb{Z}/n\mathbb{Z}$ if and only if n is coprime with a_i , we deduce $N = \mathbb{T}_n$ if and only if n is coprime with all the a_i , which is the desired result.

Let us now prove the claims. For all $k \geq 0$, let $\mu_k := A^k \mu$. N is by definition the subgroup generated by $\bigcup_{k \geq 0} \text{supp } \mu_k - \text{supp } \mu_k$. For $k \geq 1$, the random walk can be written

$$X_t = \sum_{i=1}^k A^{k-i} g_i \bmod n$$

with $g_i \in \text{supp } \mu$ for all $i = 1, \dots, k$, so $P^k = \mu_0 * \dots * \mu_{k-1}$, from which we deduce $\text{supp } P^k - \text{supp } P^k \subseteq N$ for all $k \geq 0$.

Conversely, notice that $0 \in \text{supp } P^k - \text{supp } P^k$ for all $k \geq 0$. Thus, given $x_k - y_k \in \text{supp } \mu_k - \text{supp } \mu_k$, we can choose $x_i = y_i \in \text{supp } \mu_i$ for $i = 0, \dots, k-1$ and write $x_k - y_k = x_k - y_k + \sum_{i \leq k-1} (x_i - y_i)$ to obtain that $\text{supp } \mu_k - \text{supp } \mu_k \subseteq \text{supp } P_k - \text{supp } P_k$. Since N is generated by the $\text{supp } \mu_k - \text{supp } \mu_k$, it suffices to prove that the set $\bigcup_{k \geq 0} \text{supp } P^k - \text{supp } P^k$ is a group.

Stability under inverse is clear, whereas stability under addition will be derived from the following observation: since $\det A = \pm 1$, A induces an automorphism of the finite group \mathbb{T}_n , a power of which is the identity. Consequently, the sequence of image measures $(\mu_k)_{k \geq 0}$ is periodic, say of period m . Now for $l \geq 0$, using the trick that $0 \in \text{supp } \mu_l - \text{supp } \mu_l$ for all $l \geq 0$, we see that $\text{supp } P^k - \text{supp } P^k \subseteq \text{supp } P^l - \text{supp } P^l$ for all $l \geq k$. We can consider in particular $l = -1 \bmod m$, so that for all $k' \geq 0$, $P^{l+k'} = P^l * \mu^0 * \dots * \mu^{k'-1} = P^l * P^{k'}$. All in all, this shows that $\text{supp } P^k - \text{supp } P^k + \text{supp } P^{k'} - \text{supp } P^{k'} \subseteq \text{supp } P^{l+k'} - \text{supp } P^{l+k'}$, which is the desired addition property.

Finally, the second claim is easily deduced from the first. Consider any sequence $(x_t)_{t \geq 0}$ such that $x_t \in \text{supp } P^t$ for all $t \geq 0$. Then by the characterization of N , for all

$t \geq 0$, X_t is included in the coset $x_t + N$. Hence if $N \neq \mathbb{T}_n$ the random walk cannot be irreducible and aperiodic. □

3.5 An upper bound for almost all n : proof of Theorem 3.3

In this section we reuse the results of Section 3.3. The proof follows the same argumentation as in [49, 95]. For all $n \geq 1$, let U_n be the uniform measure on \mathbb{T}_n and P_n^t the law at time t of the random walk on \mathbb{T}_n defined by (3.1).

Fix an integer $k \geq 2$, $t = \Omega(k)$ and a_i be defined as in Lemma 3.4. We will consider at once all integers $n \in N_k := \{c\tilde{\lambda}^{k-1} \leq m < c\tilde{\lambda}^k, m \text{ coprime with all } a_i\}$, where $c > 0$ is some constant which can be explicitated and $\tilde{\lambda} = \lambda^{-1/c_2}$. The choice of N_k (and thus of c) is made so that

$$\epsilon_c \lambda^k < \frac{c_1}{n^{c_2}} \leq \epsilon_c \lambda^{k-1}$$

whenever $n \in N_k$, c_1, c_2 being the constants appearing in Lemma 3.4. In the sequel, n is implicitly taken in this set. By the upper bound Lemma (3.4) and (3.14), we can upper bound $\sum_n \|P_n^{t+d} - U_n\|_{\text{TV}}$ by

$$S := \sum_n \sum_{\rho \in \mathbb{T}_n \setminus \{0\}} \prod_{j=0}^{t-1} f((A^\top)^j \rho/n)$$

Since n is not fixed anymore, it may happen that $\rho/n = \rho'/n'$ for distinct pairs $(\rho, n), (\rho', n')$. We regroup these terms as follows.

Given $r \in \mathbb{N}^d$ a vector with integer entries and an integer $s > 0$, define $\text{gcd}(r, s)$ as $\text{gcd}(r_1, \dots, r_d, s)$. Any vector ρ/n can be rewritten uniquely under the form r/s with $\text{gcd}(r, s) = 1$. Furthermore s is necessarily coprime with the a_i if n is. Given r and s such that $\text{gcd}(r, s) = 1$, let $M(r/s)$ be the number of pairs (ρ, n) such that $\rho/n = r/s$. Such ρ/n are obtained simply by multiplying numerators and denominators of r/s by an integer factor necessarily smaller than $c\tilde{\lambda}^k/s$, hence $M(r/s) \leq c\tilde{\lambda}^k/s$. Then rewrite the sum as

$$\begin{aligned} S &= \sum_{1 \leq s < c\tilde{\lambda}^k} \sum_{\substack{r \in \mathbb{T}_s \setminus \{0\} \\ \text{gcd}(r, s) = 1}} M(r/s) \prod_{j=0}^{t-1} f((A^\top)^j r/s) \\ &= \sum_{l=1}^k \sum_{c\tilde{\lambda}^{l-1} \leq s < c\tilde{\lambda}^l} \sum_{\substack{r \in \mathbb{T}_s \setminus \{0\} \\ \text{gcd}(r, s) = 1}} M(r/s) \prod_{j=0}^{t-1} f((A^\top)^j r/s) \\ &\leq c\tilde{\lambda}^k \sum_{l=1}^k \sum_{c\tilde{\lambda}^{l-1} \leq s < c\tilde{\lambda}^l} \sum_{\substack{r \in \mathbb{T}_s \setminus \{0\} \\ \text{gcd}(r, s) = 1}} \frac{1}{s} \prod_{j=0}^{t-1} f((A^\top)^j r/s). \end{aligned}$$

As we did for n , the integers s considered will now be implicitly assumed to be coprime with the a_i .

For every pair (r, s) the product can be bounded by $\gamma^{g(r/s)}$, where we recall $g(r/s)$ is the maximal number of rectangles in \mathcal{R}_1 appearing in the t first letters of some symbolic representation of r/s . What follows now is the same application of Lemma 3.6 as we did in Section 3.3 but with l and s in place of respectively k and n . Define for all $l \geq 1$,

$$Q_l := \{(r, s) : c\tilde{\lambda}^{l-1} \leq s < c\tilde{\lambda}^l, r \in \mathbb{T}_s \setminus \{0\}, \gcd(r, s) = \gcd(s, a_i) = 1\}.$$

Then

$$S \leq c\tilde{\lambda}^k \sum_{l=1}^k \sum_{(r,s) \in Q_l} \frac{1}{s} \gamma^{g(r/s)}.$$

Decompose the symbolic representations of r/s into blocks of $l_1 = Cl$ letters and set $t' = \lfloor t/l_1 \rfloor$. Since s is coprime with the a_i , we can apply Lemma 3.6 with s instead of n basically in the same way. The only difference concerns point (ii), as the distance between distinct points is now lower bounded by

$$d(r/s, r'/s') \geq \frac{c'}{ss'}$$

for some constant c' . Arguing as in the proof of Lemma 3.4, this implies a lower bound on the unstable component v_u of $r/s - r'/s'$, namely $\|v_u\| \geq c'/(ss')^{c_2}$ for some possibly different constant c' . The definition of Q_l is made so that $(r, s) \in Q_l$ implies $\epsilon_c \lambda^l < c_1/s^{c_2}$. Thus choosing C large enough in the definition of blocks, one has $\epsilon_c \lambda^{l_1} < c_1/(ss')^{c_2}$ so by Proposition 3.2 symbolic representations of $r/s, r'/s'$ must have distinct blocks.

Then use the interchange inequality (3.21) to obtain

$$S \leq c\tilde{\lambda}^k \sum_{l=1}^k \sum_{(r,s) \in Q_l} \frac{1}{s} \gamma^{t'g(B_1(r/s))}.$$

For $(r, s) \in Q_l$, $\frac{1}{s} \leq c^{-1}\tilde{\lambda}^{-l+1} = O(\tilde{\lambda}^{-l})$. On the other hand, we can upper bound the sum over Q_l by the sum over all blocks of length l_1 , thus

$$\begin{aligned} \sum_{(r,s) \in Q_l} \frac{1}{s} \gamma^{t'g(B_1(r/s))} &\leq \tilde{\lambda}^{-l} \sum_{j=1}^{l_1} \binom{l_1}{j} \gamma^{t'j} (m_0 m_1)^j \\ &= \tilde{\lambda}^{-l} \left((1 + m_0 m_1 \gamma^{t'})^{l_1} - 1 \right). \end{aligned}$$

We now consider two different regimes for l .

If $l \leq k/\log k$, then $t' \rightarrow \infty$ as $k \rightarrow \infty$. Now for large t' we can use Taylor's theorem to bound $(1 + m_0 m_1 \gamma^{t'})^{l_1} - 1$ by $O(l_1 \gamma^{t'})$. As $l_1 = C \log l$ this yields

$$\begin{aligned} \sum_{l \leq k/\log k} \tilde{\lambda}^{-l} \left((1 + m_0 m_1 \gamma^{t'})^{l_1} - 1 \right) &\leq O \left(\sum_{l \leq k/\log k} \tilde{\lambda}^{-l} l \gamma^{t/(Cl)} \right) \\ &\leq O \left(\sum_{l \geq 0} \tilde{\lambda}^{-l} l \gamma^{t \log k / (Ck)} \right) \\ &\leq O(\gamma^{t \log k / (Ck)}) \end{aligned}$$

so for $t \geq C'k$ with a sufficiently large constant C' , $S_1 \rightarrow 0$ as $k \rightarrow \infty$.

On the other hand, for $k/\log k < l \leq k$, choose $t \geq \tilde{C}k$ with $\tilde{C} > 0$ large enough so that $\tilde{\gamma} := \tilde{\lambda}^{-1}(1 + m_0 m_1 \gamma^{\tilde{C}})^C < 1$. Then as $t' \geq t/l \geq \tilde{C}$ we can bound

$$\begin{aligned} \sum_{k/\log k < l \leq k} \tilde{\lambda}^{-l} \left((1 + m_0 m_1 \gamma^{t'})^{l_1} - 1 \right) &\leq k \max_{k/\log k < l \leq k} \tilde{\lambda}^{-l} \left((1 + m_0 m_1 \gamma^{\tilde{C}})^{Cl} - 1 \right) \\ &\leq k \max_{k/\log k < l \leq k} \left(\tilde{\lambda}^{-1} (1 + m_0 m_1 \gamma^{\tilde{C}})^C \right)^l \\ &= O(\tilde{\gamma}^{k/\log k}). \end{aligned}$$

All in all, we have proved there exists a constant C' such that for large enough k and $t = C'k$,

$$\sum_{n \in N_k} \|P_n^{t+d} - U_n\|_{\text{TV}} \leq S \leq o(\tilde{\lambda}^k).$$

As a consequence the number of integers n in the interval N_k for which the random walk has not mixed by time $t + d$ is $o(\tilde{\lambda}^k)$.

On the other hand, for every integer $a \geq 1$, the number of multiples of a in the interval $[c\tilde{\lambda}^{k-1}, c\tilde{\lambda}^k]$ is of order $O(\tilde{\lambda}^{k-1})$. Consequently it must contain $\Omega(\tilde{\lambda}^k)$ integers coprime with all a_i , hence $|N_k| \geq \Omega(\tilde{\lambda}^k)$. Ultimately this shows that the fraction of integers in N_k for which the random walk has mixed at time $C'k$ can be made arbitrarily large by choosing k large enough.

3.6 Markov partitions

In this section we define Markov partitions and prove Propositions 3.3, 3.4 and 3.5. The first use of Markov constructions for toral automorphisms goes back to the work of Berg [23] and Adler and Weiss [1] for dimension 2. Since, many other constructions, with different degrees of explicitness, have been proposed for dynamical systems more general than automorphisms of the torus. This section is based on the general construction for Axiom A diffeomorphisms by Bowen [37]; see also [21]. For the reader interested in more explicit partitions, relating to arithmetic properties of the map A , see [153, 158].

We note that these partitions are related to the β -ary expansions used in [106], but do not apply in the same generality as the Bowen construction.

3.6.1 Hyperbolic dynamics

We identify A with the induced map on \mathbb{T}^d . Recall p denotes the natural projection $\mathbb{R}^d \rightarrow \mathbb{T}^d$, and that d is the quotient metric induced by a norm $\|\cdot\|$ on \mathbb{R}^d adapted to A .

For $x \in \mathbb{T}^d$ and $\epsilon > 0$ define

$$\begin{aligned} W^s(x) &= \{y \in \mathbb{T}^d, d(A^n x, A^n y) \rightarrow 0\} \\ W_\epsilon^s(x) &= \{y \in \mathbb{T}^d, d(A^n x, A^n y) \leq \epsilon \forall n \geq 0\} \\ W^u(x) &= \{y \in \mathbb{T}^d, d(A^{-n} x, A^{-n} y) \rightarrow 0\} \\ W_\epsilon^u(x) &= \{y \in \mathbb{T}^d, d(A^{-n} x, A^{-n} y) \leq \epsilon \forall n \geq 0\}. \end{aligned} \tag{3.23}$$

$W^s(x)$, resp. $W^u(x)$ are the stable, resp. unstable manifold going through x . $W_\epsilon^s(x)$, resp. $W_\epsilon^u(x)$ are the local stable, resp. unstable manifold going through x .

They can be described as: $W^s(x) = p(x + E_s)$, $W^u(x) = p(x + E_u)$ and $W_\epsilon^s(x) = p(B(x, \epsilon) \cap (x + E_s))$, $W_\epsilon^u(x) = p(B(x, \epsilon) \cap (x + E_u))$ where we recall $B(x, \epsilon)$ denotes the ball of radius ϵ and center x in \mathbb{R}^d . The previous definitions readily imply the following.

Lemma 3.7. *For all sufficiently small $\epsilon > 0$ such that p is injective on $B(0, \epsilon)$, if $d(x, y) \leq \epsilon$ then $W_\epsilon^s(x) \cap W_\epsilon^u(y)$ consists of a single point, denoted $[x, y]$. The map $(x, y) \mapsto [x, y]$ is continuous.*

The construction of Markov partitions that we will present is quite general, for it is essentially based on compactness arguments combined with the following property, called shadowing of orbits.

Definition 3.2. Given $\alpha > 0$, a sequence $(x_k)_{k \in \mathbb{Z}}$ is called an α -pseudo-orbit if for all $k \in \mathbb{Z}$, $d(Ax_k, x_{k+1}) < \alpha$. Given $\beta > 0$, a point $x \in \mathbb{T}^d$ β - shadows the pseudo-orbit $(x_k)_{k \in \mathbb{Z}}$ if for all $k \in \mathbb{Z}$, $d(x_k, A^k x) < \beta$.

Proposition 3.7. *For all $\beta > 0$ small enough, there exists $\alpha > 0$ such that every α -pseudo-orbit $(x_k)_k$ is β -shadowed by a unique point x of \mathbb{T}^d .*

Proof. Let $\beta > 0$. Uniqueness is provided by expansiveness (Proposition 3.2), provided $\beta \leq \epsilon_c$. For the existence, it suffices to prove the analog result in \mathbb{R}^d : if $(x_k)_{k \in \mathbb{Z}}$ an α -pseudo-orbit in \mathbb{R}^d , then it can be β -shadowed by a true orbit, provided α is small enough. Indeed, lifting a pseudo-orbit of the torus to \mathbb{R}^d , then we can project the shadowing orbit back to \mathbb{T}^d to prove the statement for the torus.

Decompose $x_k = s_k + u_k$ with $s_k \in E_s, u_k \in E_u$ for all $k \in \mathbb{Z}$. Since the metric is adapted to A , it is easily seen that the sequences $(s_k)_{k \in \mathbb{Z}}$ and $(u_k)_{k \in \mathbb{Z}}$ are both α -pseudo-orbits. Write $S = A|_{E_s}$ and $T = A^{-1}|_{E_u}$. Since $(s_k)_{k \in \mathbb{Z}}$ is an α -pseudo-orbit

$$\|S^{k+1}s_{-(k+1)} - S^k s_{-k}\| \leq \|S^k\| \alpha \leq \lambda^k \alpha$$

hence the sequence $(S^k s_{-k})_{k \in \mathbb{N}}$ is a Cauchy sequence and we can define $s := \lim_{k \rightarrow \infty} S^k s_{-k}$. Then for all $m \geq 1 - k$,

$$\begin{aligned} \|s_k - S^k s\| &\leq \sum_{i=0}^{k+m-1} \|S^i s_{k-i} - S^{i+1} s_{k-i-1}\| + \|S^{k+m}(s_{-m}) - S^k s\| \\ &\leq \sum_{i=0}^{k+m-1} \lambda^i \alpha + \lambda^k \|S^m s_{-m} - s\| \\ &\leq \frac{\alpha}{1 - \lambda}. \end{aligned}$$

by taking $m \rightarrow \infty$.

Repeating the argument with T , the sequence $(T^k u_{-k})_{k \in \mathbb{N}}$ converges to a point $t \in \mathbb{R}^d$ so that $\|u_k - T^k u\| \leq \alpha/(1 - \lambda)$. Since the metric is adapted, $x := s + u$ β -shadows the pseudo orbit $(x_k)_{k \in \mathbb{Z}}$ for $\beta = \alpha/(1 - \lambda)$. \square

3.6.2 Markov Partitions

Let $\epsilon > 0$ be sufficiently small so that the conclusion of Lemma 3.7 holds.

Definition 3.3. A set $R \subset \mathbb{T}^d$ is called a rectangle if it has diameter at most ϵ and, for all $x, y \in R$, $[x, y] \in R$. A rectangle is said to be proper if $R = \overline{\text{int}(R)}$.

Given a rectangle R and $x \in R$, let

$$W^s(x, R) := W_\epsilon^s(x) \cap R \quad W^u(x, R) := W_\epsilon^u(x) \cap R.$$

Definition 3.4. A Markov partition is a finite covering $\mathcal{R} = \{R_1, \dots, R_m\}$ of \mathbb{T}^d by proper rectangles such that

- (i) $\text{int}(R_i) \cap \text{int}(R_j) = \emptyset$ for all $i \neq j$
- (ii) for all $x \in \text{int}(R_i)$ if $Ax \in \text{int}(R_j)$, then

$$A(W^s(x, R_i)) \subseteq W^s(Ax, R_j) \quad \text{and} \quad A(W^u(x, R_i)) \supseteq W^u(Ax, R_j). \quad (3.24)$$

Notice that the second part of (ii) above is equivalent to

$$A^{-1}(W^u(x, R_i)) \subseteq W^u(A^{-1}(x), R_j).$$

It is thus a dual statement of the first part obtained by replacing A with A^{-1} , which has the effect of exchanging stable and unstable directions.

Condition (ii) is really the Markov property of the partition. It ensures that if $\text{int}R_i \cap A^{-1}\text{int}R_j \neq \emptyset$ and $\text{int}R_j \cap A^{-1}\text{int}R_k \neq \emptyset$ then $\text{int}R_i \cap A^{-1}\text{int}R_j \cap \text{int}A^{-2}R_k \neq \emptyset$: if x, y are respectively in the first and second intersection, consider $A^{-1}[y, Ax]$.

We have all the tools for the proofs. We give all the essential arguments, but only sketch the technical details. We refer to [37] for a detailed proof, which extends to the general case of Axiom A diffeomorphisms.

Proof of Proposition 3.5. (i) Let R be a rectangle and consider the compact set $U := \{\omega \in \Omega \mid \omega_0 = R\}$. Since rectangles of a Markov partition are proper and of disjoint interior, one has $\pi^{-1}(\text{int}R) \subset U$. Because π is surjective we deduce $\text{int}R \subset \pi(U)$. Finally continuity implies $\pi(U)$ is compact, hence $R \subset \pi(U)$.

(ii) Recall that π is finite-to-one. Thus if x is periodic of period k , the shift θ induces a permutation of the finite set of symbolic representations $\pi^{-1}(\{x, Ax, \dots, A^{k-1}x\})$. For the integer m such that the permutation induced by θ^m is the identity, one gets exactly that all sequences in $\pi^{-1}(x)$ are m -periodic. \square

Proof of Proposition 3.4. Given two non-empty rectangles R, S of a Markov partition, S will be called a stable, resp. unstable subrectangle of R if $S \subset R$, S is proper and for all $x \in S$, $W^s(x, S) = W^s(x, R)$, resp. $W^u(x, S) = W^u(x, R)$.

Consider a Markov partition $\mathcal{R} = \{R_1, \dots, R_m\}$, $\mathcal{A} = \mathcal{A}(R)$ and $\Omega := \Omega_{\mathcal{A}}$ the associated shift.

If $S \subseteq R_i$ is a non-empty unstable subrectangle and $\mathcal{A}_{ij} = 1$, Lemma 3.17 in [37] establishes that $A(S) \cap R_j$ is a non-empty unstable rectangle of R_j . Similarly, $A^{-1}(S) \cap R_j$ is a non-empty stable rectangle of R_i .

Let $\omega \in \Omega$. By the previous statement, $R_{\omega_0} \cap A^{-1}(R_{\omega_1})$ is a non-empty stable subrectangle of R_{ω_0} . By an immediate induction, every intersection $\bigcap_{k=0}^l A^{-k}R_{\omega_k}$ is a non-empty stable subrectangle of R_{ω_0} . By compactness, $\bigcap_{k=0}^{\infty} A^{-k}R_{\omega_k}$ is then also a non-empty stable subrectangle. Then reiterate the argument with A^{-1} to deduce that $\pi(\omega) := \bigcap_{k \in \mathbb{Z}} A^{-k}R_{\omega_k}$ is non-empty. By construction, the orbits of two points in this intersection must remain at distance ϵ from each other. If ϵ is taken smaller than the expansiveness constant, the intersection is thus a singleton.

Let $\pi(\omega)$ be the unique point of this singleton. From the construction, it is immediate that $\pi \circ \theta = A \circ \pi$. By what precedes the diameter of the intersection $\bigcap_{k=-l}^l A^{-k}R_{\omega_k}$ tends to 0 as l tends to infinity, which implies the continuity of the map π .

We prove surjectivity. Let $\partial\mathcal{R} := \bigcup_{R \in \mathcal{R}} \partial R$ denote the union of all boundaries of rectangles. Its complement $\text{int}(\mathcal{R})$ is an open dense subset of \mathbb{T}^d . If $x \in \bigcap_k A^k \text{int}(\mathcal{R})$, x

is the image of any sequence $\omega \in \Omega$ such that $A^k x \in R_{\omega_k}$ for all k . By Baire's theorem, an intersection of open dense subsets is dense, hence the image of π contains an open dense subset of \mathbb{T}^d . However $\pi(\Omega)$ has to be compact by continuity, whence $\pi(\Omega) = \mathbb{T}^d$.

Finally, let us prove that any point of \mathbb{T}^d cannot have more than m^2 pre-images under π . Suppose there exist $\omega^{(1)}, \dots, \omega^{(m^2+1)} \in \Omega$ with the same image x . Then one can find $k_1 \leq k_2$ such that the indices between k_1 and k_2 of these sequences do not coincide. However there is at most m^2 possibilities when choosing the k_1 -th and k_2 -th index, therefore there exists among these sequences a pair (ω, ω') such that $\omega_{k_1} = \omega'_{k_1}$, $\omega_{k_2} = \omega'_{k_2}$ and $\omega_l \neq \omega'_l$ for $k_1 < l < k_2$.

Now from the Markov property of the partition, the intersection $\bigcap_{k=k_1}^{k_2} A^{-k} \text{int} R_{\omega_k}$ must be non-empty. Take u in this open set and consider $\tau \in \pi^{-1}(u)$. Since u is an interior point τ and ω coincide on indexes between k_1 and k_2 .

On the other hand, since indexes k_1 and k_2 of ω and ω' coincide, one can build a new sequence τ' by replacing the block of τ between k_1 and k_2 by that of ω' , that is we define

$$\tau'_k := \begin{cases} \tau_k & \text{for } k \leq k_1 \text{ or } k \geq k_2 \\ \omega'_k & \text{if } k_1 \leq k \leq k_2 \end{cases}.$$

Consider finally the point $v := \pi(\tau')$. By construction $A^k u$ and $A^k v$ are contained in the same rectangles for all $k \leq k_1$ and $k \geq k_2$, whereas for $k_1 \leq k \leq k_2$, they are contained in the rectangles R_{ω_k} and $R_{\omega'_k}$ respectively. However the latter both contain the point $A^k x$, so $A^k u, A^k v$ are at distance at most 2η from each other by triangle inequality. Thus $d(A^k u, A^k v) \leq 2\eta$ for all k . Consequently if $2\eta \leq \epsilon_c$ expansiveness ensures that $A^l u \in W^s(A^l, R_{\omega_l})$. In the end we obtain $A^l u \in \text{int} R_{\omega_l} \cap R_{\omega'_l}$, which contradicts the fact that rectangles have disjoint interior. \square

Proof of Proposition 3.3. Let $\beta > 0$ and consider $\eta > 0$ such that all $2\eta(1 + \|A\|)$ -pseudo-orbits are β -shadowed by a true orbit (Proposition 3.7). Consider a finite cover of \mathbb{T}^d with balls $B(x_i, \eta)$, $i = 1, \dots, m$. Let \mathcal{A} be the matrix defined by $\mathcal{A}_{ij} = 1$ if and only if $Ax_i \in B(x_j, 2\eta(1 + \|A\|))$ and consider the shift $\Omega := \Omega_{\mathcal{A}}$ associated.

For all $\omega \in \Omega$, the sequence $(x_{\omega_n})_n$ is a $2\eta(1 + \|A\|)$ -pseudo-orbit and thus can be β -shadowed by points of a set $\pi(\omega) \subset \mathbb{T}^d$ (which is a singleton in the invertible case).

By construction, $A \circ \pi = \pi \circ \theta$. π is surjective: every point $x \in \mathbb{T}^d$ is an image $\pi(\omega)$ for any sequence ω such that $A^n x \in B(x_{\omega_n}, \eta)$ for all $n \in \mathbb{Z}$. These sequences are indeed in Ω , since

$$d(Ax_{\omega_n}, x_{\omega_{n+1}}) \leq d(Ax_{\omega_n}, A^{n+1}x) + d(A^{n+1}x, x_{\omega_{n+1}}) \leq (\|A\| + 1)\eta < 2(1 + \|A\|)\eta.$$

π is also continuous when giving Ω the product topology. The argument is similar to the one given in the previous proof.

For $i \in [m]$, let

$$T_i := \pi(\{\omega \in \Omega, \omega_0 = i\}).$$

By construction for all $\omega \in \Omega$ $d(\rho(\omega), x_{\omega_0}) \leq \beta$, so the sets T_i have diameter at most 2β , which can be made arbitrarily small. From the continuity of π , the sets T_i are closed sets.

Consider now the local product on Ω defined for $\omega, \omega' \in \Omega$ with $\omega_0 = \omega'_0$ by

$$[\omega, \omega']_k = \begin{cases} \omega_k & \text{if } k \geq 0 \\ \omega'_k & \text{if } k \leq 0 \end{cases}$$

Then $d(A^k \pi(\omega), A^k \pi([\omega, \omega'])) \leq 2\beta$ for all $k \geq 0$, and $d(A^k \pi(\omega), A^k \pi([\omega, \omega'])) \leq 2\beta$ for all $k \leq 0$. Thus $\pi([\omega, \omega']) \in W_{2\beta}^s(\pi(\omega)) \cap W_{2\beta}^u(\pi(\omega'))$, so that $\pi([\omega, \omega']) = [\pi(\omega), \pi(\omega')]$ by Lemma 3.7. This immediately implies that the sets T_i are rectangles.

Suppose then that $x = \pi(\omega)$, with $\omega_0 = i, \omega_1 = j$. From $\pi \circ \theta = A \circ \pi$ we deduce $W^s(x, T_i) = \pi(\{\omega' \in \Omega, \forall k \geq 0, \omega_k = \omega'_k\})$. Thus

$$\begin{aligned} AW^s(x, T_i) &= \pi(\{\omega' \in \Omega, \forall n \geq -1 \omega'_n = \omega_{n+1}\}) \\ &\subset \pi(\{\omega' \in \Omega, \forall n \geq 0 \omega'_n = \omega_{n+1}\}) \\ &= W^s(Ax, T_j), \end{aligned}$$

hence

$$A(W^s(x, T_i)) \subseteq W^s(Ax, T_j).$$

Using A^{-1} in place of A , we can prove similarly $A^{-1}(W^u(x, T_i)) \subseteq W^u(A^{-1}x, T_j)$. Therefore the rectangles T_i satisfy almost all the properties of a Markov partition, except there may not be proper or of disjoint interiors.

To get rectangles satisfying these properties, the idea is to divide the T_i further into smaller rectangles. Namely if $T_j \cap T_k \neq \emptyset$, let

$$\begin{aligned} T_{j,k}^1 &= \{x \in T_j, W^u(x, T_j) \cap T_k \neq \emptyset, W^s(x, T_j) \cap T_k \neq \emptyset\} = T_j \cap T_k \\ T_{j,k}^2 &= \{x \in T_j, W^u(x, T_j) \cap T_k \neq \emptyset, W^s(x, T_j) \cap T_k = \emptyset\} \\ T_{j,k}^3 &= \{x \in T_j, W^u(x, T_j) \cap T_k = \emptyset, W^s(x, T_j) \cap T_k \neq \emptyset\} \\ T_{j,k}^4 &= \{x \in T_j, W^u(x, T_j) \cap T_k = \emptyset, W^s(x, T_j) \cap T_k = \emptyset\} \end{aligned}$$

Then for all $x, y \in T_j$, the fact that $W^s([x, y], T_j) = W^s(x, T_j)$ and $W^u([x, y], T_j) = W^u(y, T_j)$ proves that the $T_{j,k}^i$ are rectangles. For any $x \in \mathbb{T}^d$ let

$$R(x) := \bigcap \{\text{int}(T_{j,k}^i) : T_j \cap T_k \neq \emptyset \text{ and } x \in T_{j,k}^i\}$$

and $\mathcal{R} := \{\overline{R(x)}, x \in \mathbb{T}^d\}$. It can be easily checked that $R(x) \cap R(x') \neq \emptyset$ implies $R(x) = R(x')$ so the set \mathcal{R} is finite. Therefore \mathcal{R} defines a finite covering by proper

rectangles of disjoint interiors. It remains to prove the second property of Markov partitions.

Suppose $y \in W^s(x, \overline{R(x)})$. Suppose that $x \in \text{int}(T_i)$ and $Ax \in \text{int}(T_j)$. From the property proved previously for the T_i , $Ay \in W^s(Ax, T_j)$. We will argue by contradiction to prove that $R(Ax) = R(Ay)$. If this does not hold, it implies the existence of a rectangle T_k intersecting T_j such that Ax, Ay are not in the same $T_{j,k}$. From the definition of $T_{j,k}$ we can suppose without loss of generality that $W^u(Ax, T_j) \cap T_k \neq \emptyset$ and $W^u(Ay, T_j) \cap T_k = \emptyset$. Since $AW^u(x, T_i) \supset W^u(Ax, T_j)$, we can thus find $z \in W^u(x, T_i)$ such that $Az \in T_j \cap T_k = T_{j,k}^1$. Consider l such that $z \in T_i \cap T_l = T_{i,l}^1$. Since $R(x) = R(y)$, x, y are in the same subrectangle $T_{i,l}$, so there exists a point $z' \in W^u(y, T_i) \cap T_l$. Then $z'' := [z, z'] = [z, y]$ is in $W^s(z, T_l) \cap W^u(y, T_i)$ and $Az'' = [Az, Ay] \in W^s(Az, T_k) \cap W^u(Ay, T_j)$, which contradicts $W^u(Ay, T_j) \cap T_k = \emptyset$. Hence $R(Ax) = R(Ay)$ which implies $AW^s(x, \overline{R(x)}) \subset W^s(Ax, \overline{R(Ax)})$.

□

Chapter 4

Concentration for low-degree functions on the symmetric group

This chapter is based on a work not yet published. We give a concentration inequality for functions on the symmetric group that was developed as a part of the work on the cutoff phenomenon presented in the next chapter.

4.1 Introduction

4.1.1 Main results

In [43][Prop. 1.1], Chatterjee established an exponential concentration inequality for random variables of the form

$$Z = \sum_i A_{i\sigma(i)} \quad (4.1)$$

where $A \in M_n(\mathbb{R}_+)$ and σ is a uniform permutation of n elements. Namely, for all $t \geq 0$

$$\mathbb{P}[|Z - \mathbb{E}[Z]| \geq t] \leq 2 \exp\left(\frac{-t^2}{2\|A\|(t + 2\mathbb{E}[X])}\right), \quad (4.2)$$

where $\|A\| = \max_{i,j \in [n]} A_{i,j}$. The random variable Z can be seen as arising from a linear function on the symmetric group, in the sense that it is a linear combination of indicators $\mathbb{1}_{\sigma(i)=j}$, which are the entries of the matrix representation of σ . With this point of view, it seems natural to inquire about more general polynomial functions. Of course any function on the symmetric group can be represented as a polynomial of degree n and in fact even $n-1$. Furthermore there already exist concentration results for generic functions on the symmetric group, regardless of the degree: Maurey's inequality [133] (see also Thm. 2.14 in [41]), Talagrand's inequality [156][Thm 5.1], and Proposition 4.8 in [41] are such examples. Our motivation is thus mainly to investigate whether an additional assumption of "low degree" can yield better concentration inequalities.

Some notations are necessary to state our result. Let \mathfrak{S}_n denote the symmetric group on n elements. Permutations $\sigma \in \mathfrak{S}_n$ are identified with permutation matrices S defined by $S_{ij} := \mathbb{1}_{\sigma(i)=j}$. Let us remark first that there is no unique representation of a function on \mathfrak{S}_n as a polynomial. In the following result, particular representations are considered, but some more intrinsic notion of degrees will be discussed in Remark 4.3. We can without loss of generality suppose that the constant term is zero. Furthermore, since we are restricting to permutation matrices it is enough to consider

$$\mathfrak{F} := \{ \phi : M_n(\mathbb{R}) \rightarrow \mathbb{R} \mid \forall i, j, k \in [n] : \partial_{ij} \partial_{ik} \phi \equiv \partial_{ji} \partial_{ki} \phi \equiv 0 \}, \quad (4.3)$$

where ∂_{ij} denotes the partial derivative with respect to the entry (i, j) . Functions of \mathfrak{F} are called multilinear, as the degree is at most one in each entry. In particular, these are polynomial functions on $M_n(\mathbb{R})$, which we identify with polynomials in indeterminates $X_{ij}, i, j \in [n]$. By restriction to the set of permutations, each function $\phi \in \mathfrak{F}$ induces a map on \mathfrak{S}_n . Conversely, any map on \mathfrak{S}_n , can be written as $\phi|_{\mathfrak{S}_n}$ for some $\phi \in \mathfrak{F}$ of degree at most $n - 1$, however this representation is in general not unique.

The vector space \mathfrak{F} can be decomposed as

$$\mathfrak{F} = \bigoplus_{d \geq 0} \mathfrak{F}_{=d}$$

where for each $d \geq 0$, $\mathfrak{F}_{=d}$ is the vector space of homogeneous polynomials of degree d . Let $\mathfrak{F}_{\leq d} = \bigoplus_{k=0}^d \mathfrak{F}_{=k}$ be the vector space of polynomials of degree at most d . If $d \geq 1$, a convenient way to write $\phi \in \mathfrak{F}_{=d}$ is given by Euler's theorem:

$$\phi(X) = \frac{1}{d} \sum_{i, j \in [n]} X_{ij} \partial_{ij} \phi(X).$$

In particular, evaluating ϕ at $\sigma \in \mathfrak{S}_n$ yields

$$\phi(\sigma) = \frac{1}{d} \sum_{i, j \in [n]} \partial_{i, \sigma(i)} \phi(\sigma), \quad (4.4)$$

which can be seen as a generalization of (4.1). To state our result, we need the following linear operators on \mathfrak{F} . Given an homogeneous function $\phi \in \mathfrak{F}_{=d}$, let

$$D\phi := \frac{1}{dn} \sum_{i, j \in [n]} \partial_{ij} \phi \quad U\phi(X) := \frac{1}{dn} \sum_{i, j, k, l \in [n]} X_{il} X_{kj} \partial_{ij} \partial_{kl} \phi(X). \quad (4.5)$$

Finally, for a function $\phi \in \mathfrak{F}$, let

$$\|\phi\|_{\infty} := \max_{\sigma \in \mathfrak{S}_n} |\phi(\sigma)|, \quad \|\nabla \phi\|_{\infty} := \max_{\sigma \in \mathfrak{S}_n} \max_{i, j \in [n]} |\partial_{ij} \phi(\sigma)|.$$

In this chapter, we write $\mathbb{E}[\phi] := \mathbb{E}[\phi(\sigma)]$ for the expectation with respect to a uniformly distributed $\sigma \in \mathfrak{S}_n$.

Theorem 4.1. *Let $n \geq 1$, $d \in [n]$, σ a uniform permutation of n elements and $\phi \in \mathfrak{F}_d$ a polynomial map of degree d with non-negative coefficients. Suppose there exist $C_D, C'_D, C_U \geq 0$ such that for all $k \in [0, d]$,*

$$\|D^k \phi\|_\infty \leq C_D, \quad \|\nabla D^k \phi\|_\infty \leq C'_D \quad \text{and} \quad \|D^k U \phi\|_\infty \leq C_U. \quad (4.6)$$

Then for all $t \geq 0$,

$$\mathbb{P}[\phi(\sigma) - \mathbb{E}[\phi] \geq t] \leq \exp\left(\frac{-t^2}{2(\gamma_\phi + \beta_\phi t)}\right), \quad \text{and} \quad \mathbb{P}[\phi(\sigma) - \mathbb{E}[\phi] \leq -t] \leq \exp\left(\frac{-t^2}{2\gamma_\phi}\right)$$

where

$$\begin{aligned} \beta_\phi &:= 6dC'_D \left(\log\left(\frac{4C_D n}{C'_D}\right)^+ + \frac{(2/n)(2 - e^{-2/n})}{1 - e^{-2/n}} \right) \\ \gamma_\phi &:= \frac{2\beta_\phi}{3} (2\mathbb{E}[\phi] + C_U) \end{aligned} \quad (4.7)$$

Remark 4.1. If $\phi(X) = \sum_{i,j} A_{i,j} X_{i,j}$, then $\nabla \phi \equiv A^\top$. Therefore, the condition required on $\|\nabla \phi\|_\infty$ generalizes the control needed on $\|A\|$ in (4.2). Up to the logarithmic term and multiplicative constants, Theorem 4.1 thus really aims to generalize Chatterjee's inequality to higher degrees.

Remark 4.2. As was mentioned earlier, the finiteness of \mathfrak{S}_n implies that any function can in fact be seen as a polynomial function of degree at most $n-1$, so Theorem 4.1 provides in theory a concentration inequality for any non-negative random variable defined by a uniform permutation. It is likely however that this bound becomes irrelevant when $d/n \rightarrow 0$. In the next chapter, Theorem 4.1 will be applied with $d = O(\sqrt{\log n})$.

Remark 4.3. There are intrinsic notions of degree for functions on \mathfrak{S}_n . A natural notion of degree is as follows: given $k \geq 1$, a k -coset of \mathfrak{S}_n is a subset of the form

$$E_{\substack{i_1 \dots i_k \\ j_1 \dots j_k}} := \{\sigma \in \mathfrak{S}_n \mid \forall m = 1, \dots, k \quad \sigma(i_m) = j_m\}.$$

where $\mathbf{i} = (i_1, \dots, i_k)$, $\mathbf{j} = (j_1, \dots, j_k) \in [n]^k$ are multi-indices of length k . If $k = 0$, we consider the whole set \mathfrak{S}_n to be a 0-coset. Given a function $f : \mathfrak{S}_n \rightarrow \mathbb{R}$, the degree of f can be defined as the least integer $d \geq 0$ such that f writes as a linear combination of k -coset indicator functions with $k \leq d$. With this definition the degree of f is the minimal degree of a polynomial representation of ϕ .

The previous notion of degree is also intimately related to Fourier analysis: from [69][Thm. 7], it coincides with the *Fourier degree* of f , which is the least integer $d \geq 0$ such that all Fourier coefficients of f corresponding to irreducible representations of dimension $k > d$ are zero. We not know whether the considerations of these notions of degree could lead to better concentration inequalities. The use of Fourier analysis, in

particular character theory could definitely be of help at some point in the proof but other arguments seem to require more than the sole use of the characters. See Remark 4.4.

The presence of the operators D and U is a consequence of the proof. It follows the method of exchangeable pairs used by Chatterjee in the one-dimensional case to bound the log-Laplace transform, which can be done in terms of the functions $D^k\phi, UD^k\phi$. In the degree one case, these operators are trivial and thus need not be considered, but in the more general case an induction argument on the degree seems necessary, which is the reason why one needs to bound $\|D^k\phi\|_\infty, \|\nabla D^k\phi\|_\infty, \|D^kU\phi\|_\infty$ for all $k \leq d$. These quantities seem to lack good monotonicity properties that could simplify the bounds. However if we can control the coefficients of the polynomial directly as well as the number of monomials that do not evaluate to zero, the following proposition provides some control on the constants C_D, C'_D, C_U . These are rough bounds but can be sufficient for small degree functions, which is how the theorem is intended to be used.

Proposition 4.1. *Let ϕ be a polynomial in the indeterminates $(X_{ij})_{i,j=1}^n$ of degree $d \geq 1$. Let $M(\phi)$ be the maximal coefficient of ϕ . Given $\sigma \in \mathfrak{S}_n$ let $N(\phi, \sigma)$ be the number of monomials in ϕ which are non-zero when evaluated at σ and $N(\phi) := \max_{\sigma \in \mathfrak{S}_n} N(\phi, \sigma)$. For all $k \geq 0$*

$$\|D^k\phi\|_\infty \leq 2^k M(\phi)N(\phi), \quad \|U\phi\|_\infty \leq \frac{d-1}{n} M(\phi)N(\phi).$$

Plugging the previous estimates into Theorem 4.1 yields the following corollary.

Corollary 4.1. *Let $\phi \in \mathfrak{F}_d$ be a polynomial function of degree at most $d \geq 1$. Using the notations of the previous proposition, let $A_\phi := M(\phi)N(\phi)$ and $A_{\nabla\phi} := \max_{i,j \in [n]} A_{\partial_{ij}\phi}$. For all $t \geq 0$,*

$$\mathbb{P}[|\phi(\sigma) - \mathbb{E}[\phi]| \geq t] \leq 2 \exp\left(\frac{-t^2}{2\alpha_\phi(\frac{4}{3}\mathbb{E}[\phi] + \frac{2^{d+1}(d-1)}{3n}A_\phi + t)}\right). \quad (4.8)$$

with

$$\alpha_\phi := 6d2^d A_{\nabla\phi} \left(\log\left(\frac{4A_\phi n}{A_{\nabla\phi}}\right)^+ + \frac{(2/n)(2 - e^{-2/n})}{1 - e^{-2/n}} \right)$$

4.1.2 Related works

Random variables of the form (4.1) were introduced in the "combinatorial central limit theorem" by Hoeffding [101]. Since then, they have been one of the main motivations for the developments of Stein's method, which is at the basis of the concentration inequality (4.2) and our result. Early applications of Stein's method for such random

variables are found in the works of Bolthausen and Goetze [25, 26] who obtained error bounds on the normal approximation.

Introduced by Stein in [154] to give a new proof of the classical CLT, Stein's method of exchangeable pairs rapidly became an important tool to prove limit theorems that go way beyond the setting of the CLT, we refer to Stein's paper [155] and to the survey [44] of Chatterjee and [46] of Chatterjee, Diaconis and Meckes which focuses more closely on the subject of Poisson approximation. The first concentration inequalities using Stein's method are due to Chatterjee in his PhD. Thesis [41], see also [43]. The inequality (4.2) can also be interpreted as the concentration of the uniform measure on the symmetric group. Generally, the subject of concentration for the Haar measure is considered by Chatterjee in [42], where he establishes a connection between concentration and the rate of convergence of some random walks on groups.

Organization of the chapter

We start proving Proposition 4.1 in Section 4.2 then prove the main Theorem in Section 4.3.

4.2 Control of the smoothness parameters: proof of Proposition 4.1

Let ϕ be a multilinear polynomial in the indeterminates $(X_{ij})_{i,j=1}^n$ of degree d . The operator D has the effect of replacing one indeterminate in each monomial by 1 and take the average. When doing so, some coefficients coming from different monomials of ϕ can be regrouped together. To avoid this, we can introduce new indeterminates instead of evaluating them at 1. This is formalized by the following construction: for all $m \geq 1$ write $X^{(m)} = (X_{ij}^{(m)})_{i,j=1}^n$. Let \tilde{D} be the linear operator on $\mathcal{R} = \mathbb{R}[X^{(1)}, X^{(2)}, \dots]$ defined on monomials as

$$\tilde{D} : X_{i_1 j_1}^{(1)} \dots X_{i_k j_k}^{(1)} X_{i_{k+1} j_{k+1}}^{(m_{k+1})} X_{i_d j_d}^{(m_d)} \mapsto \frac{1}{kn} \sum_{l=1}^k X_{i_1 j_1}^{(1)} \dots X_{i_{l-1} j_{l-1}}^{(1)} X_{i_l j_l}^{(m_d+1)} X_{i_{l+1} j_{l+1}}^{(m_{l+1})} \dots X_{i_d j_d}^{(m_d)},$$

where $m_1 = \dots m_k = 1 < m_{k+1} \leq \dots \leq m_d$. Let $T : \mathcal{R} \rightarrow \mathbb{R}[X_{ij}]$ be the operator that identifies $X^{(1)}$ with the X_{ij} and evaluates all other indeterminates $X^{(2)}, X^{(3)}, \dots$ at 1. Then by construction for all $l \geq 1$,

$$D^l \phi = T \tilde{D}^l \phi. \quad (4.9)$$

Extend the definitions of M, N to \mathcal{R} as follows: if $\psi \in \mathcal{R}$, let $M(\psi)$ be the maximal coefficient in absolute value of ψ and $N(\psi, \sigma)$ the number of non-zero monomials when

evaluating $X_{ij}^{(1)} = \mathbb{1}_{\sigma(i)=j}$ and indeterminates $X^{(m)}$ at 1 for all $m \geq 2$. Note these are not $M(T\psi), N(T\psi, \sigma)$: the point is precisely to distinguish monomials that would otherwise be regrouped together when applying T . Write $N(\psi) := \max_{\sigma \in \mathfrak{S}_n} N(\psi, \sigma)$. If $\psi = \sum_{k=1}^d \psi_k \in \mathcal{R}$ decomposes as the sum of functions that are homogeneous in $X^{(1)}$ of degree k , then for all $\sigma \in \mathfrak{S}_n$

$$|T\psi(\sigma)| \leq \sum_{k=1}^d M(\psi_k) N(\psi_k, \sigma) \quad (4.10)$$

On the other hand we claim that if ψ is homogeneous of degree $d \geq 1$, then

$$M(\tilde{D}\psi) \leq \frac{1}{dn} M(\psi), \quad \text{and} \quad \forall \sigma \in \mathfrak{S}_n : N(\tilde{D}\psi, \sigma) \leq 2dn N(\psi, \sigma). \quad (4.11)$$

The first inequality is obvious from the definition of \tilde{D} . For the second inequality, observe that every monomial in $\tilde{D}\phi$ comes from a unique monomial of ϕ . The issue is that when evaluating $X^{(1)}$ at σ and other indeterminates $X^{(m)}$ at 1, a non-zero contribution in $\tilde{D}\phi(\sigma)$ may arise from a monomial that was zero in $\phi(\sigma)$. This occurs if this monomial contains exactly one indeterminate $X_{ij}^{(1)}$ with $\sigma(i) \neq j$, whereas all other indeterminates evaluate to 1. This leaves at most $2dn$ choices for each monomial, from which we deduce the claim.

Combining (4.9), (4.10) and (4.11), we obtain that for all $\sigma \in \mathfrak{S}_n$ and $l \leq d$

$$\begin{aligned} |D^l \phi(\sigma)| &\leq \sum_{k=1}^d M(\tilde{D}^l \phi_k) N(\tilde{D}^l \phi_k, \sigma) \\ &\leq 2^l \sum_{k=1}^d M(\phi_k) N(\phi_k, \sigma) \\ &\leq 2^l M(\phi) N(\phi) \end{aligned}$$

using that $M(\phi) = \max_{k \leq d} M(\phi_k)$ and $N(\phi, \sigma) = \sum_{k=1}^d N(\phi_k, \sigma)$.

The inequality for $U\phi$ is proved similarly. Clearly $M(U\phi) \leq (dn)^{-1} M(\phi)$ if ϕ is homogeneous of degree d . On the other hand, monomials of $U\phi$ are obtained by picking a monomial of ϕ , two indeterminates X_{ij}, X_{kl} in this monomial and replace them by X_{il}, X_{kj} . Consequently every monomial of ϕ gives rise to at most $d(d-1)$ monomials in $U\phi$. Thus we deduce that

$$M(U\phi) N(U\phi) \leq \frac{d-1}{n} M(\phi) N(\phi)$$

Then for a general function decomposing $\phi = \sum_{k=1}^d \phi_k$ as a sum of homogeneous func-

tions implies that for all $\sigma \in \mathfrak{S}_n$

$$\begin{aligned} |U\phi(\sigma)| &\leq \sum_{k=1}^d M(U\phi_k)N(U\phi_k, \sigma) \\ &\leq \sum_{k=1}^d \frac{k-1}{n} M(\phi_k)N(\phi_k, \sigma) \\ &\leq \frac{d-1}{n} M(\phi)N(\phi, \sigma). \end{aligned}$$

4.3 Proof of the concentration inequality

4.3.1 Induction with the method of exchangeable pairs

The proof of Theorem 4.1 follows the original argument of Chatterjee for the $d = 1$ case, based on Stein's method of exchangeable pairs. Eventually, the argument consists in establishing an inequality involving the moment generating function of the random variable and its derivatives, so that it can be integrated to give a bound in the tail probability by Chernoff's inequality. In our case, we will obtain a differential inequality relating the moment generating functions of ϕ and that of the functions $D^k\phi$, which are of lower degree. A bound is thus proved for the function ϕ only by induction on the degree, which is made possible by the condition (4.6). To that end, we reindex the quantities β_ϕ, γ_ϕ in terms of d :

$$\begin{aligned} \beta_d &:= 6dC'_D \left(\log \left(\frac{4C_D n}{C'_D} \right)^+ + \frac{(2/n)(2 - e^{-2/n})}{1 - e^{-2/n}} \right) \\ \gamma_d &:= \frac{2\beta_d}{3} (2\mathbb{E}[\phi] + C_U) \end{aligned} \quad (4.12)$$

Theorem 4.1 is thus the consequence of the following lemma.

Lemma 4.1. *Let $\phi \in \mathfrak{F}_d, d \geq 1$ and $m(\theta) := \log \mathbb{E}[e^{\theta(\phi(\sigma) - \mathbb{E}[\phi])}]$. Suppose ϕ satisfies the assumptions of Theorem 4.1. Then for all $\theta \in [0, 1/\beta_d]$*

$$m(\theta) \leq \frac{\gamma_d \theta^2}{2(1 - \beta_d \theta)}. \quad (4.13)$$

and for all $\theta \in (-1/\beta_d, 0]$,

$$m(\theta) \leq \frac{\gamma_d \theta^2}{2}. \quad (4.14)$$

Proof of Theorem 4.1. For $\theta := t/(\gamma_d + \beta_d t)$, Chernoff's inequality and (4.13) give for all $t \geq 0$

$$\mathbb{P}[\phi(\sigma) - \mathbb{E}[\phi(\sigma)] \geq t] \leq e^{-\theta t + \log m(\theta)} \leq e^{\frac{-t^2}{2(\gamma_d + \beta_d t)}}.$$

Taking $\theta = -t/\gamma_d$ and (4.14) yields the lower tail. \square

For the case $d = 1$, a better inequality is established in the proofs of Proposition 1.1 and Theorem 1.5 in [41] using the method of exchangeable pairs. An exchangeable pair is a couple of random variables (X, X') invariant by permutation, so it has the same distribution as (X', X) . We prove Lemma 4.1 by induction, using the following rephrasing of Chatterjee's method of exchangeable pairs to prove concentration.

Proposition 4.2 ([43][Proof of Thm. 1.5]). *Let \mathcal{X} be a separable metric space, (X, X') be an exchangeable pair of \mathcal{X} -valued random variables. Let $f : \mathcal{X} \rightarrow \mathbb{R}$ such that $\mathbb{E}[f] := \mathbb{E}[f(X)] = 0$. Suppose $F : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a square-integrable antisymmetric function which satisfies*

$$f(X) := \mathbb{E}[F(X, X') \mid X]$$

and $\mathbb{E}|e^{\theta f(x)} F(X, X')| < \infty$ for all $\theta \in \mathbb{R}$. Then for all $\theta \in \mathbb{R}$

$$\left| \mathbb{E}[f(X)e^{\theta f(X)}] \right| \leq |\theta| \mathbb{E}[e^{\theta f(X)} \Delta(X)] \quad (4.15)$$

where

$$\Delta(X) := \frac{1}{2} \mathbb{E}[|F(X, X')(f(X) - f(X'))| \mid X].$$

As explained in [41, 42], defining an exchangeable pair (X, X') on $\mathcal{X} \times \mathcal{X}$ is equivalent to the consideration of a reversible Markov kernel P on $\mathcal{X} \times \mathcal{X}$ defined by $Ph(x) = \mathbb{E}[h(X') \mid X = x]$ for all function h . Given a function f of zero mean, the antisymmetric function F can then be obtained as $F(X, X') = g(X) - g(X')$ where g satisfies the Poisson equation

$$g - Pg = f$$

which can generally be constructed as $g = \sum_{k \geq 0} P^k f$. If \mathcal{X} is a finite group and P is the kernel of an ergodic random walk on \mathcal{X} which puts constant mass on conjugacy classes, the previous infinite sum converges and [42][Theorem 1.2] provides a concentration result in terms of what is essentially the mixing time of the random walk. Lemma 4.1 is proved by refining this result, to handle the case of a non-uniform bound on the quantity $\Delta(\sigma)$ and get a Bernstein-like inequality. This is analog to proving concentration bounds for self-bounding functions or weakly self-bounding functions [33].

On the symmetric group, one obvious candidate of a Markov kernel is that of random transpositions. The corresponding exchangeable pair is $(\sigma, \sigma\tau)$ where $\tau = (IJ)$ is a random transposition with I, J uniform and independent in $[n]$ (τ can be the identity). The random transposition chain puts constant mass on the identity and on transpositions, which are conjugacy classes, and thus can be applied the previously mentioned argumentation, its mixing time having been completely determined from the work of Diaconis and Shahshahani [61]. Adapting arguments of [41][Chapt. 4], we arrive at the following lemma, proved in Section 4.3.2.

Lemma 4.2. *Let $f : \mathfrak{S}_n \rightarrow \mathbb{R}$ have zero mean. There exists a function $F : \mathfrak{S}_n \times \mathfrak{S}_n \rightarrow \mathbb{R}$ such that $\mathbb{E}[F(\sigma, \sigma\tau) \mid \sigma] = f(\sigma)$. Furthermore, if $C \geq 0$ is a constant such that $|f(\sigma) - f(\sigma\tau)| \leq C$ for all $\sigma \in \mathfrak{S}_n$ and transposition τ then*

$$|F(\sigma, \sigma\tau)| \leq \frac{Cn}{2} \left(\log \left(\frac{24 \|f\|_\infty n}{C} \right) + \frac{(2/n)(2 - e^{-2/n})}{1 - e^{-2/n}} \right). \quad (4.16)$$

If f has degree 1, $F = (n/2)f$, so one has actually $|F(\sigma, \sigma\tau)| \leq Cn/2$.

From the previous lemma, upper bounding $\Delta(\sigma)$ essentially comes down to various quantities involving ϕ only. This will be done using a tensor representation of the function ϕ , namely we represent the function ϕ as $\phi(\sigma) = \text{tr}(AS^{\otimes d})$ for some $A \in M_n(\mathbb{R}_+)^{\otimes d}$. Restricting first to homogeneous functions, we can follow most of the arguments of Chatterjee in the one-dimensional case, which leads to the following lemmas, proved in Section 4.3.3.

Lemma 4.3. *Let $\phi \in \mathfrak{F}$. A.s.*

$$|\phi(\sigma) - \phi(\sigma\tau)| \leq \begin{cases} 2 \|\nabla\phi\|_\infty & \text{if } d = 1 \\ 6 \|\nabla\phi\|_\infty & \text{if } d \geq 2 \end{cases} \quad (4.17)$$

Lemma 4.4. *Let $\phi \in \mathfrak{F}_{=d}$. Then*

(i)

$$\frac{n}{2d} \mathbb{E}[\phi(\sigma) - \phi(\sigma\tau) \mid \sigma] = \left(1 - \frac{d-1}{2n}\right) \phi(\sigma) - D\phi(\sigma) - U\phi(\sigma).$$

(ii)

$$\frac{n}{2d} \mathbb{E}[|\phi(\sigma) - \phi(\sigma\tau)| \mid \sigma] \leq \left(1 - \frac{d-1}{2n}\right) \phi(\sigma) + D\phi(\sigma) + U\phi(\sigma).$$

(iii) $D\phi \in \mathfrak{F}_{=(d-1)}$ satisfies

$$\mathbb{E}[D\phi] = \left(1 - \frac{d-1}{n}\right) \mathbb{E}[\phi].$$

Remark 4.4. The degree 1 case is made much simpler as in this case the function $\phi - \mathbb{E}[\phi]$ is actually an eigenfunction of the random transposition kernel. In general, decomposing the function into a basis of eigenfunctions or using representation theory can provide a neat expression of the function F in Lemma 4.2, but is not clear how to relate the projections onto eigenspaces to the hypotheses made on ϕ , in particular the non-negativity of the coefficients. This seems however to be the good strategy if one wants to get rid of the log factor, and could provide further improvements in the proof of Theorem 4.1, allowing perhaps to get rid of the consideration of the operators D and U . For instance, it is always possible to replace the function ϕ by another representative $\psi \in \mathfrak{F}$, which yields the same function on \mathfrak{S}_n but has the property that

$D\phi = 0$. The issue is of course that we lose the non-negativity of the coefficients, which seems essential to get a self-bounding property like in the lemma. Note that writing $\phi(\sigma) = \text{tr}(AS^{\otimes d})$ can already be seen as the use of a specific representation of the symmetric group, the d -fold tensor product of the standard representation (by permutation matrices).

Proof of Lemma 4.1. The result is proved by induction on the degree. The case $d = 1$ is proved in [41] but can be recovered from the following arguments.

Let $\phi \in \mathfrak{F}_d$ satisfy the assumptions of Theorem 4.1. Let $\psi := \phi - \mathbb{E}[\phi]$ denote the centered version of ϕ and $m(\theta) := \log \mathbb{E}[e^{\theta\psi(\sigma)}]$.

Consider the function F obtained from Lemma 4.2 for the case $f = \psi$ and set $\Delta(\sigma) := 1/2 \mathbb{E}[|F(\sigma, \sigma\tau)| |f(\sigma) - f(\sigma\tau)| | \sigma]$. Equation (4.17) gives an upper bound on the constant C appearing in (4.16), and the definition (4.12) of β_d is made to give the a.s. bound $|F(\sigma, \sigma\tau)| \leq \beta_d(n/2d)$. On the other hand, decompose $\phi = \sum_{l=0}^d \phi_l$ as a sum of homogeneous functions with non-negative coefficients. We can suppose that $\phi_0 = 0$. From Point (ii) in Lemma 4.4, one gets the bound in (conditional) expectation

$$\begin{aligned} \frac{n}{2d} \mathbb{E}[|\phi(\sigma) - \phi(\sigma\tau)| | \sigma] &\leq \sum_{l=1}^d \frac{n}{2l} \mathbb{E}[|\phi_l(\sigma) - \phi_l(\sigma\tau)| | \sigma] \\ &\leq \sum_{l=1}^d \phi_l(\sigma) + D\phi_l(\sigma) + U\phi_l(\sigma) \\ &= \phi(\sigma) + D\phi(\sigma) + U\phi(\sigma). \end{aligned}$$

By assumption (4.6), $U\phi(\sigma) \leq C_U$. From Point (iii) of Lemma 4.4 $\mathbb{E}[D\phi] \leq \mathbb{E}[\phi]$, hence letting $\psi_2 := D\phi - \mathbb{E}[D\phi]$, one has

$$\begin{aligned} \Delta(\sigma) &\leq \frac{\beta_d}{2} (\psi(\sigma) + \psi_2(\sigma) + 2\mathbb{E}[\phi] + C_U) \\ &\leq \frac{\beta_d}{2} (\psi(\sigma) + \psi_2(\sigma) + \tilde{\gamma}_d) \end{aligned}$$

where we write $\tilde{\gamma}_d := 2\mathbb{E}[\phi] + C_U$. Apply (4.15) to obtain

$$\left| \mathbb{E}[\psi(\sigma)e^{\theta\psi(\sigma)}] \right| \leq \frac{\beta_d}{2} |\theta| \mathbb{E}[e^{\theta\psi(\sigma)} (\psi(\sigma) + \psi_2(\sigma) + \tilde{\gamma}_d)]. \quad (4.18)$$

Now by (iii), ψ_2 is the centered version of a degree $d - 1$ function with non-negative coefficients. In the case $d = 1$, $\psi_2 = 0$ so the previous inequality can easily be integrated to give (4.13) (see the proof of Thm. 1.5 in [41]). In the general case, we make use of the following duality formula for the entropy functional, to relate the moment generating functions of ψ and that of ψ_2 .

Proposition 4.3 ([34][Thm. 4.13]). *Let Y be a non-negative random variable. Define the entropy of Y as*

$$\text{Ent}(Y) := \mathbb{E}[Y \log Y] - \mathbb{E}[Y] \log \mathbb{E}[Y]. \quad (4.19)$$

The entropy satisfies the variational relation

$$\text{Ent}(Y) = \sup_W \mathbb{E} \left[(W - \log \mathbb{E} [e^W]) Y \right] \quad (4.20)$$

where the supremum is over all random variables with finite exponential moment.

Notice that

$$\frac{\text{Ent}(e^{\theta\psi(\sigma)})}{\mathbb{E}[e^{\theta\psi}]} = \theta m'(\theta) - m(\theta),$$

which is at the basis of the well known Herbst argument to prove concentration inequalities (see [34]). Letting $m_2(\theta) := \log \mathbb{E}[e^{\theta\psi_2(\sigma)}]$, (4.20) implies

$$\mathbb{E} \left[|\theta| e^{\theta\psi(\sigma)} \psi_2(\sigma) \right] \leq \mathbb{E} \left[e^{\theta\psi(\sigma)} \right] m_2(|\theta|) + \text{Ent}(e^{\theta\psi(\sigma)}).$$

Dividing by $\mathbb{E}[e^{\theta\psi(\sigma)}]$ in (4.18) thus yields

$$\begin{aligned} |m'(\theta)| &\leq \frac{\beta_d}{2} \left(|\theta| m'(\theta) + \tilde{\gamma}_d |\theta| + m_2(|\theta|) + \frac{\text{Ent}(e^{\theta\psi(\sigma)})}{\mathbb{E}[e^{\theta\psi(\sigma)}]} \right) \\ &= \frac{\beta_d}{2} \left((|\theta| + \theta) m'(\theta) - m(\theta) + \tilde{\gamma}_d |\theta| + m_2(|\theta|) \right). \end{aligned}$$

The term $m(\theta)$ can be neglected, observing that e^m is a convex function, which at $\theta = 0$ takes value 1 and derivative $\mathbb{E}[\psi] = 0$, implying $m'(\theta)$ has the sign of θ and $m(\theta) \geq 0$ for all $\theta \in \mathbb{R}$. Hence for $\theta \in [0, 1/\beta_d)$

$$m'(\theta) \leq \frac{(\beta_d \tilde{\gamma}_d / 2) \theta}{1 - \beta_d \theta} + \frac{(\beta_d / 2) m_2(\theta)}{1 - \beta_d \theta}.$$

Now assumption (4.6) was made so that $D\phi$ satisfies the same conditions as ϕ and can thus be applied the induction hypothesis:

$$m_2(\theta) \leq \frac{\gamma_{d-1} \theta^2}{2(1 - \beta_{d-1} \theta)}.$$

For all $\theta \in [0, 1/\beta_d)$, we can bound further

$$\frac{\beta_d \theta}{1 - \beta_{d-1} \theta} \leq \frac{\beta_d}{\beta_d - \beta_{d-1}} = 1.$$

As $m(0) = 0$ we deduce that for all $\theta \in [0, 1/\beta_d)$

$$\begin{aligned} m(\theta) &\leq \int_0^\theta \frac{(\beta_d \tilde{\gamma}_d / 2 + \gamma_{d-1} / 4) u}{1 - \beta_d u} du \\ &\leq \frac{(\beta_d \tilde{\gamma}_d / 2 + \gamma_{d-1} / 4) \theta^2}{2(1 - \beta_d \theta)}. \end{aligned}$$

Finally, (4.12) implies that $\beta_d \tilde{\gamma}_d / 2 + \gamma_{d-1} / 4 \leq \gamma_d$, which proves (4.13).

For the lower tail, consider $\theta \in (-1/\beta_d, 0]$. Using the same computations as above, bound

$$\begin{aligned} |m'(\theta)| &\leq \frac{\beta_d}{2} (\tilde{\gamma}_d |\theta| + m_2(|\theta|)) \\ &\leq (\beta_d \tilde{\gamma}_d / 2 + \gamma_{d-1} / 4) |\theta| \\ &\leq \gamma_d |\theta| \end{aligned}$$

which shows $m(\theta) \leq \gamma_d \theta^2 / 2$. \square

4.3.2 Almost sure bound on F : proof of Lemma 4.2

Let μ be the probability measure on \mathfrak{S}_n which puts mass $1/n$ on the identity and $2/n^2$ on transpositions. The random transposition Markov chain is the random walk on \mathfrak{S}_n defined by iid increments of law μ . These are symmetric hence the uniform distribution unif is stationary. The Markov chain is ergodic and its mixing properties have been thoroughly investigated in [61]. In particular, it was proved that for all $k \geq 0$,

$$\|\mu^{*k} - \text{unif}\|_{\text{TV}} \leq 6ne^{-2k/n}. \quad (4.21)$$

Let P denote the transition matrix of random transpositions and f be a function on \mathfrak{S}_n with zero mean under the uniform measure. Then the function F given by

$$F(\sigma, \sigma') := \sum_{k \geq 0} (P^k f(\sigma) - P^k f(\sigma'))$$

is well defined by the total variation convergence above and satisfies $\mathbb{E}[F(\sigma, \sigma') \mid \sigma] = f(\sigma)$. We refer to [41, 42] for details.

Lemma 4.2 is obtained by bounding F in two ways. On the one hand (4.21) implies

$$|P^k f(\sigma) - \mathbb{E}[f]| \leq 12 \|f\|_{\infty} ne^{-2k/n}$$

and thus

$$|P^k f(\sigma) - P^k f(\sigma\tau)| \leq 24 \|f\|_{\infty} ne^{-2k/n}.$$

The second bound is based on the fact that μ puts constant mass on conjugacy classes. As observed by Chatterjee in [41, 42], this implies that

$$|P^k f(\sigma) - P^k f(\sigma\tau)| \leq \max_{\sigma', \tau'} |f(\sigma') - f(\sigma'\tau')| \leq C$$

for all $k \geq 0$, $\sigma \in \mathfrak{S}_n$ and transposition τ . Combine the two bounds as

$$\begin{aligned} |F(\sigma, \sigma\tau)| &\leq \sum_{k \geq 0} \min\left(C, 24 \|f\|_{\infty} ne^{-2k/n}\right) \\ &\leq C \sum_{k \geq 0} \min\left(1, 24 C^{-1} \|f\|_{\infty} ne^{-2k/n}\right) \\ &\leq C \left(\frac{n}{2} \log(24C^{-1} \|f\|_{\infty} n) + 1 + \frac{1}{1 - e^{-2/n}}\right) \end{aligned}$$

4.3.3 Tensor representation

We now prove Lemmas 4.3 and 4.4.

Lemma 4.5. *Let $\phi \in \mathfrak{F}$ and $\tau = (IJ)$. For all $\sigma \in \mathfrak{S}_n$,*

$$\begin{aligned} \phi(\sigma\tau) &= \phi(\sigma) + \partial_{I\sigma(J)}\phi(\sigma) + \partial_{J\sigma(I)}\phi(\sigma) - \partial_{I\sigma(I)}\phi(\sigma) - \partial_{J\sigma(J)}\phi(\sigma) \\ &\quad + 2\partial_{I\sigma(I)}\partial_{J\sigma(J)}\phi(\sigma) + 2\partial_{I\sigma(J)}\partial_{J\sigma(I)}\phi(\sigma). \end{aligned} \quad (4.22)$$

Proof. Let $\phi \in \mathfrak{F}$. Decomposing ϕ into homogeneous components, it suffices to consider the case of a homogeneous function. Suppose therefore that ϕ is homogeneous of degree $d \geq 1$. It can be realized as $\phi(\sigma) = \text{tr}(AS^{\otimes d})$ for some $A \in M_n(\mathbb{R}_+)^{\otimes d}$. We start with a simple computation relating derivatives of ϕ with the tensor A .

Let E_{ij} denote the matrix which has entry (i, j) equal to 1 and all other entries equal to 0. For all $M \in M_n(\mathbb{R})$, $i, j \in [n]$,

$$\partial_{ij}\phi(M) = \sum_{k=1}^d \text{tr}(A M^{\otimes(k-1)} \otimes E_{ij} \otimes M^{\otimes(d-k)}). \quad (4.23)$$

Indeed, for all $t \in \mathbb{R}$, expanding the tensor product yields

$$(M + tE_{ij})^{\otimes d} = M^{\otimes d} + t \sum_{k=1}^d M^{\otimes(k-1)} \otimes E_{ij} \otimes M^{\otimes(d-k)} + O(t^2)$$

Dividing by $t \neq 0$ and taking the limit $t \rightarrow 0$ gives the result.

For the sequel, we make use of the multilinearity of ϕ . Given multi-indices $\mathbf{i}, \mathbf{j} \in [n]^d$, write $A_{\mathbf{i}, \mathbf{j}} = A_{i_1 \dots i_d, j_1 \dots j_d}$. The multilinearity implies we can suppose that $A_{\mathbf{i}, \mathbf{j}} = 0$ whenever \mathbf{i} or \mathbf{j} has two identical coordinates. Consequently the computation of $\text{tr}(AM)$ does not depend either on the entries $M_{\mathbf{i}, \mathbf{j}}$ when \mathbf{i} or \mathbf{j} has some identical coordinates. More precisely, the kernel of the linear map $M \mapsto \text{tr}(AM)$ contains the subspace H of tensors whose only non-zero entries are such multi-indices. Therefore when computing $\text{tr}(AM)$, one can freely replace M with any of its representative modulo H , which allows in particular to get rid of potential dependency properties between entries of S . The permutation matrix of τ is $T = I + M_{IJ}$ with

$$M_{IJ} := E_{IJ} + E_{JI} - E_{II} - E_{JJ}.$$

Note that if we write products of permutation from right to left, so $\sigma\tau$ applies τ first and then σ , the permutation matrix of $\sigma\tau$ is TS , hence $\phi(\sigma\tau) = \text{tr}(AT^{\otimes d}S^{\otimes d})$. By expanding the tensor product,

$$\begin{aligned} T^{\otimes d} &= I + \sum_{k=1}^d I^{\otimes(k-1)} \otimes M_{IJ} \otimes I^{\otimes(d-k)} \\ &\quad + \sum_{k_1 < k_2 \in [d]} I^{\otimes(k_1-1)} \otimes M_{IJ} \otimes I^{\otimes(k_2-k_1)} \otimes M_{IJ} \otimes I^{\otimes(d-k_2)} \pmod{H}. \end{aligned}$$

Expanding the expression of M_{IJ} in the second sum yields terms like $E_{IJ} \otimes E_{II}$ which are also in H and can be discarded. Now by (4.23),

$$\sum_{k=1}^d \operatorname{tr}(A(I^{\otimes(k-1)} \otimes E_{IJ} \otimes I^{\otimes(d-k)})S^{\otimes d}) = \partial_{I\sigma(J)}\phi(\sigma)$$

and a similar observation can be made for order 2 derivatives. This proves Lemma 4.5. \square

Lemma 4.3 can be deduced easily, provided one can control second order derivatives. This requires no additional assumption, for taking partial derivatives can only give a smaller function, as proved by the following lemma.

Lemma 4.6. *Let $\phi \in \mathfrak{F}$. For all $k \geq 1$ and $\mathbf{i}, \mathbf{j} \in [n]^k$,*

$$\|\partial_{\mathbf{i}\mathbf{j}}^k \phi\|_{\infty} \leq \|\phi\|_{\infty}. \quad (4.24)$$

Proof. The general case follows from the $k = 1$ case by an easy induction. Let $i, j \in [n]$. Note that by multilinearity, the partial derivative $\partial_{ij}\phi$ cannot contain any indeterminate X_{ik} or X_{kj} , $k \in [n]$, so $\partial_{ij}\phi(\sigma) = \partial_{ij}\phi((\sigma(i)j)\sigma)$. Hence the maximum is always realized for a permutation σ such that $\sigma(i) = j$, but then for such permutations ϕ actually coincides with the partial derivative $\partial_{ij}\phi$. Consequently

$$\begin{aligned} \max_{\sigma \in \mathfrak{S}_n} |\partial_{ij}\phi(\sigma)| &= \max_{\sigma: \sigma(i)=j} |\partial_{ij}\phi(\sigma)| \\ &\leq \max_{\sigma: \sigma(i)=j} |\phi(\sigma)| \\ &\leq \|\phi\|_{\infty}. \end{aligned}$$

\square

Proof of Lemma 4.3. If ϕ is assumed to have non-negative coefficients, (4.22) shows that

$$\begin{aligned} |\phi(\sigma) - \phi(\sigma\tau)| &\leq |\partial_{I\sigma(J)}\phi(\sigma) + \partial_{J\sigma(I)}\phi(\sigma) - \partial_{I\sigma(I)}\phi(\sigma) - \partial_{J\sigma(J)}\phi(\sigma)| \\ &\quad + 2\partial_{I\sigma(I)}\partial_{J\sigma(J)}\phi(\sigma) + 2\partial_{I\sigma(J)}\partial_{J\sigma(I)}\phi(\sigma) \\ &\leq 2\|\nabla\phi\|_{\infty} + 4 \max_{i,j,k,l \in [n]} |\partial_{ij,kl}^2\phi(S)|, \end{aligned}$$

which establishes (4.17) thanks to the previous lemma. \square

Proof of Lemma 4.4. Restricting to a homogeneous function $\phi \in \mathfrak{F}_{=d}$, (4.22) gives by

averaging over I, J

$$\begin{aligned} \frac{n}{2d} \mathbb{E}[\phi(\sigma) - \phi(\sigma\tau) \mid \sigma] &= \frac{1}{2dn} \sum_{i,j \in [n]} (\partial_{i\sigma(i)}\phi(\sigma) + \partial_{j\sigma(j)}\phi(\sigma) - \partial_{i\sigma(j)}\phi(\sigma) - \partial_{j\sigma(i)}\phi(\sigma) \\ &\quad - 2\partial_{i\sigma(i)}\partial_{j\sigma(j)}\phi(\sigma) - 2\partial_{i\sigma(j)}\partial_{j\sigma(i)}\phi(\sigma)) \\ &= \left(1 - \frac{d-1}{2n}\right)\phi(\sigma) - \frac{1}{dn} \sum_{i,j \in [n]} \partial_{i\sigma(j)}\phi(\sigma) - \frac{1}{dn} \sum_{i,j \in [n]} \partial_{i\sigma(j)}\partial_{j\sigma(i)}\phi(\sigma), \end{aligned}$$

which gives Point (i) of the Lemma. The second equality arises from the relation (4.4).

The bound in absolute value (ii) is obtained similarly, using first triangle inequality in (4.22). Finally Point (iii) is proved easily. \square

Chapter 5

Cutoff for mixtures of permuted Markov chains

5.1 Models and main results

Like the previous chapter, the content of this chapter is yet to be published. We establish a cutoff phenomenon at entropic time for a model of Markov chain in random environment.

5.1.1 Cutoff for mixtures of permuted Markov Chains

Let $n \geq 1$ be an integer and P_1, P_2 two $n \times n$ stochastic matrices. This work is basically concerned with the mixing time of the Markov chain with transition matrix

$$p_1 P_1 + p_2 S P_2 S^{-1}$$

where $p_1 \in (0, 1)$, $p_2 = 1 - p_1$ and S is a uniform $n \times n$ permutation matrix. This model is inspired by the work [92] of Hermon, Sousi and Sly, who proved the cutoff phenomenon at entropic time for the simple random walk on a sequence of deterministic graphs to which is added a random uniform matching of the vertices. The model above aims to look at more general Markov chains, possibly non reversible, obtained by superposing two deterministic ones with a layer of randomness in between taking the form of a permutation of the states. Our model will in fact allow the probabilities p_1, p_2 to depend on the current state of the chain and on the permutation S , which is necessary to recover the case of the simple random walk on the superposition of graphs if these are not regular. For this case however, and more generally for reversible chains, there is a more adapted model tailored to preserve reversibility, corresponding to a specific choice of p_1, p_2 , which we will describe afterwards.

Let us introduce or recall notations that will be used throughout this chapter. Given two measures μ, ν on a countable set S , their total variation distance is defined as

$$\|\mu - \nu\|_{\text{TV}} := \sup_{A \subseteq S} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)|.$$

If P is the transition kernel of a positive recurrent, irreducible and aperiodic Markov chain on S , it admits a unique invariant measure π . In that case, given a starting vertex $x \in S$ and $\varepsilon \in (0, 1)$, the mixing time is defined as

$$t_{\text{mix}}(x, \varepsilon) := \inf\{t \geq 0 : \|P^t(x, \cdot) - \pi\|_{\text{TV}} < \varepsilon\}.$$

and the worst-case mixing time is

$$t_{\text{mix}}(\varepsilon) := \sup_{x \in S} t_{\text{mix}}(x, \varepsilon).$$

If the chain is not irreducible or aperiodic, we consider the mixing time to be infinite. Given two functions $f, g : \mathbb{N} \rightarrow \mathbb{R}$, $f(n) = O(g(n))$ if there exists a constant $C > 0$ such that $|f(n)| \leq Cg(n)$ for all n , $f = o(g)$ if $f(n)/g(n) \xrightarrow[n \rightarrow \infty]{} 0$ and $f = \Theta(g)$ if $f = O(g)$ and $g = O(f)$. An event $A = A(n)$ is said to occur with high probability if $\mathbb{P}(A) = 1 - o(1)$. Given an integer $n \geq 1$, we write $[n] := \{1, \dots, n\}$.

Theorem 5.1. *Let P_1, P_2 be two $n \times n$ stochastic matrices, $p_1, p_2 \in M_n([0, 1])$ be $n \times n$ matrices with entries in $[0, 1]$ such that $p_1 + p_2 = 1$ entry-wise and σ a permutation of n elements chosen uniformly at random. Consider the Markov chain with transition probabilities*

$$\forall x, y \in [n] : \mathcal{P}_0(x, y) := p_1(x, \sigma(x)) P_1(x, y) + p_2(x, \sigma(x)) P_2(\sigma(x), \sigma(y)). \quad (5.1)$$

Suppose

(H1) for each $i \in \{1, 2\}$ the non-zero entries of P_1, P_2 are $\Theta(1)$ and $\max_{y \in V} \sum_{x \in V} P_i(x, y) = O(1)$.

(H2) the entries of p_1, p_2 are $\Theta(1)$.

(H3) For all $x \in [n]$ and $i = 1, 2$:

$$|\{y \in V : \exists k \geq 0 : P_i^k(x, y) > 0\}| \geq \begin{cases} 3 & \text{if } i = 1 \\ 2 & \text{if } i = 2 \end{cases},$$

(H4) There exists a constant $l \geq 1$ independent of n such that

$$|\{x \in V \mid \forall y \in V : P(x, y) > 0 \Leftrightarrow (I + P)^l(y, x) > 0\}| = n - o(n)$$

Then the chain defined by (5.1) is irreducible, aperiodic and there exists $h = \Theta(1)$ for which the following holds. For all $\varepsilon \in (0, 1)$, there exists $C(\varepsilon)$ such that for all $x \in [n]$ with high probability,

$$\begin{aligned} t_{\text{mix}}(x, \varepsilon) &\leq \frac{\log n}{h} + C(\varepsilon)\sqrt{\log n}, \\ t_{\text{mix}}(x, 1 - \varepsilon) &\geq \frac{\log n}{h} - C(\varepsilon)\sqrt{\log n}. \end{aligned}$$

Our proof adapts the strategy used to prove cutoff for non-backtracking chains [19, 28, 29] to handle backtracking chains as well, including reversible ones, but reversibility does not need to be assumed. This is undoubtedly the most salient aspect of our contribution, as earlier works proving cutoffs in random environment generally relied on reversibility or a non-backtracking property, which can be thought of as the opposite of reversibility. Thus an important goal of our work is to provide a unified approach to prove cutoff for Markov chains in random environment using the "entropic method", with or without reversibility.

Nevertheless, the absence of reversibility brings real additional difficulties which in turn yield a weaker result than in the reversible model studied in [92]: note that in Theorem 5.1, the starting state of the chain is fixed before generating the random environment. This implicitly makes the state x *typical*: from the result for a fixed x one can for instance replace x with a state chosen uniformly at random. However it cannot be chosen arbitrarily after the permutation σ is fixed. Consequently our result does not establish uniform cutoff in the sense that with high probability $t_{\text{mix}}(x, \varepsilon) = \log n/h + o(\log n)$ simultaneously for all $x \in [n]$. It turns out that with the level of generality considered here, the lack of a uniform result is not an artifact of the proof but is characteristic of the model. In whole generality, there is no uniform cutoff at entropic time, as proved by the following result.

Theorem 5.2. *For all $a > 1$, there exist transition matrices P_1, P_2 satisfying Assumptions (H1) - (H4) such that with high probability, the Markov chain $\mathcal{P}_0 = 1/2(P_1 + SP_2S^{-1})$ has worst-case mixing time $\max_x t_{\text{mix}}(x, 1/2) \geq (\log n)^a$.*

The lack of uniform cutoff is similar for instance to the case of random walks on the giant component of Erdős-Renyi graphs: with high probability these contain segments of length of order $\log n$, resulting in the worst-case mixing time being of order $(\log n)^2$ [20, 73], while the typical mixing time is $O(\log n)$ [22].

In our model, the potential slowdown of the chain is a consequence of irreversibility: having too much bias on the transition probabilities, the environment can with positive probability contain *trapping neighbourhoods*, from which the chain escapes in a time that scales exponentially with the size of the neighbourhood. The previous proposition

is based on an example where these neighbourhoods are of size $O(\log \log n)$, hence the polylogarithmic order in the result. Our next result establishes a matching upper bound of polylogarithmic order on the worst-case mixing time.

Theorem 5.3. *In the setting of Theorem 5.1 under the same assumptions there exists $a \geq 1$ such that for all $\varepsilon \in (0, 1)$, with high probability*

$$t_{\text{mix}}(\varepsilon) \leq (\log n)^a.$$

Remark 5.1. Let us comment on the assumptions made above. Hypotheses (H1) and (H2) bound uniformly in n the transition probabilities and the number of ingoing and outgoing transitions at every state, while (H3) imposes a minimal size on the set of states reachable by a given state. These are not restrictive: it is easy to find examples where these assumptions are not satisfied and the mixing time scales differently as $\Theta(\log n)$, without cutoff. The most restrictive assumption is thus (H4). It can be interpreted as a constraint on the degree of irreversibility of the chains P_1, P_2 , which for instance prevents them from having too many transient states. It is not clear whether it is really needed for our result but from the proof it does not seem superfluous either, as having too many transient states may amplify the trapping phenomenon mentioned above. Nonetheless, irreversibility is still permitted to a considerable extent.

5.1.2 A reversible model

Taking a mixture of transition matrices seems a most natural way to blend together two Markov chains. However if the chains are reversible there is another, which preserves reversibility, while a convex combination of reversible chains fails to be reversible in general.

Recall that a chain on state space S with transition kernel P is reversible if there exists a measure π on S such that $\pi(x)P(x, y) = \pi(y)P(y, x)$ for all $x, y \in S$. An electrical network is a pair (G, c) consisting in a weighted non-directed graph $G = (V, E)$ equipped with non-negative weights $c = (c(e))_{e \in E}$, called conductances, on the edges. Any reversible Markov chain can be represented as a random walk on an electrical network, defining conductances as

$$c(x, y) := \pi(x)P(x, y) \tag{5.2}$$

for all $x, y \in V$, which are symmetric by assumption. Conversely, any electrical network gives rise to a reversible Markov chain whose transition probabilities are proportional to conductances. We refer to [128] for a detailed account of this theory. The particular case of the simple random walk on a multi-graph is obtained by taking all conductances

equal to 1. For more general reversible chains, the electrical network theory provides a natural and generic way to mix together two reversible chains by superpositioning the corresponding electrical networks, that is taking a linear positive combination of the conductances. As was explained before, the lack of a uniform bound in Theorem 5.1 is a consequence of irreversibility, so we may hope for stronger results in the reversible case. The following theorem confirms this intuition.

Theorem 5.4. *Let $n \geq 1$ be an integer, $(G_1, c_1), (G_2, c_2)$ two electrical networks with common vertex set $[n]$, $\alpha, \beta > 0$ and σ a permutation of n elements chosen uniformly at random. Consider the Markov chain on $[n]$ defined by the electrical network (G^*, c) with conductances*

$$\forall x, y \in [n] : c^*(x, y) := \alpha c_1(x, y) + \beta c_2(\sigma(x), \sigma(y)). \quad (5.3)$$

Suppose

(H1) *The degrees and non-zero conductances of (G_1, c_1) and (G_2, c_2) are all $\Theta(1)$.*

(H2) *α, β are constants independent of n .*

(H3) *The connected components of G_1 have size at least 3 and that of G_2 size at least 2.*

Then there exists $h = \Theta(1)$ for which the following holds. For all $\varepsilon \in (0, 1)$, there exists a constant $C(\varepsilon)$ such that with high probability,

$$\begin{aligned} \min_{x \in [n]} t_{\text{mix}}(x, 1 - \varepsilon) &\geq \frac{\log n}{h} - C(\varepsilon) \sqrt{\log n}, \\ \max_{x \in [n]} t_{\text{mix}}(x, \varepsilon) &\leq \frac{\log n}{h} + C(\varepsilon) \sqrt{\log n}. \end{aligned}$$

In particular, the chain is irreducible and aperiodic with high probability and exhibits a uniform cutoff phenomenon: for all $\varepsilon \in (0, 1)$

$$\lim_{n \rightarrow \infty} \frac{\max_{x \in [n]} t_{\text{mix}}(x, \varepsilon)}{\min_{x \in [x]} t_{\text{mix}}(x, 1 - \varepsilon)} = 1$$

in probability.

Remark 5.2. Being defined with conductances, the chain considered in the previous theorem is thus automatically reversible. It is a particular case of (5.1): supposing P_1, P_2 are reversible and correspond to electrical networks $(G_1, c_1), (G_2, c_2)$ respectively, $P_i(x, y) = c_i(x, y)/c_i(x)$ where $c_i(x) := \sum_{z \in V} c_i(x, z)$ for $i = 1, 2$. Hence the reversible model is realized as (5.1) taking

$$p_1(x, y) := \frac{\alpha c_1(x)}{\alpha c_1(x) + \beta c_2(y)}.$$

Remark 5.3. Letting $c_1, c_2 \equiv 1$ take the value 1 on every edge and setting $\alpha = \beta = 1$ as well, we obtain the particular case of the superposition of two multi-graphs. The case of a graph with an added random matching analog to the case studied in [92] can then be obtained by taking for G_2 a sequence of edges if n is even. A little difference lies however in the fact that the authors consider there simple graphs while we consider multi-graphs, thus edges of G_1, G_2 that align under the permutation σ would result in a transition with a higher probability. To obtain rigorously the case of simple graphs, it would in theory be necessary to adjust α, β to make them $1/2$ when edges of the two networks are aligned. However a close inspection at the proof shows this makes no difference (see Remark 5.7), so our result is also true when superpositionning simple graphs and discarding multi-edges in the resulting graph.

5.1.3 Related works

To this date, it remains an open question to determine a characterization of cutoff. A sufficient condition was given by Salez in [149], which is certainly the closest result one has ever been from a full characterization. This condition expresses an entropic concentration phenomenon, which has been at the heart of many important achievements on cutoff over the last decade. While the initial focus was on specific, explicit Markov chains with high degrees of symmetry, like random walks on groups (see [57, 150] for global references on the subject), the seminal work of Lubetzky and Sly [122] on random walks on random regular graphs initiated a series of papers studying instead generic Markov chains. In that regard, a lot of attention was drawn on Markov chains in random environment, which showed that cutoff is actually quite common and often the result of an entropic concentration phenomenon, leading to an *entropic mixing time* $\log n/h$, where n is the size of the state space and h can be interpreted as a entropy rate.

After Lubetzky and Sly proved the cutoff for simple and non-backtracking random walks on random regular graphs, Ben Hamou and Salez proved the cutoff for the non-backtracking walk on random graphs with a given degree sequence, ie the configuration model [19]. This case is also considered in [22], in which Berestycki, Lyons, Peres and Sly prove the cutoff, for both the simple and non-backtracking walks on the giant component of an Erdős Renyi random graph as well as for the configuration model. In the case of the simple random walk, the starting point is not uniform but needs to be typical. For Erdős Renyi random graphs, the worst-case mixing time had been established to be $\Theta((\log n)^2)$ by Fountoulakis and Reed [73] and Benjamini, Kosma and Wormald [20], while in the configuration model cutoff at entropic time was obtained subsequently by Ben Hamou, Lubetzky and Peres in [17]. Models of random

non-backtracking chains have also been considered by Bordenave, Caputo and Salez: [28] considers the non-backtracking walk on directed configuration models, while [29] considers the case of a stochastic matrix in which the entries of each row are permuted uniformly at random. In [50], Conchon-Kerjan considers random walks on random lifts of weighted graphs which are not reversible. Let us also mention the work of Hermon and Olesker-Taylor [90, 91] on random walks on random Cayley graphs of Abelian groups. A model similar to ours is the *PS* model, introduced by Chatterjee and Diaconis [45] and shown to exhibit cutoff at entropic time by Ben Hamou and Peres [18].

Some recent works investigate the case of random graphs with community structures: in [16], Ben Hamou proves a phase transition for the cutoff of the non-backtracking random walk on a random graph with two communities. An extension of this result for the simple random walk was obtained by Hermon, Šarković and Sousi [93] who also consider a second model of random graphs allowing more communities.

While the previously cited works consider the case of an essentially totally random Markov chains, cutoff was also shown to occur when randomizing a given chain, where the final environment still keeps a lot of the structure of the initial chain. We already cited the work of Hermon, Sly and Sousi [92] as the main inspiration of this work, where cutoff is proved for the simple random walk on a graph to which is added a uniform matching. This model was extended by Baran, Hermon, Šarković and Sousi [13] who prove a phase transition when weights are added on the random matching.

Another direction of recent works is the competition between different mechanisms such as having a dynamic environment: see the papers by Avena, Güldaş, van der Hofstad and den Hollander [11] and Caputo and Quattropani [39]. A related work is [38] by Caputo and Quattropani on PageRank random walks on random digraphs.

Let us mention that cutoff at entropic phenomenon is not bound to random Markov chains: it can also arise in deterministic settings, such as Ramanujan graphs [121, 143] or environments with a Ramanujan property [31]. In fact from the work of Friedman [75] (see also [27]) it is known that that random regular graphs are Ramanujan with high probability, so the above results can be interpreted as cutoff phenomena in pseudo-deterministic settings. The paper by Eberhard and Varjù [68] falls in that category, which can also be interpreted as a deterministic realization of the *PS* model studied in [18]. Finally, the sufficient condition of Salez [149] allows him to deduce cutoff for a large family of chains satisfying a non-negative curvature criterion, not necessarily random.

5.1.4 Proof outline

Dealing with an unknown invariant measure: first the absence of a reversibility assumption prevents any knowledge of the invariant measure. To establish mixing, we follow the strategy of [28, 29] of proving convergence towards an approximate invariant measure $\hat{\pi}$. If $\|\mathcal{P}_0^t(x, \cdot) - \hat{\pi}\|_{\text{TV}} \leq \varepsilon$ holds uniformly in x for a given time t , then the invariance property implies that a true invariant measure π satisfies $\|\pi - \hat{\pi}\|_{\text{TV}} \leq \varepsilon$, establishing in particular the uniqueness of the invariant measure. The additional issue here is that Theorem 5.1 is valid only for typical starting states, so we will have to prove that mixing towards $\hat{\pi}$ occurs both at $t \simeq \log n/h$ for a typical starting state and at larger times for an arbitrary starting state, namely $t = (\log n)^a$. Note that since $\hat{\pi}$ is not a priori a genuine invariant measure, there is no monotonicity in the distance to equilibrium that would maintain the chain distributed as $\hat{\pi}$ once it has become close to it. We note that Proposition 5.2 below gives an expression for the proxy measure $\hat{\pi}$ we use in our proofs.

The entropic method: the entropic method has become over the last years quite a standard technique to prove cutoff, proof being that it is common to almost all previous works on cutoff for randomized Markov chains cited above. It is essentially made of two arguments: first an entropic concentration is shown to occur at the entropic time $\log n/h$, which is shown to imply cutoff in the second part of the proof. These arguments come with different variations, depending on whether reversibility is supposed or not. We follow the approach originally designed for non-backtracking random walks [19, 28, 29]. First we prove a quenched entropic concentration property for trajectories, namely we prove a statement of the form $\mathcal{P}(X_0 \cdots X_t) \simeq e^{-th + O(\sqrt{t})}$, where we write $\mathcal{P}(X_0 \cdots X_t) = \prod_{i=0}^{t-1} \mathcal{P}(X_i, X_{i+1})$. This statement is given here as a rough heuristic, which would be correct in the non-backtracking case. In general, what we prove in practice is rather concentration of the probability to follow the *loop-erased trace* of X , so we are back at studying non-backtracking trajectories. This concentration phenomenon is proved by a coupling with another Markov chain lying on an infinite random state space. We call this environment a quasi-tree in reference to [92], as this object is similar to theirs. Basically, it is designed to be a stationary approximation of the universal cover of the finite chain, in the same way Erdős-Renyi random graphs can be approximated by infinite Galton-Watson trees for instance.

Analysis on the quasi-tree: this part is essentially the only one where reversibility plays a role to obtain a uniform cutoff result. At the basis of the analysis for reversible chains in [92, 93, 13] is the property of uniform transience, which asserts that at any

vertex, the probability to escape to infinity along the neighbouring "branch" of the quasi-tree, conditional on the environment, is lower bounded away from 0 uniformly in n . This property extends to any reversible chain on the quasi-tree thanks to a comparison argument, after which the analysis can be conducted pretty much as is done in [92]. On the other hand, such a uniform lower bound may not be achieved in the general model of Theorem 5.1, which is ultimately the reason why cutoff cannot occur for a uniform starting state. Incidentally, the lack of uniform transience is an issue that could already be present in the case of a Galton-Watson tree, leading to a non-uniform cutoff for Erdős-Renyi graphs [22]. Our analysis is thus largely inspired by this paper and the references cited therein. The lack of uniform transience is bypassed by proving that the probability to escape at infinity along is in expectation lower bounded uniformly in n . This will again be proved thanks to a comparison argument, namely the chain is going to be compared with a reversible chain, which is the main reason why Hypothesis (H4) comes into play. The lower bound on the expected escape probabilities implies then that such a lower bound is also achieved when restricting to typical points. Very roughly, the difference with the reversible case will thus consist in arguing that the trajectories of the chain remain up to the mixing time on typical parts of the quasi-tree. While uniform transience allows to make many "quenched" statements, ie conditional on the environment, we will have to replace these by "annealed" statements, which average over the environment.

As in [22, 92, 93, 13], the proof of the entropic concentration relies on the use of *regeneration times*, which are times at which the chain in the quasi-tree crosses a long-range edge for the first and last time in the quasi-tree. In fact, the use of these times goes beyond the entropic concentration phenomenon, so we will spend some time studying the regeneration process. In particular, as in [93], our model requires the consideration of an underlying Markov chain. We thought useful to introduce the setting of Markov renewal processes to formalize these arguments in a a generic way. An important requirement for the proof is the fast mixing of the Markov chain underlying the regeneration process. In our case, the law of the environment presents enough independence so that the regeneration chain satisfies Doeblin's condition, which implies the mixing time is $O(1)$.

Argument for the lower bound: the lower bound in the mixing time is based on a simple coverage argument. From the concentration of entropy, at time $t = O(\log n)$ the chain is necessarily confined in a set of size $e^{th+O(\sqrt{t})}$ at most. This becomes $o(n)$ for $t \leq \log n/h - C\sqrt{\log n}$ and a large constant $C > 0$, which is sufficient to conclude provided the invariant measure is well spread-out across the state space. Since the

latter is unknown, we use the proxy measure $\hat{\pi}$ and its expression which is explicit enough for our purpose.

Concentration of nice trajectories: the second part of the argument for the upper bound uses the entropic concentration to show a second concentration phenomenon. Specifically, for all $x, y \in V$, we define a set $\mathfrak{N}^t(x, y)$ of "nice" trajectories between x, y , of length t , which will have the property that: 1. they are typical, that is the trajectory of the chain is likely to be nice, 2. the probability $\mathcal{P}_{\mathfrak{N}}^t(x, y)$ of following a nice trajectory concentrates around its mean, which is shown to be independent of the starting point. From these properties we deduce that $\mathcal{P}^t(x, \cdot)$ mixes towards a proxy stationary distribution $\hat{\pi}$.

Nice paths are basically defined by splitting a path in three components $\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{p}_3$ and restricting $\mathfrak{p}_1, \mathfrak{p}_3$ to be contained in a nice quasi-tree-like neighbourhood around x and y respectively. Thanks to the entropic concentration, paths can be restricted to have probability $e^{-th+O(\sqrt{t})}$, which allows to use the concentration result of Theorem 4.1 conditional on the two neighbourhoods. The notable difference with the case of non-backtracking chains [19, 28, 29] comes from the path \mathfrak{p}_2 . In the case of non-backtracking chains, it merely consisted in one edge so only the concentration for degree 1 functions was needed. When backtracking is allowed, trajectories could bounce back and forth between the two neighbourhoods, so taking just one edge is not sufficient. Our strategy is here to reduce as much as possible the setting to the non-backtracking case. To that end, we can use concentration of the speed (or drift), which is proved in the same time as for the entropy: the distance traveled by the chain at time t is shown to be $dt + O(\sqrt{t})$. On the other hand, using the lower bound on escape probabilities, we can argue that the chain is unlikely to backtrack over more than $L = \Theta(\log \log n)$ steps on a time scale $t = O(\log n)$. Thus by allowing the distance between the two neighbourhoods to be of order \sqrt{t} , we can ensure the chain does not backtrack from one neighbourhood to the other and discard trajectories which have intermediate length much smaller or much larger than \sqrt{t} . In the end, the estimate on the traveled distance leads us to consider neighbourhoods of variable size, with a variable distance between them, but with bounded windows on these parameters. Concentration is thus proved using Theorem 4.1 for fixed parameters and extended by union bound. When realized as a multilinear function, the probability to follow a nice path between x and y will have degree the length of the path \mathfrak{p}_2 between the two neighbourhoods. Since this is $O(\sqrt{t}) = O(\sqrt{\log n})$, the exponential factors in Corollary 4.1 remain sufficiently small to obtain exponential concentration bounds.

Computation of $\hat{\pi}$: another difficulty is to compute the expectation precisely enough to get the probability measure $\hat{\pi}$. To that end, we use regeneration times for the finite chain, essentially defined to be times at which the chain makes a transition and does not backtrack before it has traveled distance $L = \Theta(\log \log n)$. From the non-backtracking property mentioned above, this essentially amounts to consider transitions which are made only once on the time scale $t = O(\log n)$. Furthermore, such regeneration times can of course be coupled with those defined in the quasi-tree setting. From this, we can use the mixing results proved for the regeneration chain in the quasi-tree to deduce similar mixing properties in the finite setting, which yields the expression of the limiting measure $\hat{\pi}$ (Proposition 5.2).

Getting a bound on the worst-case mixing time : nice paths are in fact really defined only from typical states, so the previous arguments will ensure $\mathcal{P}_0^t(x, \cdot) \simeq \hat{\pi}$ at time $t = \log n/h + O(\sqrt{\log n})$ only for a typical x . To obtain an upper bound on the worst-case mixing time, which as remarked in the first paragraph above is crucial to justify $\hat{\pi}$ is close to being invariant, we will essentially argue that for some $a \geq 1$, at time $s = (\log n)^a$ the chain is likely to reach a typical point and have its t subsequent steps following a nice trajectory, regardless of its starting state. Thus mixing at time $t = O(\log n)$ from typical states implies mixing at time $s + t = O((\log n)^a)$ from any starting state. This argument is also used in the reversible case, where $s = \Theta(\log \log n)$ steps are enough, which is the consequence of the quenched uniform lower bound on escape probabilities in quasi-trees. In the general case, a heuristic is as follows: we will get from the estimates on escape probabilities (Prop. 5.4) that these are so unlikely to fall below $(\log n)^{-b}$ for some $b > 0$ that we can basically take $(\log n)^{-b}$ as a quenched lower bound. Thus we can expect it takes $(\log n)^{bl}$ steps at worst for the chain to escape potentially trapping neighbourhoods of size l , after which the chain is likely to be back in a typical neighbourhood if l is chosen adequately.

Organization of the chapter

In Section 5.2 we give material that will be used throughout the paper, including the definition of quasi-trees. In Section 5.3, we give in detail the main technical arguments of the paper that sum up the arguments of the entropic method, from which we will be able to deduce Theorems 5.1, 5.3 and 5.4. In Section 5.4, we provide a counter-example to uniform entropic mixing time, proving Theorem 5.2. The three next sections deal with the analysis of the quasi-tree: Section 5.5 establish the lower bound on escape probabilities, Section 5.6 proves technical results about the regeneration structure of quasi-trees while Section 5.7 basically establishes "nice properties" in the quasi-tree

setting, including in particular the concentration of the speed and entropy. These nice properties are then transferred back to the finite setting in Section 5.8, which also contains the arguments that will lead to the upper bound on the worst-case mixing time. Finally, these properties are used in Section 5.9 where nice paths are properly defined and shown to have their probability concentrate around the mean.

5.2 First moment argument, quasi-trees

Basic notations: throughout the paper, all quantities involved may have an implicit dependency in n and the term constant will refer to quantities that are independent of n and of the randomness. Given a set S , $|S|$ denotes its cardinality, $\mathbb{1}_S$ is the indicator of S . Given $x, y \in \mathbb{R}$, we write $x \wedge y := \min(x, y)$, $x \vee y := \max(x, y)$. We use the standard Landau notations o, O for deterministic sequences: $f(n) = O(g(n))$ if there exists a constant $C > 0$ such that $|f(n)| \leq Cg(n)$ for all n , $f = o(g)$ if $f(n)/g(n) \xrightarrow[n \rightarrow \infty]{} 0$. We may write $O_\varepsilon(\cdot)$ to precise a potential dependence of the implicit constant in ε . Given two functions, $f = \Theta(g)$ if $f = O(g)$ and $g = O(f)$, and $f \gg g$ if $g = o(f)$. All the random variables considered in this paper are defined on an implicit probability space with measure \mathbb{P} . If Y_n, Z_n, Z are random variables, we write $Z_n \xrightarrow{\mathbb{P}} Z$ for convergence in probability and $Z_n = o_{\mathbb{P}}(Y_n)$ if $Z_n/Y_n \xrightarrow{\mathbb{P}} 0$. In particular $Z_n = o_{\mathbb{P}}(1)$ if $Z_n \xrightarrow{\mathbb{P}} 0$. An event $A = A(n)$ is said to occur with high probability if $\mathbb{P}(A) = 1 - o(1)$.

5.2.1 Quenched vs annealed probability, first moment argument

Consider $(X_t)_{t \geq 0}$ a Markov chain in random environment, ie with random transition probabilities, such as the one defined by (5.1). There are two laws naturally associated with the process $(X_t)_{t \geq 0}$. The probability \mathbb{P} , which averages over both the random walk and the environment, gives rise to the *annealed* law of the process $(X_t)_{t \geq 0}$ under which it is not a Markov chain. It is however a Markov chain under the *quenched* law, which conditions on the environment. To emphasize the distinction, it is written using a different font, namely \mathbf{P} will denote the quenched distribution. For all state x , we write $\mathbb{P}_x := \mathbb{P}[\cdot | X_0 = x]$, $\mathbf{P}_x := \mathbf{P}[\cdot | X_0 := x]$. Of course taking the expectation of the quenched law gives back the annealed law, which is the basis of the following first moment argument that will be used throughout the article. To prove a trajectorial event A has quenched probability vanishing to 0 in probability as $n \rightarrow \infty$, it suffices to prove A has annealed vanishing probability, as Markov's inequality implies for all $\varepsilon > 0$,

$$\mathbb{P}[\mathbf{P}[A] \geq \varepsilon] \leq \frac{\mathbb{P}[A]}{\varepsilon}. \quad (5.4)$$

In most of this paper we will first prove statements valid for fixed starting states, chosen independently of the environment, such as $\mathbb{P}_x[A] = o(1)$, to obtain $\mathbf{P}_x[A] = o_{\mathbb{P}}(1)$. These can be interpreted as conditional statements for the case where the starting state is also random, with a law which independent of the environment. Results for fixed starting states thus extend automatically to *typical states*, for instance taken uniformly at random in the case the state space is $[n]$. To obtain stronger results valid for all states simultaneously, union bound shows it suffices to improve the annealed error bounds to $\mathbb{P}_x[A] = o(1/n)$. This strategy will be used in Section 5.8 to extend results from typical to all states.

5.2.2 The two-lift chain and half-integer time steps

Let us start by rewriting and actually generalizing our model further, which lacks symmetry in the roles of P_1 and P_2 . To obtain a model that is more symmetric, we construct the chain as a projection of its two-lift: this consists in seeing P_1, P_2 as two disjoint components of one Markov chain on a twice larger state space, which can be projected back to obtain the original model. This is essentially a matter of unifying notation and recover a setting which is somewhat similar to that of [92], but it can also prove useful on its own. All in all, the main model we will work with in this paper is the following.

Let $V := [2n], V_1 := [n], V_2 := [n+1, 2n]$. Let σ be now a uniform bijection from V_1 to V_2 , which can be extended as a matching, or involution, η on V by

$$\eta|_{V_1} := \sigma, \quad \eta|_{V_2} := \sigma^{-1}.$$

This matching defines an equivalence relation, namely $x \sim \eta(x)$. For all $x \in V$, define $V(x) = V_1 \mathbb{1}_{x \in V_1} + V_2 \mathbb{1}_{x \in V_2}$. Let P be a stochastic matrix on V with block form $P := \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$ and $p : V \times V \rightarrow [0, 1]$ such that $p(x, y) + p(y, x) = 1$ for all $x, y \in V$ and let $q(x, y) := 1 - p(x, y) = p(y, x)$. Consider the Markov chain on V given by the transition probabilities:

$$\forall x, y \in V : \mathcal{P}(x, y) := \begin{cases} p(x, \eta(x)) P(x, y) & \text{if } V(x) = V(y) \\ q(x, \eta(x)) P(\eta(x), y) & \text{if } V(x) \neq V(y). \end{cases} \quad (5.5)$$

We will work with this Markov chain most of the time, but the main object of interest of this paper is rather its projection to the quotient $V / \sim = [n]$. As is easily checked, the condition $q(x, y) = p(y, x)$ implies that $\mathcal{P}(x, y) + \mathcal{P}(x, \eta(y)) = \mathcal{P}(\eta(x), y) + \mathcal{P}(\eta(y), \eta(y))$. This condition ensures the projection is itself a Markov chain, with transition matrix given by

$$\forall x, y \in [n], \bar{\mathcal{P}}(x, y) := \mathcal{P}(x, y) + \mathcal{P}(x, \eta(y)) \quad (5.6)$$

This model contains (5.1) as a particular case: taking $p(x, y) := p_1(x, y - n)$ for all $x \in V_1, y \in V_2$ suffices to define p completely. Furthermore $\sigma - n$ is a uniform permutation on $[n]$. From these it is easily checked that the projected chain $\bar{\mathcal{P}}$ coincides with \mathcal{P}_0 . This applies in particular to the reversible model by Remark 5.2. Let us remark that in this case the two-lift chain may not be reversible, however its projection to V/\sim is as projecting gives back the chain defined by (5.3).

Given a probability measure μ on V , its projection $\bar{\mu}$ on $[n]$ is defined by $\bar{\mu}(x) = \mu(x) + \mu(\eta(x))$ for all $x \in [n]$. Notice that if π is invariant for \mathcal{P} , then $\bar{\pi}$ is invariant for $\bar{\mathcal{P}}$. Furthermore, total variation distance is non-increasing under projections, so

$$\|\bar{\mathcal{P}}^t(x, \cdot) - \bar{\pi}\|_{\text{TV}} \leq \|\mathcal{P}^t(x, \cdot) - \pi\|_{\text{TV}}. \quad (5.7)$$

Therefore to obtain the upper bound in Theorem 5.1, which is the most difficult part of the argument, it suffices to prove an upper bound for the two-lift. In fact, our results hold for both chains.

Now let us introduce another characteristic of the two-lift. The two-lift is by construction made of two disjoint subspaces V_1, V_2 : when at x the Markov chain stays in the same subspace with probability $p(x, \eta(x))$, and change with probability $q(x, \eta(x))$, after which it takes a step independent of η . Because of this, it may be convenient to actually distinguish between these two steps, adding transitions at half integer times, defining transition probabilities

$$\begin{aligned} \mathbf{P}[X_{t+1/2} = y \mid X_t = x] &= \begin{cases} p(x, \eta(x)) & \text{if } y = x \\ q(x, \eta(x)) & \text{if } y = \eta(x) \\ 0 & \text{otherwise} \end{cases} \\ \mathbf{P}[X_{t+1} = y \mid X_{t+1/2} = x] &= P(x, y) \end{aligned} \quad (5.8)$$

for all $t \in \mathbb{N}$. As is easily checked, the Markov chain $(X_t)_{t \in \mathbb{N}}$ evaluated at integer times exactly has the transition matrix \mathcal{P} (5.5).

5.2.3 The graph point of view

To emphasize the underlying tree structure of the Markov chains considered in this paper, these can be visualized as evolving on directed graphs.

Definition 5.1. If K is a Markov kernel on a countable state space S , the underlying graph $G = (S, E)$ is the oriented graph with vertex set S and edge set

$$E := \{(x, y) \in S \mid K(x, y) > 0\}.$$

We consider in fact multi-graphs, so that loops and multi-edges are allowed in G . In quasi-trees, we will use graphs of Markov chains as parts of a larger structure. To that

end, define a morphism of directed graphs as a map between two graphs that preserves the adjacency relations as well as the orientations of the edges. If weighted graphs are considered, such as underlying graphs of Markov chains, we may also implicitly require that weights are preserved. An isomorphism is as usual a bijective morphism.

Given an oriented edge $e = (x, y) \in E$, write $e^- := x$ and $e^+ := y$ for its initial and terminal vertex respectively. A path \mathbf{p} in G can be made either of vertices or oriented edges, which can be identified. Given a path \mathbf{p} , we write $K(\mathbf{p})$ for the product of transition probabilities along the edges of this path.

The set S is equipped with the following metric: for all $x, y \in S$, $d_K(x, y) := \inf\{l \geq 0 : K^l(x, y) > 0\}$. It is not symmetric in general because of irreversibility. It thus gives rise to several notions of balls, depending on the orientation: given $x \in V$, $r \in \mathbb{N} \cup \{+\infty\}$, the forward and backward (closed) K -balls of radius r around x are respectively the sets

$$B_K^+(x, r) := \{y \in S \mid d_K(x, y) \leq r\} \quad B_K^-(x, r) := \{y \in S \mid d_K(y, x) \leq r\}.$$

Without further precision, the K -ball of radius r around x is $B_K(x, r) := B_K^-(x, r) \cup B_K^+(x, r)$, which forgets about the orientation of edges. Note that for any $x \in S$, $B_K^+(x, \infty)$ is the set that can be reached by the chain from the state x . If x is recurrent, this is the communicating class of x .

Apply the previous definition to the case at hand: from now on, let $G = (V, E)$ be the underlying graph of the Markov chain P : the graphs $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ spanned by V_1 and V_2 respectively are the underlying graphs of P_1 and P_2 and form two disconnected components of G . Adding the edges $(x, \eta(x))$ of the uniform matching yields a random graph $G^* := (V, E^*)$, which can be seen as the underlying graph of the Markov chain $(X_t)_{t \geq 0}$, provided we take into account half-integer time steps (5.8). We introduce a terminology inspired by [92] to distinguish between these edges: an edge $(x, y) \in E^*$ is called a *small-range* edge if $(x, y) \in E$, that is if $P(x, y) > 0$ or $P(y, x) > 0$. An edge $(x, \eta(x))$ added by the random matching will be called a *long-range* edge. Note that long-range edges appear in E with both orientations. When restricting to integer times only, the Markov chain X moves along a long-range and a small-range edge at once or along a small-range edge only. Thus two vertices at \mathcal{P} -distance 1 from each other may in fact be separated by two edges in G^* . Similarly, a path in G^* without further precision will be referring to a possible trajectory of the chain \mathcal{P} . Finally, let us remark that in the reversible model edges appear with both orientations so graphs can actually be considered non-oriented.

5.2.4 Quasi-trees

We now define quasi-trees, which are designed to be an infinite approximation of the graph G^* . The same terminology and notation as in the finite setting will be used to emphasize analogies. This abuse is justified by the fact that later, both settings will be coupled so that quantities with matching terminology or notation will be equal. If necessary, we will introduce distinct notation.

Definition 5.2. A rooted quasi-tree is a 4-tuple $(\mathcal{G}, O, \eta, \iota)$ where $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an oriented graph, $O \in \mathcal{V}$ is a distinguished vertex and ι is a map $\iota : \mathcal{V} \rightarrow V$ that labels vertices of \mathcal{G} with states in V . We call $\iota(x)$ the type of the vertex $x \in V$. Finally η is a map $\eta : \mathcal{V} \rightarrow \mathcal{V}$, which satisfies:

- (i) η is an involution either of \mathcal{V} or of $\mathcal{V} \setminus \{O\}$ with no fixed point. The quasi-tree is called respectively *two-sided* or *one-sided* in these cases.
- (ii) For all $x \in \mathcal{V}$, the edge $(x, \eta(x))$ is present in \mathcal{E} with both orientations. Such edges are called long-range edges, others are called small-range edges.
- (iii) for all $x, y \in \mathcal{V}$, there is a unique family, possibly empty, of long-range edges e_1, \dots, e_k such that all paths between x and y of minimal length contain the edges e_1, \dots, e_k .

In particular, there must exist a unique sequence of long-range edges joining the root O to x . If this sequence is non-empty let x° be the the endpoint of these long-range edges which is the furthest from O . We call such a vertex a *center*. If this sequence is empty, let $x^\circ := O$, however the root is not considered a center.

The two types of edges lead in turn to two additional types of paths and distances.

1. A *small-range path* is a path made exclusively of small-range edges. Given $x, y \in V$, the *small-range distance* $d_{\text{SR}}(x, y)$ is the minimal number of edges in small-range path from x to y . The corresponding *small-range balls* are written $B_{\text{SR}}(x, r)$.
2. A *long-range path* is a sequence $e = (e_i)_{i=1}^k$ of long-range edges such that for all $i \leq k - 1$ $d_{\text{SR}}(e_i, e_{i+1}) < \infty$ (so e could be completed with small-range paths between long-range edges to obtain a genuine path in \mathcal{G}). The length $|e|$ of e is the number of edges it contain. It joins two vertices $x, y \in \mathcal{V}$ if $d_{\text{SR}}(x, e_1) < \infty$ and $d_{\text{SR}}(e_k, y) < \infty$. Given $x, y \in V$, the *long-range distance* $d_{\text{LR}}(x, y)$ is the minimal length of a long-range path between x and y . We write $B_{\text{LR}}(x, r)$ for the corresponding *long-range balls*.

As for the distance in Markov chain underlying graphs, these distances are not symmetric, so we may add + or - sign in exponent to denote forward and backward balls. From the definition, for any $x \in \mathcal{V}$, $B_{\text{SR}}^+(x, \infty) = B_{\text{LR}}^+(x, 0)$ is the set of vertices which can be joined from x by a small-range path. We call the subgraph spanned by this set the *small-range component* of x . We can now state a fourth property we require for quasi-trees:

(iv) for all $x \in \mathcal{V}$, $B_{\text{LR}}^+(x, 0)$ is isomorphic to $B_P^+(\iota(x^\circ), \infty)$,

that is, the small-range component of x is the whole set of vertices in V than can be reached from $\iota(x^\circ)$ in the graph G .

In the sequel we will refer to quasi-trees by their graph component only while keeping the other parameters implicit.

Property (iii) above clearly gives quasi-trees a tree structure. For this reason, let us introduce a few extra notations. Given $k \geq 0$, the k -th level \mathcal{V}_k is the set of vertices which are at long-range distance k from the root. Set also $\mathcal{V}_{\leq k} := \bigcup_{i \leq k} \mathcal{V}_i$. The requirement that η has no fixed point implies that every vertex x is joined to vertex $\eta(x)$ in another level. If x is a center, $\eta(x)$ is in a level below, whereas if x is not a center $\eta(x)$ is in a level above.

One-sided quasi-trees and subquasi-trees We introduced one-sided quasi-trees to consider subquasi-trees. If x is a center, the subquasi-tree \mathcal{G}_x of x is the graph spanned by the vertices y for which all paths between O and y pass through x . If x is not a center, $\eta(x)$ is and we set $\mathcal{G}_x := \mathcal{G}_{\eta(x)}$, so \mathcal{G}_x does not in fact contain x in that case. Finally the complement subgraph $\mathcal{G} \setminus \mathcal{G}_x$ is the graph spanned by vertices which can be reached from O by a long-range path that does not use the long-range edge $(x, \eta(x))$.

Non-backtracking paths, loop-erased paths, deviation, regeneration One main interest of having a genuine tree structure is the consideration of loop-erased trajectories. We thus introduce the following definitions.

Definition 5.3. Let \mathcal{G} be a quasi-tree. A long-range path $e = (e_1, \dots, e_k)$ backtracks over a distance $l \geq 1$ if it contains a subpath of distinct edges and of length l immediately followed by the reversed path, that is there exists $i \leq k - 2l + 1$ such that $e_{i+l+j} = \overline{e_{i+l-1-j}}$ for all $j \in [0, l - 1]$, where $\overline{e_j}$ denotes the edge e_j with reversed orientation. The loop-erased path $\xi(e)$ is the path obtained after erasing all backtracking steps. The long-range path e is called *non-backtracking* if $\xi(e) = e$.

To previous definition is extended to general paths by extracting the long-range path: given a general path \mathbf{p} , let $\xi(\mathbf{p})$ denote the loop-erased path formed by the long-range edges crossed by \mathbf{p} . This is a non-backtracking path, which we call the loop-erased path or loop-erased trace of \mathbf{p} . The long-range distance crossed by \mathbf{p} is the length of $\xi(\mathbf{p})$, or equivalently the long-range distance between its endpoints.

The last two definitions require integer parameters. Given $R \geq 1$, let $\mathcal{G}^{(R)}$ be the connected component of O of the subgraph of \mathcal{G} spanned by the set

$$\mathcal{V}^{(R)} := \{x \in \mathcal{V} \mid d_{\text{SR}}(x^\circ, x) < R\}. \quad (5.9)$$

A path \mathbf{p} is said to *deviate* from a small-range distance R if it is not included in $\mathcal{G}^{(R)}$.

Finally, the following notion of *regeneration edges* will be used in several places: let $L \geq 1$ be an integer and consider a path \mathbf{p} . A long-range edge e crossed by \mathbf{p} is said to be a regeneration edge for \mathbf{p} with horizon L if after the first time going through e the path crosses a long-range distance L or ends before going back to the endpoint of e which was first visited by \mathbf{p} .

Markov chains on quasi-trees Let \mathcal{G} be a quasi-tree. It is naturally the underlying graph of the Markov chain $(\mathcal{X}_t)_{t \geq 0}$ on \mathcal{G} which has transition probabilities:

$$\begin{aligned} \mathbf{P}[\mathcal{X}_{t+1/2} = y \mid \mathcal{X}_t = x] &= \begin{cases} p(\iota(x), \iota(\eta(x))) & \text{if } y = x \\ q(\iota(x), \iota(\eta(x))) & \text{if } y = \eta(x) \\ 0 & \text{otherwise} \end{cases} \\ \mathbf{P}[\mathcal{X}_{t+1} = y \mid \mathcal{X}_{t+1/2} = x] &= \begin{cases} P(\iota(x), \iota(y)) & \text{if } y^\circ = x^\circ \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (5.10)$$

for all $t \in \mathbb{N}$. The kernel of this Markov chain will be written \mathcal{P} . As for the chain X in finite environment, this chain may not be reversible but can be projected to the quotient \mathcal{G}/\sim , which identifies each $x \sim \eta(x)$. The projection is then a reversible chain in the reversible model.

5.2.5 The covering quasi-tree

Among all quasi-trees, one is very natural to consider: this is the quasi-tree obtained by using the random matching $\eta : V \rightarrow V$ to define the matching on the quasi-tree. Given $x \in V$, this is the quasi-tree $(\mathcal{G}^*(x), O, \iota, \tilde{\eta})$ defined by $\iota(O) = x$ and

$$\forall y \in \mathcal{V} : \iota(\tilde{\eta}(y)) = \eta(\iota(y)).$$

The fact that this defines a unique quasi-tree is the consequence from property (iv) in Definition 5.2, which imply the quasi-tree can be built iteratively. Starting from

the small-range of O , which is necessarily $B_P^+(x, \infty)$, the previous equation uniquely determines the long-range edges at long-range distance 0 from O and thus the whole ball $B_{\text{LR}}(O, 1)$. The process iterates to infinity to yield an infinite quasi-tree \mathcal{G}^* .

This quasi-tree has the property that it covers exactly G^* : the map ι is a surjective morphism of \mathcal{G}^* onto G^* which preserves the transition probabilities, so the Markov chain \mathcal{X} defined above projects exactly onto the chain X : for all $t \geq 0$ $\iota(\mathcal{X}_t) = X_t$ in distribution, conditional on $X_0 = x, \mathcal{X}_0 = O$.

The chain \mathcal{X} on \mathcal{G}^* will not be studied per se. We introduced it mainly to obtain easier definitions in the finite setting of objects and quantities that are naturally considered in the idealized setting of quasi-trees. Namely we extend notions of small-range, long-range paths, backtracking, etc. defined above to G^* by taking their projections under ι . An exception is the long-range distance, which in the finite setting will make sense only if restricting the quasi-tree: let $R \geq 1$ and recall the definition of the restricted quasi-tree $\mathcal{G}^{(R)}$ (5.9). Given $x \in V$ and $l \geq 0$, we define $B_{\text{LR}}^{(G,R)}(x, r)$ in G^* as the projection

$$B_{\text{LR}}^{(G,R)}(x, r) := \iota\left(B_{\text{LR}}(O, r) \cap \mathcal{G}^{(R)}(x)\right).$$

The exponent (G, R) is used to distinguish between the two settings. When not necessary, we may drop it from notation. Note from the definition that in G^* , for any vertices $x, y \in V$, $d_{\text{LR}}(x, y) = 0$ if and only if $d_{\text{SR}}(x, y) < R$.

Finally, we introduce a last definition which is specific to G^* : a *long-range cycle* is a non-deviating non-backtracking long-range path whose starting and ending point are at small-range distance at most R from each other. A subgraph of G^* is said to be *quasi-tree-like* if it does not contain any long-range cycle. A quasi-tree-like subgraph can thus be identified with a neighbourhood of the root in the covering quasi-tree. Lemma 5.1 below will establish that most vertices have in fact a quasi-tree-like neighbourhood. This relies on a bounded degree property, which is the object of the following paragraph.

Bounded degrees and transition probabilities: Assumption (H1) imply that the graph G and consequently G^*, \mathcal{G} as well, have degrees bounded uniformly in n . Together with Assumption (H2) it also implies all transition probabilities are lower bounded uniformly in n . Consequently, let Δ denote a uniform bound on all the degrees and $\delta > 0$ a uniform bound on transition probabilities, which will serve throughout the paper. It gives in particular the growth of balls in G^* and any quasi-tree \mathcal{G} : for all $x \in V$ and $l \geq 0$

$$|B_{\text{SR}}(x, l)| \leq |B_{\mathcal{G}}(x, l)| \leq \frac{\Delta^{l+1} - 1}{\Delta - 1}, \quad \left|B_{\text{LR}}^{(G,R)}(x, l)\right| \leq \Delta^R \frac{\Delta^{R(l+1)} - 1}{\Delta^R - 1}. \quad (5.11)$$

In particular, Δ^R can be thought of as the "long-range" degree.

5.2.6 Random setting

The above definitions should make pretty clear that the forthcoming proofs are based on a coupling between the finite chain X and the chain \mathcal{X} on an infinite quasi-tree. The quasi-tree in question is similar to the covering quasi-tree but contains much more independence, allowing basically to resample the matching η at each new long-range edge. It can be constructed iteratively as follows: $\iota(O)$ is taken uniformly at random in V , which determines the small-range component around O by point (iv) of the definition. Then to every vertex x whose type $\iota(x)$ is known, adjoin a long-range edge $(x, \eta(x))$ to x , with $\iota(\eta(x))$ being taken uniformly in $V \setminus V(\iota(x))$.

The following formalisation of random quasi-trees will be useful to deal with subquasi-trees and their independency properties. It is analog to the Ulam-Harris labelling for Galton-Watson trees. Let

$$\mathcal{U} := \bigcup_{k \geq 0} \{((u_i)_{i=1}^k, v) \mid (u, v) \in V^{k+1}, \forall i \in [2, k], u_i \notin V(u_{i-1}), v \notin V(u_k)\}. \quad (5.12)$$

The elements in \mathcal{U} corresponding to $k = 0$ have their first coordinate empty and are identified with V . Given two sequences u, v , write uv for their concatenation. If $u = (u_i)_{i=1}^k$ is non-empty, write $u^\flat := (u_i)_{i=1}^{k-1}$ and $\bar{u} := u_k$. A rooted quasi-tree can then be represented as a pair (\mathcal{V}, O) where $\mathcal{V} \subset \mathcal{U}$, $O \in V$ satisfy:

- (i) $\mathcal{V} \cap V = B_{\text{SR}}^+(O, \infty)$
- (ii) $\forall (u, v) \in \mathcal{V}, (u^\flat, \bar{u}) \in \mathcal{V}$.
- (iii) there exists an involution $\eta: \mathcal{V} \rightarrow \mathcal{V}$ such that for all $(u, v) \in \mathcal{V}$,
 - either $\eta(u, v) = (u^\flat, \bar{u})$, $((u, v)$ is then a center),
 - or $\eta(u, v) = (uv, \rho(u, v))$ for some $\rho(u, v) \in V \setminus V(v)$.

In the second case, for all $w \in V$, $(uv, w) \in \mathcal{V}$ if and only if $w \in B_{\text{SR}}^+(\rho(u, v), \infty)$.

It is easily checked that with this definition the set \mathcal{V} identifies with the vertex set of a quasi-tree as defined in 5.2. Note the type of a vertex $(u, v) \in \mathcal{U}$ is $\iota(u, v) = v$.

Consider now a family $(\zeta_{(u,v)})_{(u,v) \in \mathcal{U}}$ of independent random variables, where for every $(u, v) \in \mathcal{U}$, $\zeta_{(u,v)}$ is uniform in $V \setminus V(v)$. Letting U be independent and uniform in V , let \mathcal{V} be the random subset of \mathcal{U} obtained by taking $O = U$ and $\rho(u, v) := \zeta(u, v)$ in the case (u, v) is not a center. This yields the same random quasi-tree as in the iterative definition above.

The previous construction is a bit heavy and can be avoided most of the time, however it has the advantage that it allows to consider subquasi-trees in a rigorous

way: for every $x = (u, v) \in \mathcal{U}$, x being a center or not is measurable with respect to $\mathcal{G} \setminus \mathcal{G}_x$. Then recall that if $x \in \mathcal{V}$ is not a center \mathcal{G}_x does not contain x , so conditional on $\mathcal{G} \setminus \mathcal{G}_x$ and $x \in \mathcal{V}$ not being a center, \mathcal{G}_x has the law of $\mathcal{G}'_{O'}$, where (\mathcal{G}', O') is an independent copy of (\mathcal{G}, O) . If x is a center then the root of \mathcal{G}_x is already determined by $\iota(x) = v$, thus conditional and $\mathcal{G} \setminus \mathcal{G}_x$ and $x \in \mathcal{V}$ being a center, \mathcal{G}_x has the law of $\mathcal{G}'_{O'}$ conditional on $\iota(\eta(O')) = v$.

5.2.7 Coupling and sequential generation

We now explain how one can couple the Markov chain $(X_t)_{t \geq 0}$ on G^* with the Markov chain $(\mathcal{X}_t)_{t \geq 0}$ on the random infinite quasi-tree \mathcal{G} defined in the previous section.

The following procedure explores the neighbourhood of a vertex in either G^* or \mathcal{G} up to some given long-range distance.

Let $x \in V$ be the point whose long-range neighbourhood is to be explored up to long-range distance $L \geq 0$. For $t \geq 1$ we write EQ_t for the exploration queue, that is the set of vertices which remain to be explored, and D_t for the set of explored vertices. The initiation is similar for G^* and quasi-trees: $D_0 := \emptyset$; $EQ_0 := B_{\text{SR}}^+(x, R)$ in G , $EQ_0 := B_{\text{SR}}^+(x, \infty)$ in \mathcal{G} . The procedure repeats then the following steps. If one wants to explore a neighbourhood in G^* sampling is performed without replacement:

Sequential generation of G^* , sampling without replacement: for $t \geq 0$,

1. pick $y \in EQ_t$: sample $\eta(y)$ uniformly at random in $(V \setminus V(y)) \setminus D_t$,
2. add $y, \eta(y)$ to D_{t+1} and remove them from EQ_{t+1} ,
3. add all vertices $z \in B_{\text{SR}}^+(\eta(y), R) \setminus \{\eta(y)\}$ such that $z \notin D_t$ and $d_{\text{LR}}(x, z) < L$ to EQ_{t+1} .

If performed with replacement, each new value $\eta(y)$ is picked uniformly in $V \setminus V(y)$ independently of previous draws and considered a new vertex in a set \mathcal{V} . Specifically the procedure becomes:

Sequential generation of the quasi-tree \mathcal{G} , sampling with replacement: for $t \geq 0$,

- 1'. pick $y \in EQ_t$: sample $\eta(y)$ uniformly at random in $V \setminus V(y)$,
- 2'. add $y, \eta(y)$ to D_{t+1} and remove y from EQ_{t+1} ,
- 3'. add all vertices $z \in B_{\text{SR}}^+(\eta(y), R) \setminus \{\eta(y)\}$ such that $d_{\text{LR}}(x, z) < L$ to EQ_{t+1} .

Since the constraint $z \notin D_t$ has been removed in step 3', vertices which would in G be already explored are here added to the exploration queue and thus considered new vertices. The environment explored is the L long-range ball of a quasi-tree as in Definition 5.2. Taking $L = \infty$ would thus yield a realization of the infinite random quasi-tree \mathcal{G} . Finally, the procedures can be adapted to explore balls $B_{\mathcal{P}}$ in G^* or $B_{\mathcal{P}}$ in \mathcal{G} .

Sequential generation along trajectories Under the annealed law, the two processes can be generated together with the environment. Later our goal will be to couple weights, which require the exploration of the whole L -long range neighbourhood around the trajectory. The sequential generation of the environment along trajectories thus consists in the following steps. Consider for instance the finite setting G : let the chain be started at $X_0 := x \in V$, and $D_0 := \emptyset$. Then for all $t \geq 0$:

- a. Explore the long-range L -neighbourhood of X_t with the first procedure described above.
- b. This determines completely the transition probabilities (5.8) at the state X_t , allowing to sample X_{t+1} .

If one considers the second procedure one obtains instead $(\mathcal{X}_t)_{t \geq 0}$. One can very naturally couple the two procedures and hence both processes using rejection sampling: for each $y \in EQ_t$ sample $\eta(y)$ uniformly in $V \setminus V(y)$ and use it for step 1' in the generation of \mathcal{G} . If in addition $B_{\text{SR}}(\eta(y), R) \cap D_t = \emptyset$, it can also be used for step 1 in the generation of G^* . Otherwise, make a second draw with the first procedure. This rejection sampling scheme yields a coupling of $(X_t)_{t \geq 0}$ and $(\mathcal{X}_t)_{t \geq 0}$ until the first time a newly revealed small-range ball $B_{\text{SR}}(\eta(y), R)$ contain vertices that have already been explored, that is when a long-range cycle appears around the trajectory:

$$\tau_{\text{coup}} := \inf \left\{ t \geq 0 \mid \bigcup_{k=0}^t B_{\text{LR}}(X_k, L) \text{ contains a long-range cycle} \right\}. \quad (5.13)$$

As a first application of the coupling and the first moment argument, we give the following Lemma.

Lemma 5.1. *There exists C_R, C_L large enough such that for $R = C_R \log \log n, L = C_L \log \log n$, for all $x \in V, \varepsilon \in (0, 3/10)$ and $t = O(n^{3/10-\varepsilon})$,*

- (i) $\mathbf{P}_x [\exists s \leq t : B_{\mathcal{P}}(X_s, \lfloor \log n / (5 \log \Delta) \rfloor)]$ is not quasi-tree-like $= o_{\mathbb{P}}(n^{-\varepsilon}) = o_{\mathbb{P}}(1)$.
- (ii) $\mathbf{P}_x [\exists s \leq t : B_{\text{LR}}^{(G,R)}(X_s, L)$ is not quasi-tree-like $] = o_{\mathbb{P}}(n^{-\varepsilon}) = o_{\mathbb{P}}(1)$. Thus

$$\mathbf{P}_x [\tau_{\text{coup}} \leq t] = o_{\mathbb{P}}(n^{-\varepsilon}) = o_{\mathbb{P}}(1).$$

Proof. Let the chain be started at $x \in V$ and $t = o(n^{3/10-\varepsilon})$ for $\varepsilon \in (0, 3/10)$. Write $k := \lfloor \log n / (5 \log \Delta) \rfloor$. By (5.11), the number of vertices contained in a \mathcal{P} -ball of radius k is $O(\Delta^k)$. This implies $\cup_{s \leq t} B_{\mathcal{P}}(X_s, k)$ contains $m = O((t+1)\Delta^k)$ vertices. Therefore, the exploration procedure along the path $X_0 \cdots X_t$ requires at most m draws of values $\eta(y)$, each having probability at most $m\Delta^R/n$ that the small-range ball $B_{\text{SR}}(\eta(y), R)$ contains already explored vertices. Hence the number of long-range cycles in $\cup_{s \leq t} B_{\mathcal{P}}(X_s, k)$ is stochastically upper bounded by a binomial $\text{Bin}(m, m\Delta^R/n)$ so that by Markov's inequality

$$\begin{aligned} \mathbb{P}_x [\exists s \leq t : B(X_s, k) \text{ is not quasi-tree-like}] &\leq \frac{m^2 \Delta^R}{n} = O\left(\frac{(t+1)^2 \Delta^R \Delta^{2k}}{n}\right) \\ &= O\left(t^2 n^{-3/5+o(1)}\right) = o(n^{-2\varepsilon+o(1)}) = o(n^{-\varepsilon}) \end{aligned}$$

by the choice of $t = O(n^{3/10-\varepsilon})$ and $R = O(\log \log n)$. This bound on the annealed probability implies the quenched result (i) by the first moment argument (5.4). The result (ii) is proved similarly, noting that Δ^{RL} is still $n^{o(1)}$ if $R, L = O(\log \log n)$ (in this case we can take t up to $n^{1/2-\varepsilon}$). \square

5.3 The entropic method: main arguments

5.3.1 Nice trajectories

Our application of the entropic method consists in finding a definition of nice trajectories designed to be typical trajectories and have their probability concentrating around the mean. The latter will come from a constraint about an entropy-like quantity for the chain, which arises from comparing the trajectories of the finite chain X with the loop-erased trace of \mathcal{X} in the (infinite) quasi-tree setting. Other properties will thus be required from nice paths, whose goal is basically to make them close to being non-backtracking trajectories in a quasi-tree. All in all, the defining properties of a nice trajectory will essentially be that:

- (i) it is contained in a quasi-tree-like portion of the graph, the endpoint having a quasi-tree-like neighbourhood up to \mathcal{P} -distance $\lfloor \alpha \log n \rfloor$ for some $\alpha > 0$ (Lemmas 5.1 and 5.2),
- (ii) it does not deviate or backtrack too much, and contains sufficiently many regeneration edges (Lemma 5.3),
- (iii) the drift, ie the long-range distance travelled, and entropy concentrate on this trajectory (Proposition 5.1).

As mentioned in the proof outline, nice trajectories will be likely to be followed by the chain only if it is started at a typical state. To derive the upper bound on the worst-case mixing time, we will thus establish that for an arbitrary starting state the trajectory becomes nice only after some time, namely after $s = (\log n)^a$ steps for some $a > 0$ or $s = O(\log \log n)$ steps in the reversible case. This is why in the sequel we state different results with a "typical, non-shifted" version valid from time 0 and a "worst-case, shifted" version valid for arbitrary starting states from time s . For the first important property of nice trajectories, the typical version has in fact already been proved in Lemma 5.1. The worst-case version is given in the following.

Lemma 5.2. (i) *There exists $a \geq 1$ such that for all $s \geq (\log n)^a$ and $t = o(n^{1/16})$, for all $C_R, C_L > 0$, letting $R := C_R \log \log n, L := C_L \log \log n$,*

$$\max_{x \in V} \mathbf{P}_x \left[\exists t' \in [s, s+t] : B_{\text{LR}}^{(G^*, R)}(X_{t'}, L) \text{ is not quasi-tree-like} \right] = o_{\mathbb{P}}(1),$$

In the reversible model, the conclusion holds for $s \geq C \log \log n$ and a large enough constant $C > 0$.

(ii) *There exists $a \geq 1, \alpha > 0$ such that for all $s \geq (\log n)^a$,*

$$\max_{x \in V} \mathbf{P}_x [B_{\mathcal{D}}(X_s, [\alpha \log n]) \text{ is not quasi-tree-like}] = o_{\mathbb{P}}(1). \quad (5.14)$$

In the reversible model: for all $C > 0$, there exists $\alpha > 0$ such that for any $t \geq C \log n$,

$$\max_{x \in V} \mathbf{P}_x [B_{\mathcal{D}}(X_t, [\alpha \log n]) \text{ is not quasi-tree-like}] = o_{\mathbb{P}}(1)$$

For the second property we recall the notions of deviation, backtracking and regeneration edges are given in Definition 5.3.

Lemma 5.3. *Let $\Gamma(R, L, M)$ denote the set of paths \mathbf{p} in G^* such that \mathbf{p} does not deviate from a small-range distance more than R , backtrack over a long-range distance L or contain a subpath of length M without a regeneration edge. There exist constants $\kappa \geq 1, a \geq 1, C_R, C_L > 0$ such that for $L = C_L \log \log n, R = C_R \log \log n$ and $M = (\log \log n)^\kappa$, for all $t = O(\log n)$, the following holds:*

(i) *for all $x \in V$:*

$$\mathbf{P}_x [(X_0 \cdots X_t) \in \Gamma(R, L, M)] = 1 - o_{\mathbb{P}}(1).$$

Furthermore, for any $b > 0$, this probability is even $1 - o_{\mathbb{P}}((\log n)^{-b})$ provided the constants C_R, C_L are allowed to depend on b .

(ii) for all $s \geq (\log n)^a$:

$$\min_{x \in V} \mathbf{P}_x [(X_s \cdots X_{s+t}) \in \Gamma(R, L, M)] = 1 - o_{\mathbb{P}}(1).$$

In the reversible model, the conclusion holds for $s \geq C \log \log n$ and a large enough constant $C > 0$.

Remark 5.4. Notice that for any trajectorial event A

$$\begin{aligned} \mathbf{P}_x [(X_s, X_{s+1}, \dots) \in A] &= \sum_{y \in V} \mathbf{P}_x [X_s = y] \mathbf{P}_y [(X_0, X_1, \dots) \in A] \\ &\leq \max_{y \in V} \mathbf{P}_y [(X_0, X_1, \dots) \in A]. \end{aligned}$$

Thus if we prove A holds from time t with probability $1 - o_{\mathbb{P}}(1)$ uniformly over the starting state this automatically extends to larger times $t \geq s$. Thus in the above lemmas it will be sufficient to prove the case $s = (\log n)^a$ to establish the case $s \geq (\log n)^a$ for instance.

5.3.2 Concentration of drift and entropy

The third and main property of nice trajectories consists in the concentration of an entropy-like quantity for the finite chain. To that end, we define weights as follows.

Throughout the paper, we will use the letter τ for a variety of stopping times. It should be clear from the context what the notation refers to. If x is an element or a set τ_x will generally denote the hitting time of x . If $l \geq 1$, τ_l will generally denote the time a certain distance l is reached. For all $l \geq 0$, consider in this section

$$\tau_l := \inf\{t \geq 0 \mid |\xi(X_0 \cdots X_t)| = l\}$$

and fix $R, L \geq 1$ for the rest of this section. Let

$$\begin{aligned} \tau_{\text{SR}}^{(R)} &:= \inf\{t \geq 0 \mid (X_0 \cdots X_t) \text{ deviates from a small-range distance } R\} \\ \tau_{\text{NB}}^{(L)} &:= \inf\{t \geq 0 \mid (X_0 \cdots X_t) \text{ backtracks over a long-range distance } L\}. \end{aligned} \tag{5.15}$$

These are stopping times which depend on R, L . By construction if \mathbf{p} is a path in $\Gamma(R, L, M)$ as defined in Lemma 5.3, the stopping time $\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)}$ does not occur on the trajectory \mathbf{p} . Given a long-range edge e and $x \in V$ such that $d_{\text{SR}}(x, e^-) < R$, define the weights

$$\begin{aligned} w_{x,R,L}(e) &:= \mathbf{P}_x \left[\xi(X_0 \cdots X_{\tau_L})_1 = e, \tau_L < \tau_{\text{SR}}^{(R)} \right] \\ w_{R,L}(e \mid x) &:= \mathbf{P}_x \left[\xi(X_0 \cdots X_{\tau_L})_1 = e \mid \tau_L < \tau_{\eta(x)} \wedge \tau_{\text{SR}}^{(R)} \right] \end{aligned} \tag{5.16}$$

Here $\xi(X_0 \cdots X_{\tau_L})_1$ denotes the first edge of $\xi(X_0 \cdots X_{\tau_L})$. Then for a non-backtracking long-range path $\xi = \xi_1 \cdots \xi_k$, set

$$w_{x,R,L}(\xi) := w_{x,R,L}(\xi_1) \prod_{i=2}^k w_{R,L}(\xi_i \mid \xi_{i-1}^+) \quad (5.17)$$

where empty products are by convention equal to 1. The notation is consistent with the the identification of edges with paths of length 1.

Remark 5.5. Note that for fixed $x \in V$, $\sum_e w_{x,R,L}(e) \leq 1$ and $\sum_e w_{R,L}(e \mid x) \leq 1$ where the sum is over all long-range edges. By extension the sum of weights over all non-backtracking paths starting from x is at most 1.

Given a sequence $u = (u_i)_{i \leq l}$ of length l and $k \geq 1$, we write $(u)_{\leq k} := (u_i)_{i \leq k}$ for the sequence truncated at length k . The following lemma show that weights are good proxies for measuring the probability that the loop-erased trace follows a given non-backtracking path.

Lemma 5.4. *Let $x \in V$ and $k \geq 1$ be an integer. Suppose that $B_{\text{LR}}^{(G,R)}(x, k)$ is quasi-tree-like. Then for all non-backtracking long-range path ξ of length k , started in $B_{\text{SR}}^+(x, R)$*

$$\mathbf{P}_x \left[\begin{array}{l} \xi(X_0 \cdots X_{\tau_{k+L-1}})_{\leq k} = \xi, \\ \tau_{k+L-1} < \tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \end{array} \right] \leq w_{x,R,L}(\xi) \leq \mathbf{P}_x \left[\begin{array}{l} \xi(X_0 \cdots X_{\tau_{k+L-1}})_{\leq k} = \xi, \\ \tau_{k+L-1} < \tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \end{array} \right] + u(\xi) \quad (5.18)$$

where $u(\xi) \geq 0$ is such that

$$\sum_{\xi} u(\xi) \leq \mathbf{P}_x \left[\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq \tau_{k+L-1} \right],$$

the sum being over non-backtracking long-range paths of length k from x .

Proof. The proof is by induction on $k \geq 1$. For $k = 1$ the inequalities are in fact equalities by definition, since $\tau_{\text{NB}}^{(L)} > \tau_L$ necessarily.

To ease notation, drop the parameters R, L from the stopping times and weights for the rest of the proof. Suppose that the result holds for $k \geq 1$ and let ξ be of length $k+1$. Let $e := \xi_k, f := \xi_{k+1}$ and write L_k for the last time t that $|\xi(X_0 \cdots X_t)| = k$ after τ_k and before τ_{k+L} or coming back to e^- . Then having $\xi(X_0 \cdots X_{\tau_{k+L-1}})_{\leq k} = (\xi)_{\leq k}$ requires that $X_{L_k} = e^+$, after which the chain crosses a long-range distance $L-1$ without backtracking to e^- . Thus by the induction hypothesis,

$$\begin{aligned} w_x((\xi)_{\leq k}) &\geq \mathbf{P}_x [\xi(X_0 \cdots X_{\tau_{k+L-1}})_{\leq k} = \xi, \tau_{k+L-1} < \tau_{\text{SR}} \wedge \tau_{\text{NB}}] \\ &= \mathbf{P}_x [X_{L_k} = e^+, L_k < \tau_{\text{SR}} \wedge \tau_{\text{NB}}] \mathbf{P}_{e^+} [\tau_{L-1} < \tau_{e^-} \wedge \tau_{\text{SR}}]. \end{aligned}$$

On the other hand,

$$w(f | e^+) = \frac{\mathbf{P}_{e^+} [\xi(X_0 \cdots X_{\tau_L})_1 = f, \tau_L < \tau_{e^-} \wedge \tau_{\text{SR}}]}{\mathbf{P}_{e^+} [\tau_L < \tau_{e^-} \wedge \tau_{\text{SR}}]}.$$

Since $\mathbf{P}_{e^+} [\tau_{L-1} < \tau_{e^-} \wedge \tau_{\text{SR}}] \geq \mathbf{P}_{e^+} [\tau_L < \tau_{e^-} \wedge \tau_{\text{SR}}]$, we deduce that

$$\begin{aligned} w_x(\xi) &= w_x((\xi)_{\leq k}) w(f | e^+) \\ &\geq \mathbf{P}_x [X_{L_k} = e^+, \tau_{\text{SR}} \wedge \tau_{\text{NB}} > L_k] \mathbf{P}_{e^+} [\xi(X_0 \cdots X_{\tau_L})_1 = f, \tau_L < \tau_{\text{SR}} \wedge \tau_{e^-}]. \end{aligned}$$

This is the probability that after L_k , the chain directly crosses a long-range distance L using the edge f , without reaching the boundary of a small-range ball of radius R and without coming back to e^- . Hence on this event $\xi(X_0 \cdots X_{\tau_{k+L}})_{k+1} = f$, with $\tau_{k+L} < \tau_{\text{SR}} \wedge \tau_{\text{NB}}$, which proves the lower bound.

For the upper bound, the induction hypothesis yields this time

$$w_x((\xi)_{\leq k}) \leq \mathbf{P}_x [X_{L_k} = e^+, L_k < \tau_{\text{SR}} \wedge \tau_{\text{NB}}] \mathbf{P}_{e^+} [\tau_{L-1} < \tau_{e^-} \wedge \tau_{\text{SR}}] + u((\xi)_{\leq k}).$$

Then use that

$$\mathbf{P}_{e^+} [\tau_{L-1} < \tau_{e^-} \wedge \tau_{\text{SR}}] \leq \mathbf{P}_{e^+} [\tau_L < \tau_{e^-} \wedge \tau_{\text{SR}}] + \mathbf{P}_{e^+} [\tau_{L-1} < \tau_{e^-} \wedge \tau_{\text{SR}} \leq \tau_L]$$

to bound

$$\begin{aligned} w_x(\xi) &\leq \mathbf{P}_x [X_{L_k} = e^+] \mathbf{P}_{e^+} [\xi(X_0 \cdots X_{\tau_L})_1 = f, \tau_L < \tau_{\text{SR}} \wedge \tau_{e^-}] \\ &\quad + \mathbf{P}_x [X_{L_k} = e^+] \mathbf{P}_{e^+} [\tau_{L-1} < \tau_{e^-} \wedge \tau_{\text{SR}} \leq \tau_L] w(f | e^+) + u((\xi)_{\leq k}) w(f | e^+). \end{aligned}$$

The first term is that of the lower bound. Regroup the two other terms as $u(\xi)$. Since weights sum to 1, summing over ξ , which involves in particular summing over e and f , yields that

$$\sum_{\xi} u(\xi) \leq \sum_e \mathbf{P}_x [X_{L_k} = e^+] \mathbf{P}_{e^+} [\tau_{L-1} < \tau_{e^-} \wedge \tau_{\text{SR}} \leq \tau_L] + \mathbf{P}_x [\tau_{\text{SR}} \wedge \tau_{\text{NB}} \leq \tau_{k+L-1}].$$

Observe now the first term corresponds to backtracking or deviating after reaching level $k+L-1$ but before reaching level $k+L$ from x . Thus the two terms correspond to disjoint events which both imply $\tau_{\text{SR}} \wedge \tau_{\text{NB}} \leq \tau_{k+L}$, hence the upper bound. \square

The first part of the proof of Theorem 5.1 consists in proving the following quenched concentration phenomenon.

Proposition 5.1. *There exist $d, h = \Theta(1)$ and constants $a \geq 1, C_R, C_L > 0$ for which the following holds. Letting $R := C_R \log \log n, L := C_L \log \log n$, for all $\varepsilon \in (0, 1)$, there exist $C_{\text{LR}}(a, \varepsilon), C_h(a, \varepsilon)$ such that for all $t \gg 1, t = O(\log n)$:*

(i) for all $x \in V$, with high probability

$$\mathbf{P}_x \left[\begin{array}{l} |\xi(X_0 \cdots X_t) - dt| \leq C_{\text{LR}} \sqrt{t} \\ |-\log w_{X_0, R, L}(\xi(X_0 \cdots X_t)) - ht| \leq C_h \sqrt{t} \end{array} \right] \geq 1 - \varepsilon.$$

(ii) for all $s \geq (\log n)^a$, with high probability,

$$\min_{x \in V} \mathbf{P}_x \left[\begin{array}{l} |\xi(X_s \cdots X_{s+t}) - dt| \leq C_{\text{LR}} \sqrt{t} \\ |-\log w_{X_s, R, L}(\xi(X_s \cdots X_{s+t})) - ht| \leq C_h \sqrt{t} \end{array} \right] \geq 1 - \varepsilon.$$

In the reversible model, the conclusion holds for $s \geq C \log \log n$ and a large enough constant $C > 0$.

5.3.3 Concentration of nice paths

The second part of the argument consists in finding a set of "nice trajectories", defined in particular so that they satisfy the entropic and drift requirements given by Proposition 5.1. Eventually the drift and entropic concentration are used to show that the probability of following a nice trajectory concentrates around its mean. The latter can be computed, providing an approximate stationary distribution $\hat{\pi}$ for \mathcal{P} on V .

Given $u \in V$, $L \geq 1$ consider again $\tau_L := \inf\{t \geq 0 \mid |\xi(X_0 \cdots X_t)| = L\}$ and

$$\mathbf{Q}_u^{(L)} := \mathbf{P}[\cdot \mid X_{1/2} = u, \tau_{\eta(u)} > \tau_L] \quad (5.19)$$

In words, this measure considers trajectories immediately after a regeneration time. By Lemma 5.3, if one takes $L = \Theta(\log \log n)$, the conditioning by $\tau_L < \tau_{X_0}$ essentially forbids the chain to come back at all to u on a time scale $O(\log n)$. If ν is a probability measure on V , write $\mathbf{Q}_\nu^{(L)} := \sum_{u \in V} \nu(u) \mathbf{Q}_u^{(L)}$ and $\mathbf{E}_{\mathbf{Q}_\nu}^{(L)}$ for the expectation with respect to this measure. All in all, the second part of the argument is summarized in the following proposition.

Proposition 5.2. *There exist a deterministic probability measure ν on V , a deterministic $s_0 = \Theta(\log n)$, constants $C_L, C_0 > 0, a, \kappa \geq 1$ and for all $x, y \in V, t \in \mathbb{N}$ a set $\mathfrak{N}^t(x, y)$ of length t paths between x and y for which the following holds. Let $L := C_L \log \log n, M := (\log \log n)^\kappa$ and write $\mathcal{P}_{\mathfrak{N}}^t(x, y) := \sum_{\mathfrak{p} \in \mathfrak{N}^t(x, y)} \mathcal{P}(\mathfrak{p})$. Consider the random probability measure*

$$\hat{\pi}(v) := \frac{1}{\mathbf{E}_{\mathbf{Q}_\nu^{(L)}}[T_1 \wedge M]} \sum_{r=0}^M \mathbf{Q}_\nu^{(L)} [X_{r+s_0} = v, r < T_1 \leq M] \quad (5.20)$$

where T_1 denotes the first regeneration time with horizon $L = C_L \log \log n$. For all $\varepsilon \in (0, 1)$ there exists $C(\varepsilon) > 0$ such that for $t = \log n/h + C(\varepsilon)\sqrt{\log n}$,

(i) for all $x \in V$, with high probability,

$$\sum_{y \in V} \mathcal{P}_{\mathfrak{N}}^t(x, y) \geq 1 - \varepsilon,$$

(ii) for all $s \geq (\log n)^a$ or $s \geq C \log \log n$ in the reversible model, with high probability

$$\min_{x \in V} \sum_{y \in V} \mathcal{P}^s \mathcal{P}_{\mathfrak{N}}^t(x, y) \geq 1 - \varepsilon.$$

(iii) there exists $c = (c_v)_{v \in V}$ such that $\sum_{v \in V} c_v = o_{\mathbb{P}}(1)$ and with high probability, for all $x, y \in V$,

$$\mathcal{P}_{\mathfrak{N}}^t(x, y) \leq (1 + \varepsilon)\hat{\pi}(y) + c(y) + \frac{\varepsilon}{n},$$

(iv) the measure $\hat{\pi}$ can be decomposed as $\hat{\pi} = \hat{\pi}_1 + \hat{\pi}_2$ with

$$\|\hat{\pi}_1\|_2^2 = o_{\mathbb{P}}((\log n)^b/n), \quad \hat{\pi}_2(V) = o_{\mathbb{P}}(1)$$

for some $b > 0$.

Proof of Theorems 5.1, 5.3 and 5.4. Recall \mathcal{P} is the transition matrix of the two-lift chain on V , which projects to a transition matrix $\bar{\mathcal{P}}$ on $[n]$.

Start with the upper bound on the mixing time. Let $\varepsilon > 0$ and $t := \log n/h + C(\varepsilon)\sqrt{\log n}$. Since $\mathcal{P} \geq \mathcal{P}_{\mathfrak{N}}$ entry-wise, for all $x \in V$ and $s \geq 0$,

$$\begin{aligned} \|\hat{\pi} - \mathcal{P}^{s+t}(x, \cdot)\|_{\text{TV}} &= \sum_{z \in V} [\hat{\pi}(z) - \mathcal{P}^{s+t}(x, z)]_+ \leq \sum_{y, z \in V} \mathcal{P}^s(x, y) [\hat{\pi}(z) - \mathcal{P}^t(y, z)]_+ \\ &\leq \sum_{y, z \in V} \mathcal{P}^s(x, y) \left[(1 + \varepsilon)\hat{\pi}(z) + c(z) + \frac{\varepsilon}{n} - \mathcal{P}_{\mathfrak{N}}^t(y, z) \right]_+. \end{aligned}$$

Point (iii) of the above proposition implies that with high probability, the right hand side summands are non-negative for all $x \in V$, so the sum can be computed to obtain that with high probability,

$$\|\hat{\pi} - \mathcal{P}^{s+t}(x, \cdot)\|_{\text{TV}} \leq 1 - \mathcal{P}^s \mathcal{P}_{\mathfrak{N}}^t(x, y) + 3\varepsilon.$$

Using point (ii), if $s \geq (\log n)^a$ or $s \geq C \log \log n$ in the reversible model, with high probability

$$\max_{x \in V} \|\mathcal{P}^{s+t}(x, \cdot) - \hat{\pi}\|_{\text{TV}} \leq 4\varepsilon$$

which projects by (5.7) to

$$\max_{x \in V} \|\bar{\mathcal{P}}^{s+t}(x, \cdot) - \bar{\pi}\|_{\text{TV}} \leq 4\varepsilon.$$

with $\bar{\pi}$ the projection onto $[n]$ of $\hat{\pi}$. Since this estimate is uniform in the starting state, it extends to any starting distribution and particularly to a stationary distribution. Thus for any stationary distribution π of $\bar{\mathcal{P}}$,

$$\|\pi - \bar{\pi}\|_{\text{TV}} \leq 4\varepsilon. \quad (5.21)$$

and from triangular inequality we obtain that with high probability

$$\max_{x \in [n]} \|\bar{\mathcal{P}}^{s+t}(x, \cdot) - \pi\|_{\text{TV}} \leq 8\varepsilon.$$

Since this is valid for any invariant measure, the latter must be unique. This proves in particular the irreducibility and aperiodicity of the chain and establishes Theorem 5.3 as well as the upper bound of Theorem 5.4, since $O(\log \log n) = o(\sqrt{\log n})$.

On the other hand for a fixed $x \in V$, taking $s = 0$ in the above bounds yields and using point (i) that with high probability,

$$\|\bar{\mathcal{P}}^t(x, \cdot) - \pi\|_{\text{TV}} \leq 8\varepsilon,$$

proving the upper bound in Theorem 5.1.

Let us prove the lower bound. For all $t \geq 0, \theta > 0$ and $x, y \in [n]$,

$$\begin{aligned} \bar{\mathcal{P}}^t(x, y) &= \mathcal{P}^t(x, y) + \mathcal{P}^t(x, \eta(y)) \\ &\geq \mathbf{P}_x [X_t \in \{y, \eta(y)\}, w_{x,R,L}(\xi(X_0 \cdots X_t)) \leq \theta]. \end{aligned}$$

If equality holds, then

$$\bar{\pi}(y) - \mathbf{P}_x [X_t \in \{y, \eta(y)\}, w_{x,R,L}(\xi(X_0 \cdots X_t)) \leq \theta] \leq [\bar{\pi}(y) - \bar{\mathcal{P}}^t(x, y)]_+$$

If equality does not hold, there must exist a non-backtracking long-range path ξ between x and y or x and $\eta(y)$ for which $w(\xi) > \theta$, in which case

$$\begin{aligned} \bar{\pi}(y) - \mathbf{P}_x [X_t \in \{y, \eta(y)\}, w_{x,R,L}(\xi(X_0 \cdots X_t)) \leq \theta] &\leq \hat{\pi}(y) \mathbb{1}_{\exists \xi: w_{x,R,L}(\xi) > \theta} \\ &\quad + \hat{\pi}(\eta(y)) \mathbb{1}_{\exists \xi: w_{x,R,L}(\xi) > \theta}. \end{aligned}$$

We did not precise where ξ lies to ease notation in the indicator functions but it depends on y and $\eta(y)$ respectively. Combining the two inequalities we obtain that in either case

$$\begin{aligned} \bar{\pi}(y) - \mathbf{P}_x [X_t \in \{y, \eta(y)\}, w_{x,R,L}(\xi(X_0 \cdots X_t)) \leq \theta] &\leq [\bar{\pi}(y) - \bar{\mathcal{P}}^t(x, y)]_+ + \hat{\pi}(y) \mathbb{1}_{\exists \xi: w_{x,R,L}(\xi) > \theta} \\ &\quad + \hat{\pi}(\eta(y)) \mathbb{1}_{\exists \xi: w_{x,R,L}(\xi) > \theta}. \end{aligned}$$

Decompose now $\hat{\pi} = \hat{\pi}_1 + \hat{\pi}_2$ as in Proposition 5.2. From (iv), summing over $y \in [n]$ in the previous inequality yields

$$\mathbf{P}_x [w_{x,R,L}(\xi(X_0 \cdots X_t)) > \theta] \leq \|\bar{\mathcal{P}}^t(x, \cdot) - \bar{\pi}\|_{\text{TV}} + \sum_{y \in V} \hat{\pi}_1(y) \mathbb{1}_{\exists \xi: w_{x,R,L}(\xi) > \theta} + o_{\mathbb{P}}(1).$$

By Cauchy-Schwarz inequality,

$$\sum_{y \in V} \hat{\pi}_1(y) \mathbb{1}_{(\exists \xi : w_{x,R,L}(\xi) > \theta)} \leq \left(\sum_{y \in V} \hat{\pi}_1(y)^2 \right)^{1/2} \left(\sum_{\xi} \mathbb{1}_{(w_{x,R,L}(\xi) > \theta)} \right)^{1/2},$$

where the second sum is over non-backtracking long-range paths from x . By Remark 5.5 weights sum up to at most 1 hence so this sum contains at most θ^{-1} positive terms and

$$\mathbf{P}_x [X_t = y, w_{x,R,L}(\xi(X_0 \cdots X_t)) > \theta] \leq \|\bar{\mathcal{P}}^t(x, \cdot) - \bar{\pi}\|_{\text{TV}} + \sqrt{\frac{1}{\theta} \sum_{y \in V} \hat{\pi}_1(y)^2} + o_{\mathbb{P}}(1). \quad (5.22)$$

To complete the proof, let $\varepsilon \in (0, 1)$ and specialize to $t := \log n/h - C_1 \sqrt{\log n}$ and $\theta := n^{-1} \exp(C_2 \sqrt{\log n})$ for some $C_1(\varepsilon), C_2(\varepsilon) > 0$. Choosing the constant C_1 large enough, $\exp(-th - C_h(\varepsilon)\sqrt{t}) = n^{-1} \exp((C_1 - C_h/h)\sqrt{\log n} - o(\sqrt{\log n})) > \theta$ for large enough n , hence Proposition 5.1 implies that the left hand-side of (5.22) is at least $1 - \varepsilon$ with high probability. On the other hand, (iv) shows that the square-root in the right hand side is $o_{\mathbb{P}}(1)$. All in all, this proves that with high probability

$$\|\bar{\mathcal{P}}^t(x, \cdot) - \bar{\pi}\|_{\text{TV}} \geq 1 - 2\varepsilon.$$

Finally by (5.21) we deduce

$$\|\bar{\mathcal{P}}^t(x, \cdot) - \bar{\pi}\|_{\text{TV}} \geq 1 - 6\varepsilon$$

with high probability. □

Remark 5.6. From the previous proof, the decomposition of $\hat{\pi}$ given by (iv) also extends to the unique invariant measures of \mathcal{P} and $\bar{\mathcal{P}}$: $\pi = \pi_1 + \pi_2$ with $\|\pi_1\|_2^2 = o_{\mathbb{P}}((\log n)^b/n)$ and $\pi_2([n]) = o_{\mathbb{P}}(1)$.

Remark 5.7. Let us also comment on how the proof accomodates the superposition of simple graphs in the reversible model. Observe that if two edges are aligned under the matching η , ie if there exist $x, y \in V$ such that $P(x, y) > 0$ and $P(\eta(x), \eta(y)) > 0$, these constitute an obstruction to the quasi-tree likeness of the neighbourhood of x . As the nice trajectories considered in Proposition 5.2 will require having a quasi-tree-like neighbourhood, they avoid in particular these edges, so it makes no difference in the end to adjust the transition probabilities of these edges or not.

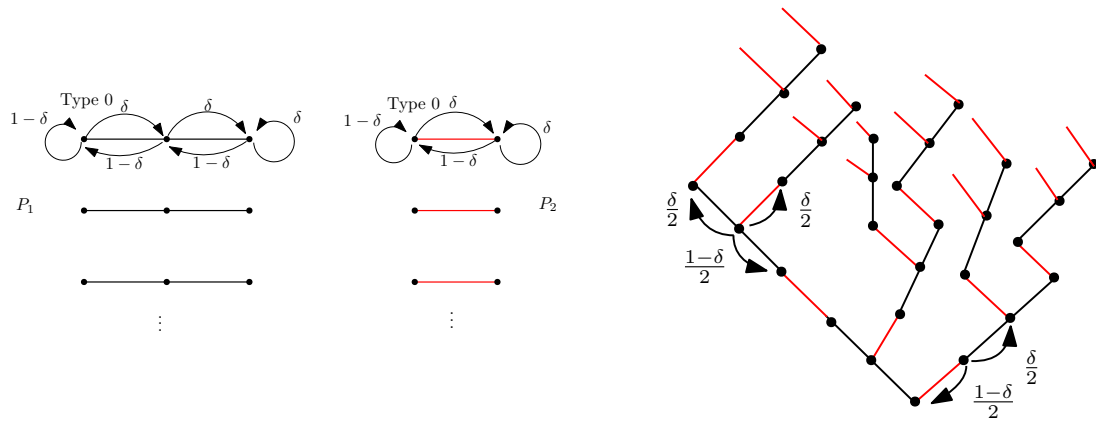


Figure 5.1: The finite chains P_1, P_2 and the quasi-tree T with all centers of type 0, after identifying each $x \simeq \eta(x)$. The probability to remain at a vertex is not represented.

5.4 Counter-example to uniform entropic time

In this section we prove Theorem 5.2. Let $n \geq 1$ and $\delta \in (0, 1)$. Consider the reversible Markov chains Q_1, Q_2 on the segments $\{0, 1, 2\}$ and $\{0, 1\}$ which go from left to right with probability δ , from right to left with probability $1 - \delta$ and stay at the extremities with remaining probability. Let P_1 be the transition matrix on $[6n]$ corresponding to $2n$ disjoint copies of Q_1 , that is $P_1(i, j) = Q_1(i \bmod 3, j \bmod 3)$ for all $i, j \in [6n]$. Similarly let P_2 be the matrix corresponding to $3n$ copies of Q_2 . Our counter-example to uniform mixing time is obtained with the matrix $\bar{\mathcal{P}} := 1/2(P_1 + SP_2S^{-1})$ and small δ . On this example, it will be exceptionally be clearer to forget about the two-lift and half-integer time transitions. We thus identify each x with $\eta(x)$, both in the finite and quasi-tree setting, which has the effect of pruning the long-range edges in the graphs, and we use the notations $G^*, \mathcal{G}, \mathcal{X}$ to refer to the objects obtained after this operation.

Consider first the quasi-tree \mathcal{G} that would be obtained from P_1 and P_2 : because the underlying graphs G_1, G_2 are forests, all realizations of quasi-trees are in fact genuine trees. On the other hand, if δ is small, the Markov chains on segments are heavily biased towards 0. Say a vertex $x \in \mathcal{G}$ has type 0 if the map ι identifies x with the leftmost vertex of a segment. Then consider the quasi-tree obtained T if all centers in \mathcal{G} have type 0. As one can check, this is a periodic weighted tree T where except the root all vertices have their transition probabilities depending only on whether the degree is 2 or 3, see Figure 5.1. Notice that for sufficiently small δ , this chain becomes recurrent. Of course this configuration occurs with probability 0, however a finite portion of T can arise with positive probability. Coming back to the finite graph G^* , which resembles locally a quasi-tree, this probability is in fact sufficiently large that G^* also contains such neighbourhoods, which trap the chain for a long time.

Lemma 5.5. *There exists $\beta > 0$ such that for all $\delta \in (0, 1)$, with high probability there exists $x \in [6n]$ for which*

$$\mathbf{P}_x [X_t \notin B_{\mathcal{P}}(x, \beta \log \log n)] = o(1)$$

for all $t \leq (\log n)^{\beta \log(2\delta/(1-\delta))}$.

Proof. Let $l := \beta \log \log n$. In the present setting, the degrees of G^* and \mathcal{G} are clearly bounded by $\Delta = 3$. Consequently the exploration of a \mathcal{P} -ball of radius l around any $x \in V$ reveals at most $O(\Delta^l) = O((\log n)^{\beta \log 3})$ vertices. Suppose we reveal such neighbourhoods around vertices x_0, \dots, x_k , each time picking a vertex disjoint from the previously explored neighbourhoods. As explained in Section 5.2.7, this exploration can be performed with replacement at a total variation cost which is $o(1)$ as long as $k\Delta^l = o(\sqrt{n})$. If this holds we can thus make the neighbourhoods of all x_k coincide with trees with probability $1 - o(1)$. Now in the quasi-tree \mathcal{G} , each center has type 0 with probability at least $1/3$, so the probability that all centers up to depth l have type 0 is at least $3^{-\Delta^{l+1}}$. Consequently, the probability that no vertex among x_0, \dots, x_k has its neighbourhood isomorphic to $T_{\leq l}$, the subtree of T up to depth l , is upper bounded by

$$(1 - 3^{-\Delta^{l+1}})^k \leq \exp(-k3^{-\Delta^{l+1}}) = \exp(-k9^{-(\log n)^{\beta \log 3}})$$

For $\beta > 0$ small enough, we can take for instance $k = n^{1/4}$ so that the previous probability is $o(1)$ and in the same time $k\Delta^l = o(\sqrt{n})$ so the approximation with the quasi-tree is justified. This proves that with high probability, there exists $x_0 \in [6n]$ such that $B_{\mathcal{P}}(x_0, l)$ coincides with $T_{\leq l}$.

It remains to show that the Markov chain $(X_t)_{t \geq 0}$ is very unlikely to escape $B_{\mathcal{P}}(x, l)$ by time e^{cl} . It suffices to prove the statement for the Markov chain $(\mathcal{X}_t)_{t \geq 0}$ on T . This is done by considering the maximal drift of the chain: the vertices which maximize the probability to go away from the root are vertices of degree 3. If \mathcal{X}_t is at such a vertex, it goes away from the root with probability δ , moves towards the root with probability $(1 - \delta)/2$ and stays put with remaining probability. Consequently, the distance between \mathcal{X}_t and the root is stochastically upper bounded by a lazy random walk on the segment $[0, l]$ which goes left with probability $(1 - \delta)/2$ and right with probability δ . By the classical Gambler's ruin problem, this random walk has probability at most $O((2\delta/(1 - \delta))^l)$ to go from 1 to l before returning to 0. In turn this implies that the hitting time τ_l of level l by \mathcal{X}_t dominates stochastically a geometric random variable

with parameter $r := (2\delta/(1-\delta))^l$. In the end we deduce that if $t = t(n) = o(1/r)$

$$\begin{aligned} \mathbb{P}[X_t \notin B_{\bar{\mathcal{P}}}(x_0, l)] &\leq \mathbb{P}[\tau_l \leq t] \\ &\leq 1 - (1-r)^t \\ &= rt + o(rt) = o(1). \end{aligned}$$

Since $1/r = (\log n)^{\beta \log(2\delta/(1-\delta))}$, this proves the result. \square

Proof of Theorem 5.2. Let $a > 1$, $t := (\log n)^a$ and $\varepsilon \in (0, 1)$. By Theorem 5.1 the chain is irreducible and aperiodic with high probability. Let π be the unique stationary measure of $\bar{\mathcal{P}}$. Thanks to the lemma, we can find $\beta > 0$ and $\delta > 0$ small enough for which there exists with high probability a state $x \in [n]$ satisfying $\mathbf{P}_x[X_t \notin B] = o(1)$, where $B := B_{\bar{\mathcal{P}}}(x, \beta \log \log n)$. Then recall the decomposition of $\pi = \pi_1 + \pi_2$ given by Remark 5.6. Since the ball B contains at most $O(\Delta^{\beta \log \log n}) = n^{o(1)}$ vertices, Cauchy-Schwarz inequality implies

$$\pi_1(B) \leq \sqrt{|B| \sum_{u \in B} \pi_1(u)^2} = n^{-1/2+o_{\mathbb{P}}(1)},$$

while $\pi_2(B) \leq \pi_2([n]) = o_{\mathbb{P}}(1)$. Consequently

$$\bar{\mathcal{P}}^t(x, B) - \pi(B) = 1 - o_{\mathbb{P}}(1)$$

which implies $t_{\text{mix}}(x, \varepsilon) \geq (\log n)^a$. \square

Remark 5.8. As was noticed above for δ sufficiently small the Markov chain $(\mathcal{X}_t)_{t \geq 0}$ on the tree T becomes sufficiently biased towards the root to be recurrent. On the other hand, the Borel-Cantelli lemma implies that the quasi-tree \mathcal{G} contains a.s. infinitely many copies of $T_{\leq l}$ for any l . The existence of such "trapping neighbourhoods" which could potentially slow down the chain by an important amount is a problem that will arise in the next section where we study asymptotic properties of $(\mathcal{X}_t)_{t \geq 0}$. In particular the spectral radius of this Markov chain may be equal to 1, despite the tree-like aspect of the graph.

5.5 Analysis on the quasi-tree on quasi-trees I: escape probabilities

The objective of the three following sections is to prove the concentration of the drift and entropy along with the other nice properties of Section 5.3 for the Markov chain $(\mathcal{X}_t)_{t \geq 0}$ on a random infinite quasi-tree \mathcal{G} . This will allow us later to deduce the corresponding statements for the chain X thanks to the coupling presented in Section 5.2.7. As in [22, 92], the argument is based on the existence of regeneration times,

which are times at which \mathcal{X}_t visits a long-range edge for the first and last time. A first step towards this objective is to lower bound the probability of escaping to infinity in the quasi-tree. Because of the potential trapping phenomenon mentioned in Remark 5.8, this requires non trivial arguments, which is the object of this whole section.

5.5.1 Escape probabilities

We recall \mathcal{G} denotes a random rooted quasi-tree which under \mathbb{P} has the law of the random quasi-tree described in Section 5.2.6. Its vertex set is \mathcal{V} .

Definition 5.4. Given a non-center vertex $x \in \mathcal{V}$, let

$$q_{\text{Esc}}(x) := \mathbf{P}_x [\forall t \geq 1 : \mathcal{X}_t \in \mathcal{G}_x].$$

be the quenched probability that the chain enters the subquasi-tree of x and never leaves it. If x is a center, it is useful to also consider starting at time $1/2$ and let

$$q_{\text{Esc}}(x) := \mathbf{P}_x [\forall t \geq 0 : \mathcal{X}_t \in \mathcal{G}_x] \wedge \mathbf{P} [\tau_{\eta(x)} = \infty \mid \mathcal{X}_{1/2} = x].$$

We call these quantities the *escape probability* at x .

Remark 5.9. Note that if x is not a center,

$$q_{\text{Esc}}(x) \geq q(x, \eta(x)) \mathbf{P} [\tau_x = \infty \mid \mathcal{X}_{1/2} = \eta(x)].$$

Similarly, if x is a center, Assumption (H3) asserts there exists $y \neq x$ in the same small-range component as x , which is thus not a center, so that

$$q_{\text{Esc}}(x) \geq p(x, \eta(x)) P(x, y) q_{\text{Esc}}(y).$$

By Assumption (H2) the entries of p and q are bounded. Thus to lower bound escape probabilities, it matters little to consider a center vertex or not, and to start at integer or half-integer time.

Furthermore, recall the Ulam labelling of Section 5.2.6. For all $x \in \mathcal{U}$, if $x \in \mathcal{V}$ is not a center then the probability $q_{\text{Esc}}(x)$ is measurable with respect to \mathcal{G}_x and $\iota(x)$ only. Hence

$$\mathbb{P} [q_{\text{Esc}}(x) \in \cdot \mid x \in \mathcal{V}, \mathcal{G} \setminus \mathcal{G}_x] = \mathbb{P} [q_{\text{Esc}}(O) \in \cdot \mid \iota(O) = \iota(x)]. \quad (5.23)$$

If x is a center,

$$\mathbb{P} [q_{\text{Esc}}(x) \in \cdot \mid x \in \mathcal{V}, \mathcal{G} \setminus \mathcal{G}_x] = \mathbb{P} [q_{\text{Esc}}(\eta(O)) \in \cdot \mid \iota(O) = \iota(\eta(x)), \iota(\eta(O)) = \iota(x)].$$

Thus the law of escape probabilities is entirely determined by the case of the root, conditional on $\iota(O)$ and $\iota(\eta(O))$. To ease the notation, we will sometimes keep ι implicit in the sequel and write only $O, \eta(O)$ when conditioning on the types of these vertices.

In [92], the authors prove that the escape probability is lower bounded uniformly in n , conditional on \mathcal{G} . This comes from the fact that the momentum of the chain, that is the expected change in distance to the root, is everywhere non-negative, and even positive for most vertices. This argument is very specific to the case of the simple random walk: with biases on the transition probabilities the momentum can be negative. In the reversible model, we can still prove a uniform lower bound on escape probabilities thanks to a comparison argument.

Proposition 5.3. *In the reversible model of Section 5.1.2,*

$$\inf_{x \in \mathcal{V}} q_{\text{Esc}}(x) = \Theta(1).$$

In the general model, a uniform lower bound on escape probabilities may not be achievable. The counter example of Section 5.4 and Remark 5.8 show that in the corresponding quasi-tree,

$$\inf_{x \in \mathcal{V}} q_{\text{Esc}}(x) = 0.$$

This phenomenon, which was already present in [22], is ultimately what prevents uniform cutoff. Cutoff from typical vertices remains possible as we will argue that "trapping neighbourhoods", where escape probabilities tend to 0, are sufficiently rare that they are essentially ignored from the chain, provided it is started at a typical vertex. The starting point is the following lemma, also proved with comparison arguments. The result is stated for the root but from Remark 5.9 it extends to other subquasi-trees.

Lemma 5.6. *For all types of $O, \eta(O)$,*

$$\mathbb{E}[q_{\text{Esc}}(O) \mid O, \eta(O)] = \Theta(1).$$

Equivalently, there exists a constant $q_0 > 0$ such that for all types of $O, \eta(O)$,

$$\mathbb{P}[q_{\text{Esc}}(O) \geq q_0 \mid O, \eta(O)] \geq q_0. \quad (5.24)$$

Establishing this lemma is the main difficulty so the proof is deferred to the next sections. Using the tree structure of \mathcal{G} , the previous lemma can be bootstrapped to yield a much better control on the behaviour of the escape probability at typical vertices.

Proposition 5.4. *There exist constants $q_0, \delta \in (0, 1)$ such that for all types of $O, \eta(O)$ and $k \geq 0$,*

$$\mathbb{P}[q_{\text{Esc}}(x) < q_0 \delta^{4k} \mid O, \eta(O)] \leq q_0^{2^k}. \quad (5.25)$$

Proof of Proposition 5.4. Recall that from assumptions (H1) and (H2) we suppose all transition probabilities are lower bounded by a constant δ . This is also the constant δ

of the proposition. By Assumption (H3), every vertex $y \in \mathcal{V}$ has at least one vertex at small-range distance 1 and two vertices if $\iota(y) \in V_1$. Consequently, for any realization of \mathcal{G}_x and $k \geq 0$, there are at least 2^k centers at distance at long-range distance $2k$ from x in \mathcal{G}_x which can be joined from x in less than $4k$ steps by \mathcal{X}_t . Since transition probabilities are bounded below by δ , it suffices that one of these vertices y has $q_{\text{Esc}}(y) \geq q_0$ to obtain that $q_{\text{Esc}}(O) \geq q_0 \delta^{4k}$. Now since these vertices are centers and at the same long-range distance from x their respective subtrees are disjoint and thus independent. The probability that all y have $q_{\text{Esc}}(y) < q_0$ is thus upper bounded by $q_0^{2^k}$ by Lemma 5.6. \square

From the previous proposition, we also deduce the following. Below, a path refers to a possible trajectory of the chain \mathcal{P} and its \mathcal{P} -length is the corresponding number of transitions.

Proposition 5.5. *For all $C > 0$ there exist constants $q_0, \alpha \in (0, 1]$ such that the following holds: for all $r \geq 0$, with probability at least $1 - \exp(-Cr)$ conditional on the types of $O, \eta(O)$, every path in \mathcal{G} of \mathcal{P} -length r starting from O contains a proportion at least α of vertices x such that $q_{\text{Esc}}(x) \geq q_0$.*

The proof is concluded with an argument similar in spirit to Lemma 2.2 in [52] (initially due to Grimmett and Kesten [81]) and used to prove Lemma 2.3 in [22].

Proof of Prop 5.5. We reason conditional on $\iota(O)$ and $\iota(\eta(O))$ so we can assume these are fixed. We distinguish the case of large and small paths, for which it suffices to find one state with lower bounded escape probability. Namely if $r < 10$, Lemma 5.6 shows that $q_{\text{Esc}}(O) \geq q_0$ with positive probability, in which case every path of length at most 10 contains a proportion at least $1/10$ vertices with escape probability lower bounded by q_0 .

For larger paths, since \mathcal{G} is random, we use the Ulam labelling to consider vertices independently of the realization of \mathcal{G} . Given $q_0 > 0$ to be determined and $x \in \mathcal{U}$, let $A_x = \mathbb{1}_{x \in \mathcal{V}, q_{\text{Esc}}(x) \geq q_0}$. Consider now a self-avoiding path \mathbf{p} in \mathcal{U} started at $\iota(O)$ which can be the realization of a path of \mathcal{P} -length $r \geq 10$ in \mathcal{G} , starting at O . If \mathbf{p} is indeed a path in \mathcal{G} , let $K_{\mathbf{p}}$ be the set of vertices which are at small-range distance at most 1 from \mathbf{p} but are not endpoints of a long-range edge crossed by \mathbf{p} . Let us give an example for clarity: a path in \mathcal{U} may for instance contain vertices $u_0, u_1, (u_1, v_1), (u_1, v_2) \in \mathcal{U}$ such that $u_0 = \iota(O)$, $P(u_0, u_1) > 0$, $V(v_1) \neq V(v_2)$ and $P(v_1, v_2) > 0$. If $\eta(u_1)$ is such that $P(\eta(u_1), v_1) > 0$, this path becomes a genuine path in \mathcal{G} of \mathcal{P} -length 3, with $K(\mathbf{p}) = \{u_0, (u_1, v_1), (u_1, v_2)\}$ unless $\eta(u_1) \in \{v_1, v_2\}$ in which case one needs to discard the corresponding vertex. Define $A(\mathbf{p}) := \sum_{x \in K_{\mathbf{p}}} A_x$. Then by construction vertices

which are in $K_{\mathbf{p}}$ must have disjoint subquasi-trees, so conditional on $\mathbf{p} \in \mathcal{G}$ $A(\mathbf{p})$ is a sum of independent Bernoulli variables. Let us lower bound the number of these variables. Since the path is self-avoiding \mathbf{p} contains $r + 1$ vertices. Supposing k is the number of long-range edges crossed by \mathbf{p} , $r + 1 - 2k$ vertices do not belong to a long-range edge of \mathbf{p} . $k + 1$ is also the number of small-range components crossed by \mathbf{p} . Now every other of these small-range component is isomorphic to a component of V_1 , which from assumption (H3) contains at least three vertices. By the self-avoiding property, there is at least one vertex of such a small-range component which is in $K_{\mathbf{p}}$. Hence

$$m := |K_{\mathbf{p}}| \geq \max(r + 1 - 2k, \lfloor (k + 1)/2 \rfloor) \geq \frac{r}{5} - 1 \geq 1.$$

On the other hand, for all $q \in (0, 1)$ Proposition 5.4 and Remark 5.9 imply the existence of a value of $q_0 > 0$, independent of n , such that conditional on $x \in \mathcal{V}$, the Bernoulli variable A_x has parameter lower bounded by q . Thus for any path \mathbf{p} in \mathcal{U} , conditional on it being in \mathcal{G} , $A(\mathbf{p})$ dominates a Binomial(m, q). From Chernoff's bound, if Z is such a binomial variable, for any $\alpha, \theta > 0$

$$\begin{aligned} \mathbb{P}[Z < \alpha m] &= \mathbb{P}[e^{-\theta Z} > e^{-\theta \alpha m}] \\ &\leq e^{\theta \alpha m + \log \mathbb{E}[e^{-\theta Z}]} \\ &= e^{m(\theta \alpha + \log(1 - q + qe^{-\theta}))}. \end{aligned}$$

The base of the exponential can be made arbitrarily small by letting $q \rightarrow 1$, $\theta \rightarrow \infty$ and $\alpha \rightarrow 0$, namely for all $\varepsilon \in (0, 1)$, for all $q > 1 - \varepsilon$, there exists $\alpha \in (0, 1)$ such that for all $m \geq 0$,

$$\mathbb{P}[Z < \alpha m] \leq \varepsilon^m. \quad (5.26)$$

Hence, combining the previous observations, for all $\varepsilon > 0$ one can find $q_0 > 0$ and $\alpha > 0$, all independent of n , so that for all self-avoiding path \mathbf{p} of length r starting from $\iota(O)$ in \mathcal{U} .

$$\mathbb{P}[A(\mathbf{p}) < \alpha r \mid \iota(O), \iota(\eta(O)), \mathbf{p} \in \mathcal{G}] \leq \varepsilon^r. \quad (5.27)$$

Since the number of self-avoiding paths of \mathcal{P} -length r in \mathcal{G} is bounded by $O(\Delta^r)$ thanks to (5.11), we can bound

$$\begin{aligned} &\mathbb{P}[\exists \mathbf{p} \subset \mathcal{G} : |\mathbf{p}|_{\mathcal{P}} = r, A(\mathbf{p}) < \alpha r \mid O, \eta(O)] \\ &= \mathbb{E} \left[\sum_{\substack{\mathbf{p} \subset \mathcal{U} \\ \mathbf{p} \subset \mathcal{G}, |\mathbf{p}|_{\mathcal{P}} = r}} \mathbb{1}_{\mathbf{p} \subset \mathcal{G}} \mathbb{P}[A(\mathbf{p}) < \alpha r \mid \iota(O), \iota(\eta(O)), \mathbf{p} \in \mathcal{G}] \mid O, \eta(O) \right] \\ &\leq O((\Delta \varepsilon)^r) \end{aligned}$$

where $\mathbf{p} \subset \mathcal{U}, \mathcal{G}$ denotes a self-avoiding path of \mathcal{U} or \mathcal{G} respectively. Thus given $C > 0$, $A(\mathbf{p}) \geq \alpha r$ holds simultaneously for all paths with probability $1 - (\varepsilon \Delta)^r = 1 - e^{-Cr}$

choosing ε and then q_0, α accordingly. This almost proves the result, except that the vertices considered in the random variables $A(\mathbf{p})$ may not be on the path \mathbf{p} . However even if the proportion αr of vertices with lower bounded escape probability may lie partly outside \mathbf{p} , this consists of vertices which can be reached in at most one step from \mathbf{p} . Hence up to modifying the constant q_0 we can in the end deduce that every path \mathbf{p} contains a proportion αr of vertices with lower bounded escape probability. \square

Remark 5.10. The proof actually shows a further property: with probability $1 - e^{-Cr}$ every path \mathbf{p} contains a constant proportion of vertices for which the probability to escape to infinity outside \mathbf{p} is lower bounded. This can be useful to argue that the chain is unlikely to follow a given path.

5.5.2 Escape probabilities: reversible model

The goal is now to establish Proposition 5.3. Some of the results proved here will also be used for the general model, when we will compare $(\mathcal{X}_t)_{t \geq 0}$ with a reversible chain. The proof is based on methods to prove transience of a Markov chain, which is fundamental for our purpose and is already not clear in our model. The reversibility assumption augments considerably the available toolbox. In particular Rayleigh's monotonicity principle is a well-known result which states that the effective conductance to infinity increases monotonically with individual conductances (see [128][Chapter 2]). While this result is generally used qualitatively to establish transience or recurrence of a given chain, it is used here to obtain a quantitative comparison between escape probabilities.

Let us recall briefly the notions we are going to use. We refer to [128, Chapters 2, 3, 5] for a detailed account. Let $(\mathcal{Y}_t)_{t \geq 0}$ be an irreducible, reversible Markov chain on a state space W , with reversible measure μ . As explained in Section 5.1.2 it can be represented as a random walk on an electrical network (G_W, c) , with G_W a graph with vertex set W and c is a family of conductances on the edges of G_W . We write hitting times as τ , return times as τ^+ , with the corresponding subset or vertex as index. A classical result for reversible chains is that for $a \in W$ and $Z \subset W$, the probability to reach before Z before returning to a can be expressed as

$$\mu(a)\mathbb{P}_a[\tau_Z < \tau_a^+] = \mathcal{C}(a \leftrightarrow Z)$$

where $\mathcal{C}(a \leftrightarrow Z)$ is the effective conductance between a and Z , which can be computed using network reductions. If W is infinite, we can very well take $Z = \{\infty\}$ thanks to a limit argument, to obtain

$$\mu(a)\mathbb{P}_a[\tau_a^+ = \infty] = \mathcal{C}(a \leftrightarrow \infty) \tag{5.28}$$

In particular the chain is transient if and only if $\mathcal{C}(a \leftrightarrow \infty) > 0$. Effective conductances satisfy a simple yet powerful monotonicity property, in that they are monotonous with respect to individual conductances. This is called Rayleigh's monotonicity principle, see p.35 of [128]. This monotonicity can be made quantitative as follows. The proof is identical to that of [128, p.35] of the same reference.

Lemma 5.7. *Let G_W be an infinite connected graph with vertex set W and two sets of conductances c_1, c_2 . Suppose there exists $\lambda > 0$ such that $c_1 \geq \lambda c_2$ edge-wise. Then for all $a \in W$*

$$\mathcal{C}_1(a \leftrightarrow \infty) \geq \lambda \mathcal{C}_2(a \leftrightarrow \infty).$$

Remark 5.11. If $G' \subset G_W$ is a subgraph, one can always consider the Markov chain on G' to have state space W as well by setting zero conductances outside G' . The lemma thus gives in particular quantitative comparisons between a reversible chain and the chain restricted to a subnetwork.

The last required notion is that of branching number. The branching number $\text{br } T$ is a parameter that can be associated to any tree T that basically counts the average number of children per vertex. For instance for a d -regular tree, $d \geq 1$, the branching number is $d - 1$. We refer to [128] for a more general definition. We will only need the following facts, which can be found in Chapters 3 and 5:

- Proposition 5.6.**
1. *Given a tree T and $k \geq 1$, let $T^{(k)}$ be the tree obtained by keeping only the vertices of T which are at depth a multiple of k from the root, and where vertices are joined by an edge if one is the ancestor of the other in T . Then $\text{br } T^{(k)} = (\text{br } T)^k$.*
 2. *The simple random walk on a tree T is transient if and only if $\text{br } T > 1$*
 3. *Conditional on non-extinction, the branching number of a supercritical Galton-Watson tree is a.s. equal to its average offspring.*

Proof of Proposition 5.3. We prove that the root has uniformly lower bounded escape probability, but the same arguments can be applied to any vertex of the quasi-tree. Recall the equivalence relation $x \simeq \eta(x)$ on \mathcal{V} which identifies the endpoints of each long-range edge. Note that the chain \mathcal{X} is not reversible but only its projection $\bar{\mathcal{X}}$ to \mathcal{V}/\sim . Clearly, lower bounding the escape probabilities of \mathcal{X} or $\bar{\mathcal{X}}$ is equivalent. Consider the graph obtained by pruning the long-range edges of \mathcal{G}_O and adding an edge (\dagger, O) between an extra vertex \dagger and the root. Then consider the Markov chain $\tilde{\mathcal{X}}$ which goes from \dagger to O with probability 1 and otherwise has the same transition probabilities as $\bar{\mathcal{X}}$, with the probability of going from O to \dagger being that of $\bar{\mathcal{X}}$ leaving the image of \mathcal{G}_O

in the quotient. Since $\bar{\mathcal{X}}$ is reversible, the chain $\tilde{\mathcal{X}}$ is also a reversible chain. Letting $\tau_{\dagger}, \tau_{\dagger}^+$ denote the hitting and return time to \dagger of the chain $\tilde{\mathcal{X}}$, (5.28) implies

$$\mathbf{P}_O[\tau_{\dagger} = \infty] = \mathbf{P}_{\dagger}[\tau_{\dagger}^+ = \infty] = \frac{\mathcal{C}_{\tilde{\mathcal{X}}}(\dagger \leftrightarrow \infty)}{\mu(\dagger)}$$

where $\mathcal{C}_{\tilde{\mathcal{X}}}(\dagger \leftrightarrow \infty)$ is the effective conductance for the chain $\tilde{\mathcal{X}}$, and μ is the invariant measure defined by the conductances \tilde{c} of $\tilde{\mathcal{X}}$, that is $\mu(x) = \sum_{y \in \mathcal{V}} \tilde{c}(x, y)$ (μ does not need to be a probability measure here). By assumptions (H1) and (H2), conductances and degrees are bounded uniformly in n thus so is $\mu(\dagger)$ and it suffices to lower bound the effective conductance. Now let \mathcal{V}' be the set of vertices in \mathcal{G}_O that are at small-range distance at most 2 from their centers and consider any spanning tree \mathcal{T} of the subgraph spanned by \mathcal{V}' . Let $\tilde{\mathcal{T}}$ be the graph spanned by its projection in \mathcal{V}/\sim together with the edge (\dagger, O) , which remains a tree, and $\mathcal{C}_{\tilde{\mathcal{T}}}(\dagger \leftrightarrow \infty)$ the effective conductance of the simple random walk on $\tilde{\mathcal{T}}$. Then as conductances of $\tilde{\mathcal{X}}$ are bounded Lemma 5.7 implies that

$$\mathcal{C}_{\tilde{\mathcal{X}}}(\dagger \leftrightarrow \infty) \geq \lambda \mathcal{C}_{\tilde{\mathcal{T}}}(\dagger \leftrightarrow \infty)$$

where $\lambda > 0$ is a constant independent of n .

We claim now $\mathcal{C}_{\tilde{\mathcal{T}}}(\dagger \leftrightarrow \infty)$ is bounded away from 0 by a universal constant, which will prove the result. Observe that the tree \mathcal{T} must contain every long-range edge leaving from a vertex of \mathcal{V}' . Hence \mathcal{T} has no leaves and every vertex other than O has degree at least 2. Furthermore, from Assumption (H3), each small-range component in $\tilde{\mathcal{G}}$ contains at least two vertices and every other which identifies with a component of V_1 actually contains at least three vertices. In these V_1 components, there is thus a vertex which is connected to a long-range edge and at least two other vertices. Consequently, we can see that every sequence of three consecutive vertices in $\tilde{\mathcal{T}}$ contains at least one vertex with degree more than 3. This implies that the power tree $\tilde{\mathcal{T}}^{(3)}$, as defined in Proposition 5.6, contains the 2-forward regular tree T_2 , ie the tree where every vertex has 2 children. From Proposition 5.6 $\tilde{\mathcal{T}}$ contains thus a subtree with branching number $2^{1/3} > 1$, so the simple random walk on $\tilde{\mathcal{T}}$ is transient. Using again Lemma 5.7, we can in the end lower bound $\mathcal{C}_{\tilde{\mathcal{T}}}(O \leftrightarrow \infty)$ by a universal constant. \square

Remark 5.12. The proof shows more generally that any reversible Markov chain supported by a quasi-tree, (whether it makes half-integer time steps or not) is transient and has uniformly lower bounded escape probabilities, provided the quasi-tree is sufficiently branching.

5.5.3 Escape probabilities: general model

The proof for the general model also relies on a comparison with a reversible chain. For this comparison to be valid, we will need Assumption (H4). We start with the simpler case where the property of (H4) is satisfied by all vertices, ie there exists $l \geq 1$ such that

$$\forall x, y \in V \quad P(x, y) > 0 \Leftrightarrow (I + P)^l(y, x) > 0. \quad (\text{H4}')$$

Note this implies in particular that all states are recurrent.

Comparison of Green Functions

The comparison will only be possible on a random subgraph of \mathcal{G} . The starting point is Theorem 2.25 in [159], which proves that transience of a non-reversible Markov chain can be deduced from that of a reversible chain, provided their transition probabilities and their invariant measures can be compared up to multiplicative factors. As for Rayleigh's monotonicity principle, this qualitative result can be made quantitative as it based on a comparison between Green functions, which is the consequence of the following lemma.

Lemma 5.8 ([159, Lemma 2.24]). *Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and let T_1, T_2 be two invertible linear operators on H such that:*

- (i) T_1 is self-adjoint,
- (ii) $\langle T_2 f, f \rangle \geq \langle T_1 f, f \rangle \geq 0$ for all $f \in H$.

Then for all $f \in H$,

$$\langle T_1^{-1} f, f \rangle \geq \langle T_2^{-1} f, f \rangle.$$

By Remark 5.9 it suffices to lower bound $\mathbf{P}[\tau_O = \infty \mid \mathcal{X}_{1/2} = \eta(O)]$. Consider $\tilde{\mathcal{G}} = \{O\} \cup \mathcal{G}_O$ and let \mathcal{P}_0 denote the transition kernel of the chain on $\tilde{\mathcal{G}}$ which goes from O to $\eta(O)$ with probability 1 at integer time steps, stays at O at half-integer time steps, and otherwise has the transitions of \mathcal{X} . Let

$$G_{\mathcal{P}_0}(x, y) := \sum_{t \in \mathbb{N}} \mathcal{P}_0^t(x, y)$$

be the Green function of \mathcal{P}_0 . $G_{\mathcal{P}_0}(O, O)$ is the expected number of visits to O , which has a geometric distribution of parameter $q_{\text{Esc}}(\eta(O))$ since the chain automatically goes from O to $\eta(O)$. Hence

$$G_{\mathcal{P}_0}(O, O) = \mathbf{P}[\tau_O = \infty \mid \mathcal{X}_{1/2} = \eta(O)]^{-1} \quad (5.29)$$

so it is equivalent to lower bound escape probabilities and upper bound the Green function. Lemma 5.8 will allow to do so by comparison, expressing the Green function as $G_{\mathcal{P}_0} = \lim_{z \rightarrow 1} (I - z\mathcal{P}_0)^{-1}$.

We will apply the lemma to compare the Markov chain $(\mathcal{X}_t)_{t \geq 0}$ with basically the simple random walk on a subgraph of the quasi-tree. To this end, we will need to show that the transition probabilities as well as an invariant measure can be lower bounded pointwise by those of the simple random walk, up to multiplicative factor. Since the graph \mathcal{G} has bounded degrees by (5.11), the transition probabilities and invariant measure of the simple random walk are both of constant order. By Assumption (H1) transition probabilities of $(\mathcal{X}_t)_{t \geq 0}$ are also of constant order however these transitions may occur in only one direction. From the hypothesis done in this section (H4'), this issue is solved by considering a bounded power of a lazy version of the chain. However when it comes to the invariant measure, the counter example from Section 5.4 shows that we can expect the infimum of the invariant measure to be 0, which is why comparison is only possible on a subgraph. Since the invariant measure can be taken up to multiplicative constant, comparison with the simple random walk requires finding a subgraph where the invariant measure is of the same order or larger than the measure of the root. In particular we need to find explicitly one such invariant measure for $(\mathcal{X}_t)_{t \geq 0}$. This can be done thanks to the tree structure of \mathcal{G} .

Invariant measure of $(\mathcal{X}_t)_{t \geq 0}$

The idea is based on Kolmogorov's criterion for reversibility: for a reversible chain, an invariant measure can be determined explicitly from the detailed balance equation, which allows to propagate the value of the measure from a state x to another state y by considering any path between x and y . Since the invariant measure, which does not need to be a probability, is determined up to multiplicative constant anyway, this gives an algorithmic way to compute an invariant measure. For a non-reversible chain this procedure does not work as multiple paths between vertices may give different values of the measure (the absence of this phenomenon is precisely Kolmogorov's condition for reversibility). However if there are no multiple paths, in other words if the underlying graph is a tree, and transitions can be reversed, then we can find such a reversible measure. Now in our setting, quasi-trees have a large-scale tree structure arising from the long-range edges, while cycles can only be present in small-range components. We can thus combine invariant measures of small-range components to obtain invariance with respect to transitions inside a component, with the previous idea to obtain invariance along long-range edges as well.

Let π_P be an invariant measure of P which puts positive mass on every commu-

nicating class. Assumption (H4') implies that P has no transient states so $\pi_P(x) > 0$ for all $x \in V$. Extend π_P to \mathcal{G} by setting $\pi_P(x) := \pi_P(\iota(x))$ for all x . The invariant measure ν is described in terms of the following weights: for every center x , set

$$a_x := \frac{\pi_P(\eta(x))q(\eta(x), x)}{\pi_P(x)q(x, \eta(x))} \quad (5.30)$$

Then define

$$Z(x) := \prod_{i=1}^k a_{x_i} \quad (5.31)$$

where $x_1, \dots, x_k = x$ is the unique sequence of centers that lie on any path in \mathcal{G} from O to x , and $Z(O) := 1$. We extend Z to non-center vertices by $Z(x) := Z(x^\circ)$.

Lemma 5.9. *Under Assumption (H4') the measure ν on \mathcal{V} defined by*

$$\nu(x) := Z(x)\pi_P(x) \quad (5.32)$$

is an invariant measure of the Markov chain $(\mathcal{X}_t)_{t \geq 0}$.

Proof. Recall \mathcal{P} denotes the transition kernel of the Markov chain $(\mathcal{X}_t)_{t \geq 0}$ and consider $x, y \in \mathcal{V}$. Having $\mathcal{P}(x, y) > 0$ requires that either x is in the same small-range component as y , which occurs if and only if $x^\circ = y^\circ$, or $x = \eta(y^\circ)$ is in a lower level, or x is in a higher level, that is $x = \eta(z)$ for $z \neq y^\circ$ in the small-range component of y . In the second case, (5.30) implies

$$\begin{aligned} \nu(\eta(y^\circ))\mathcal{P}(\eta(y^\circ), y) &= Z(\eta(y^\circ))\pi_P(\eta(y^\circ))q(\eta(y^\circ), y^\circ)P(y^\circ, y) \\ &= Z(y^\circ)\pi_P(y^\circ)q(y^\circ, \eta(y^\circ))P(y^\circ, y) \end{aligned}$$

Similarly if $z \neq y^\circ$ is in the small-range component of y ,

$$\begin{aligned} \nu(\eta(z))\mathcal{P}(\eta(z), y) &= Z(y^\circ)a_{\eta(z)}\pi_P(\eta(z))q(\eta(z), z)P(z, y) \\ &= Z(y^\circ)\pi_P(z)q(z, \eta(z))P(z, y). \end{aligned}$$

Thus these two cases can be regrouped together as

$$\sum_{x: x^\circ \neq y^\circ} \nu(x)\mathcal{P}(x, y) = \sum_{x: x^\circ = y^\circ} Z(y^\circ)\pi_P(x)q(x, \eta(x))P(x, y).$$

Then using that $p + q \equiv 1$ and the invariance of π_P with respect to P , we deduce

$$\begin{aligned} \sum_x \nu(x)\mathcal{P}(x, y) &= \sum_{x: x^\circ = y^\circ} Z(y^\circ)\pi_P(x)p(x, \eta(x))P(x, y) + Z(y^\circ)\pi_P(x)q(x, \eta(x))P(x, y) \\ &= Z(y^\circ) \sum_{x \in V(y)} \pi_P(x)P(x, y) \\ &= Z(y^\circ)\pi_P(y) \\ &= \nu(y). \end{aligned}$$

□

Lower bounding the invariant measure

Let us rearrange weights in the product (5.31) to obtain ratios that involve vertices in the same small-range component. Let $x \in \mathcal{V}$ and $(x_i)_{i=1}^k$ be the unique sequence of centers joining O to x , possibly empty. Then set

$$Z'(x) := \prod_{i=1}^{k-1} \frac{\pi_P(\eta(x_{i+1}))q(\eta(x_{i+1}), x_{i+1})}{\pi_P(x_i)q(x_i, \eta(x_i))}. \quad (5.33)$$

taking by convention empty products to be equal to 1. By construction

$$Z(x) = \frac{\pi_P(\eta(x_1))q(\eta(x_1), x_1)}{\pi_P(x_k)q(x_k, \eta(x_k))} Z'(x) \quad (5.34)$$

Thus to lower bound $Z(x)$ is essentially suffices to bound $Z'(x)$. The idea now is that the process $Z'(x)$ can be described as a branching random walk on \mathbb{R}_+^* with symmetric increments. We do not need to dive to much into the theory of branching random walks, so we only define a branching random walk informally as a Markov field defined on a Galton-Watson tree.

In our case, the underlying tree \mathcal{T} can be obtained by collapsing all vertices of a small-range component into one node and keeping only long-range edges. We call this tree the skeleton tree of \mathcal{G} . This tree has naturally a root which identifies with O in \mathcal{G} , while other vertices identify with centers. Thanks to this identification, Z' can be seen as a random field on \mathcal{T} . Write \mathcal{T}_O for the subtree of \mathcal{T} spanned by centers of \mathcal{G}_O .

Lemma 5.10. *Suppose P satisfies (H4') with constant $l \geq 1$. Then for all realization of \mathcal{G} there exists a connected subgraph $\mathcal{G}' \subset \mathcal{G}_O$ containing $\eta(O)$ such that*

- (i) *all vertices of \mathcal{G}' are at \mathcal{P} -distance at most $2l$ from their center,*
- (ii) *$Z'(x) = \Theta(1)$ for every center $x \in \mathcal{G}'$,*
- (iii) *with probability $\Theta(1)$, any reversible chain with underlying graph \mathcal{G}' has uniformly lower bounded escape probabilities.*

Proof. The starting point is that the tree \mathcal{T} and the identification with centers of the quasi-tree can be generated in two steps. Suppose for instance that P_1 and P_2 are irreducible. Then all small-range components of \mathcal{G} are isomorphic to the unique communicating class of either P_1 or P_2 , so the tree \mathcal{T} is just a n -regular tree. In general the tree \mathcal{T} can be described as a two-type Galton-Watson tree. For $i = 1, 2$ and $k \geq 0$, let $M_i(k)$ be the number of communicating classes in V_i with exactly k vertices. The tree \mathcal{T} has the law of the random tree where every vertex of type $i \in \{1, 2\}$ gives rise to $k - 1$ children with probability $kM_{\bar{i}}(k)/n$, all of types \bar{i} , where $\bar{i} = 2$ if $i = 1$ and vice

versa. By Assumption (H3), $M_1(1) = M_1(2) = M_2(1) = 0$, so \mathcal{T} always contains the deterministic tree that alternates between degree 2 and degree 3 vertices.

Suppose now \mathcal{T} is a tree generated according to the previous law. Suppose the type of the root is fixed, which determines the types of all other vertices. Then assign labels to edges of \mathcal{T} , consisting of pairs (u, v) such that: $v \in B_P^+(u, \infty), v \neq u$. They need to satisfy the two constraints: 1. the labels on edges incident to a same vertex x share the same first coordinate; 2. if this coordinate is u , then for all $v \in B_P^+(u, \infty), v \neq u$ there is an edge incident to x with label (u, v) . Because of the first constraint, we extend labels to vertices, calling u the label of x . In addition, call the label (u, v) a backbone label if $d_{\text{SR}}(u, v) \leq 2$ or $d_{\text{SR}}(v, u) \leq 2$, which is symmetric in (u, v) . Remark that every vertex x has always one backbone label, nay two if x has type $i = 1$. If \mathcal{T} is a tree with such a labelling of the edges, we can construct a quasi-tree \mathcal{G} such that \mathcal{T} is the skeleton tree of \mathcal{G} : labels of vertices define the type ι of centers, while labels of edges detail how the corresponding small-range components are linked to each other. Backbone labels correspond to vertices which are close to their center: if $d_{\text{SR}}(u, v) \leq 2$ then (H4') implies that $d_{\text{SR}}(v, u) \leq 2l$ so u, v are at bounded distance from each other in both directions.

We then define weights on labelled edges. If e has label (u, v) , set

$$b_e = b_{(u,v)} := \frac{\pi_P(v)q(v, \eta(v))}{\pi_P(u)q(u, \eta(u))}. \quad (5.35)$$

Note that $b_{(u,v)} = b_{(v,u)}^{-1}$. Given a vertex $x \in \mathcal{T}$, we have by construction that $Z'(x) := \prod_{i=1}^k b_{e_i}$ where $(e_i)_{i=1}^k$ is the unique path of edges from the root to x and empty product is equal to 1. From what precedes, if (u, v) is a backbone label then $P(u, v)^k > 0$ for some $k \in [1, 2l]$. Recall that from Assumption (H1) and (H2) there is a constant lower bound δ on all transition probabilities. Thus

$$b_{(u,v)} \geq \delta^{2l+1} =: b_{\min}$$

is bounded below uniformly in n , using that $\pi_P(v) \geq \pi_P(u)P^k(u, v)$. Consequently backbone labels imply bounded weights uniformly in n .

Now labels can be assigned randomly: suppose $x \in \mathcal{T}$ has type $i \in \{1, 2\}$ and $k - 1$ vertices descendant vertices. Pick a communicating class \mathcal{C} of size k in V_i uniformly at random, pick a uniform $u \in \mathcal{C}$ and distribute labels $((u, v))_{v \in \mathcal{C} \setminus \{u\}}$ uniformly over the descendant edges of x . With this procedure, the quasi-tree obtained from the random labels has the distribution of \mathcal{G} . Furthermore, labels are independent over edges that are not incident to a same vertex.

In particular Z'_x is a product of independent weights over the path from the root to x . Furthermore, since every choice in this construction is uniform, the labels are

also uniform over all possible pairs (u, v) . In particular an edge is given a label (u, v) or (v, u) with the same probability, which implies it is given a weight $b_{(u,v)}$ or $b_{(v,u)}$ with same probability. Since the latter are inverse of each other we deduce that the distribution of the weight assigned to an edge is symmetric, that is b_e has the same law as b_e^{-1} . Then for all $x \in \mathcal{T}$, Z'_x needs also be symmetric as it is a product of independent weights.

We conclude the proof by a comparison with a branching process. We can reason conditional on \mathcal{T} , so only labels and the associated weights are random. Consider the power tree $\mathcal{T}_O^{(4)}$ as defined in Proposition 5.6. By definition an edge of $\mathcal{T}_O^{(4)}$ identifies to a 4-tuple of edges in \mathcal{T}_O . Let \mathcal{T}'' be the tree defined as the connected component of $\eta(O)$ of all edges in $\mathcal{T}_O^{(4)}$ which identify to 4-tuple of edges with backbone labels, and such that the product of weights is above 1. What precedes shows that except $\eta(O)$, every vertex in $\mathcal{T}_O^{(4)}$ has always 4 edges with backbone labels, and the product of weights is above 1 with probability $1/2$. Thus the tree \mathcal{T}'' dominates stochastically a Galton Watson tree whose offspring law is binomial $\text{Bin}(4, 1/2)$ and is thus supercritical. This gives back a subtree $\mathcal{T}' \subset \mathcal{T}_O$ of the original tree, which is the one we are looking for. The labels on these edges give rise to a set of vertices which span the desired graph \mathcal{G}' . Since all labels on \mathcal{T}' are backbone labels, these vertices are all at distance at most $2l$ from their center, and $Z'(x) \geq b_{\min}^4$ for all $x \in \mathcal{T}'$. It remains to prove the last statement. This is the consequence of the arguments used in the reversible model: from the above comparison with a supercritical Galton-Watson tree, if the latter is infinite, which occurs with probability $\Theta(1)$, then the skeleton tree \mathcal{T}' is sufficiently branching to make any reversible chain on \mathcal{G}' have uniformly lower bounded escape probabilities by Remark 5.12. \square

Proof of Lemma 5.6 with assumption (H4'). By (5.29), it suffices to bound the Green function $\mathcal{G}_{\mathcal{P}_0}$ at O . The measure ν is easily tweaked to give an invariant measure for \mathcal{P}_0 : the graph \mathcal{G}_O^* is still a quasi-tree which can be seen as having $\mathcal{C}_O := \{O\} \cup \text{SR}(\eta(O))$ as the small-range component of the root. Then replace π_P locally on that small-range component by the invariant measure π_P^* of \mathcal{C}_O . This modification does not affect the field $Z'(x)$ (5.33) thus from (5.34) $Z(x)$ gets only modified by one factor. In the end we obtain an invariant measure ν_0 for \mathcal{P}_0 given by $\nu_0(x) := \pi_P^*(x)$ if $x \in \mathcal{C}_O$ and otherwise

$$\nu_0(x) := \frac{\pi_P^*(\eta(x_1))q(\eta(x_1), x_1)}{\pi_P(x_k)q(x_k, \eta(x_k))} Z'(x) \pi_P(x)$$

if $(x_i)_{i=1}^k$ is the sequence of centers from O to x . In particular, $\eta(x_1) \in \mathcal{C}_O$ and $x_k \in B_{\text{SR}}(x, +\infty)$. Consider now the graph $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ of Lemma 5.10. For all $x \in \mathcal{V}'$, $Z'(x) \geq c$ and all vertices of \mathcal{V}' are at bounded distance from their centers. As was

argued in the proof of Lemma 5.10, this implies that the ratios between the measures of vertices inside the same small-range component can be bounded uniformly in n . Thus

$$\frac{\pi_P^*(\eta(x_1))}{\nu_0(O)} \frac{\pi_P(x)}{\pi_P(x_k)} = \Theta(1).$$

Since the invariant measure is defined up to multiplicative constant, let us suppose π_P^* was taken in order to have $\nu_0(O) = 1$. Using also the fact that the probabilities q are bounded by (H2), we deduce that there exist a constant c independent of n such that for all $x \in \mathcal{V}'$

$$\nu_0(x) \geq c.$$

On the other hand, let K denote the transition kernel of the simple random walk on \mathcal{G}' . The measure μ defined by $\mu(x) := \deg_{\mathcal{G}'}(x)$ for all $x \in \mathcal{V}'$ is an invariant measure for K . Since the graph \mathcal{G} has degrees bounded by Δ , the entries of both K and μ can be lower bounded by a constant depending only on Δ . We can thus bound

$$\sup_{x \in \mathcal{V}'} \mu(x)/\nu_0(x) \leq C$$

for some other constant $C > 0$ independent of n . Consider now the transition kernels $\bar{\mathcal{P}} := (I + \mathcal{P}_0)^{l+1}/2^{l+1}$ and $\bar{K} := (1 - \frac{1}{2C} \frac{\mu}{\nu_0})I + \frac{1}{2C} \frac{\mu}{\nu_0} K$, where l is given by the assumption (H4'). The measure ν_0 is still invariant for $\bar{\mathcal{P}}$ but now one also has that \bar{K} is reversible with respect to ν_0 .

Let $z \in (0, 1)$. We apply Lemma 5.8 to operators $T_1 := (I - z\bar{K})$ and $T_2 := (I - z\bar{\mathcal{P}})$ on the Hilbert space

$$\ell^2(\mathcal{V}') := \{f : \mathcal{V}' \rightarrow \mathbb{R} \mid \langle f, f \rangle_{\nu_0} < \infty\}, \quad \langle f, g \rangle_{\nu_0} := \sum_{x \in \mathcal{V}'} \nu_0(x) f(x) g(x).$$

By (H4') every small-range edge of \mathcal{G}' , regardless of the orientation, has positive transition probability under $\bar{\mathcal{P}}$, bounded below by $(\delta/2)^{l+1}$. Similarly, if $x, y \in \mathcal{V}'$ and $P(\eta(x), y) > 0$, then $\bar{\mathcal{P}}(x, \eta(x)) \geq 2^{-l-1} \mathcal{P}_0(x, y) \mathcal{P}_0(y, \eta(x)) \geq (\delta/2)^{l+1}$ so long-range edges too have a lower bounded probability. Thus for all $x, y \in \mathcal{V}'$

$$\bar{\mathcal{P}}(x, y) \geq \epsilon \bar{K}(x, y)$$

for some constant ϵ . As a consequence $\frac{1}{1-\epsilon}(\bar{\mathcal{P}} - \epsilon\bar{K})$ defines a transition kernel with invariant measure ν_0 . This defines thus a contracting operator on $\ell^2(\mathcal{V}')$, which implies

$$\langle (\bar{\mathcal{P}} - \epsilon\bar{K})f, f \rangle_{\nu_0} \leq (1-\epsilon) \langle f, f \rangle_{\nu_0}$$

and

$$\langle T_2 f, f \rangle_{\nu_0} \geq \epsilon \langle T_1 f, f \rangle_{\nu_0}$$

for all $f \in \ell^2(\mathcal{V}')$. The reversibility of \bar{K} implying also the self-adjointness of T_2 all the assumptions of Lemma 5.8 are satisfied, hence

$$\langle T_2^{-1} f, f \rangle_{\nu_0} \leq \epsilon^{-1} \langle T_1^{-1} f, f \rangle_{\nu_0}.$$

Taking f as the indicator function at $O \in \mathcal{V}'$ we obtain

$$G_{\bar{\mathcal{P}}}(O, O | z) \leq \epsilon^{-1} G_{\bar{K}}(O, O | z)$$

with $G_{\bar{\mathcal{P}}}(O, O | z) = \sum_{t \geq 0} \bar{\mathcal{P}}^t(O, O) z^t$ and similarly for \bar{K} . Take the limit $z \rightarrow 1$ from below to obtain

$$G_{\bar{\mathcal{P}}}(O, O) \leq \epsilon^{-1} G_{\bar{K}}(O, O)$$

Now we claim that Green functions of $\bar{\mathcal{P}}$ and \mathcal{P}_0 can be related as

$$G_{\mathcal{P}_0}(O, O) \leq 2^{l-1} (l+1) G_{\bar{\mathcal{P}}}(O, O).$$

Indeed, $\bar{\mathcal{P}}$ is obtained by applying the successive operations of "lazification" and taking the power $l+1$. It can be checked that the "lazification" multiplies the Green function by 2. Then let $(\mathcal{Y}_t)_{t \geq 0}$ be the lazy chain with matrix $(I + \mathcal{P}_0)/2$. Taking the power $l+1$ of this chain comes down to consider the chain only at times which are a multiple of $l+1$. Now for all $m \in [0, l]$, for all $t \geq 1$, laziness implies

$$\mathbb{P}_O [\mathcal{Y}_{(l+1)t} = O] \geq \mathbb{P}_O [\mathcal{Y}_{(l+1)t-m} = O] / 2^m$$

so we can average and get

$$\mathbb{P}_O [\mathcal{Y}_{(l+1)t} = O] \geq \sum_{m=0}^l \frac{1}{2^m (l+1)} \mathbb{P}_O [\mathcal{Y}_{(l+1)t-m} = O]$$

Then summing over t yields:

$$\begin{aligned} G_{\bar{\mathcal{P}}}(O, O) &\geq \sum_{m=0}^l \frac{1}{2^m (l+1)} \sum_{t=-m \pmod{l+1}} \mathbb{P}_O [\mathcal{Y}_t = O] \\ &\geq \frac{1}{2^l (l+1)} G_{(I+\mathcal{P}_0)/2}(O, O) \\ &= \frac{1}{2^{l-1} (l+1)} G_{\mathcal{P}_0}(O, O). \end{aligned}$$

Finally it remains to upper bound the Green function of \bar{K} . Since it is a reversible chain, Point (iii) of Lemma 5.10 shows there is an event of probability $\Theta(1)$ on which it has uniformly lower bounded escape probabilities which from (5.29) implies its Green function is uniformly upper bounded at the root. \square

5.5.4 General case: $o(n)$ transient or close to transient states

We now consider the general case where we suppose only (H4). In this case there can be $o(n)$ vertices around which it is not possible to revert the transitions in l steps. Transient states are such vertices, so we call these vertices *close to transient*. There are two issues that prevent the previous proof to apply directly: first if P contains transient states, π_P takes the value 0 so the measure ν as defined in Lemma 5.9 cannot be invariant, and the chain $(\mathcal{X}_t)_{t \geq 0}$ is incidentally not irreducible. This problem is fixed by the fact that we need only a "sub-invariant" measure, satisfying $\nu \mathcal{P} \leq \nu$, to apply the comparison argument, which allows for a loss of mass at transient states. Secondly, even without genuine transient states, the comparison with a reversible chain fails at close to transient states. We will argue that since these are in number $o(n)$, they are sufficiently rare to be discarded.

Quasi-stationary measures and transient classes Our goal is to define a sub-invariant measure which puts mass on transient states if there are any. We call such measures quasi-stationary, by analogy to classical quasi-stationary measures, which describe the limiting behaviour of absorbing Markov chains before absorption. Transitioning along a non-backtracking edge, that is an edge that can only be crossed in one direction, is indeed very similar to absorption.

Let $(Y_t)_{t \geq 0}$ be a Markov chain on a countable state space W with transition kernel Q . A set $A \subset W$ is called absorbing if for all $x \in A$, $Q(x, A) = 1$. Suppose the chain contains a non-empty absorbing set A and write B for its complement. A measure μ on B is called quasi-stationary if for all $t \geq 0$, for all $y \in B$,

$$\mu(y) = \sum_{x \in B} \mu(x) \mathbb{P}_x [Y_t = y, \tau_A > t]$$

As for classical stationary measures, quasi-stationary measures can be obtained from the Perron-Frobenius theorem. Let Q_B the restriction of Q to B : this is a sub-stochastic matrix, with non-negative entries, hence it admits a maximal positive eigenvalue $\lambda \leq 1$, which in turn admits left and right eigenvectors with all positive entries. Since $Q_B^t(x, y) = \mathbb{P}_x [Y_t = y, \tau_A > t]$, we deduce that left eigenvectors μ with positive entries define quasi-stationary measures on B . Note also that $\mu Q_B = \lambda \mu \leq \mu$ is a sub-invariant measure.

Remark 5.13. Our notion of quasi-stationary measures differs slightly from the classical one which not only intersects but conditions by the event that $\tau_A > t$.

Now, if we suppose that W contains transient and recurrent states, the latter form an absorbing set. A quasi-stationary measure in this setting consists thus of a quasi-

stationary measure on the set of transient states, with respect to the set of recurrent states.

Escape probability Let us move back to lower bounding the escape probability. We can in fact prove a slightly stronger result, namely allowing for a positive fraction αn of close to transient states, with α small enough. P contains in particular at most αn genuine transient states. If there are such transient states, there exist edges (x, y) between the transient and recurrent classes such that $P(x, y) > 0$ but $P^k(y, x) = 0$ for all $k \geq 0$. Call these non-backtracking edges. Non-backtracking edges in G obviously give non-backtracking edges in \mathcal{G} . We apply the argument with a version of the Green function $G_{\mathcal{P}_0}$ (5.29) killed at non-backtracking edges. Let \mathcal{G}' denote the connected component of the root in \mathcal{G} after deletion of the non-backtracking edges and consider the transition kernel \mathcal{P}'_0 given by the restriction of \mathcal{P}_0 to \mathcal{G}' , which is a sub-Markov operator. The corresponding Green function $G_{\mathcal{P}'_0}(O, O)$ counts the average number of visits to O before crossing a non-backtracking edge. Crossing a non-backtracking edge of \mathcal{G}_O obviously prevents the chain from going back to O , hence the number of visits to O remains a geometric variable of parameter $q_{\text{Esc}}(\eta(O))$ and thus $q_{\text{Esc}}(\eta(O)) = G_{\mathcal{P}'_0}(O, O)^{-1}$.

We thus aim to upper bound this Green function uniformly in n . After truncation, the remaining long-range edges in \mathcal{G}' connect small-range components which are either both isomorphic to a communicating class of G , or the component furthest from the root is isomorphic to a transient class. We can apply the same idea as in Lemma 5.9 to find a sub-invariant measure, considering quasi-stationary measures on transient classes. Let π_P be a positive combination, with bounded coefficients, of invariant measures on communicating classes and quasi-stationary measures on transient states. Then the exact same formula as in Lemma 5.9 yields a sub-invariant measure.

Next we truncate \mathcal{G}' a second time to get rid of close to transient states as well. Let \mathcal{G}'' be the subgraph of \mathcal{G}' spanned by the set \mathcal{V}'' of vertices which do not identify under ι to such states. By construction, the operator $(I + \mathcal{P}')^l/2^l$ can be compared edge-wise with the simple random walk on \mathcal{G}'' . We can thus apply Lemma 5.8 provided we have a comparison between sub-invariant measures. This can be done similarly by studying the random field Z , the only difference with the previous section is that we need we only look at vertices in \mathcal{V}'' . This can be done by modifying the definition of the "backbone labels" in the proof of Lemma 5.10, adding the requirement that a label (u, v) is a backbone label only if u and v are not close to transient. Provided α is small enough, the expected number of backbone labels can be made strictly larger than one, so that the comparison with a supercritical Galton-Watson tree still holds. One last

think to be aware of is that we want a result conditional on $\iota(\eta(O))$, which can be close to transient and thus $\eta(O) \notin \mathcal{G}''$. In this case the argument is to be applied to another subquasi-tree of \mathcal{G}_O at bounded distance from $\eta(O)$.

5.6 Analysis on the quasi-tree II: Markovian regeneration structure

Proposition 5.5 implies in particular that the Markov chain $(\mathcal{X}_t)_{t \geq 0}$ is almost surely transient. As a consequence, the shortest path from O to \mathcal{X}_t eventually has to go through a unique sequence of long-range edges $(\xi_i)_{i=1}^\infty$, which is called the *loop-erased chain*. Among the edges of the loop-erased chain, some have the property to be crossed only once. Thanks to this property, these so-called *regeneration edges* yield a Markov decomposition of the quasi-tree which we will use in the next section to prove concentration of the drift and entropy of the harmonic measure. The regeneration process will also be used later to compute the approximate invariant measure $\hat{\pi}$.

5.6.1 Markov renewal processes

We start with general results about Markov renewal processes that will be necessary in the sequel. The theory of Markov renewal processes is certainly not new, however we could not find references proving the results established in this section.

Definition 5.5. Let S be a countable state space and E a Polish space. Consider a process $(Y, Z) = (Y_k, Z_k)_{k \geq 0}$ taking values in $S \times E$, satisfying for all $k \geq 1$, $x, y \in S$, $z \in E$,

$$\mathbb{P}[Y_k = y, Z_k = z \mid Y_{k-1} = x, Z_{k-1}, \dots, Y_0, Z_0] = \mathbb{P}[Y_1' = y, Z_1' = z \mid Y_0 = x], \quad (\text{M1})$$

where (Y_1', Z_1') is an independent copy of (Y_1, Z_1) . This process is thus a Markov chain with a stronger Markov property than the usual one, in that the time dependence occurs only through the Y -coordinate. In particular $Y = (Y_k)_{k \geq 0}$ is a Markov chain on S .

A *Markov renewal process* is a process $(Y_k, T_k)_{k \geq 0}$ is a process taking values in $S \times \mathbb{N}$ such that $(Y_k, T_k - T_{k-1})_{k \geq 0}$ satisfies (M1) and $T_k - T_{k-1} \geq 1$ a.s. for all $k \geq 1$, taking $T_{-1} := 0$. The delay T_0 can have arbitrary distribution. We call transition kernels of (Y, T) the family of kernels $(Q_t)_{t \geq 1}$, where for each $t \geq 1$,

$$Q_t(x, y) := \mathbb{P}[Y_1 = y, T_1 = t \mid Y_0 = x, T_0 = 0] = \mathbb{P}[Y_1 = y, T_1 - T_0 = t \mid Y_0 = x].$$

Notice then that Y has transition kernel $Q := \sum_{t \geq 1} Q_t$.

Remark 5.14. If (Y, Z) satisfies (M1), the initial pair (Y_0, Z_0) can have an arbitrary law. As usual, if ν is a probability measure on $S \times E$ we write

$$\mathbb{P}_\nu := \sum_{y \in S} \int \mathbb{P}[\cdot \mid Y_0 = y, Z = z] \nu(y, dz).$$

On the other hand, the (M1) implies that $(Y_k, Z_k)_{k \geq 1}$ conditional on Y_0 is independent of Z_0 . Thus if we are interested in a quantity that is measurable only with respect to $(Y_k, Z_k)_{k \geq 1}$, we will slightly abuse notation by writing $\mathbb{P}_u := \mathbb{P}[\cdot \mid Y_0 = u]$ (and write similarly for expectation), and by extension $\mathbb{P}_\mu := \sum_{y \in S} \mu(y) \mathbb{P}_y$ for a measure μ on S . In particular, note that if μ is an invariant measure for Y , the process $(Y_k, Z_k)_{k \geq 1}$ becomes stationary under \mathbb{P}_μ . In the sequel if Y is positive recurrent, we only consider invariant measures which are probability distributions.

Lemma 5.11. *Let S be a countable set and consider a process $(Y_k, Z_k)_{k \geq 0}$ taking values in $S \times \mathbb{R}_+$ which satisfies (M1). Suppose there exists $\alpha \in (0, 1], \theta > 0$ such that $\mathbb{E}[e^{\theta Z_0^\alpha}] < \infty$ and $\max_{u \in S} \mathbb{E}_u[e^{\theta Z_1^\alpha}] < \infty$. Then for all $k \geq 0$,*

$$\mathbb{E}\left[e^{\theta(\sum_{i=0}^k Z_i)^\alpha}\right] \leq \left(\max_{u \in S} \mathbb{E}_u[e^{\theta Z_1^\alpha}]\right)^k. \quad (5.36)$$

As a consequence, there exists $C, C' > 0$ depending on $\theta, \mathbb{E}[e^{\theta Z_0^\alpha}]$ and $\max_{u \in S} \mathbb{E}_u[e^{\theta Z_1^\alpha}]$ such that for all $k \geq 0$,

$$\mathbb{P}\left[\sum_{i=0}^k Z_i > Ck^{1/\alpha}\right] \leq e^{-C'k}. \quad (5.37)$$

Proof. Let $\alpha \in (0, 1], \theta > 0$ such that $\mathbb{E}[e^{\theta Z_0^\alpha}] < \infty$ and $\max_{u \in S} \mathbb{E}_u[e^{\theta Z_1^\alpha}] < \infty$. Let $(\mathcal{F}_k)_{k \geq 0}$ denote the standard filtration of (Y, Z) . Since $\alpha \leq 1$, $(x + y)^\alpha \leq x^\alpha + y^\alpha$ for all $x, y \geq 0$. Consequently Markov's property used with (M1) implies

$$\begin{aligned} \mathbb{E}\left[e^{\theta(\sum_{i=0}^k Z_i)^\alpha}\right] &= \mathbb{E}\left[e^{\theta(\sum_{i=0}^{k-1} Z_i)^\alpha} \mathbb{E}\left[e^{\theta Z_k^\alpha} \mid \mathcal{F}_{k-1}\right]\right] \\ &= \mathbb{E}\left[e^{\theta(\sum_{i=0}^{k-1} Z_i)^\alpha} \mathbb{E}_{Y_{k-1}}\left[e^{\theta(Z_1')^\alpha}\right]\right], \end{aligned}$$

where Z_1' denotes an independent copy of Z_1 . Thus

$$\mathbb{E}\left[e^{\theta(\sum_{i=0}^k Z_i)^\alpha}\right] \leq \max_{u \in S} \mathbb{E}_u[e^{\theta Z_1^\alpha}] \mathbb{E}\left[e^{\theta(\sum_{i=0}^{k-1} Z_i)^\alpha}\right]$$

from which the first inequality follows by induction. The second is an application of Chernoff's bound. \square

The next results specify to the setting where the process Z takes integer values. In particular we state analogs of classical renewal theorems in the context of a Markov renewal process (Y, T) . The following generalizes the so-called elementary renewal theorem.

Proposition 5.7. *Let (Y, T) be a Markov renewal process with state space S such that $\max_{u \in S} \mathbb{E}_u [T_1] < \infty$ and Y is positive recurrent with stationary distribution μ . Given $t \geq 0$ let*

$$N_t := \sup\{k \geq 0, T_k \leq t\}.$$

Then a.s.

$$\lim_{k \rightarrow \infty} \frac{T_k}{k} = \mathbb{E}_\mu [T_1] \quad (5.38)$$

and

$$\lim_{t \rightarrow \infty} \frac{N_t}{t} = \frac{1}{\mathbb{E}_\mu [T_1]}. \quad (5.39)$$

Proof of Proposition 5.7. The laws of large numbers are consequences of the ergodic theorem applied to two different Markov chains.

For T_k , consider the pair $(Y_k, T_k - T_{k-1})_{k \geq 0}$, with $T_{-1} := 0$, which by definition satisfies (M1). It was noted in Remark 5.14 that \mathbb{P}_μ is a stationary distribution for this chain. Hence the law of large numbers follows from the ergodic theorem applied to the ergodic averages of the projection onto the second coordinate.

Let us move to N_t . Consider the Markov chain $(U_k)_{k \geq 0}$ on $S \times \mathbb{N}$ defined by the transition probabilities

$$K((x, 0), (y, t-1)) = Q_t(x, y) \quad K((x, t), (x, t-1)) = 1$$

for all $t \geq 1$ and $x, y \in S$. This chain can be thought of as representing Y together with the waiting time until the next renewal: letting τ_k be the k -th successive hitting time of $S \times \{0\}$, the S -coordinate of $(U_{\tau_k})_{k \geq 1}$ has the distribution of $(Y_k)_{k \geq 0}$, given $Y_0 = U_0$. The number of jumps N_t made before time $t \geq 0$ is thus the number of such hitting times that occurred before t . By the ergodic theorem for Markov chains, one immediately gets

$$\lim_{t \rightarrow \infty} \frac{N_t}{t} = \tilde{\mu}(S \times \{0\})$$

where $\tilde{\mu}$ is the unique stationary measure of the Markov chain $(U_k)_{k \geq 0}$. As can be checked easily, it is given by:

$$\tilde{\mu}(x, t) = \frac{\mathbb{P}_\mu [X_1 = y, T_1 > t]}{\mathbb{E}_\mu [T_1]} \quad (5.40)$$

for all $x \in S, t \geq 0$. In particular $\tilde{\mu}(S \times \{0\}) = \mathbb{E}_\mu [T_1]^{-1}$. \square

The two following Propositions establish mixing results for a Markov renewal process (Y, T) . These are not necessary for the asymptotic analysis on the quasi-tree but will be used to essentially compute the annealed laws of \mathcal{X} and X and obtain the value of the limiting measure $\hat{\pi}$. Proposition 5.9 is an analog of the classical renewal theorem, with a quantitative bound on the speed of convergence. To prove it, we establish first

a stronger result in Proposition 5.8, namely a mixing property for the whole process (Y, T) , where T is allowed here to take negative values. Since T is unbounded mixing is here understood as the fact that the process forgets about its starting state.

Proposition 5.8. *Let $(Y_k, T_k - T_{k-1})_{k \geq 0}$ satisfy (M1), with $T_{-1} := 0$ and $T_k - T_{k-1} \in \mathbb{Z}$ a.s. for all $k \geq 0$. Suppose Y is positive recurrent, irreducible and aperiodic. Let $Q_t(x, y) := \mathbb{P}[Y_1 = y, T_1 - T_0 = t \mid Y_0 = x]$ for all $t \in \mathbb{Z}$ and*

$$\alpha := \min_{\substack{x \in S \\ t \in \mathbb{Z} \\ y \in S}} \sum (Q_t(x, y) \wedge Q_{t+1}(x, y)).$$

Consider two starting probability distributions ν_1, ν_2 on $S \times \mathbb{Z}$. Given $\varepsilon \in (0, 1)$, let $t_{\text{mix}}^{(Y)}(\varepsilon)$ denote the ε -mixing time of $(Y_k)_{k \geq 0}$ and $K(\varepsilon) \geq 0$ be the minimal integer such that

$$\sup_{\nu} \mathbb{P}_{\nu} [|T_0 - T'_0| > K(\varepsilon)] \leq \varepsilon,$$

where the supremum is over all couplings $((Y_0, T_0), (Y'_0, T'_0))$ of ν_1 and ν_2 . Then for all $\varepsilon \in (0, 1)$, there exists $C(\varepsilon) \geq 0$ such that

$$\| \mathbb{P}_{\nu_1} [(Y_k, T_k) = \cdot] - \mathbb{P}_{\nu_2} [(Y_k, T_k) = \cdot] \|_{\text{TV}} \leq \varepsilon$$

for all

$$k \geq \frac{C(\varepsilon)}{\alpha} \left(t_{\text{mix}}^{(Y)}(\varepsilon) \max_{u \in S} \mathbb{E}_u |T_1| + K(\varepsilon) \right)^2. \quad (5.41)$$

Proof of Proposition 5.8. The proof is based on coupling arguments. If the Y_k are iid, T_k becomes the sum of iid random variables on \mathbb{Z} , ie a random walk on \mathbb{Z} . Two random walks on \mathbb{Z} identically distributed but started apart from a distance A can be coupled to meet at a random time τ which satisfies $\mathbb{P}[\tau > k] \leq CA/\sqrt{k}$ for some constant $C > 0$. The argument can be found in [120] and [119][II.14], where the coupling is called Mineka coupling. In the general case where Y_k are not iid, the idea is to first couple the chains Y_k started at different states to make them coincide, after which one can adapt the Mineka coupling to make the subsequent variables T_k coalesce.

Suppose $((Y, T), (Y', T'))$ is a coupling of two versions of the process $(Y_k, T_k)_{k \geq 0}$ started at ν_1 and ν_2 respectively. Let $\tau := \inf\{k \geq 0 : (Y_k, T_k) = (Y'_k, T'_k)\}$. If the coupling is such that the two processes coincide after the coalescence time τ , one has

$$\| \mathbb{P} [(Y_k, T_k) = \cdot] - \mathbb{P} [(Y'_k, T'_k) = \cdot] \|_{\text{TV}} \leq \mathbb{P}[\tau > k].$$

We now precise such a coupling. Let $\varepsilon \in (0, 1)$ and $k_0 := t_{\text{mix}}^{(Y)}(\varepsilon)$. The coupling is actually started at k_0 : couple (Y_{k_0}, T_{k_0}) and (Y'_{k_0}, T'_{k_0}) in order to have an optimal coupling of Y_{k_0} and Y'_{k_0} . Thus

$$\mathbb{P}[Y_{k_0} \neq Y'_{k_0}] = \| \mathbb{P}[Y_{k_0} = \cdot] - \mathbb{P}[Y'_{k_0} = \cdot] \|_{\text{TV}} \leq 2 \| \mathbb{P}[Y_{k_0} = \cdot] - \mu \|_{\text{TV}} \leq 2\varepsilon, \quad (5.42)$$

where μ is the stationary measure of Y . Then for all $k \geq k_0$, conditional on $Y_k = Y'_k$ draw $Y_{k+1}, Y'_{k+1}, S_{k+1}, S'_{k+1}$ according to the distribution:

$$\begin{aligned} \mathbb{P}[Y_{k+1} = Y'_{k+1} = y, S_{k+1} = t-1, S'_{k+1} = t \mid Y_k = Y'_k = x] &= \alpha_{t-1}(x, y) \\ \mathbb{P}[Y_{k+1} = Y'_{k+1} = y, S_{k+1} = t, S'_{k+1} = t-1 \mid Y_k = Y'_k = x] &= \alpha_{t-1}(x, y) \\ \mathbb{P}[Y_{k+1} = Y'_{k+1} = y, S_{k+1} = t, S'_{k+1} = t \mid Y_k = Y'_k = x] &= Q_t(x, y) - \alpha_{t-1}(x, y) - \alpha_t(x, y) \end{aligned}$$

writing $\alpha_t(x, y) := Q_t(x, y) \wedge Q_{t+1}(x, y)$. It is readily seen that

$$\mathbb{P}_x[Y_{k+1} = y, S_{k+1} = t \mid Y_k = x] = Q_t(x, y) = \mathbb{P}_x[Y'_{k+1} = y, S'_{k+1} = t \mid Y'_k = x].$$

Therefore if $Y_k = Y'_k$ setting $T_{k+1} := T_k + S_{k+1}$, $T'_{k+1} := T'_k + S'_{k+1}$ yields a coupling of $(Y_{k+1}, T_{k+1}), (Y'_{k+1}, T'_{k+1})$ with $Y_{k+1} = Y'_{k+1}$. Consequently, this can be used to couple the two processes $(Y, T), (Y', T')$ for every step after the coalescence of Y and Y' , so that the Y -coordinate stays identical. Once the T -coordinate coalesce, we couple the two processes so that they coincide indefinitely. We now bound the tail of the stopping time τ under this coupling.

Let $\tau_1 := \inf\{k \geq k_0 : Y_k = Y'_k\}$ and for all $k \geq 0$, $Z_k := \sum_{i=\tau_1}^k S_k - S'_k$. Observe that $(Y_{\tau_1+k}, Z_k - Z_{k-1})_{k \geq 0}$ (with $Z_{-1} := 0$) satisfies the (M1) property. In addition, for every $k \geq 0$, conditional on $Y_{\tau_1+k} = x$, $Z_{k+1} - Z_k$ has symmetric distribution in $\{-1, 0, 1\}$ a.s. and

$$\mathbb{P}[Z_{k+1} - Z_k = 1 \mid Y_k = x] = \sum_{t \geq 1} \sum_{y \in S} \alpha_t(x, y) \geq \alpha. \quad (5.43)$$

On the event $\{\tau_1 = k_0\}$, for all $k \geq 0$ we can decompose

$$T_{k_0+k} - T'_{k_0+k} = T_{k_0} - T'_{k_0} + Z_k$$

which implies that $\tau = k_0 + \inf\{k \geq 0 : Z_k = T'_{k_0} - T_{k_0}\}$. Letting $\tilde{\tau}_a$ be the hitting time of $a \in \mathbb{Z}$ by the process Z , we deduce from Markov's property that for all $k \geq 0$ and $A > 0$,

$$\mathbb{P}[\tau > k_0 + k] \leq \mathbb{P}[Y_{k_0} \neq Y'_{k_0}] + \mathbb{P}[|T_{k_0} - T'_{k_0}| > A] + \max_{a \in [-A, A]} \max_{u \in S} \mathbb{P}_{(u, 0)}[\tilde{\tau}_a > k].$$

By (5.42) the first term is smaller than 2ε . For the second term, bound

$$\mathbb{P}[|T_{k_0} - T'_{k_0}| > K(\varepsilon) + A] \leq \mathbb{P}[|T_0 - T'_0| > K(\varepsilon)] + \mathbb{P}[|(T_{k_0} - T_0) - (T'_{k_0} - T'_0)| > A].$$

By definition of $K(\varepsilon)$ the first term is bounded by ε . The second can be bounded by triangle inequality and Markov's inequality to obtain

$$\mathbb{P}[|T_{k_0} - T'_{k_0}| > K(\varepsilon) + A] \leq \varepsilon + \frac{2k_0 \max_{u \in S} \mathbb{E}_u[T_1]}{A} \leq 2\varepsilon$$

for $A := 2k_0 \max_{u \in S} \mathbb{E}_u[T_1] / \varepsilon$. The last term is bounded as follows. First, notice that since the variables Z_k are bounded by 1 in absolute value, the maximal probability is

obtained for $a = \pm A$, and from the symmetry of $Z_{k+1} - Z_k$ we can suppose $a = A$. Then we claim that there exists a constant $C \geq 1$, such that for all $k \geq 0$

$$\max_{u \in S} \mathbb{P}_{(u,0)} [\tilde{\tau}_A > k] \leq \frac{CA}{\sqrt{\alpha k}}.$$

Provided the claim holds, we get that the right-hand side is below ε for $k \geq C^2 A^2 / (\alpha \varepsilon^2)$. Combining with the previous choices of k_0 and A , this yields eventually that

$$\|\mathbb{P}[(Y_k, T_k) = \cdot] - \mathbb{P}[(Y'_k, T'_k) = \cdot]\|_{\text{TV}} \leq 5\varepsilon$$

for

$$k \geq t_{\text{mix}}^{(Y)}(\varepsilon) + \frac{C^2(K(\varepsilon) + 2t_{\text{mix}}^{(Y)}(\varepsilon) \max_{u \in S} \mathbb{E}_u [T_1])^2}{\alpha \varepsilon^4},$$

proving the result.

Let us now prove the claim: for all $k \geq 1$, let $N_k := |\{i \leq k : Z_i - Z_{i-1} \neq 0\}|$ and define \tilde{Z}_k as the sum of the first k non zero variables $Z_i - Z_{i-1}$. By (5.43), for all $u \in S$, N_k dominates stochastically a binomial random variable $\text{Bin}(k, 2\alpha)$ and from the symmetry of the increments the process \tilde{Z} is the simple random walk on \mathbb{Z} . Letting $\tau_A^{(\text{SRW})}$ denote the hitting time of A by \tilde{Z} , [71][III 7.5] shows

$$\mathbb{P}_0 [\tau_A^{(\text{SRW})} = k] = \frac{A}{n} \binom{k}{\frac{A+k}{2}} 2^{-k} \leq C \frac{A}{k^{3/2}},$$

for some constant $C > 0$ independent of k and A . Summing over k implies

$$\mathbb{P}_0 [\tau_A^{(\text{SRW})} > k] \leq \frac{C'A}{\sqrt{k}}$$

for some other constant $C' > 0$. Thus

$$\begin{aligned} \mathbb{P}_{(u,0)} [\tilde{\tau}_A > k] &\leq \mathbb{P}_u [\tilde{\tau}_A > k, N_k \geq \alpha k] + \mathbb{P}_u [N_k < \alpha k] \\ &\leq \mathbb{P}_0 [\tau_A^{(\text{SRW})} > \lfloor \alpha k \rfloor] + \mathbb{P} [|\text{Bin}(k, 2\alpha) - 2\alpha k| > \alpha k] \\ &\leq \frac{C'A}{\sqrt{\alpha k}} + \frac{2(1-2\alpha)}{\alpha k} \end{aligned}$$

using Chebychev's inequality, which proves the claim. \square

Proposition 5.9. *Let (Y, T) be a Markov renewal process with state space S , such that $\max_{u \in S} \mathbb{E}_u [T_1] < \infty$ and Y is positive recurrent with stationary distribution μ , irreducible and aperiodic. Let α be as in Proposition 5.8 and suppose that $\alpha > 0$. For all probability distribution ν on $S \times \mathbb{N}$,*

$$\sum_{y \in S} \left| \mathbb{P}_\nu [\exists k \geq 0 : Y_k = y, T_k = t] - \frac{\mu(y)}{\mathbb{E}_\mu [T_1]} \right| \xrightarrow[t \rightarrow \infty]{} 0. \quad (5.44)$$

More precisely, given $\varepsilon \in (0, 1)$, let $K_\nu(\varepsilon)$ be the minimal integer such that

$$\mathbb{P}_\nu [T_0 > K_\nu(\varepsilon)] \leq \varepsilon.$$

There exists $C(\varepsilon) > 0$ such that for all probability distribution ν on $S \times \mathbb{N}$,

$$\sum_{y \in S} \left| \mathbb{P}_\nu [\exists k \geq 0 : Y_k = y, T_k = t] - \frac{\mu(y)}{\mathbb{E}_\mu [T_1]} \right| \leq \varepsilon$$

for all

$$t \geq \frac{C(\varepsilon)}{\alpha} \left(t_{\text{mix}}^{(Y)}(\varepsilon) \max_{u \in S} \mathbb{E}_u [T_1] + K_\nu(\varepsilon) + \frac{\mathbb{E}_\mu [T_1^2]}{\mathbb{E}_\mu [T_1]} \right)^2 \max_v \mathbb{E}_v [T_1] \quad (5.45)$$

where $t_{\text{mix}}^{(Y)}(\varepsilon)$ denotes the ε -mixing time of Y .

Proof of Proposition 5.9. Consider the Markov chain $(U_k)_{k \geq 0}$ on $S \times \mathbb{N}$ considered in the proof of Proposition 5.7. Then for all starting measure ν and $t \geq 0$

$$u_t(\nu, y) := \mathbb{P}_\nu [\exists k \geq 0 : Y_k = y, T_k = t] = \mathbb{P}_\nu [U_t = (y, 0)] \quad (5.46)$$

It was proved in the proof of Proposition 5.7 that U has unique invariant measure given by $\tilde{\mu}(x, t) = \mathbb{P}_\mu [Y_1 = y, T_1 > t] / \mathbb{E}_\mu [T_1]$. Thus if one can prove that U is aperiodic, the convergence theorem for Markov chains directly implies that $u_t(\nu, y) \rightarrow \tilde{\mu}(y, 0) = \mu(y) / \mathbb{E}_\mu [T_1]$. It is easily proved that U is aperiodic from the assumptions that Y converges to μ and $\alpha > 0$. Actually, we do not need even to check aperiodicity. The coupling argument that we use afterwards implies the convergence, which implies in turn aperiodicity.

Let $\varepsilon \in (0, 1)$. Let $k_0 = k_0(\varepsilon)$ be the right-hand side of 5.41. By Proposition 5.8 there exists a coupling of two versions of (Y, T) started at ν and $\tilde{\mu}$, with coalescence time κ such that

$$\mathbb{P}[\kappa > k_0] \leq \varepsilon.$$

These processes can in turn be coupled in an obvious way with two versions $U^{(\nu)}, U^{(\tilde{\mu})}$ of U started with distributions ν and $\tilde{\mu}$ respectively. Then notice that the two processes $U^{(\nu)}, U^{(\tilde{\mu})}$ coincide after time T_κ , therefore T_κ is a coupling time and we deduce

$$\|\mathbb{P}_\nu [U_t = \cdot] - \tilde{\mu}\|_{\text{TV}} \leq \mathbb{P}[T_\kappa > t].$$

Since the sequence $(T_i)_{i \geq 0}$ is increasing, the previous tail probability can be bounded as

$$\mathbb{P}[T_\kappa > t] \leq \mathbb{P}[\kappa > k_0] + \mathbb{P}_\nu [T_{k_0} > t].$$

The first term is bounded by ε . The second term can be bounded by Markov's inequality as

$$\mathbb{P}_\nu [T_{k_0} > t] \leq \frac{k_0 \max_{u \in S} \mathbb{E}_u [T_1]}{t} \leq \varepsilon$$

for $t \geq k_0 \max_{u \in S} \mathbb{E}_u [T_1] / \varepsilon$. Let us now explicit the value of k_0 . We first bound for the worst possible coupling of ν and $\tilde{\mu}$,

$$\begin{aligned} \mathbb{P} [|T_0 - T'_0| > K_\nu(\varepsilon) + k] &\leq \mathbb{P}_\nu [T_0 > K_\nu(\varepsilon)] + \mathbb{P}_{\tilde{\mu}} [T_0 > k] \\ &\leq \varepsilon + \mathbb{E}_\mu [(T_1 - k - 1)_+] / \mathbb{E}_\mu [T_1] \end{aligned}$$

using the definition of $K_\nu(\varepsilon)$ and the expression of $\tilde{\mu}$. Then by Markov's inequality

$$\begin{aligned} \mathbb{E}_\mu [(T_1 - k - 1)_+] &= \sum_{t \geq k+1} \mathbb{P}_\mu [T_1 > t] \\ &\leq \sum_{t \geq k+1} \frac{\mathbb{E}_\mu [T_1^2]}{t^2} \leq \frac{C \mathbb{E}_\mu [T_1^2]}{k} \end{aligned}$$

for some constant $C > 0$. Thus $\mathbb{P} [|T_0 - T'_0| > K_\nu(\varepsilon) + k] \leq 2\varepsilon$ for all $k \geq C \mathbb{E}_\mu [T_1^2] / (\mathbb{E}_\mu [T_1] \varepsilon)$, which gives the value of $K(\varepsilon)$ in (5.41) and

$$k_0 \leq \frac{C(\varepsilon)}{\alpha} \left(t_{\text{mix}}^{(Y)}(\varepsilon) \max_{u \in S} \mathbb{E}_u |T_1| + K_\nu(\varepsilon) + \frac{\mathbb{E}_\mu [T_1^2]}{\mathbb{E}_\mu [T_1]} \right)^2.$$

We deduce eventually that taking t as (5.45) yields

$$\|\mathbb{P}_\nu [U_t = \cdot] - \tilde{\mu}\|_{\text{TV}} \leq \varepsilon$$

for large enough $C(\varepsilon)$. Finally using (5.46) yields

$$\frac{1}{2} \sum_{y \in S} \left| u_t(\nu, y) - \frac{\mu(y)}{\mathbb{E}_\mu [T_1]} \right| \leq \varepsilon.$$

□

The last tool we introduce about Markov renewal processes is a variance bound that uses spectral arguments. Let us gather a few known facts about spectral theory for Markov chains. Consider a Markov chain $Y = (Y_k)_{k \geq 0}$ on countable state space Y with transition kernel Q , which is irreducible, positive recurrent with stationary distribution μ .

Let $\ell^2(\mu)$ be the Hilbert space of real-valued functions on S which are square-integrable with respect to the measure μ , equipped with the inner product

$$\langle f, g \rangle_\mu := \sum_{x \in S} \mu(x) f(x) g(x).$$

Q defines a contracting linear operator on $\ell^2(\mu)$ by $Qf(x) := \sum_{y \in S} Q(x, y)f(y)$. Its spectrum is defined as

$$\text{Spec}(Q) := \{\lambda \in \mathbb{C} \mid \lambda I - Q \text{ is not invertible as a bounded linear operator}\}.$$

The adjoint operator of Q^* is given by

$$Q^*(y, x) = \frac{\mu(x)Q(x, y)}{\mu(y)}$$

for all $x, y \in S$. Reversibility of Q with respect to μ is equivalent to self-adjointness of Q . In this case, the spectrum of Q is included the interval $[-1, 1]$. Writing $\mathbb{1}$ for the constant function equal to 1, the fact that μ is a probability measure implies that $\mathbb{1} \in \ell^2(S)$ and is an eigenvector of Q associated with the eigenvalue 1. The absolute spectral gap of Q is then defined as

$$\gamma := 1 - \sup\{|\lambda|, \lambda \in \text{Spec}(Q), \lambda \neq 1\}$$

if the eigenvalue 1 has multiplicity 1 and $\lambda := 0$ otherwise. It is well known that the absolute spectral gap is related to mixing properties of Y , see for instance [117, Chap. 12]. In the non-reversible case, spectral arguments can be applied by considering reversibilizations of the chain. From a theoretical point of view, the optimal parameter to consider is the so-called *pseudo spectral gap*, defined as

$$\gamma_{ps} := \max_{k \geq 1} \left\{ \frac{\gamma((Q^*)^k Q^k)}{k} \right\}.$$

If Q is reversible obviously $\gamma_{ps} = \gamma^2$. The pseudo spectral gap is a rather natural quantity introduced in [145] but it may have appeared before-hand in other places under different names. In [31], the quantity $1 - \gamma_{ps}$ is considered under the name singular radius.

As for the classical spectral gap, the pseudo spectral gap is intimately related with the mixing properties of the chain, as shown by the following proposition.

Proposition 5.10 ([145, Prop. 3.4]). *Let $(Y_k)_{k \geq 0}$ be an irreducible, positive recurrent Markov chain on a countable state space S , with stationary distribution μ and ε -mixing time $t_{\text{mix}}(\varepsilon)$. Suppose it is uniformly ergodic, in the sense that there exists $C > 0$ and $\rho \in (0, 1)$ such that*

$$\sup_{x \in S} \|\mathbb{P}_x [Y_t = \cdot] - \mu\|_{\text{TV}} \leq C\rho^t.$$

Then for all $\varepsilon \in [0, 1)$,

$$\gamma_{ps} \geq \frac{1 - \varepsilon}{t_{\text{mix}}(\varepsilon)}. \quad (5.47)$$

Furthermore if S is finite,

$$t_{\text{mix}}(\varepsilon) \leq \frac{1 + 2 \log((2\varepsilon)^{-1}) + \log(\mu_{\min}^{-1})}{\gamma_{ps}} \quad (5.48)$$

where $\mu_{\min} := \min_{x \in S} \mu(x)$.

The pseudo spectral gap can also be used to precisely handle correlations between different steps of a Markov chains. Thus we obtain the following variance bound for Markov processes with the (M1) Markov property.

Proposition 5.11. *Let S be countable and E a Polish space. Let $(Y_k, Z_k)_{k \geq 0}$ be a process on $S \times E$ satisfying (M1). Suppose that Y is irreducible, positive recurrent, with invariant measure μ and pseudo spectral gap γ_{ps} . Let $(f_i)_{i \geq 1}$ be a family of functions such that for all $i \geq 1$, $f_i : S \times E \rightarrow \mathbb{R}$ and $\max_{u \in S} \mathbb{E}_u [f_i(Y_1, Z_1)^2] < \infty$. Then for all $k \geq 1$*

$$\text{Var}_\mu \left[\sum_{i=1}^k f_i(Y_i, Z_i) \right] \leq \frac{6}{\gamma} \sum_{i=1}^k \text{Var}_\mu [f_i(Y_1, Z_1)]. \quad (5.49)$$

The previous result can be applied to the case of a Markov renewal process $(Y_k, T_k)_{k \geq 0}$, to get that

$$\text{Var}_\mu [T_k] \leq \frac{6}{\gamma_{ps}} k \text{Var}_\mu [T_1]. \quad (5.50)$$

for all $k \geq 1$.

Proof. First, observe that if (Y, Z) satisfies (M1) so does $(Y, (Y, Z))$. Thus up to changing the second coordinate of the process considered it suffices to prove the result for functions f_i of Z_i only.

We use the same arguments as in [145]. Let Q denote the transition kernel of the chain Y and given $i \in [m]$, $x, y \in S$ let

$$Q_i(x, y) := \mathbb{E}_x [f_i(Z_1) \mathbb{1}_{Y_1=y}].$$

Using matrix notations, for all $i \in [m]$, $x \in S$, $Q_i \mathbb{1}(x) = \sum_{y \in S} Q_i(x, y) = \mathbb{E}_x [f_i(Z_1)]$. We can suppose without generality that $\mathbb{E}_\mu [f_i(Z_i)] = \mathbb{E}_\mu [f_i(Z_1)] = 0$, which can be written matricially as

$$\mu Q_i \mathbb{1} = 0. \quad (5.51)$$

The random variable $\sum_{i=1}^m f_i(Z_i)$ is thus also centered and from stationarity one has

$$\begin{aligned} \text{Var}_\mu \left[\sum_{i=1}^m f_i(Z_i) \right] &= \sum_{i,j=1}^m \mathbb{E}_\mu [f_i(Z_i) f_j(Z_j)] \\ &= \sum_{i,j=1}^m \mathbb{E}_\mu [f_i(Z_1) f_j(Z_{|j-i|+1})]. \end{aligned}$$

For any $l \geq 1$

$$\begin{aligned}
\mathbb{E}_\mu [f_i(Z_1)f_j(Z_{l+1})] &= \sum_{x,y \in S} \mu(x)Q_i(x,y)\mathbb{E}_y [f_j(Z_{l+1})] \\
&= \sum_{x,y,z \in S} \mu(x)Q_i(x,y)Q^l Q_j(y,z) \\
&= \langle Q_i Q^l Q_j \mathbb{1}, \mathbb{1} \rangle_\mu \\
&= \langle Q_i (Q^l - \mathbb{1}\mu) Q_j \mathbb{1}, \mathbb{1} \rangle_\mu
\end{aligned}$$

where in the last line we used (5.51). For any $l \geq 1$, since μ is stationary one has $Q^l - \mathbb{1}\mu = (Q - \mathbb{1}\mu)^l$, thus by Cauchy-Schwarz inequality

$$\begin{aligned}
\langle Q_i (Q - \mathbb{1}\mu)^l Q_j \mathbb{1}, \mathbb{1} \rangle_\mu &= \langle (Q - \mathbb{1}\mu)^l Q_j \mathbb{1}, Q_i^* \mathbb{1} \rangle_\mu \\
&\leq \|(Q - \mathbb{1}\mu)\|^l \|Q_j \mathbb{1}\| \|Q_i^* \mathbb{1}\|.
\end{aligned}$$

Then by Jensen's inequality,

$$\|Q_j \mathbb{1}\|^2 = \sum_{x \in S} \mu(x) |\mathbb{E}_x [f_j(Z_1)]|^2 \leq \mathbb{E}_\mu [f_j(Z_1)^2] = \text{Var}_\mu [f_j(Z_1)].$$

Similarly,

$$Q_i^* \mathbb{1}(y) = \sum_{x \in S} \frac{\mu(x)Q_i(x,y)}{\mu(y)} = \mathbb{E}_\mu [f_i(Z_1) \mid Y_1 = y]$$

so

$$\|Q_i^* \mathbb{1}\|^2 \leq \sum_{y \in S} \mu(y) \mathbb{E}_\mu [f_i(Z_1)^2 \mid Y_1 = y] = \mathbb{E}_\mu [f_i(Z_1)^2] = \text{Var}_\mu [f_i(Z_1)^2].$$

Hence

$$\begin{aligned}
\mathbb{E}_\mu [f_i(Z_1)f_j(Z_{l+1})] &\leq \|(Q - \mathbb{1}\mu)^l\| \text{Var}_\mu [f_i(Z_1)]^{1/2} \text{Var}_\mu [f_j(Z_1)]^{1/2} \\
&\leq \frac{1}{2} \|(Q - \mathbb{1}\mu)^l\| (\text{Var}_\mu [f_i(Z_1)] + \text{Var}_\mu [f_j(Z_1)]).
\end{aligned}$$

It remains to sum over i, j . Let $k \geq 1$ be such that $\gamma(Q^k(Q^k)^*) = \gamma_{ps}k$. Then by observing that $\|(Q - \mathbb{1}\mu)^k\|^2 = 1 - \gamma(Q^k(Q^k)^*)$ one can deduce

$$\begin{aligned}
\|(Q - \mathbb{1}\mu)^l\| &\leq \|(Q - \mathbb{1}\mu)^k\|^{[l/k]} \\
&= (1 - k\gamma_{ps})^{[l/k]/2}
\end{aligned}$$

Consequently

$$\begin{aligned}
 \sum_{i \neq j=1}^m \mathbb{E}_\mu [f_i(Z_1) f_j(Z_{|j-i|+1})] &\leq \frac{1}{2} \sum_{i \neq j=1}^m \left\| (Q - \mathbb{1}\mu)^{|j-i|-1} \right\| (\text{Var}_\mu [f_i(Z_1)] + \text{Var}_\mu [f_j(Z_1)]) \\
 &\leq 2 \sum_{i=1}^m \sum_{l=0}^{\infty} \left\| (Q - \mathbb{1}\mu)^l \right\| \text{Var}_\mu [f_i(Z_1)] \\
 &\leq 2 \sum_{i=1}^m \text{Var}_\mu [f_i(Z_1)] \sum_{l=0}^{\infty} (1 - k\gamma_{ps})^{\lfloor l/k \rfloor / 2} \\
 &\leq 4 \sum_{i=1}^m \text{Var}_\mu [f_i(Z_1)] \sum_{l=0}^{\infty} (1 - k\gamma_{ps})^{\lfloor l/k \rfloor} \\
 &= \frac{4}{\gamma_{ps}} \sum_{i=1}^m \text{Var}_\mu [f_i(Z_1)]
 \end{aligned}$$

and combining with the diagonal terms $i = j$ the result follows as $\gamma_{ps} \leq 2$. □

5.6.2 Regeneration structure

Let us come back to the setting of quasi-trees.

Definition 5.6. A time $t \in \mathbb{N}/2$ is called a regeneration time if $(\mathcal{X}_{t-1/2}, \mathcal{X}_t)$ is a long-range edge crossed for the first and last time at time t .

Lemma 5.12. *A.s. there exists an infinity of regeneration times.*

The proof is similar to that of Lemma 3.3 in [127]. It can also be deduced from the tail estimates proved later on.

From now on, let $T_0 := 0, Y_0 := \iota(O), L_0 := 0$ and $(T_k)_{k \geq 1}$ be the sequence of successive regeneration times of \mathcal{X} . For all $k \geq 1$, define

$$Y_k = \iota(\mathcal{X}_{T_k-1/2}) \quad L_k := d(O, \mathcal{X}_{T_k}).$$

$(Y_k)_{k \geq 1}$ is the Markov chain that dictates the law of the environments between successive regeneration times. To describe this Markov chain, introduce the measures

$$\mathbb{Q}_u = \mathbb{P}[\cdot \mid \mathcal{X}_{1/2} = \eta(O), \tau_O = \infty, \iota(O) = u] \tag{5.52}$$

for all $u \in V$.

Remark 5.15. By Lemma 5.6, there exists a constant $q_0 > 0$ such that for all $u \in V$,

$$\mathbb{P}[\tau_O = \infty \mid \mathcal{X}_{1/2} = \eta(O), \iota(O) = u] \geq \mathbb{E}[q_{\text{Esc}}(\eta(O)) \mid \iota(O) = u] \geq q_0.$$

which implies

$$\mathbb{P}[\cdot, \tau_O = \infty \mid \mathcal{X}_{1/2} = \eta(O), \iota(O) = u] \leq \mathbb{Q}_u[\cdot] \leq q_0^{-1} \mathbb{P}[\cdot \mid \mathcal{X}_{1/2} = \eta(O), \iota(O) = u].$$

Consequently the laws of the quasi-tree \mathcal{G} under $\mathbb{P}[\cdot \mid \iota(O) = u]$ and \mathbb{Q}_u are within constant factors of each other for events that do not contradict $\tau_O = \infty$. In particular, up to a change of constant the uniform lower bounds of Proposition 5.4 and 5.5 for the random variables $q_{\text{Esc}}(O), q_{\text{Esc}}(\eta(O))$ also hold under the probability \mathbb{Q}_u , for any $u \in V$. Furthermore, the law of the centers added to the quasi-trees remains essentially uniform in the sense that for all $v \in V$ such that $V(v) \neq V(u)$

$$\mathbb{Q}_u[\iota(\eta(O)) = v] = \Theta(1/n). \quad (5.53)$$

Indeed by what precedes

$$\mathbb{Q}_u[\iota(\eta(O)) = v] \leq q_0^{-1} \mathbb{P}[\eta(O) = v \mid \iota(O) = u] = q_0^{-1}/n$$

and

$$\begin{aligned} \mathbb{Q}_u[\iota(\eta(O)) = v] &\geq \mathbb{E}[\mathbb{1}_{\iota(\eta(O))=v} \mathbb{1}_{\tau_O=\infty} \mid \iota(O) = u, \mathcal{X}_{1/2} = \eta(O)] \\ &\geq \mathbb{E}[\mathbb{1}_{\iota(\eta(O))=v} \mathbb{E}[q_{\text{Esc}}(\eta(O)) \mid \iota(O), \iota(\eta(O))] \mid \iota(O) = u] \\ &\geq q_0/n. \end{aligned}$$

Remark 5.16. Section 5.6.1 only considered integer valued processes for the time component of Markov renewal processes. To apply the results of this section we will thus implicitly identify regeneration times with an integer-valued process.

The following lemma is the analog of Lemma 3.6 in [92] and is proved in a similar way.

Lemma 5.13. • *The sequences $(Y_k, T_k)_{k \geq 1}$ and $(Y_k, L_k)_{k \geq 1}$ are Markov renewal processes.*

- *The sequence $(Y_{k+1}, \mathcal{G}_{\mathcal{X}_{T_k}} \setminus \mathcal{G}_{\mathcal{X}_{T_{k+1}}}, (\mathcal{X}_t)_{T_k \leq t < T_{k+1}})_{k \geq 0}$ is a Markov chain whose transition probabilities only depend on the first coordinate Y_k , ie the law of this triplet at time $k+1$ conditional on time k is only measurable with respect to Y_k .*
- *For all $k \geq 1$, conditional on Y_k , the pair $(\mathcal{G}_{\mathcal{X}_{T_k}}, (\mathcal{X}_t)_{t \geq T_k})$ has the law of $(\mathcal{G}_O, \mathcal{X})$ under the probability \mathbb{Q}_{Y_k} .*

Remark 5.17. Note that although (Y, T) starts at time 0, it is a Markov renewal process from time 1, and the delay is thus T_1 . However by the lemma, under the measure \mathbb{Q}_u , for any $u \in V$, T_1 has distribution given by a transition probability and thus (Y, T) becomes a Markov renewal process already from time 0, with 0 delay. We will use this observation often in the sequel.

Proof. Let

$$E := \{(g, (x_i)_{i \in \mathbb{N}/2}) \mid g \text{ quasi-tree, } (x_i)_{i \in \mathbb{N}/2} \text{ path in } g\},$$

where paths refer here to potential trajectories of the chain $(\mathcal{X}_t)_{t \in \mathbb{N}/2}$. Let $k \geq 1$ and $A, B \subset E$ two possible events for $(\mathcal{G}_{\mathcal{X}_{T_k}}, (\mathcal{X}_t)_{t \geq T_k})$. For all $x \in \mathcal{V}$ let $\tau_x^{(s)} := \inf\{t \geq s \mid \mathcal{X}_t \notin \mathcal{G}_x\}$. Notice that

$$\begin{aligned} & \mathbb{P} \left[\left(\mathcal{G}_{\mathcal{X}_{T_k}}, (\mathcal{X}_t)_{t \geq T_k} \right) \in A, \left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_{T_k}}, (\mathcal{X}_t)_{t < T_k} \right) \in B \right] \\ &= \sum_{s \geq 0} \mathbb{P} \left[T_k = s, \left(\mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t \geq s} \right) \in A \cap \{\tau_{\mathcal{X}_{s-1/2}} = \infty\}, \left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t < s} \right) \in B \right] \end{aligned}$$

Call a time $t \geq 0$ fresh if $(\mathcal{X}_{t-1/2}, \mathcal{X}_t)$ is a long-range edge visited for the first time at time t , and let $A_{k,s}$ be the event that there are k regeneration times before time s . Using the dependence properties between subquasi-trees and their complement described in Section 5.2.6,

$$\begin{aligned} & \mathbb{P} \left[T_k = s, \left(\mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t \geq s} \right) \in A \cap \{\tau_{\mathcal{X}_{s-1/2}} = \infty\}, \left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t < s} \right) \in B \right] \\ &= \mathbb{P} \left[s \text{ fresh}, A_{k-1,s}, \left(\mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t \geq s} \right) \in A \cap \{\tau_{\mathcal{X}_{s-1/2}} = \infty\}, \left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t < s} \right) \in B \right] \\ &= \mathbb{P} \left[s \text{ fresh}, A_{k-1,s}, \left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t < s} \right) \in B \right] \\ & \quad \times \mathbb{P} \left[\left(\mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t \geq s} \right) \in A \cap \{\tau_{\mathcal{X}_{s-1/2}}^{(s)} = \infty\} \mid s \text{ fresh}, A_{k-1,s}, \left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t < s} \right) \in B \right] \\ &= \mathbb{P} \left[s \text{ fresh}, A_{k-1,s}, \left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t < s} \right) \in B \right] \\ & \quad \times \mathbb{P} \left[\left(\mathcal{G}'_{O'}, (\mathcal{X}'_t)_{t \geq 1/2} \right) \in A \cap \{\tau_{O'} = \infty\} \mid \mathcal{X}'_{1/2} = \eta(O'), \iota(O') = \iota(\mathcal{X}_{s-1/2}) \right] \end{aligned}$$

where $(\mathcal{G}', O', \mathcal{X}')$ is an independent copy of $(\mathcal{G}, O, \mathcal{X})$. Also

$$\begin{aligned} \mathbb{P} \left[s \text{ fresh}, A_{k-1,s}, \left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t < s} \right) \in B \right] &= \frac{\mathbb{P} \left[T_k = s, \left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t < s} \right) \in B \right]}{\mathbb{P} \left[\tau_{\mathcal{X}_{s-1/2}}^{(s)} = \infty \mid \mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_s} \right]} \\ &= \frac{\mathbb{P} \left[T_k = s, \left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_s}, (\mathcal{X}_t)_{t < s} \right) \in B \right]}{\mathbb{P} \left[\tau_{O'} = \infty \mid \mathcal{X}'_{1/2} = \eta(O'), \iota(O') = \iota(\mathcal{X}_{s-1/2}) \right]} \end{aligned}$$

Summing over $s \geq 0$ yields eventually

$$\begin{aligned} & \mathbb{P} \left[\left(\mathcal{G}_{\mathcal{X}_{T_k}}, (\mathcal{X}_t)_{t \geq T_k} \right) \in A, \left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_{T_k}}, (\mathcal{X}_t)_{t < T_k} \right) \in B \right] \\ &= \mathbb{Q}_{Y_k} \left[\left(\mathcal{G}_{\mathcal{X}_{T_k}}, (\mathcal{X}_t)_{t \geq T_k} \right) \in A \right] \mathbb{P} \left[\left(\mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_{T_k}}, (\mathcal{X}_t)_{t < T_k} \right) \in B \right]. \end{aligned}$$

□

A non-negative random variable Z has *stretched exponential tail* if there exists $\alpha \in (0, 1)$ such that $\mathbb{E}[\exp Z^\alpha] < \infty$. A stretched exponential tail implies the existence of polynomial moments of all orders.

Lemma 5.14. *For all $u \in V$, under both measures \mathbb{Q}_u and $\mathbb{P}_O[\cdot \mid \iota(O) = u]$:*

- L_1 has exponential tail,
- T_1 has stretched exponential tail.

Furthermore, given $L, t \geq 0$, let $\tilde{T}_1^{(L)}$ denote the first time a regeneration occurs outside the ball $B_{\text{LR}}(O, L)$ and $U_t(L) = \sum_{s \leq t} \mathbb{1}_{\mathcal{X}_s \in B_{\text{LR}}(O, L)}$ the total time spent by the chain in $B_{\text{LR}}(O, L)$ before t . There exists $\alpha \in (0, 1]$ such that for all $L, m, k \in \mathbb{N}, m \geq 1$, for all $x \in B_{\text{LR}}(O, L)$ such that $d_{\text{LR}}(O, x) = L$, conditional on either $\mathcal{X}_0 = x$ or $\mathcal{X}_{1/2} = x$

$$\mathbb{P} \left[\tilde{T}_1^{(L)} > k + m \mid B_{\text{LR}}(O, L), U_{k+m}(L) \leq m \right] \leq e^{-k^\alpha}.$$

Remark 5.18. Proposition 2.2 in [2] shows that in the case of a randomly biased random walk on a Galton-Watson tree, stretched exponential can be the right order of magnitude for the tail of the first regeneration time. From the proof, it can be seen that this slowdown is precisely caused by the same trapping phenomenon as mentioned in Remark 5.8, when the walk can move from a part of the tree where it goes very fast to infinity to a part where it tends to go back to the root. This suggests that for some choices P_1, P_2 and p , stretched exponential tails give the right tail for the regeneration times in our model too.

Proof. For the tails, Remark 5.15 shows it suffices to prove the result for the law $\mathbb{P}_O[\cdot \mid \iota(O)]$. Suppose $\iota(O)$ is fixed. The proof of the exponential tail for regeneration levels is taken from that of Lemma 4.2 in [52]. It actually proves exponential tails for a stronger notion of regeneration, which looks at times when a long-range distance is reached for the first and last time. Namely let

$$\tilde{T}_1 := \inf\{t \in \mathbb{N}/2 \mid \forall s < t : d_{\text{LR}}(O, \mathcal{X}_s) < d_{\text{LR}}(O, \mathcal{X}_t), \forall s \geq t : d_{\text{LR}}(O, \mathcal{X}_s) \geq d_{\text{LR}}(O, \mathcal{X}_t)\}.$$

and $\tilde{L}_1 := d_{\text{LR}}(O, \mathcal{X}_{\tilde{T}_1})$. Obviously $T_1 \leq \tilde{T}_1$ and $L_1 \leq \tilde{L}_1$, so it suffices to prove an exponential tail for \tilde{L}_1 .

Suppose the chain is started at O . Consider the "ladder times" $(S_i)_{i=0}^K$, where $K \in \mathbb{N} \cup \{\infty\}$ is a random integer, possibly infinite, defined recursively as follows. Let $S_0 := 0$, and define τ_0 as the first return time to O . If $\tau_0 = \infty$ set $K := 0$. Otherwise, let $M_0 := \max\{d_{\text{LR}}(O, \mathcal{X}_t), 0 \leq t \leq \tau_0\}$ be the maximal distance reached until τ_0 and $S_1 := \min\{t \in \mathbb{N}/2 : d_{\text{LR}}(O, \mathcal{X}_t) > M_0\}$. Define the rest of the sequence recursively: let τ_i the first return time to level $d_{\text{LR}}(O, \mathcal{X}_{S_i}) - 1$. If $\tau_i = \infty$, $K := i$, and $S_j := \infty$ for all $j \geq i + 1$, otherwise set $M_i := \max\{d_{\text{LR}}(O, \mathcal{X}_t), 0 \leq t \leq \tau_i\}$ and $S_{i+1} := \min\{t \in \mathbb{N}/2 : d_{\text{LR}}(O, \mathcal{X}_t) > M_i\}$. Notice that $S_K = \tilde{T}_1$ if $K \in (0, \infty)$ and $\tilde{T}_1 = 1/2$ if $K = 0$.

Consequently for any $C > 0$, if $l \geq 1$ is large enough,

$$\begin{aligned} \mathbb{P}_O[\tilde{L}_1 > Cl \mid \iota(O)] &\leq \mathbb{P}_O[K < \infty, d(O, \mathcal{X}_{S_K}) > Cl \mid \iota(O)] \\ &= \mathbb{P}_O\left[K < \infty, \sum_{i=1}^K (d_{\text{LR}}(O, \mathcal{X}_{S_i}) - d_{\text{LR}}(O, \mathcal{X}_{S_{i-1}})) > Cl \mid \iota(O)\right] \\ &= \sum_{k=1}^{\infty} \mathbb{P}_O\left[\sum_{i=1}^k (d_{\text{LR}}(O, \mathcal{X}_{S_i}) - d_{\text{LR}}(O, \mathcal{X}_{S_{i-1}})) > Cl, \{\tau_i < \infty\}_{i=1}^{k-1}, \tau_k = \infty \mid \iota(O)\right] \\ &\leq \mathbb{P}_O[K > l \mid \iota(O)] + \mathbb{P}_O\left[\sum_{i=0}^{l-1} Z_i > Cl \mid \iota(O)\right]. \end{aligned}$$

where we write $Z_0 := d_{\text{LR}}(O, \mathcal{X}_{S_1}) \mathbb{1}_{\tau_0 < \infty}$, $Z_i := (d_{\text{LR}}(O, \mathcal{X}_{S_{i+1}}) - d_{\text{LR}}(O, \mathcal{X}_{S_i})) \mathbb{1}_{\tau_i < \infty}$ for all $i \geq 1$. To prove this is exponentially small in l , it thus suffices to establish that K has exponential tail and that for fixed $l \geq 1$, the second probability in the right-hand side is also exponentially small in l .

It was already observed in Remark 5.15 that $\mathbb{P}[\tau_O = \infty \mid \mathcal{X}_{1/2} = \eta(O), \iota(O)] = \Theta(1)$. We deduce that K is stochastically dominated by a geometric variable of constant parameter and thus has exponential tail, even conditional on $\iota(O)$.

On the other hand, by construction, for all $i \geq 1$ if $S_i < \infty$ the trajectory $(\mathcal{X}_t)_{S_i \leq t < \tau_i}$ is entirely contained in the subquasi-tree $\mathcal{G}_{\mathcal{X}_{S_i}}$. Furthermore, $d_{\text{LR}}(O, \mathcal{X}_{S_{i+1}}) = M_i + 1$ is determined by this trajectory. Thus for $i \geq 1$, $Z_i = (M_i - M_{i-1}) \mathbb{1}_{\tau_i < \infty}$ depends only on $\eta(\mathcal{X}_{S_{i-1/2}})$ and $\mathcal{G}_{\mathcal{X}_{S_i}}$. Letting $\tilde{Y}_i := \iota(\eta(\mathcal{X}_{S_{i-1/2}}))$ for $i \geq 0$, we can deduce that if $i \geq 1, l \geq 1$ and $S_i < \infty$,

$$\begin{aligned} \mathbb{P}_O[Z_i = l \mid \mathcal{G} \setminus \mathcal{G}_{\mathcal{X}_{S_i}}, (\mathcal{X}_s)_{s \leq S_i}] \\ = \mathbb{P}[\tau_{O'} < \infty, M'_0 = l \mid \mathcal{X}_{1/2} = \eta(O'), \iota(O') = \tilde{Y}_i] \end{aligned}$$

with (\mathcal{G}', O', M'_0) an independent copy of (\mathcal{G}, O, M_0) . We do not prove this in detail, which can be established as Lemma 5.13. See also the proof of Lemma 4.4 of [52]. If we extend the process (\tilde{Y}, Z) by considering that \tilde{Y}_i takes an arbitrary value \dagger if $S_i = \infty$ and consider that $Z_{i+1} = 0$ a.s. given $\tilde{Y}_i = \dagger$, the previous equation implies in particular that $(\tilde{Y}_i, Z_i)_{i \geq 0}$ is a Markov chain that satisfies the (M1) property. Furthermore for all $l \geq 1$,

$$\begin{aligned} \mathbb{P}[\tau_O < \infty, M_0 \geq l \mid \mathcal{X}_{1/2} = \eta(O), \iota(O)] \\ \leq \mathbb{P}[\exists t > s \geq l : d_{\text{LR}}(\eta(O), \mathcal{X}_s) = l - 1, \mathcal{X}_t = O \mid \mathcal{X}_{1/2} = \eta(O), \iota(O)]. \end{aligned}$$

This is an annealed probability of backtracking over a long-range distance L . Let us here anticipate on the sequel and use Lemma 5.18 which will prove this probability is exponentially small in L . With what precedes, this proves that there exists constants $C', \theta > 0$ such that for any possible state u of the chain \tilde{Y} , $\mathbb{E}[e^{\theta Z_1} \mid \tilde{Y}_0 = u] < C'$.

Exponential moments of Z_0 can be bounded similarly. The reader will then have recognized the condition to apply Lemma 5.11, which proves that for some constant $C > 0$, $\mathbb{P}_O \left[\sum_{i=0}^{l-1} Z_i > Cl \mid \iota(O) \right]$ is exponentially small in l , with all constants being independent of n .

Let us move to the proof of stretched exponential tails for T_1 . It is inspired from the proof of Prop. 2.2 in [2]. For ease of notation we will here keep the conditioning by $\iota(O)$, but all probabilities below should be considered conditioning on it. Given $k, t \geq 0$, let τ_k denote the hitting time of level k and K_t denote the time spent before t at a center, at half-integer times steps. For any $C > 0$ and $k \geq 0$,

$$\begin{aligned} \mathbb{P}_O [T_1 > Ck^3] &\leq \mathbb{P}_O [T_1 > \tau_k] + \mathbb{P}_O [Ck^3 < T_1 \leq \tau_k] \\ &\leq \mathbb{P}_O [T_1 > \tau_k] + \mathbb{P}_O [K_{Ck^3} \leq k^3] + \mathbb{P}_O [K_{\tau_k} > k^3]. \end{aligned}$$

The first term can be rewritten as $\mathbb{P} [T_1 > \tau_k] = \mathbb{P} [L_1 > k]$ which is exponentially small by the first part. For the second term, note that when the chain is at a non-center vertex x at integer times, it has probability $\Theta(1)$ to go to the center $\eta(x)$ by (H2). Thus the first hitting time of a center is stochastically dominated by a geometric variable, even under the quenched law. In turn we deduce

$$\mathbf{P}_O [K_{Ck^3} \leq k^3] \leq \mathbb{P} \left[\sum_{i=1}^{k^3} Z_i \geq Ck^3 \right],$$

where the Z_i are independent geometric variables of positive constant parameter. By Chernoff's inequality, the right-hand side is exponentially small in k^3 for a good choice of $C > 0$. The third term can be further decomposed as follows. Given a center $x \in \mathcal{V}$, let $N(x)$ denote the number of visits to x at half-integer times. Given $i \geq 1$, write z_i for the i -th distinct center visited by the chain and M_i for the number of distinct centers visited in \mathcal{V}_i at half-integer times. Then

$$\begin{aligned} \mathbb{P}_O [K_{\tau_k} > k^3] &\leq \mathbb{P}_O \left[\begin{array}{c} \text{more than } k^2 \text{ distinct centers are visited} \\ \text{before } \tau_k \text{ at half-integer times} \end{array} \right] \\ &\quad + \mathbb{P}_O [\exists i \leq k^2 : N(z_i) > k]. \end{aligned}$$

By union bound

$$\mathbb{P}_O \left[\begin{array}{c} \text{more than } k^2 \text{ distinct centers are visited} \\ \text{before } \tau_k \text{ at half-integer times} \end{array} \right] \leq \sum_{i=1}^k \mathbb{P}_O [M_i \geq k].$$

Let $\tau_k^{(i)}$ denote here the hitting time of the k -th distinct center of \mathcal{V}_i . By the strong Markov property

$$\begin{aligned} \mathbf{P}_O [M_i \geq k] &= \mathbf{P}_O \left[\tau_k^{(i)} < \infty \right] \\ &\leq \mathbf{E}_O \left[\mathbb{1}_{\tau_{k-1}^{(i)} < \infty} (1 - q_{\text{Esc}}(\mathcal{X}_{\tau_{k-1}^{(i)}}^{(k)})) \right]. \end{aligned}$$

Since the centers visited at $\tau_{k-1}^{(i)}$ and $\tau_k^{(i)}$ are distinct and at the same level, their subtrees are disjoint and independent. Hence

$$\mathbb{P}_O [M_i \geq k] \leq \mathbb{E} \left[\mathbf{P}_O \left[\tau_{k-1}^{(i)} < \infty \right] \left(1 - \min_{u,v} \mathbb{E} \left[q_{\text{Esc}}(\eta(O')) \mid \iota(O') = u, \iota(\eta(O')) = v \right] \right) \right],$$

where (\mathcal{G}', O') is as usual an independent copy of (\mathcal{G}, O) and $V(u) \neq V(v)$. By Lemma 5.6, the expected escape probability can be lower bounded by some constant $q_0 \in (0, 1]$ hence by induction

$$\mathbb{P} [M_i \geq k] \leq (1 - q_0)^{k-1}$$

and

$$\mathbb{P}_O \left[\begin{array}{c} \text{more than } k^2 \text{ distinct centers are visited} \\ \text{before } \tau_k \text{ at half-integer times} \end{array} \right] \leq k(1 - q_0)^{k-1}$$

is exponentially small in k .

To bound the last term, let

$$q'_{\text{Esc}}(x) := \mathbf{P} \left[\tau_x = \tau_{\eta(x)} = \infty \mid \mathcal{X}_{1/2} = x \right].$$

for all center $x \in \mathcal{V}$. Like $q_{\text{Esc}}(x)$, this quantity depends only on \mathcal{G}_x and $\eta(x)$. Furthermore, if y is a vertex at small-range distance 1 from x , then

$$\begin{aligned} q'_{\text{Esc}}(x) &\geq p(x, \eta(x))P(x, y)q_{\text{Esc}}(y) \\ &\geq \delta q_{\text{Esc}}(y), \end{aligned}$$

thus lower bounding q'_{Esc} is essentially the same as lower bounding q_{Esc} . For all $i \geq 1$,

$$\begin{aligned} \mathbf{P}_O [N(z_i) > k] &= \sum_{x \in \mathcal{V}} \mathbf{E} \left[\mathbb{1}_{z_i=x, N(x) > k} \right] \\ &\leq \sum_{x \in \mathcal{V}} \mathbf{E} \left[\mathbb{1}_{z_i=x} (1 - q'_{\text{Esc}}(x))^k \right] \\ &= \sum_{x \in \mathcal{V}} \mathbf{P}_O [z_i = x] (1 - q'_{\text{Esc}}(x))^k \end{aligned}$$

To average over the environment, use the "Ulam labelling" of Section 5.2.6:

$$\mathbb{P}_O [N(z_i) > k] \leq \sum_{x \in \mathcal{U}} \mathbb{E} \left[\mathbb{1}_{x \in \mathcal{V}} \mathbf{P}_O [z_i = x] \right] \mathbb{E} \left[(1 - q'_{\text{Esc}}(x))^k \mid x \in \mathcal{V}, \mathcal{G} \setminus \mathcal{G}_x \right]$$

From Proposition 5.4, one can easily obtain that

$$\max_{x \in \mathcal{U}} \mathbb{E} \left[(1 - q'_{\text{Esc}}(x))^k \mid x \in \mathcal{V}, \mathcal{G} \setminus \mathcal{G}_x \right] \leq e^{-k^c}$$

for some constant $c > 0$, hence by union bound

$$\mathbb{P}_O \left[\exists i \leq k^2 : N(z_i) > k \right] \leq k^2 e^{-k^c},$$

which is stretched exponential, as desired.

Finally, for the last statement note that if the chains is started precisely at the boudnary of the ball $B_{\text{LR}}(O, L)$, then a regeneration can occur within a number of steps which is bounded in L , thus from the lower bound on escape probabilities (Lemma 5.6) it is easy to lower bound $\mathbb{P}_x[U_{k+m}(L) \leq m \mid B_{\text{LR}}(O, L)]$ uniformly in n, L, m . Thus the conditioning by $U_{k+m}(L) \leq m$ only modifies upper bounds under the usual law of the quasi-tree by a factor of constant order. The upper bound for the conditional probability then follows from the same arguments as above. \square

5.6.3 Mixing of the regeneration chain

We now investigate the mixing properties of the Markov chain Y underlying the regeneration process, which we call the regeneration chain. Lemma 5.15 below will establish that Y has constant mixing time, allowing us later to get moment bounds similar to the iid case. The lemma actually proves a stronger mixing property of both the chain Y and the time process T that will be used to derive approximation of the invariant measure on X .

Lemma 5.15. *Let $(Q_t)_{t \geq 1}$ denote the transition kernels of the Markov renewal process (Y, T) and $Q := \sum_{t \geq 1} Q_t$, after identification of T with a process in \mathbb{N} . Recall that from Hypothesis ((H4)) there exists a constant $l \geq 1$ such that the set*

$$S = S(l) := \{x \in V \mid \forall y \in V : P(x, y) > 0 \Leftrightarrow (I + P)^l(y, x) > 0\}.$$

satisfies $|S| = 2n - o(n)$.

(i) For all $u \in V$ and $v \in S$,

$$Q(u, v) = \Theta(1/n). \quad (5.54)$$

As a consequence for all $\varepsilon \in (0, 1)$, the chain Y has mixing time $O_\varepsilon(1)$ as $n \rightarrow \infty$.

(ii) for all $u \in V$,

$$\sum_{t \geq 1} \sum_{v \in V} Q_t(u, v) \wedge Q_{t+1}(u, v) = \Theta(1). \quad (5.55)$$

Proof of Lemma 5.15. Let $\varepsilon \in (0, 1)$. The statement about the mixing time of Y is easily deduced from (5.54) and the fact that $|S| = 2n - o(n)$. Summing on v yields the following Doeblin's condition for Y : there exists a constant $c > 0$ such that

$$\|Q(u, \cdot) - Q(u', \cdot)\|_{\text{TV}} \leq 1 - c - o(1),$$

for all $u, u' \in V$. It is then easy to obtain from Doeblin's condition that Y has mixing time $t_{\text{mix}}^{(Y)}(\varepsilon) \leq (c^{-1} + o(1)) \log \varepsilon^{-1}$.

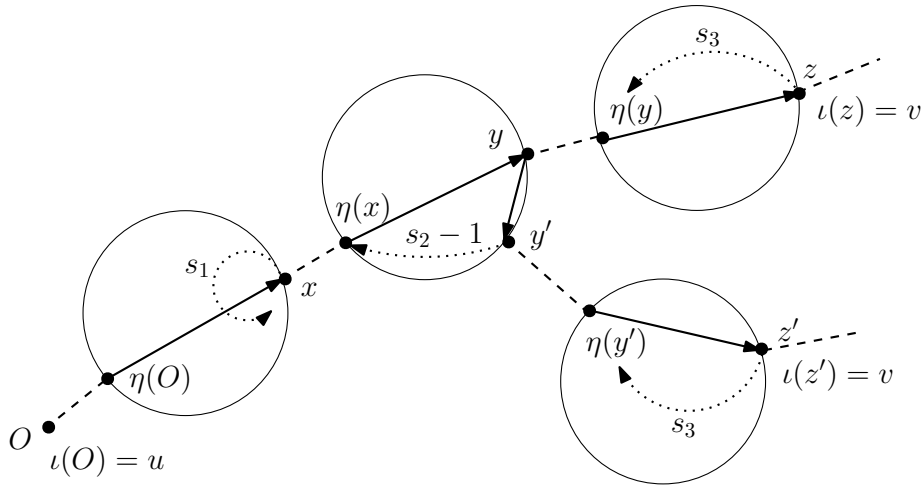


Figure 5.2: Argument of the proof of Lemma 5.15

Let us now prove (5.54). Consider $u, v \in V$ with $v \in S$. By definition for all $t \geq 1$

$$Q_t(u, v) = \mathbb{Q}_u [Y_1 = v, T_1 = t] = \mathbb{P} [T_1 = t, \iota(\mathcal{X}_{t-1/2}) = v \mid \tau_O = \infty, \iota(O) = u, \mathcal{X}_{1/2} = \eta(O)].$$

Probabilistic statements below will thus be made with respect to \mathbb{Q}_u unless stated otherwise.

Observe that if $V(u) \neq V(v)$, it is possible to realize $Y_1 = v$ with regeneration occurring at the first long-range edge crossed by the chain, which can be reached in half a step from $\eta(O)$ by Assumption (H3). Let $w \in V$ such that $P(w, v) > 0$. From Remark 5.15, $\mathbb{Q}_u [\iota(\eta(O)) = w] = \Theta(1/n)$. Using that transition probabilities are bounded uniformly in n by (H1), (H2), we deduce that $\mathbb{Q}_u [Y_1 = v] \geq \Theta(1/n)$. On the other hand, if $V(u) = V(v)$ the alternation between V_1 and V_2 forbids a regeneration at the first long-range edge. This is why we need to suppose $v \in S$ to make sure the chain can backtrack and prevent regeneration at the first long-range. Since S occurs with probability $1 - o(1)$ under the uniform law on V , centers of the quasi-tree \mathcal{G}_O also have probability $\Theta(1)$ to be in S under \mathbb{Q}_u . Thus using the above arguments we can lower bound by $\Theta(1/n)$ the probability that the chain goes from $\eta(O)$ to a vertex y with $\iota(y) = v$ using one long-range edge, comes back to $\eta(O)$, which ensures no regeneration occurred on the long-range edge, and returns to y , all that within a bounded number of steps. The chain can then escape in \mathcal{G}_y with probability $\Theta(1)$ by Lemma 5.6, which will imply $Y_1 = v$.

The proof of (5.55) is similar but requires taking time into account. Suppose $u \in V_1$ and let $v \in S \cap V_2$. Our argument is illustrated in Figure 5.2. The idea is to use an intermediary component of V_1 to make the regeneration time shift by 1. Consider the

set

$$S_1 := \{x \in V \cap S \mid \exists y \neq x \in V, s \leq l : P^2(x, y) > 0 \text{ and } P^s(y, x) > 0\}.$$

From Assumption (H3), $n - o(n)$ vertices of $S \cap V_1$ are recurrent and contained in a communicating class of size at least 3. Consequently if $x \in S \cap V_1$, either there exists $y \neq x$ on a loop around x of length less than l such that $P^2(x, y) > 0$, in which case $x \in S_1$, or x is in a communicating class $\{x, y, z\}$ of size exactly 3 with $P^2(x, x) = 1$. However in that case $P^2(y, z), P^2(z, y) > 0$, so $y, z \in S_1$. Thus S_1 has size at least $2n/3 - o(n)$ and probability $\Theta(1)$ under the uniform law on V_1 or \mathbb{Q}_w for any $w \in V_2$.

For any realization of \mathcal{G} , there exists a path $(\eta(O), x, \eta(x), y, \eta(y), z)$ with two long-range edges $(x, \eta(x))$ and $(y, \eta(y))$ whereas $(\eta(O), x)$, $(\eta(x), y)$, $(\eta(y), z)$ are small-range edges with distinct endpoints. If $\iota(\eta(O)) \in S$, which occurs with probability $\Theta(1)$, we can suppose that x was chosen so that $P^{s_1}(\iota(x), \iota(x)) > 0$, for some $s_1 \leq l + 1$. Similarly, $\iota(\eta(x)) \in S_1$ with probability $\Theta(1)$, in which case we can suppose y was chosen so that there exists a path between y and $\eta(x)$ of length $s_2 \leq l$ and a vertex $y' \neq x$ on this path with $P^2(\iota(\eta(x)), \iota(y')) > 0$. Then if $v \in S$ there exists $w \in V_2$ such that $P(w, v) > 0$ and $P(v, w)^{s_3} > 0$ for some $s_3 \leq l$. Conditional on the previous values of ι , $\iota(\eta(y)) = w$ with probability at least $\Theta(1/n)$. Gathering what precedes, make the following observation: from $\eta(O)$ the chain can go to x in half a step and then to z in two more steps, after which it can come back to x in $s_3 + s_2 + s_1$ steps going through y' , to eventually return at z in two steps and escape to infinity through $\eta(z)$. We will then have $Y_1 = v, T_1 = t + 1/2$ with $t = 4 + s_1 + s_2 + s_3$.

Now instead of y , we can apply the same arguments with y' , supposing that the chain reaches a vertex z' with $\iota(\eta(y')) = w$ and $\iota(z') = v$. By construction z' is reachable in just one more step from x , but the loop around x which passes through z' takes the same number of steps. Thus the previous argument can be applied similarly with just one time step difference to obtain $Y_1 = v, T_1 = t + 1 + 1/2$. Summing on the values of $\iota(\eta(O)), \iota(\eta(x))$, we obtain that

$$\sum_{t \geq 1} Q_t(u, v) \wedge Q_{t+1}(u, v) \geq \Theta(1/n).$$

The case where $u \in V_2$ is similar, with one less necessary long-range edge to cross to reach a component of $V_1 \cap S_1$, hence the bound applies to all $u \in V$. Then as $|V_2 \cap S| = n - o(n)$ summing over $v \in V_2 \cap S$ yields (5.55). \square

Let μ denote the stationary distribution of the Markov chain $(Y_k)_{k \geq 0}$

$$\mathbb{Q}_\mu := \sum_{u \in V} \mu(u) Q_u.$$

In the sequel $\mathbb{E}_{\mathbb{Q}_\mu}$ denotes the expectation with respect to μ . We use similar notations for variance, covariance, etc.

The fact that Y mixes in $O(1)$ steps will be sufficient for the rest of the analysis of the quasi-tree. Point (ii) in Lemma 5.15 will then be used to obtain the expression of the approximate invariant measure in (5.20).

Proposition 5.12. *Let ν be the law of $\iota(\eta(O))$ under \mathbb{Q}_μ . For all $v \in V$,*

$$\nu(v) = \Theta(1/n). \quad (5.56)$$

Furthermore given $L, t \geq 0$, let $U_t(L) = \sum_{s \leq t} \mathbb{1}_{\mathcal{X}_s \in B_{\text{LR}}(O, L)}$ denote the total time spent by the chain in $B_{\text{LR}}(O, L)$ before t . Consider here $\tilde{T}_1 = \tilde{T}_1(L)$ to be the first regeneration time outside $B_{\text{LR}}(O, L)$, while for $k \geq 2$, let \tilde{T}_k be the first regeneration time after \tilde{T}_{k-1} . For all $\varepsilon \in (0, 1)$ there exists $C(\varepsilon)$ such that for all $L, m \geq 0$ and $t \in \mathbb{N}/2$, for all $x \in B_{\text{LR}}(O, L)$, $d_{\text{LR}}(O, x) = L$, if $t \geq C(\varepsilon)m^2$ then

$$\sum_{v \in V} \left| \mathbb{P}_x \left[\exists k \geq 0 : \tilde{T}_k = t, \iota(\mathcal{X}_t) = v \mid B_{\text{LR}}(O, L), U_t(L) \leq m \right] - \frac{\nu(v)}{\mathbb{E}_{\mathbb{Q}_\mu} [T_1]} \right| \leq \varepsilon.$$

and

$$\sum_{v \in V} \left| \mathbb{P} \left[\exists k \geq 0 : \tilde{T}_k = t, \iota(\mathcal{X}_t) = v \mid \mathcal{X}_{1/2} = x, B_{\text{LR}}(O, L), U_t(L) \leq m \right] - \frac{\nu(v)}{\mathbb{E}_{\mathbb{Q}_\mu} [T_1]} \right| \leq \varepsilon.$$

Proof. The first statement is a direct consequence of (5.53).

Let us prove the second statement. Let $\varepsilon \in (0, 1), m, t \geq 1$. For all $k \geq 1$ let $\tilde{Y}_k := \iota(\mathcal{X}_{\tilde{T}_{k-1/2}})$. Apply Proposition 5.9 with the Markov renewal process of regeneration times considered here. Note that by the Markov property of (Y, T) conditioning by $B_{\text{LR}}(O, L), U_t(L) \leq m$ yields the same transition kernels as (Y, T) and only affects the law of $(\tilde{Y}_1, \tilde{T}_1)$. Lemmas 5.15 and 5.14 imply that the two quantities α and $\max_{u \in S} \mathbb{E}_u [T_1]$ in this Proposition are bounded uniformly in n , as is the mixing time $t_{\text{mix}}^{(Y)}(\varepsilon)$ of Y for all $\varepsilon \in (0, 1)$. Finally, the last statement of Lemma 5.14 implies that

$$\mathbb{P}_x \left[\tilde{T}_1 \geq m + C_1(\varepsilon) \mid B_{\text{LR}}(O, L), U_{m+C_1(\varepsilon)}(L) \leq m \right] \leq \varepsilon$$

for some $C_1(\varepsilon) > 0$. This gives the value of the quantity $K_\nu(\varepsilon)$ considered in Proposition 5.9, which thus proves that there exists $C(\varepsilon)$ such that for $t \geq C(\varepsilon)m^2$,

$$\sum_{u \in V} \left| \mathbb{P}_x \left[\exists k \geq 0 : \tilde{T}_k = t - 1/2, \tilde{Y}_k = u \mid B_{\text{LR}}(O, L), U_t(L) \leq m \right] - \frac{\mu(u)}{\mathbb{E}_{\mathbb{Q}_\mu} [T_1]} \right| \leq \varepsilon.$$

Then note that conditional on $\tilde{T}_k = t - 1/2, \tilde{Y}_k = u$, $\iota(\mathcal{X}_t)$ is distributed as $\iota(\eta(O))$ under \mathbb{Q}_u . Finally the arguments apply in the same way if the chain is started at time x at time $1/2$ instead of 0 . \square

5.7 Analysis on the quasi-tree III: concentration of drift and entropy

In this section we establish "nice properties" for the chain \mathcal{X} , proving in particular concentration of the drift and entropy.

5.7.1 Typical paths in quasi-trees

We start with the following lemma, which will basically ensure the chain \mathcal{X} only visits typical vertices by time $t = O(\log n)$, allowing to use the uniform lower bound on escape probabilities.

Lemma 5.16. *Let \mathcal{A} be a measurable set of quasi-trees. For all $t \geq 0$,*

$$\begin{aligned} \max_{u,v \in V} \mathbb{P}_O \left[\bigcup_{s \leq t} \{\mathcal{G}_{\mathcal{X}_s} \in \mathcal{A}\} \mid \iota(O) = u, \iota(\eta(O)) = v \right] \\ \leq t \max_{u,v \in V} \mathbb{P}_O [\mathcal{G}_O \in \mathcal{A} \mid \iota(O) = u, \iota(\eta(O)) = v]. \end{aligned}$$

Proof. Fix the types of $O, \eta(O)$. Let $z_0 := O$ and for $i \geq 1$ let z_i be the i -th distinct vertex visited by the chain. There are at most t distinct vertices visited up to time t , hence

$$\begin{aligned} \mathbb{P}_O \left[\bigcup_{s \leq t} \{\mathcal{G}_{\mathcal{X}_s} \in \mathcal{A}\} \mid O, \eta(O) \right] &\leq \mathbb{P} \left[\bigcup_{i \leq t} \{\mathcal{G}_{z_i} \in \mathcal{A}\} \mid O, \eta(O) \right] \\ &\leq t \max_{i \geq 0} \mathbb{P} [\mathcal{G}_{z_i} \in \mathcal{A} \mid O, \eta(O)]. \end{aligned}$$

Note that for $x \in \mathcal{U}$ the probability $\mathbf{P}_O[z_i = x]$ is measurable with respect to $\mathcal{G} \setminus \mathcal{G}_x$. Thus for all $i \geq 0$,

$$\begin{aligned} \mathbb{P}_O [\mathcal{G}_{z_i} \in \mathcal{A} \mid \iota(O), \iota(\eta(O))] &= \sum_{x \in \mathcal{U}} \mathbb{P}_O [z_i = x, \mathcal{G}_x \in \mathcal{A} \mid O, \eta(O)] \\ &= \sum_{x \in \mathcal{U}} \mathbb{E} [\mathbf{P}_O [z_i = x] \mathbb{P}_O [\mathcal{G}_x \in \mathcal{A} \mid \mathcal{G} \setminus \mathcal{G}_x] \mid O, \eta(O)]. \end{aligned}$$

Now the second factor in this sum can be bounded as

$$\mathbb{P}_O [\mathcal{G}_x \in \mathcal{A} \mid \mathcal{G} \setminus \mathcal{G}_x] \leq \max_{u,v \in V} \mathbb{P} [\mathcal{G}'_{O'} \in \mathcal{A} \mid \iota(O') = u, \iota(\eta(O')) = v]$$

with (\mathcal{G}', O') an independent copy of (\mathcal{G}, O) . The remaining terms then sum up to 1, yielding the result. \square

The previous Lemma will be combined with the following.

Lemma 5.17. *Given $\alpha, q_0 \in (0, 1]$ and $l \geq 0$, let $\mathcal{A}_l = \mathcal{A}_l(q_0, \alpha)$ be the set of quasi-trees for which every path of length l starting from O contains a proportion at least α of vertices x satisfying $q_{\text{Esc}}(x) \geq q_0$. There exists a constant C such that for all $l \geq 0$, on the event $\{\mathcal{G} \in \mathcal{A}_l\}$,*

$$\begin{aligned} \mathbf{P}_O [\exists t > s \geq 0 : \mathcal{X}_s \in \mathcal{G}_x, d_{\text{LR}}(O, \mathcal{X}_s) \geq l, d_{\text{LR}}(O, \mathcal{X}_t) = 0] &\leq e^{-Cl} \\ \mathbf{P}_O [\exists t \geq 0 : d_{\text{SR}}(O, \mathcal{X}_t) \geq l] &\leq e^{-Cl}. \end{aligned}$$

Similarly, for all $x \in \mathcal{V}$, on the event $\{\mathcal{G}_x \in \mathcal{A}_l\}$,

$$\begin{aligned} \mathbf{P}_x [\exists t > s \geq 0 : \mathcal{X}_s \in \mathcal{G}_x, d_{\text{LR}}(x, \mathcal{X}_s) \geq l, d_{\text{LR}}(x, \mathcal{X}_t) = 0] &\leq e^{-Cl} \\ \mathbf{P}_x [\exists t \geq 0 : \mathcal{X}_t \in \mathcal{G}_x, d_{\text{SR}}(\mathcal{X}_t^\circ, \mathcal{X}_t) \geq l] &\leq e^{-Cl}. \end{aligned}$$

Proof. Given $q_0 > 0$, consider the set $\text{Esc} = \text{Esc}(q_0) := \{x \in \mathcal{V} \mid q_{\text{Esc}}(x) \geq q_0\}$. For all $x \in \mathcal{G} \cap \text{Esc} \setminus \{O\}$, $\mathbf{P}_x[\tau_O < \infty] \leq 1 - q_{\text{Esc}}(x) < 1 - q_0$.

Let $l \geq 0$ and τ_l be the first time $t \in \mathbb{N}$ such that $d(O, \mathcal{X}_t) = l$ and τ_O the hitting time of O . By strong Markov's property,

$$\begin{aligned} \mathbf{P}_O [\exists t > s \geq 0 : d_{\text{LR}}(O, \mathcal{X}_s) \geq l, d_{\text{LR}}(O, \mathcal{X}_t) = 0] &= \mathbf{E}_O [\mathbb{1}_{\tau_l < \infty} \mathbf{P}_{\mathcal{X}_{\tau_l}}[\tau_O < \infty]] \\ &\leq \max_{x_1: d_{\text{LR}}(O, x_1) = l} \mathbf{P}_{x_1}[\tau_O < \infty]. \end{aligned}$$

Let $\tau_{\text{Esc}}^{(k)}$ be the k -th hitting time of $\text{Esc} \setminus \{O\}$ and suppose $x_1 \in \mathcal{G}$ is at long-range distance l from O . Then x_1 must also be at \mathcal{P} -distance at least l from O , thus if $\mathcal{G} \in \mathcal{A}_l$ and $\mathcal{X}_0 = x_1$ it is impossible that $\tau_O < \tau_{\text{Esc}}^{(\lfloor \alpha l \rfloor)}$ as this would imply the existence of a path contradicting the definition of \mathcal{A}_l . Thus when $\mathcal{G} \in \mathcal{A}_l$, applying Strong Markov's property at the successive stopping times $\tau_{\text{Esc}}^{(k)}$ yields

$$\begin{aligned} \mathbf{P}_{x_1}[\tau_O < \infty] &= \mathbf{P}_{x_1}[\tau_{\text{Esc}}^{(\lfloor \alpha l \rfloor)} \leq \tau_O < \infty] \\ &\leq (1 - q_0)^{\lfloor \alpha l \rfloor}. \end{aligned}$$

The argument for the second bound is similar. Let $\tau_{\text{Esc}}^{(k)}$ be now the k -th hitting time of Esc , so we do not discard O anymore, and let τ_l denote now the hitting time of $B_{\text{SR}}(O, l)$. As before, observe it is impossible to have $\tau_l < \tau_{\text{Esc}}^{(\lfloor \alpha l \rfloor)}$ if $\mathcal{G} \in \mathcal{A}_l$. However, to reach $B_{\text{SR}}(O, l)$ there must be no escape to infinity before τ_l . Hence on the event $\{\mathcal{G} \in \mathcal{A}_l\}$,

$$\begin{aligned} \mathbf{P}_O [\exists t \geq 0 : d_{\text{SR}}(O, \mathcal{X}_t) \geq l] &= \mathbf{P}_O[\tau_l < \infty] \\ &= \mathbf{P}_O[\tau_{\text{Esc}}^{(\lfloor \alpha l \rfloor)} \leq \tau_l < \infty] \\ &\leq (1 - q_0)^{\lfloor \alpha l \rfloor}, \end{aligned}$$

the last line being obtained by applying the strong Markov property at times $\tau_{\text{Esc}}^{(k)}$.

Finally the two last statements for a starting state $x \in \mathcal{V}$ are proved with the exact same reasoning. \square

We now establish an analog of Lemma 5.3 in the quasi-tree setting.

Lemma 5.18. *Let $\Gamma(R, L, M)$ denote the set of paths \mathbf{p} in \mathcal{G} such that \mathbf{p} does not deviate from a small-range distance more than R , backtrack over a long-range distance L or contain a subpath of length M without a regeneration edge. There exists $C > 0$ and $\alpha \in (0, 1]$ such that for all $R, L, M \geq 0$, for all $t \geq 0$, for any types of $O, \eta(O)$,*

$$\mathbb{P}_O [\mathcal{X}_s \cdots \mathcal{X}_{s+t} \notin \Gamma(R, L, M) \mid O, \eta(O)] \leq (s+t)e^{-C(R \wedge L \wedge M^\alpha)}.$$

Proof. Fix $R, L, M, s, t \geq 0$ and the types of $O, \eta(O)$. Let us start with the deviation property:

$$\begin{aligned} \mathbb{P}_O [\exists t' \in [s, s+t] : d_{\text{SR}}(\mathcal{X}_{t'}^\circ, \mathcal{X}_{t'}) \geq R \mid O, \eta(O)] \\ \leq \mathbb{P}_O [\exists t' \leq s+t : d_{\text{SR}}(O, \mathcal{X}_{t'}) \geq R \mid O, \eta(O)] \\ + \mathbb{P} [\exists t_1 \leq t_2 \leq s+t : \mathcal{X}_{t_2} \in \mathcal{G}_{\mathcal{X}_{t_1}}, d_{\text{SR}}(\mathcal{X}_{t_2}^\circ, \mathcal{X}_{t_2}) \geq R \mid O, \eta(O)] \quad (5.57) \end{aligned}$$

Given $q_0 > 0$, $\alpha \in (0, 1)$, let $\mathcal{A}_R = \mathcal{A}_R(q_0, \alpha)$ be the set of quasi-trees, for which every path of length R starting from O contains a proportion at least α of vertices x such that $q_{\text{Esc}}(x) \geq q_0$. Then the first term can be bounded as

$$\begin{aligned} \mathbb{P}_O [\exists t' \leq s+t : d_{\text{SR}}(O, \mathcal{X}_{t'}) \geq R \mid O, \eta(O)] \leq \mathbb{P} [\mathcal{G} \notin \mathcal{A}_R \mid O, \eta(O)] \\ + \mathbb{E} [\mathbb{1}_{\mathcal{G} \in \mathcal{A}_R} \mathbf{P}_O [\exists t' \leq s+t : d_{\text{SR}}(O, \mathcal{X}_{t'}) \geq R] \mid O, \eta(O)]. \end{aligned}$$

By Proposition 5.5 there exist constants $q_0, \alpha > 0$ such that the first term can be made exponentially small in R , while Lemma 5.17 shows the second term is also exponentially small. The second term of (5.57) can be bounded similarly:

$$\begin{aligned} \mathbb{P} [\exists t_1 \leq t_2 \leq s+t : \mathcal{X}_{t_2} \in \mathcal{G}_{\mathcal{X}_{t_1}}, d_{\text{SR}}(\mathcal{X}_{t_2}^\circ, \mathcal{X}_{t_2}) \geq R \mid O, \eta(O)] \\ \leq \mathbb{P} [\exists t_1 \leq s+t : \mathcal{G}_{\mathcal{X}_{t_1}} \notin \mathcal{A}_R \mid O, \eta(O)] \\ + \sum_{t_1 \leq s+t} \mathbb{E} [\mathbb{1}_{\mathcal{G}_{\mathcal{X}_{t_1}} \in \mathcal{A}_R} \mathbf{P}_{\mathcal{X}_{t_1}} [\exists t_2 \geq 0 : \mathcal{X}_{t_2} \in \mathcal{G}_{\mathcal{X}_{t_1}}, d_{\text{SR}}(\mathcal{X}_{t_2}^\circ, \mathcal{X}_{t_2}) \geq R] \mid O, \eta(O)] \\ \leq (s+t)e^{-CR} \end{aligned}$$

for some constant $C > 0$, using Lemma 5.16 and Proposition 5.5 to bound the first term and Lemma 5.17 for the second.

The probability that the trajectory backtrack over a long-range distance L between s and $s+t$ is established with the same arguments. It then remains to handle the regeneration requirement. By union bound for all $t \geq 0$ and $u \in V$

$$\begin{aligned} & \mathbb{P}_O \left[\exists s' \in [s, s+t] : [s', s'+M] \cap \{T_k, k \geq 1\} = \emptyset \mid O, \eta(O) \right] \\ & \leq \mathbb{P}_O \left[\exists k \leq s+t : T_{k+1} - T_k \geq M \mid O, \eta(O) \right] \\ & \leq (s+t) \max_{k \leq s+t} \mathbb{P}_O \left[T_{k+1} - T_k \geq M \mid O, \eta(O) \right]. \end{aligned}$$

The stretched exponential tails of regeneration times (Lemma 5.14) conclude the proof. \square

5.7.2 Concentration of the drift

Proposition 5.13. *Let $d := \frac{\mathbb{E}_{\mathbb{Q}_\mu}[L_1]}{\mathbb{E}_{\mathbb{Q}_\mu}[T_1]}$. Then for all $s \geq 0$, a.s.*

$$\frac{d_{\text{LR}}(\mathcal{X}_s, \mathcal{X}_{s+t})}{t} \xrightarrow[t \rightarrow \infty]{} d. \quad (5.58)$$

Furthermore, there exists $\alpha > 0$ for which the following holds. For all $\varepsilon > 0$ there exists $C = C(\varepsilon)$ such that for all $s, t \geq 0$, for all types of $O, \eta(O)$,

$$\mathbb{P}_O \left[|d_{\text{LR}}(\mathcal{X}_s, \mathcal{X}_{s+t}) - dt| > C\sqrt{t} \mid O, \eta(O) \right] \leq \varepsilon + C\sqrt{\varepsilon}e^{-t^\alpha}. \quad (5.59)$$

Proof. For notational simplicity we omit writing the conditioning by $\iota(O)$ and $\iota(\eta(O))$. As can be checked this conditioning does not affect the proof as the technical results that will be used hold even conditional on the long-range edge at the root. For all $t \geq 0$, let

$$N_t := \max\{k \geq 0 \mid T_k \leq t\}.$$

Then

$$L_{N_t} \leq d_{\text{LR}}(O, \mathcal{X}_t) \leq L_{N_t} + T_{N_t+1} - T_{N_t}.$$

It is easy to prove that $(T_{N_t+1} - T_{N_t})/t \rightarrow 0$, hence the law of large numbers (5.58) follows from Lemma 5.13 and Proposition 5.7 which prove $N_t/t \xrightarrow[t \rightarrow \infty]{} 1/\mathbb{E}_\mu[T_1]$ and $L_k/k \xrightarrow[k \rightarrow \infty]{} \mathbb{E}_\mu[L_1]$ a.s..

Then we establish (5.59) in the case $s = 0$. It suffices to prove that for all $\varepsilon > 0$ there exists $C > 0$ such that for all $t, k \geq 0$

$$\begin{aligned} & \mathbb{P}_O \left[\left| N_t - \frac{t}{\mathbb{E}_{\mathbb{Q}_\mu}[T_1]} \right| > C\sqrt{t} \right] \leq \varepsilon \\ & \mathbb{P}_O \left[|L_k - \mathbb{E}_{\mathbb{Q}_\mu}[L_1]k| > C\sqrt{k} \right] \leq \varepsilon. \end{aligned} \quad (5.60)$$

Indeed if this holds using that regeneration levels are non-decreasing we deduce that $|L_{N_t} - dt| > C\sqrt{t}$ with probability at most ε for some $C = C(\varepsilon)$, whereas union bound and Markov's inequality show that for all $C' > 0$

$$\begin{aligned} & \mathbb{P}_O \left[T_{N_{t+1}} - T_{N_t} > C'\sqrt{t}, \left| N_t - \frac{t}{\mathbb{E}_{\mathbb{Q}_\mu}[T_1]} \right| \leq C\sqrt{t} \right] \\ & \leq \mathbb{P}_O \left[\exists k : \left| k - \frac{t}{\mathbb{E}_{\mathbb{Q}_\mu}[T_1]} \right| \leq C\sqrt{t}, T_{k+1} - T_k > C'\sqrt{t} \right] \\ & \leq 2C\sqrt{t} \frac{\max_{u \in V} \mathbb{E}_{\mathbb{Q}_u}[T_1] \vee \mathbb{E}_O[T_1]}{C'\sqrt{t}} \\ & \leq \varepsilon \end{aligned}$$

for large enough $C' = C'(\varepsilon)$, using also Lemma 5.14 to argue the expectations are $O(1)$.

Let us now prove the claim. We only prove the upper tail of the first inequality as the other bounds are established similarly. Let

$$k = \left\lfloor \frac{t}{\mathbb{E}_{\mathbb{Q}_\mu}[T_1]} + C\sqrt{t} \right\rfloor.$$

The objective is to apply the Bienaymé Chebychev inequality with the variance bound for Markov chains (5.50), which is valid only when started at equilibrium. Let $\varepsilon \in (0, 1/2)$ and $t_0 = t_{\text{mix}}^{(Y)}(\varepsilon)$ be the mixing time of $(Y_l)_{l \geq 0}$, then decompose $T_k = Z_1 + Z_2$ with $Z_1 = \sum_{i=1}^{t_0} T_i - T_{i-1}$ and $Z_2 = \sum_{i=t_0+1}^k T_i - T_{i-1}$.

We write further $Z'_2 := Z_2 - \mathbb{E}_{\mathbb{Q}_\mu}[Z_2]$. Stationarity implies $\mathbb{E}_{\mathbb{Q}_\mu}[Z_2] = (k - t_0)\mathbb{E}_{\mathbb{Q}_\mu}[T_1]$ hence

$$\begin{aligned} \mathbb{P}_O[N_t > k] &= \mathbb{P}[T_k < t] \\ &= \mathbb{P}_O[Z_1 + Z'_2 < t - (k - t_0)\mathbb{E}_{\mathbb{Q}_\mu}[T_1]] \\ &\leq \mathbb{P}_O[Z_1 + Z'_2 < z] \end{aligned}$$

where

$$z := (t_0 + 1)\mathbb{E}_{\mathbb{Q}_\mu}[T_1] - C\mathbb{E}_{\mathbb{Q}_\mu}[T_1]\sqrt{t}.$$

By Lemmas 5.14 and 5.15, t_0 and the expectation above can be bounded by constants independent of n . Thus for some constant $C' > 0$,

$$\begin{aligned} \mathbb{P}_O[Z_1 + Z'_2 < z] &\leq \mathbb{P}_O[|Z_1 + Z'_2| > 2C'\sqrt{t}] \\ &\leq \mathbb{P}_O[Z_1 > C'\sqrt{t}] + \mathbb{P}_O[|Z'_2| > C'\sqrt{t}]. \end{aligned}$$

From Lemma 5.15 and the fact that $t_0 = O(1)$, we also deduce that $\mathbb{E}_O[Z_1] = O(1)$ so Markov's inequality ensures the first term is smaller than ε for all t by taking the

constant C' large enough. For the second term, since $t_0 = t_{\text{mix}}^{(Y)}(\varepsilon)$

$$\begin{aligned} \mathbb{P}_O [|Z'_2| > C'\sqrt{t}] &= \sum_y \mathbb{P}_O [Y_{t_0} = y] \mathbb{Q}_y [|T_{k-t_0} - \mathbb{E}_{\mathbb{Q}_\mu} [T_{k-t_0}]| > C'\sqrt{t}] \\ &\leq \|Y_{t_0} - \mu\|_{\text{TV}} + \mathbb{Q}_\mu [|T_{k-t_0} - \mathbb{E}_{\mathbb{Q}_\mu} [T_{k-t_0}]| > C'\sqrt{t}] \\ &\leq \varepsilon + \mathbb{Q}_\mu [|T_{k-t_0} - \mathbb{E}_{\mathbb{Q}_\mu} [T_{k-t_0}]| > C'\sqrt{t}]. \end{aligned}$$

Using the Bienaymé-Chebychev inequality with (5.50), we deduce

$$\mathbb{Q}_\mu [|T_{k-t_0} - \mathbb{E}_{\mathbb{Q}_\mu} [T_{k-t_0}]| > C'\sqrt{t}] \leq \frac{\text{Var}_{\mathbb{Q}_\mu}(T_{k-t_0})}{(C')^2 t} \leq \frac{C''(k-t_0) \text{Var}_{\mathbb{Q}_\mu} [T_1]}{(C')^2 t},$$

for some constant $C'' > 0$, as the regeneration chain has constant mixing time by Lemma 5.15. By Lemma 5.14, the variance of T_1 is of constant order while $k = O(t)$, hence the probability above can be made smaller than ε for all t by taking C' large enough.

For the general case $s \geq 0$, we apply the previous arguments with a shifted version of the processes. For $s \geq 0$ fixed, consider

$$T_k^{(s)} := T_{N_s+k} - s, \quad L_k^{(s)} := L_{N_s+k} - d_{\text{LR}}(O, \mathcal{X}_s)$$

if $k \geq 1$ and $T_0^{(s)} := 0, L_0^{(s)} := 0$. Then let

$$N_t^{(s)} := \max\{k \geq 0 \mid T_k^{(s)} \leq t\}.$$

Notice that

$$L_{N_t^{(s)}}^{(s)} \leq d_{\text{LR}}(\mathcal{X}_s, \mathcal{X}_{s+t}) \leq L_{N_t^{(s)}}^{(s)} + T_{N_t^{(s)}+1}^{(s)} - T_{N_t^{(s)}}^{(s)}.$$

These processes still satisfy the conclusions of Lemma 5.13 and have the same increments as the usual regeneration times. Thus the only thing to be careful when applying the above arguments is the law of the first regeneration time that now depends on s . Using the concentration (5.60) for N_s , union bound and Lemma 5.14, we can see that for all $m \geq 0$,

$$\begin{aligned} \mathbb{P} [T_1^{(s)} > m] &= \sum_{k \geq 0} \mathbb{P} [N_s = k, T_{k+1} - s > m] \\ &\leq \mathbb{P} [|N_s - s / \mathbb{E}_{\mathbb{Q}_\mu} [T_1]| > C\sqrt{s}] + \mathbb{P} [\exists k : |k - s / \mathbb{E}_{\mathbb{Q}_\mu} [T_1]| \leq C\sqrt{s}, T_{k+1} - T_k > m] \\ &\leq \varepsilon + 2C\sqrt{s}e^{-m^\alpha}. \end{aligned}$$

Taking $m = \sqrt{t}$ yields the stretched exponential term of (5.59). Then for higher regeneration times $(T_k^{(s)})_{k \geq 2}$, the above arguments apply. □

5.7.3 Concentration of the entropy

The concentration of the entropy in Proposition 5.2 is based on the convergence of the entropy for the loop-erased chain in the quasi-tree, that is the convergence of $-\log \mathbf{P}[\xi'_k = \xi_k \mid \mathcal{X}]/k$ towards a constant, the entropic rate of the chain. Such convergence is well-known in the context of groups or random walks on Galton-Watson trees, see [102, 126]. We will however not prove this result but establish concentration directly for a notion of weights similar to those of (5.16). Of course, these are designed to mimic the law of the loop-erased chain, so the argument is similar. In fact the first step is to prove the convergence and concentration of the loop-erased chain when restricted to regeneration steps.

Lemma 5.19. *Let ξ' be an independent copy of the loop-erased chain ξ . There exists $h' = \Theta(1)$ such that*

$$\lim_{k \rightarrow \infty} \frac{-\log \mathbf{P}[\xi'_{L_k} = \xi_{L_k} \mid \mathcal{X}]}{k} = h'. \quad (5.61)$$

Furthermore, for all $\varepsilon > 0$, there exists $C(\varepsilon) > 0$ such that for all $k, l \geq 1$,

$$\begin{aligned} \mathbb{P}\left[\left| -\log \mathbf{P}[\xi'_{L_k} = \xi_{L_k} \mid \mathcal{X}] - h'k \right| > C(\varepsilon)\sqrt{k} \mid O, \eta(O) \right] &\leq \varepsilon, \\ \mathbb{P}\left[\left| -\log \mathbf{P}[\xi'_{L_{k+l}} = \xi_{L_{k+l}} \mid \mathcal{X}, \xi'_{L_k} = \xi_{L_k}] - h'l \right| > C\sqrt{l} \mid O, \eta(O) \right] &\leq \varepsilon. \end{aligned} \quad (5.62)$$

Proof. Again let us omit the conditioning by $O, \eta(O)$ which does not affect the argument. Given $k \geq 1$, we write $\mathcal{G}(k) := \mathcal{G}_{\mathcal{X}_{T_k}}$, to simplify notations. Given $u \in V$ and g a possible realization of the subquasi-tree \mathcal{G}_O , let

$$\mathbf{Q}_{u,g} := \mathbb{P}[\cdot \mid \mathcal{X}_{1/2} = \eta(O), \tau_O = \infty, \iota(O) = u, \mathcal{G}_O = g].$$

Finally let $\xi'(k)$ denote a loop-erased chain on $\mathcal{G}(k-1)$, started at $\mathcal{X}_{T_{k-1}}$, independent of \mathcal{X} . Having $\xi'_{L_k} = \xi_{L_k}$ implies $\xi_{L_m} \in \xi'$ for all $m \leq k$. Thus for $k \geq 2$

$$\begin{aligned} \mathbf{P}[\xi_{L_k} \in \xi' \mid \mathcal{X}] &= \mathbf{E}\left[\mathbb{1}(\xi'_{L_{k-1}} = \xi_{L_{k-1}}) \mathbf{P}[\xi'_{L_k} = \xi_{L_k} \mid \mathcal{X}, (\xi'_i)_{i \leq L_{k-1}}] \mid \mathcal{X} \right] \\ &= \mathbf{E}\left[\mathbb{1}(\xi'_{L_{k-1}} = \xi_{L_{k-1}}) \mathbf{Q}_{Y_{k-1}, \mathcal{G}(k-1)}[\xi_{L_k} \in \xi'(k) \mid (\mathcal{X}_t)_{t \geq T_{k-1}}] \mid \mathcal{X} \right] \\ &= \mathbf{P}[\xi'_{L_{k-1}} = \xi_{L_{k-1}} \mid \mathcal{X}] \mathbf{Q}_{Y_{k-1}, \mathcal{G}(k-1)}[\xi_{L_k} \in \xi'(k) \mid (\mathcal{X}_t)_{t \geq T_{k-1}}]. \end{aligned}$$

Letting $Z_1 := -\log \mathbf{P}[\xi'_{L_1} = \xi_{L_1} \mid \mathcal{X}]$ and

$$Z_i := -\log \mathbf{Q}_{Y_{i-1}, \mathcal{G}(i-1)}[\xi_{L_i} \in \xi'(i) \mid (\mathcal{X}_t)_{t \geq T_{i-1}}]$$

for $i \geq 2$, we just proved that

$$-\log \mathbf{P}[\xi'_{L_k} = \xi_{L_k} \mid \mathcal{X}] = \sum_{i=1}^k Z_i$$

and

$$-\log \mathbf{P} \left[\xi'_{L_{k+l}} = \xi_{L_{k+l}} \mid \mathcal{X}, \xi'_{L_k} \right] = \sum_{i=k+1}^{k+l} Z_i.$$

Under \mathbb{Q}_μ , the sequence $(Y_k)_{k \geq 1}$ is stationary. Lemma 5.13 then shows that $(Z_i)_{i \geq 2}$ is stationary as well. Provided $Z_1 < \infty$ a.s., $\mathbb{E}_O [Z_1] < \infty$ and $\mathbb{E}_{\mathbb{Q}_\mu} [|Z_1|] < \infty$, which is proved afterwards, Birkhoff's ergodic theorem implies the law of large numbers (5.61).

Let us now show that for all $l \geq 1$, there exists a constant $C_l \geq 0$ such that

$$\mathbb{E}_O \left[|Z_1|^l \right] < C_l$$

and for all $i \geq 2$ and $u \in V$

$$\mathbb{E}_{\mathbb{Q}_u} \left[|Z_i|^l \right] \leq C_l. \quad (5.63)$$

This is sufficient to justify that $Z_1 < \infty$ a.s., $\mathbb{E}_{\mathbb{Q}_\mu} |Z_i| < \infty$ and will be used to prove (5.62). We only treat the case $i \geq 2$ as the case $i = 1$ is similar. Let $l \geq 1$ and $u \in V$. Let us identify here a long-range edge $(\eta(x), x)$ with its center vertex x . By Lemma 5.13

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}_u} \left[|Z_i|^l \right] &= \mathbb{E}_{\mathbb{Q}_u} \left[\sum_{x \in \mathcal{V}} \mathbf{Q}_{u, \mathcal{G}} [\mathcal{X}_{T_i} = x] \left(-\log \mathbf{Q}_{Y_{i-1}, \mathcal{G}(i-1)} [x \in \xi'(i)] \right)^l \right] \\ &= \sum_{v \in V} \mathbb{Q}_u [Y_{i-1} = v] \mathbb{E}_{\mathbb{Q}_v} \left[\sum_{x \in \mathcal{V}} \mathbf{Q}_{v, \mathcal{G}} [\mathcal{X}'_{T_1} = x] \left(-\log \mathbf{Q}_{v, \mathcal{G}} [x \in \xi'] \right)^l \right] \end{aligned}$$

where \mathcal{X}' is an independent copy of \mathcal{X} under the law $\mathbf{Q}_{v, \mathcal{G}}$, ξ' is its loop-erased trace and T_1' its first regeneration time. Thus it suffices to prove the upper bound

$$\max_{u \in V} \mathbb{E}_{\mathbb{Q}_u} \left[\sum_{x \in \mathcal{V}} \mathbf{Q}_{u, \mathcal{G}} [\mathcal{X}_{T_1} = x] \left(-\log \mathbf{Q}_{u, \mathcal{G}} [x \in \xi] \right)^l \right] \leq C_l.$$

Fix $u \in V$. For all $k \geq 1$, if $d_{\mathcal{P}}(\eta(O), x) = k$, the chain goes from $\eta(O)$ to x with probability at least δ^k , after which it escapes to infinity in \mathcal{G}_x with probability $q_{\text{Esc}}(x)$, hence

$$\begin{aligned} \sum_{x \in \mathcal{V}} \mathbf{Q}_{u, \mathcal{G}} [\mathcal{X}_{T_1} = x] \left(-\log \mathbf{Q}_{u, \mathcal{G}} [x \in \xi] \right)^l &\leq \sum_{k \geq 1} \mathbf{Q}_{u, \mathcal{G}} [d_{\mathcal{P}}(\eta(O), \mathcal{X}_{T_1}) = k] k^l (-\log \delta)^l \\ &\quad + \sum_{x \in \mathcal{V}} \mathbf{Q}_{u, \mathcal{G}} [\mathcal{X}_{T_1} = x] \left(-\log q_{\text{Esc}}(x) \right)^l \end{aligned}$$

Averaging on the environment yields

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}_u} \left[\sum_{x \in \mathcal{V}} \mathbf{Q}_{u, \mathcal{G}} [\mathcal{X}_{T_1} = x] \left(-\log \mathbf{Q}_{u, \mathcal{G}} [x \in \xi] \right)^l \right] &\leq (-\log \delta)^l \mathbb{E}_{\mathbb{Q}_u} [d_{\mathcal{P}}(\eta(O), \mathcal{X}_{T_1})^l] \\ &\quad + \mathbb{E}_{\mathbb{Q}_u} \left[\left(-\log q_{\text{Esc}}(\mathcal{X}_{T_1}) \right)^l \right] \end{aligned}$$

Now for any $k \geq 1$,

$$\mathbb{Q}_u [d_{\mathcal{P}}(\eta(O), \mathcal{X}_{T_1}) \geq k] \leq \mathbb{Q}_u [T_1 \geq k]$$

which is stretched exponential by Lemma 5.14, hence $\mathbb{E}_{\mathbb{Q}_u} [d_{\mathcal{P}}(\eta(O), \mathcal{X}_{T_1})^l] = O(1)$. On the other hand

$$\mathbb{E}_{\mathbb{Q}_u} [(-\log q_{\text{Esc}}(\mathcal{X}_{T_1}))^l] = \sum_{v \in V} \mathbb{Q}_u [Y_1 = v] \mathbb{E}_{\mathbb{Q}_v} [(-\log q_{\text{Esc}}(\eta(O)))^l].$$

By Proposition 5.4 and Remark 5.15, $\mathbb{E}_{\mathbb{Q}_v} [(-\log q_{\text{Esc}}(\eta(O)))^l] = O(1)$ for any $v \in V$, hence $\mathbb{E}_u [(-\log q_{\text{Esc}}(\mathcal{X}_{T_1}))^l] = O(1)$ as well and we deduce (5.63) for some constant C_l .

To establish (5.62), we argue as in the end of the proof of Proposition 5.13 for the drift. The bound is eventually proved using the Bienaymé-Chebychev inequality. Since Y mixes in constant time t_0 by Lemma 5.15, separating the sum $\sum_{i=1}^k Z_i$ into two sums corresponding respectively to regeneration times before or after t_0 shows that it suffices to establish the variance bound $\text{Var}_{\mathbb{Q}_\mu} [\sum_{i=2}^k Z_i] = O(k)$ under the stationary measure. Let us modify the definition of Z_1 , setting $Z_1 := -\log \mathbf{Q}_{\iota(O), \mathcal{G}_O} [\xi_{L_1} \in \xi' \mid \mathcal{X}]$ so that the whole sequence $(Z_i)_{i \geq 1}$ is stationary. It then suffices to prove that for all $k \geq 1$, $\text{Cov}_{\mathbb{Q}_\mu} (Z_1, Z_k)$ is stretched exponential in k .

Let $k > 1$. Let τ'_{k-1} be the first time the chain \mathcal{X}' reaches the L_{k-1} -th level, and consider the loop-erased trace $\xi'(1, k)$ obtained from observing \mathcal{X}' up to time τ'_{k-1} . Define $W_1 := \exp(-Z_1)$, $W_{1k} := \mathbf{Q}_{\iota(O), \mathcal{G}_O} [\xi_{L_1} \in \xi'(1, k) \mid \mathcal{X}]$, $Z_{1,k} := -\log W_{1k}$. By the inequality $|\log x - \log y| \leq |x - y| / (x \wedge y)$,

$$|Z_1 - Z_{1,k}| \leq \frac{|W_1 - W_{1k}|}{W_1 \wedge W_{1k}}$$

Observe that if ξ_{L_1} is exclusively in one of the two paths $\xi'(1, k)$, ξ' then the chain \mathcal{X}' backtracks from the L_{k-1} -th level to the $(L_1 - 1)$ -th. Letting A' denote this event, we thus have

$$|W_1 - W_{1k}| \leq \mathbf{Q}_{\iota(O), \mathcal{G}_O} [A' \mid \mathcal{X}].$$

Let $\mathcal{A}_k(q_0, \alpha)$ be the set of quasi-trees for which every length k path from the root contains a proportion at least $\alpha > 0$ of vertices x such that $q_{\text{Esc}}(x) \geq q_0 > 0$. For some constants $q_0, \alpha, \varepsilon \in (0, 1)$ and $C_1 > 0$ to determine, consider the events

$$\begin{aligned} E_0 &:= \{\mathcal{G}_O \in \mathcal{A}_k(q_0, \alpha)\} \cap \{q_{\text{Esc}}(\xi_{L_1}) \geq \varepsilon^k\} \\ E_1 &= \{d_{\mathcal{P}}(O, \mathcal{X}_{T_1}) \leq k/C_1\} \end{aligned}$$

If $\mathcal{G}_O \in \mathcal{A}_k(q_0, \alpha)$, Lemma 5.17 implies $\mathbf{Q}_{\iota(O), \mathcal{G}_O} [A' \mid \mathcal{X}]$ is exponentially small in k . On E_1 , the chain \mathcal{X}' goes from O to ξ_{L_1} with probability at least δ^{k/C_1} , after which $\xi_{L_1} \in \xi' = \xi'(1, k)$ if the chain \mathcal{X}' escapes to infinity. Thus on the event $E_0 \cap E_1$,

$$W_1 \wedge W_{1k} \geq \delta^{k/C_1} \varepsilon^k,$$

hence choosing the constant C_1 large enough and ε close enough to 1 yields eventually

$$|Z_1 - Z_{1,k}| \mathbb{1}_{E_0 \cap E_1} \leq e^{-c_1 k}$$

for some constant $c_1 > 0$. Next we claim that given this choice of C_1 and ε , there is an appropriate choice of the remaining parameters that ensures

$$\mathbb{Q}_\mu [E_0^c] \leq e^{-c_2 k}, \quad \mathbb{Q}_\mu [E_1^c] \leq e^{-c_2 k^\beta}$$

for some $c_2 > 0$ and $\beta \in (0, 1]$. By Remark 5.15 and Proposition 5.5 there exist $q_0, \alpha > 0$ such that $\mathcal{G}_O \notin \mathcal{A}_k(q_0, \alpha)$ has exponentially small probability, whereas Proposition 5.4 implies $q_{\text{Esc}}(\xi_{L_1}) < \varepsilon^k$ with doubly exponentially small probability. For E_1 observe $d_{\mathcal{P}}(O, \mathcal{X}_{T_1}) > k/C_1$ implies $T_1 > k/C_1$ which has stretched exponentially small probability by Lemma 5.14. Then by the moment bounds (5.63) for some $c_3 > 0$,

$$\begin{aligned} \text{Cov}_{\mathbb{Q}_\mu}(Z_1, Z_{2k}) &= \mathbb{E}_{\mathbb{Q}_\mu} [(Z_1 - \mathbb{E}_{\mathbb{Q}_\mu}[Z_1])(Z_{2k} - \mathbb{E}_{\mathbb{Q}_\mu}[Z_{2k}]) \mathbb{1}_{E_1}] + O(e^{-c_3 k^\beta}) \\ &= \mathbb{E}_{\mathbb{Q}_\mu} [(Z_1 - Z_{1,k})Z_{2k} \mathbb{1}_{E_1}] - \mathbb{E}_{\mathbb{Q}_\mu} [(Z_1 - Z_{1,k}) \mathbb{1}_{E_1}] \mathbb{E}_{\mathbb{Q}_\mu} [Z_{2k}] \\ &\quad + \mathbb{E}_{\mathbb{Q}_\mu} [(Z_{1,k} - \mathbb{E}_{\mathbb{Q}_\mu}[Z_1])(Z_{2k} - \mathbb{E}_{\mathbb{Q}_\mu}[Z_{2k}]) \mathbb{1}_{E_1}] + O(e^{-c_3 k^\beta}) \\ &= \mathbb{E}_{\mathbb{Q}_\mu} [(Z_1 - Z_{1,k})Z_{2k} \mathbb{1}_{E_0 \cap E_1}] - \mathbb{E}_{\mathbb{Q}_\mu} [(Z_1 - Z_{1,k}) \mathbb{1}_{E_0 \cap E_1}] \mathbb{E}_{\mathbb{Q}_\mu} [Z_{2k}] \\ &\quad + \mathbb{E}_{\mathbb{Q}_\mu} [(Z_{1,k} - \mathbb{E}_{\mathbb{Q}_\mu}[Z_1])(Z_{2k} - \mathbb{E}_{\mathbb{Q}_\mu}[Z_{2k}]) \mathbb{1}_{E_1}] + O(e^{-c_3 k^\beta}) \\ &= \mathbb{E}_{\mathbb{Q}_\mu} [(Z_{1,k} - \mathbb{E}_{\mathbb{Q}_\mu}[Z_1])(Z_{2k} - \mathbb{E}_{\mathbb{Q}_\mu}[Z_{2k}]) \mathbb{1}_{E_1}] + O(e^{-c_3 k^\beta}). \end{aligned}$$

Note that by construction $Z_{1,k} \mathbb{1}_{E_1}$ is measurable with respect to $\mathcal{G} \setminus \mathcal{G}(k)$ and $(\mathcal{X}_t)_{t \leq T_{k-1}}$, whereas Z_{2k} is measurable with respect to $\mathcal{G}(2k-1)$ and $(\mathcal{X}_t)_{T_{2k-1} \leq t}$ only. Thus by conditioning on $\{Y_{k-1}, Y_{2k-1}\}$ Markov's property implies

$$\begin{aligned} &\mathbb{E}_{\mathbb{Q}_\mu} [(Z_{1,k} - \mathbb{E}_{\mathbb{Q}_\mu}[Z_1])(Z_{2k} - \mathbb{E}_{\mathbb{Q}_\mu}[Z_{2k}]) \mathbb{1}_{E_1}] \\ &= \mathbb{E}_{\mathbb{Q}_\mu} [\mathbb{E}_{\mathbb{Q}_\mu} [(Z_{1,k} - \mathbb{E}_{\mathbb{Q}_\mu}[Z_1]) \mathbb{1}_{E_1} \mid Y_{k-1}] \mathbb{E}_{\mathbb{Q}_\mu} [Z_{2k} - \mathbb{E}_{\mathbb{Q}_\mu}[Z_{2k}] \mid Y_{2k-1}]] \end{aligned}$$

The right-hand side can be written as $\mathbb{E}_{\mathbb{Q}_\mu} [f(Y_{k-1})g(Y_{2k-1})]$ for some functions f, g . From the proof of Proposition 5.11, this expectation is exponentially small in k , provided f and g have finite moments. This is the direct consequence of (5.63) for g , while for f we need moment bounds on Z_{1k} . These can be established with the same arguments as above. \square

To relate the previous concentration with the weights (5.16), we define similar weights in the quasi-tree. Let τ_l denote here the first time t such that $d_{\text{LR}}(\mathcal{X}_0, \mathcal{X}_t) = l$. For all long-range edge $e \in \mathcal{G}$, write \mathcal{G}_e for the subquasi-tree at any endpoint of e (they give the same quasi-tree). Given $R, L \geq 0$, $x \in \mathcal{V}$ and a long-range edge e at long-range

distance 0 from x

$$\begin{aligned} w_{x,R,L}(e) &:= \mathbf{P}_x \left[\mathcal{X}_{\tau_L} \in \mathcal{G}_e, \tau_L < \tau_{\text{SR}}^{(R)} \right] \\ w_{R,L}(e \mid x) &:= \mathbf{P} \left[\mathcal{X}_{\tau_L} \in \mathcal{G}_e, \tau_L < \tau_{\text{SR}}^{(R)} \mid \mathcal{X}_{1/2} = x, \tau_L < \tau_{\eta(x)} \right] \end{aligned} \quad (5.64)$$

Then if $e = (e_i)_{i=1}^k$ is a long-range non-backtracking path starting from $B_{\text{LR}}(x, 0)$, set

$$w_{x,R,L}(e) := w_{x,R,L}(\xi_1) \prod_{i=2}^k w_{R,L}(\xi_i \mid \xi_{i-1}^+)$$

where the product is taken equal to 1 if empty.

In the proof of Lemma 5.19, we introduced the measure

$$\mathbf{Q}_{u,g} := \mathbb{P} \left[\cdot \mid \mathcal{X}_{1/2} = \eta(O), \iota(O) = u, \tau_O = \infty, \mathcal{G}_O = g \right].$$

where $u \in V$ and g is a possible realization of the subquasi-tree \mathcal{G}_O . We use a notation which may be reminiscent of (5.19) as these two measures are very similar, although note that here u is not the type of the starting state of the chain but its long-range neighbour. There should be no risk of confusion as the measure of (5.19) will not be used until Section 5.9. Using the same arguments as for the lemma, we can see that

$$\mathbf{P}_x [\xi_1 = e_1, \dots, \xi_k = e_k] = \mathbf{P}_x [\xi_1 = e_1] \prod_{i=2}^k \mathbf{Q}_{\iota(e_{i-1}^+), \mathcal{G}_{e_{i-1}}} [\xi_1 = e_i]. \quad (5.65)$$

The following Lemma establishes thus a bound on individual weights.

Lemma 5.20. *There exist constants $C, C_0, C_1, C_2 > 0$ such that for all $R, L > 0$ the following holds:*

(i) *with probability at least $1 - e^{-C(R \wedge L)}$, for all long-range edge e such that $d_{\text{SR}}(O, e^-) < R$,*

$$|\log w_{O,R,L}(e) - \log \mathbf{P}_O [\xi_1 = e]| \leq C_0 e^{-C_1 L + C_2 R},$$

(ii) *for all $x \in \mathcal{V}$, with probability at least $1 - e^{-C(R \wedge L)}$ conditional on $\mathcal{G} \setminus \mathcal{G}_x$, for all long-range edge e of \mathcal{G}_x such that $d_{\text{SR}}(x, e^-) < R$*

$$|\log w_{R,L}(e \mid x) - \log \mathbf{Q}_{\iota(e^+), \mathcal{G}_e} [\xi_1 = e]| \leq C_0 e^{-C_1 L + C_2 R}.$$

Proof. We prove the first bound in detail. Given $k \geq 1$, let $\mathcal{A}_k = \mathcal{A}_k(q_0, \alpha)$ be the set of quasi-trees for which every path of length k starting from the root contains a proportion at least α of vertices x such that $q_{\text{Esc}}(x) \geq q_0$. Let

$$E_0 := \{\mathcal{G}_O \in \mathcal{A}_R \cap \mathcal{A}_L\} \cap \{\forall x \in B_{\text{SR}}(O, R) : q_{\text{Esc}}(x) \geq c_0 e^{-c_1 R}\}.$$

We claim the event E_0^c occurs with exponentially small probability in $R \wedge L$ for an appropriate choice of parameters. First there exist $q_0, \alpha > 0$ such that $\mathcal{G} \notin \mathcal{A}_R \cap \mathcal{A}_L$ with

exponentially small probability by Proposition 5.5. Next, Proposition 5.4 implies there exists $c \in (0, 1)$ such that for a fixed $x \in B_{\text{SR}}(O, R)$, $q_{\text{Esc}}(x) < c\delta^{4R}$ with probability at most c^{2^R} . However there are $O(\Delta^R)$ vertices in this ball, hence the claim by union bound.

Notice then that to have one of the two events $e \in \xi$ or $\mathcal{X}_{\tau_L} \in \mathcal{G}_e$ realized exclusively, the chain needs to backtrack from level L to level 0. On E_0 , this occurs with probability exponentially small in L by Lemma 5.17. Thus on E_0

$$|w_{O,R,L}(e) - \mathbf{P}_O[\xi_1 = e]| \leq \mathbf{P}_O\left[\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} < \infty\right] \leq e^{-C'(R \wedge L)}$$

for some constant $C' > 0$. On the other hand, on E_0

$$w_{O,R,L}(e) \wedge \mathbf{P}_O[\xi_1 = e] \geq \delta^R q_0$$

as the right-hand side is a lower bound on the probability to go from O to e and then escape to infinity, which forces both $e \in \xi$ and $\mathcal{X}_{\tau_L} \in \mathcal{G}_e$. Using $|\log x - \log y| \leq |x - y|/(x \wedge y)$, we deduce that on E_0 ,

$$|\log w_{O,R,L}(e) - \log \mathbf{P}_O[\xi_1 = e]| \leq C_0 e^{-C_1 L + C_2 R}$$

for some constants $C_0, C_1, C_2 > 0$.

The second bound is proved similarly conditional on $\mathcal{G} \setminus \mathcal{G}_x$. The main difference is the additional conditioning by either $\tau_L < \tau_{\eta(x)}$ or $\tau_{\eta(x)} = \infty$. Since the differences implies backtracking, these can be controlled using the same arguments as above and Proposition 5.4. \square

Lemma 5.21. *There exist constants $C, C_1, \dots, C_3 > 0$ and $\beta \geq 1$ such that the following holds. For all $\varepsilon > 0$ there exist $C_0 = C_0(\varepsilon), C_h = C_h(\varepsilon)$ such that for all $s, t \geq 0$, for all $R, L > 0$, with probability at least $1 - \varepsilon - 2(s+t)e^{-C(R \wedge L)}$ conditional on O and $\eta(O)$,*

$$|-\log w_{\mathcal{X}_s, R, L}(\xi(\mathcal{X}_s \cdots \mathcal{X}_{s+t})) - ht| \leq C_h \sqrt{t} + C_0 t e^{-C_1 L + C_2 R} + C_3 R L^{\beta+1}. \quad (5.66)$$

Proof. Fix R, L for the rest of the proof. For notational simplicity, we drop subscripts R, L from the weights. Recall $\mathcal{G}^{(R)}$ is the quasi-tree truncated at the R -boundary of small-range components. Note that that weights are positive only for edges in $\mathcal{G}^{(R)}$.

Given $t \geq 0$, let

$$N_t := \max\{k \geq 0 \mid T_k \leq t\}.$$

We first argue there exists a constant $C_3 > 0$ such that for all $s, t \geq 0$ if $\mathcal{X}_0 \cdots \mathcal{X}_{s+t}$ is included in $\mathcal{G}^{(R)}$ then

$$\begin{aligned} |\log w_{\mathcal{X}_s}(\xi(\mathcal{X}_s \cdots \mathcal{X}_{s+t})) - \log w(\xi_{L_{N_{s+2}}} \cdots \xi_{L_{N_{s+t}}} \mid \xi_{L_{N_s+1}})| &\leq C_0 R L (T_{N_{s+2}} - T_{N_s} \\ &\quad + T_{N_{s+t+1}} - T_{N_{s+t}}). \end{aligned} \quad (5.67)$$

Note first that $\xi_{L_{N_s+1}} \cdots \xi_{L_{N_s+t}}$ is necessarily part of the path $\xi(\mathcal{X}_s \cdots \mathcal{X}_{s+t})$. Since weights are below one, we can easily lower bound

$$-\log w_{\mathcal{X}_s}(\xi(\mathcal{X}_s \cdots \mathcal{X}_{s+t})) \geq -\log w(\xi_{L_{N_s+2}} \cdots \xi_{L_{N_s+t}} \mid \xi_{L_{N_s+1}}).$$

On the other hand the path $\xi(\mathcal{X}_s \cdots \mathcal{X}_{s+t})$ contains at most $T_{N_s+2} - s \leq T_{N_s+2} - T_{N_s}$ edges until it reaches $\xi_{L_{N_s+2}}$ and similarly it contains at most $T_{N_s+t+1} - T_{N_s+t}$ after it leaves $\xi_{L_{N_s+t}}$. Then we bound the weights of these edges. If $x \in \mathcal{V}$ and e is a long-range edge such that $d_{\text{SR}}(x, e^-) < R$, δ^R lower bounds the probability the chain goes from x to e and leaves by this edge. As small-range components have size at least 2 (H3), δ^{RL} lower bounds the probability that the chain goes from x to e^+ and reaches long-range distance L before coming back to e^- . Consequently, for some constant $C_3 > 0$, we can always bound $-\log w_x(e) \leq C_3 RL$ for all x, e in $\mathcal{G}^{(R)}$, for any realization of the quasi-tree. We deduce that

$$\begin{aligned} -\log w_{\mathcal{X}_s}(\xi(\mathcal{X}_s \cdots \mathcal{X}_{s+t})) + \log w_{\mathcal{X}_s}(\xi_{L_{N_s+2}} \cdots \xi_{L_{N_s+t}}) &\leq C_3 RL (T_{N_s+2} - T_{N_s} \\ &\quad + T_{N_s+t+1} - T_{N_s+t}). \end{aligned}$$

Next fix $\varepsilon \in (0, 1)$ and let $s, t \geq 0$. Let \mathcal{A} be the set of quasi-trees for which every long-range edge e such that $d_{\text{SR}}(O, e^-) < R$ satisfies the bounds of Lemma 5.20 for some $C_0, C_1, C_2 > 0$. Consider the following events:

$$\begin{aligned} E_0 &:= \{\mathcal{G} \in \mathcal{A}\} \cap \bigcap_{s' \leq s+t} \{\mathcal{G}_{\mathcal{X}_{s'}} \in \mathcal{A}\}, \\ E_1 &:= \{\tau_{\text{SR}}^{(R)} > s+t\} \\ E_2 &:= \{|L_{N_s+t} - L_{N_s} - dt| \leq C_{\text{LR}} \sqrt{t}\} \end{aligned}$$

with $C_{\text{LR}} = C_{\text{LR}}(\varepsilon) > 0$ to determine, and write $E := E_0 \cap E_1 \cap E_2$. By Lemmas 5.20 and 5.16, C_0, C_1, C_2 can be chosen so that $\mathbb{P}_O[E_0^c \mid O, \eta(O)] \leq (s+t)e^{-C_4(R \wedge L)}$ for some $C_4 > 0$. By Lemma 5.18 there exists $C_5 > 0$ such that $\mathbb{P}_O[E_1^c \mid O, \eta(O)] \leq (s+t)e^{-C_5 R}$. Finally from the proof of Proposition 5.13, $C_{\text{LR}}(\varepsilon)$ can be taken so that $\mathbb{P}_O[E_2^c \mid O, \eta(O)] \leq \varepsilon$. Hence $\mathbb{P}_O[E^c \mid O, \eta(O)] \leq \varepsilon + (s+t)e^{-C(R \wedge L)}$, for some constant $C > 0$.

Then let $l_+ := \lceil dt + C_{\text{LR}} \sqrt{t} \rceil$. On the event E , using (5.65)

$$\begin{aligned} \left| \log w(\xi_{L_{N_s+2}} \cdots \xi_{L_{N_s+t}} \mid \xi_{L_{N_s+1}}) - \log \mathbf{P} \left[\xi_{L_{N_s+2}} \cdots \xi_{L_{N_s+t}} \in \xi' \mid \mathcal{X}, \xi'_{L_{N_s+1}} = \xi_{L_{N_s+1}} \right] \right| \\ \leq l_+ e^{-C_1 L + C_2 R}. \end{aligned}$$

Letting h' be the constant of Lemma 5.19, we claim then that for $h = h'/\mathbb{E}_{\mathbb{Q}_\mu}[T_1]$ and some constant $C_h = C_h(a, \varepsilon)$

$$\left| -\log \mathbf{P} \left[\xi_{L_{N_s+2}} \cdots \xi_{L_{N_s+t}} \in \xi' \mid \mathcal{X}, \xi'_{L_{N_s+1}} = \xi_{L_{N_s+1}} \right] - ht \right| \leq C_h \sqrt{t}$$

with probability at least ε . This can be done using the same arguments as in the proof of Proposition 5.13, combining the fluctuations for N_{s+t} (5.60) established in the said proof with Lemma 5.19.

Finally, we are left with bounding the right hand side of (5.67). Using again the fluctuations for N_{s+t} (5.60) with the stretched exponential tail of regeneration times, there exists $\beta \geq 1$ and $C_6 > 0$ such that

$$\mathbb{P}[T_{N_{s+1}} - T_{N_s} + T_{N_{s+t+1}} - T_{N_{s+t}} > L^\beta] \leq \varepsilon + \sqrt{s+t}e^{-C_6L}.$$

All in all, up to changing the value of C we have thus proved that with probability at least $1 - 2\varepsilon - (s+t)e^{-C(R \wedge L)}$,

$$|\log w_{\mathcal{X}_s}(\xi(\mathcal{X}_s \cdots \mathcal{X}_{s+t})) - ht| \leq C_h \sqrt{t} + C_0 l_+ e^{-C_1 L + C_2 R} + C_3 R L^{\beta+1}.$$

Since $l_+ = O_\varepsilon(t)$ this proves the result. \square

5.8 First steps towards nice paths

We now come back to the finite setting. The goal of this section is to prove Lemmas 5.2, 5.3 and Proposition 5.1, which will be used in the next section to justify the chain is likely to follow a nice trajectory.

5.8.1 The case of typical starting states

We start with the case of typical starting states, for which there is little left to prove.

Proof of (i) in Lemma 5.3. This is essentially a matter of checking definitions. For fixed $x \in V$, the choice of $s, t = O(\log n)^a, a \geq 1$, Lemma 5.1 ensures the finite chain X and the chain \mathcal{X} on the quasi-tree can be coupled until time $s+t$ with annealed probability $1 - o(n^{-\varepsilon})$ for some $\varepsilon > 0$. Provided the coupling succeeds, the property of backtracking or hitting the boundary of a small-range ball coincides for both chains, while the regeneration times of \mathcal{X} obviously give regeneration edges for the finite chain. Thus from Lemma 5.1 and Lemma 5.18 we deduce

$$\mathbb{P}_x[(X_s \cdots X_{s+t} \cdots \in \Gamma(R, L, M))] \leq (s+t)e^{-C(R \wedge L \wedge M^\alpha)} + o(n^{-\varepsilon}) \quad (5.68)$$

Since $s, t = (\log n)^a$, for any $b > 0$ the choice of $R, L = O(\log \log n)$ with large enough implicit constants depending on a, b and $M = (\log \log n)^{2/\alpha}$ makes the right hand side $o((\log n)^{-b})$. The first moment argument (5.4) implies (i) in Lemma 5.3. \square

We can repeat the same kind of arguments for point (i) in Proposition 5.1 however we obtain for the time being a slightly weaker statement than the one expected. Because

of the dependance in ε in the annealed probability of (5.59) and (5.66), applying the first moment arguments will not establish the concentration property with high probability but only $1 - \varepsilon - o(1)$. To obtain a high probability result we need to apply a higher order Markov's inequality, which is also the key argument to obtain the uniform statements.

5.8.2 Bootstrapping annealed bounds with parallel chains

Arguments used so far above establish errors on annealed error in $o(1)$, nay in $\varepsilon + o(1)$. Following ideas from [28, 29], one strategy to improve these error bounds consists in using higher order moments in Markov's inequality, which leads to an argument of "parallelizing chains" on the same environment. Namely for all $\varepsilon > 0$ and $k \geq 1$, for any trajectorial event A

$$\forall x \in V : \mathbb{P}[\mathbf{P}_x[A] > \varepsilon] \leq \frac{\mathbb{P}_x[A]^k}{\varepsilon^k} = \varepsilon^{-k} \mathbb{P}_x \left[\bigcap_{i=1}^k \{X^{(i)} \in A\} \right] \quad (5.69)$$

and

$$\begin{aligned} \mathbb{P} \left[\max_{x \in V} \mathbf{P}_x[A] > \varepsilon \right] &\leq \frac{1}{\varepsilon^k} \mathbb{E} \left[\sum_{x \in V} \mathbf{P}_x[A]^k \right] \\ &\leq \frac{n}{\varepsilon^k} \max_{x \in V} \mathbb{P}_x \left[\bigcap_{i=1}^k \{X^{(i)} \in A\} \right], \end{aligned} \quad (5.70)$$

where $X^{(1)}, \dots, X^{(k)}$ are k versions of the chain X generated independently conditional on the same environment. These k trajectories can be generated sequentially with the environment, sampling the environment only when exploring parts of the environment not already visited by the previous chains. The idea will then be to argue that the chains are unlikely to follow each other, so they end up exploring disjoint parts of the environments. If the event A depends only on what happens after the trajectories separate from each other (hence the consideration of shifted trajectories), we can then expect the annealed probability to decouple and become $\mathbb{P}[A]^k$. If the event A has probability below $f(\varepsilon)$ for some function f , this can be sufficient to beat the ε^{-k} or $\varepsilon^{-k}n$ factor.

As an example of application, we can start proving point (i) in Proposition 5.1, which also illustrates some of the arguments used in the sequel.

Proof of (i) in Proposition 5.1. Let $\varepsilon > 0$, $t = O(\log n)$ and $x \in V$. For $C_{\text{LR}} = C_{\text{LR}}(\varepsilon)$ and $C_h = C_h(\varepsilon) > 0$ to determine, consider the set of paths

$$\Gamma_{\text{Ent}} := \left\{ \mathbf{p} \left| \begin{array}{l} \|\xi(\mathbf{p}) - d|\mathbf{p}|\| \leq C_{\text{LR}}\sqrt{|\mathbf{p}|} \\ |-\log w_{\mathbf{p}_0, R, L}(\xi(\mathbf{p})) - h|\mathbf{p}|\| \leq C_h\sqrt{|\mathbf{p}|} \end{array} \right. \right\} \quad (5.71)$$

which can be paths either in G^* or \mathcal{G} . The goal is thus to show $\mathbb{P}_x [(X_0)\cdots X_t \notin \Gamma_{\text{Ent}}] = o(1)$. Consider k versions $X^{(1)}, \dots, X^{(k)}$ of the chain X generated independently conditional on the same environment, with $k \gg 1, k = O(\log \log n)$. Then using the same coupling as in Section 5.2.7 these can be coupled with k versions $\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(k)}$ of the chain \mathcal{X} . For $j \in [k]$ let $\mathcal{G}(j, t)$ denote the union of the first j trajectories up to time t , along with their long-range neighbourhoods up to depth L . Compared to the case of a single chain, the number of visited vertices gets multiplied by k at most, so the proof of Lemma 5.1 shows the chains $X^{(1)}, \dots, X^{(k)}$ can be coupled with high probability, along with their long-range neighbourhood up to length $L = O(\log \log n)$, up to time $t = O(\log n)$, to coincide with $\mathcal{G}(k, s+t)$. Now the definitions of the long-range distance and of weights ((5.16) and (5.64)), which are measurable with respect to this neighbourhood, were exactly made to coincide in the finite and quasi-tree settings. Thus if the coupling succeeds $(X_0 \cdots X_t) \in \Gamma_{\text{Ent}}$ if and only in $\mathcal{X}_0 \cdots \mathcal{X}_t \in \Gamma_{\text{Ent}}$. Letting for $j \in [k]$:

$$B_j := \bigcap_{i=1}^j \left\{ \mathcal{X}_0^{(i)} \cdots \mathcal{X}_t^{(i)} \notin \Gamma_{\text{Ent}} \right\}$$

(5.69) and the coupling imply

$$\mathbb{P}_x [X_0 \cdots X_t \notin \Gamma_{\text{Ent}}] \leq \varepsilon^{-k} \mathbb{P} [B_k \mid \iota(O) = x] + o(1),$$

so it suffices to show $\mathbb{P}_O [B_k \mid \iota(O) = x] = o(\varepsilon^k)$. We consider in fact a slightly modified event. Let $l := (\log \log n)^3$ and consider the set of quasi-trees $\mathcal{A}_l = \mathcal{A}_l(\alpha, q_0)$ for which every length l path from the root contains a proportion at least α of vertices for which the probability to escape to infinity outside \mathfrak{p} is lower bounded by q_0 . Recalling Proposition 5.5 and Remark 5.10, Lemma 5.16 implies the existence of constants $q_0, \alpha > 0$ and $C > 0$ such that all vertices in $\mathcal{G}(j, t)$ have their subquasi-tree in \mathcal{A}_l with probability at least $1 - te^{-Cl}$. Since $t = O(\log n)$ and $l = (\log \log n)^3$ this probability is $1 - o(1)$. Thus we can suppose that the event B_j also contains the event that all vertices of $\mathcal{G}(j, t)$ have their subquasi-tree in \mathcal{A}_l . Secondly, consider the l -th regeneration time $T_l^{(i)}$ of the i -th chain. By the stretched exponential tails of regeneration times 5.14 and Chernoff's bound, there exist constants $\alpha \in (0, 1]$ and $C > 0$ such that for all $i \leq k$

$$\mathbb{P}_O \left[T_l^{(i)} > t_0 \mid \iota(O) = x \right] \leq e^{-t_0^\alpha + Cl}$$

Thus considering $t_0 := (\log \log n)^\beta$, for $\beta > 3/\alpha$ large enough makes the right hand side $o(k^{-1})$. Consequently by union bound we obtain that $\mathbb{P}_O \left[\exists i \leq k : T_l^{(i)} > t_0 \right] = o(1)$, so we can also incorporate to the events B_j the event that $T_l^{(i)} \leq t_0$ for $i \leq j$.

Next we claim that $\mathbb{P}_O [B_1 \mid \iota(O) = x] \leq \varepsilon/2 + o(1)$ and $\mathbb{P}_O [B_j \mid \iota(O) = x, B_{j-1}] \leq \varepsilon/2 + o(1)$ for all $j \in [2, k]$. Noting that

$$\mathbb{P}_O [B_k \mid \iota(O) = x] = \mathbb{P}_O [B_1 \mid \iota(O) = x] \prod_{j=2}^k \mathbb{P}_O [B_j \mid \iota(O) = x, B_{j-1}],$$

the claims imply that $\mathbb{P}[B_k \mid \iota(O) = x] = (\varepsilon/2 + o(1))^k = o(\varepsilon^k)$ as desired.

Let us prove the claims. The first claim is the consequence of Proposition 5.13 and Lemma 5.21, provided the constants C_{LR}, C_h are chosen adequately. To prove the second claim, let $j \geq 2$ and suppose B_{j-1} holds. We argue that $\mathcal{X}^{(j)}$ has quenched probability at least $1 - o(1)$ to exit $\mathcal{G}(j-1, t)$ for the last time by time t_0 . Consider then the loop-erased chain $(\xi_m^{(j)})_m$ obtained from $\mathcal{X}^{(j)}$. Thanks to the first event we added to B_{j-1} , for any long-range path ζ of length l contained in $\mathcal{G}(j-1, t)$,

$$\mathbf{P}_O \left[\xi_l^{(j)} = \zeta_l \mid B_{j-1} \right] \leq e^{-Cl}$$

for some constant $C > 0$. Since $\mathcal{G}(j-1, t)$ is made of the long-range neighbourhood of the first $j-1$ trajectories, it contains at most $kt\Delta^{R(L+1)}$ long-range paths of length l . Thus by union bound

$$\mathbf{P}_O \left[\xi_l^{(j)} \in \mathcal{G}(j-1, s+t) \mid B_{j-1} \right] \leq kt\Delta^{R(L+1)} e^{-Cl} = e^{-C(\log \log n)^3 + O(\log \log n)^2} = o(1).$$

On the other hand, the second event added to B_j implies that $T_l^{(j)} \leq t_0$ so the chain must in particular have made the first l steps of its loop-erased trace before time t_0 . Combined with what precedes we deduce that conditional on B_{j-1} , the chain $\mathcal{X}^{(j)}$ leaves the graph $\mathcal{G}(j-1, t)$ before time t_0 with probability $1 - o(1)$.

Let $L^{(j)}$ denote now this last exit time. Conditional on $L^{(j)}$ and $\mathcal{X}_{L^{(j)}} = u$, the quasi-tree that contains the subsequent trajectory needs then to be generated according to the measure \mathbb{Q}_u (5.52). Using Remark 5.15 upper bounds under this law only differ by a constant factor from upper bounds with the usual law. We can thus apply Proposition 5.13 and Lemma 5.21 conditional on L_j to obtain that

$$\mathbb{P}_O \left[(X_{L_j}^{(j)} \dots X_{L_j+t}^{(j)}) \notin \Gamma_{\text{Ent}} \mid B_{j-1}, \iota(O) = x \right] \leq \varepsilon/2 + o(1).$$

Now by what precedes $d_{\text{LR}}(\mathcal{X}_0^{(j)}, \mathcal{X}_{L_j}^{(j)}) \leq l$ with probability $1 - o(1)$ conditional on B_{j-1} , hence

$$d_{\text{LR}}(\mathcal{X}_0^{(j)}, \mathcal{X}_t^{(j)}) \in [dt - C_{\text{LR}}\sqrt{t}, dt - C_{\text{LR}}\sqrt{t} + l]$$

and

$$-\log w_{\mathcal{X}_0, R, L}(\xi(\mathcal{X}_0^{(j)} \dots \mathcal{X}_t^{(j)})) \in [ht - C_h\sqrt{t}, ht + C_h\sqrt{t} + lw_{\max}],$$

where w_{\max} denotes an upper bound on all the weights contained in $\mathcal{G}(j-1, t)$. For any realization of the quasi-tree we can bound $w_{\max} \leq O(RL)$ as in the proof of Lemma 5.21. All in all, since $lw_{\max} = o(\sqrt{t})$, this ensures $(\mathcal{X}_0^{(j)}, \mathcal{X}_t^{(j)}) \in \Gamma_{\text{Ent}}$ with probability $1 - \varepsilon/2 - o(1)$, proving the claim. \square

5.8.3 From $o(1)$ to $o(1/n)$ bounds

Quasi-tree with a cycle

The rest of this section will now be devoted to the proof of the uniform, worst-case statements. The main ideas are the same as those used above. From (5.70), we only need to push them yet a little further to make error bounds on annealed probabilities $o(1/n)$. From the proof of (i) in Proposition 5.1 two arguments require a modification: the coupling with the a quasi-tree and the use of escape probabilities.

Let us focus on the first point. It seems quite obvious that the coupling with quasi-trees cannot be improved beyond to a $o(1/n)$ error as some vertices do not have a quasi-tree like neighbourhood. However from the proof of Lemma 5.1 this error can be achieved allowing just one cycle. We are thus led to consider a modified quasi-tree model which contains at most one long-range cycle.

Let $(u_i)_{i=0}^l, (v_i)_{i=0}^l, l \geq 0$ be two sequences of vertices in V such that $u_{i+1} \in B_{\text{SR}}^+(v_i, \infty)$ for all $i \in [0, l-1]$ and $B_{\text{SR}}^+(v_l, R) \cap B_{\text{SR}}^-(u_0, R) \neq \emptyset$. For every i and vertex $z \in B_{\text{SR}}^+(u_i, \infty) \setminus \{u_i, v_i\}$, add an edge $(z, \eta(z))$ and grow a one-sided quasi-tree rooted at $\eta(z)$. We call the oriented graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ obtained a quasi-tree with a cycle $((u_i, v_i))_{i=0}^l$. As a usual quasi-tree, it is given by maps ι, η such that ι identifies vertices of \mathcal{V} with vertices in V , while $\eta : \mathcal{V} \rightarrow \mathcal{V}$ is an involution. Here η is obtained by the corresponding maps in the quasi-trees outside the cycle whereas for the cycle we set $\eta(u_i) := v_{i+1}$ for all $i = 0, \dots, l-1$. A root O can be chosen, which does not to be on the cycle. The definition of the Markov chain \mathcal{X} (5.10) extends directly to this setting.

The coupling of Section 5.2.7 gives a natural way to couple \mathcal{X} on a random realization of \mathcal{G} with the Markov chain X : the rejection scheme is used until the first occurrence of a cycle, after which cycles are ignored in the construction of \mathcal{G} . The stochastic comparison in the proof of Lemma 5.1 still holds, but using now that

$$\mathbb{P}[Z \geq 2] \leq \frac{m^4 \Delta^{2R}}{n^2} \quad (5.72)$$

if Z is a binomial $\text{Bin}(m, m\Delta^R/n)$, we deduce that for $t = n^{o(1)}$, the chains X and \mathcal{X} can be coupled so that the chains coincide and $B_{\text{LR}}(X_s, L)$ and $B_{\text{LR}}(\mathcal{X}_s, L)$ are isomorphic for all $s \leq t$ with probability $1 - o(1/n)$.

More generally, let $k \geq 1$ and $X^{(1)}, \dots, X^{(k)}$, resp. $\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(k)}$, k versions of the chain X , resp. \mathcal{X} generated independently conditional on the same environment. Letting $G^*(k, t), \mathcal{G}(k, t)$ be the environments explored by these trajectories up to time t along with their L -long-range neighbourhoods, the same arguments as above implies

that

$$\mathbb{P} \left[\begin{array}{l} G^*(k, t) \text{ and } \mathcal{G}(k, t) \text{ are isomorphic} \\ \forall i \in [1, k], \forall s \leq t: X_s^{(i)} = \iota(\mathcal{X}_s^{(i)}) \end{array} \right] = 1 - o((kt)^4/n^{7/4}) = o(\varepsilon^k/n) \quad (5.73)$$

if for instance $k = \lfloor \log n / 2 \log \varepsilon^{-1} \rfloor$.

An essentially uniform bound on escape probabilities

The second argument that needs to be adapted is the use of escape probabilities, which can be expected to of constant order only for typical states. Consider \mathcal{G} to be a random realization of a quasi-tree with a cycle as described above and \mathcal{X} the associated Markov chain. Conditional on the cycle, the quasi-trees that are added to it are generated according to the model studied in Sections 5.5 - 5.7, therefore the asymptotic analysis of the chain \mathcal{X} directly extends to that case, conditional on the cycle. The chain ultimately leaves the cycle, after which it stays in a genuine quasi-tree. Given a vertex x which is not on the cycle, let $q_{\text{Esc}}(x)$ denote the probability to escape in the corresponding quasi-tree. If x is on the cycle, there is no quasi-tree at x , so let $q_{\text{Esc}}(x)$ denote the probability of escaping through one of the vertices that are in the same component as x . Using that components of V_1 have size at least 3 (H3), these escape probabilities can be bounded exactly as in Section 5.5, up to a change of constants, conditional on the cycle. Thus from Proposition 5.4, there exists $\beta \geq 1$ such that for all $u \in \mathcal{V}$ not on the cycle,

$$\mathbb{P} \left[q_{\text{Esc}}(u) < \frac{1}{(\log n)^\beta} \mid \mathcal{G} \setminus \mathcal{G}_u \right] = O(1/n^2). \quad (5.74)$$

Combined with Lemma 5.16, we see that on a time scale $t = O((\log n)^a)$, $a \geq 1$, the probability that the chain \mathcal{X} meets a vertex which does not satisfy the lower bound on escape probability above is still $o(1/n)$. This makes the chain sufficiently unlikely to meet such vertices that the chain X cannot reach such quasi-tree-like neighbourhoods either, thanks to the the coupling described above. Informally, this allows us to proceed as if escape probabilities were everywhere lower bounded by $1/(\log n)^\beta$ and reason conditional on the environment. In the reversible model, Proposition 5.3 shows we can directly work with a uniformly lower bounded escape probability.

5.8.4 Nice paths from arbitrary states

A first application of these uniform lower bounds is the following.

Proof of Lemma 5.2. We prove the first point. Let $L = C_L \log \log n$ for a fixed constant $C_L > 0$, $t \ll n^{1/16}$ and $x \in V$. By Remark 5.4 it suffices to consider the case of

$s = C(\log n)^a$ for some $a > 0$ to determine. Using the coupling described above (with only one chain here, ie $k = 1$) we can couple the trajectory of X and its whole L long-range neighbourhood with that of \mathcal{X} . By (5.73) the coupling succeeds up to time $s + t$ with probability $1 - o(1/n)$, hence for $u \in V$

$$\begin{aligned} & \mathbb{P}_u [\exists s' \in [s, s + t] : B_{\text{LR}}(X_{s'}, L) \text{ is not quasi-tree-like}] \\ &= \mathbb{P}_O [\exists s' \in [s, s + t] : \mathcal{X}_{s'} \text{ is at distance less than } L \text{ from the cycle} \mid \iota(O) = u] + o(1/n). \end{aligned}$$

Now let $\beta \geq 1$ be as in (5.74) and consider the set

$$\text{Esc} := \{x \in \mathcal{G} \mid q_{\text{Esc}}(x) \geq (\log n)^{-\beta}\}.$$

Write τ_{Esc^c} for the hitting time of the complement set of Esc . The number of vertices in the L long-range neighbourhood of the trajectory being bounded by $O(s + t)\Delta^{R(L+1)} = O(n^{1/16})$, union bound and (5.74) imply

$$\mathbb{P}_O [\tau_{\text{Esc}^c} \leq s + t \mid \iota(O) = u] = o(1/n).$$

We are being a bit sketchy here, but the argument can be formalized using the Ulam labelling. On the other hand, when the chain is at a vertex of Esc its long-range distance to the cycle has a quenched probability at least $(\log n)^{-\beta}$ to increase by 1 indefinitely. Thus if we consider the sequence of times $(S_k)_{k \geq 1}$ at which this happens, for all $x \in \mathcal{V}, s \geq 0$,

$$\mathbf{P}_x [\tau_{\text{Esc}^c} \geq s + t, S_1 > s] \leq e^{-s/(\log n)^\beta}.$$

Consequently by union bound and Strong Markov's property

$$\begin{aligned} \mathbf{P}_x [\exists s' \in [s, s + t] : \mathcal{X}_{s'} \text{ is at distance less than } L \text{ from the cycle, } \tau_{\text{Esc}^c} > s + t] \\ \leq \mathbf{P}_x [S_L > s, \tau_{\text{Esc}^c} > s + t] \leq L e^{-s/L(\log n)^\beta}. \end{aligned}$$

Since $s = (\log n)^a$, choosing $a > \beta + 1$ makes these two bounds $o(1/n)$. The first moment argument (5.70) concludes the proof of the first point in the general case.

For the reversible model, escape probabilities are $\Theta(1)$ conditional on the environment by Proposition 5.3. Hence S_L is always stochastically dominated by the sum of L geometric random variables with constant parameter. From Chernoff's bound we deduce that for some constant $C > 0$, $S_L > CL$ with exponentially small probability, thus we can take $s = \Theta(L) = \Theta(\log \log n)$.

Finally point (ii) can be proved with similar arguments. □

It remains to prove (ii) in Lemma 5.3 and Proposition 5.1. The argument is similar and uses the same ideas as above.

Proof of (ii) in Lemma 5.3 and Proposition 5.1. Consider the concentration of the drift and entropy. Let $x \in V$, $\varepsilon > 0$ and $k := \lfloor \log n / 2 \log \varepsilon^{-1} \rfloor$. As for point (i) we combine the higher Markov inequalities (5.70) and the coupling with k chains on a quasi-tree with a cycle. As a cycle is allowed the error due to the coupling is $o(1/n)$ by (5.73) so we need only focus on the infinite setting: letting

$$B_j := \bigcap_{i=1}^j \{ \mathcal{X}_s^{(i)} \dots \mathcal{X}_{s+t}^{(i)} \notin \Gamma_{\text{Ent}} \}.$$

for $j \in [k]$, with Γ_{Ent} as in (5.71) it suffices to prove $\mathbb{P}_O [B_k \mid \iota(O) = x] = o(\varepsilon^k/n)$. Then we claim that $\mathbb{P}_O [B_1 \mid \iota(O) = x] \leq \varepsilon^3/2 + o(1)$ and $\mathbb{P}_O [B_j \mid \iota(O) = u, B_{j-1}] \leq \varepsilon^3/2 + o(1)$ for all $j \in [2, k]$. With our choice of k this will imply $\mathbb{P}_O [B_k \mid \iota(O) = x] = (\varepsilon^3/2 + o(1))^k = o(\varepsilon^k/n)$ as needed.

As we mentioned already, results proved for genuine quasi-trees extend to the case where an additional cycle is present. In particular, the first regeneration time and level, defined to occur outside the cycle, have stretched exponential or exponential tail respectively, conditional on the cycle, while the remaining regenerations occurs on genuine quasi-trees. This is sufficient to establish that Proposition 5.13 and Lemma 5.21 still hold in this case, implying that $\mathbb{P}[B_1 \mid \iota(O) = u] \leq \varepsilon^3/2$ for a good choice of $C_h, C_{\text{LR}}(\varepsilon)$.

For the second claim, we argue as for (i) by showing the chains are unlikely follow each other. Let $\beta \geq 1$ be as in (5.74), consider the set

$$\text{Esc} := \{x \in \mathcal{G} \mid q_{\text{Esc}}(x) \geq (\log n)^{-\beta}\}.$$

and let τ_{Esc^c} be the first exit time by any of the chains. As we argued above in the proof of Lemma 5.2, $\mathbb{P}_O [\tau_{\text{Esc}^c} \leq s+t \mid \iota(O) = u] = o(1/n)$. On the other hand, while on this set the chains have a quenched probability at least $(\log n)^\beta$ to have their long-range distance increased by 1 permanently, hence for any $j \in [k]$, letting $(S_k^{(j)})_k$ be the sequence of times at which this occurs for the j -th chain,

$$\mathbf{P}_O \left[S_l^{(j)} > s, \tau_{\text{Esc}^c} > s+t \right] \leq l e^{-C_1 s / l (\log n)^\beta} \quad (5.75)$$

for some constant $C_1 > 0$. Taking $a > 1 + 2\beta$ in $s = (\log n)^s$ and $l = (\log n)^\beta (\log \log n)^3$ the right hand side is $o(1)$.

Consider now the loop-erased chain $(\xi_m^{(j)})_m$ obtained from $\mathcal{X}^{(j)}$. If all vertices visited by the chain are in Esc , and using that components of V_1 have size at least 3, the probability to follow a given path is exponentially small in the length. Since $\mathcal{G}(j-1, s+t)$ contains at most one long-range cycle this subgraph contains at most $2k(s+t)\Delta^{R(L+1)}$ long-range paths, so by union bound

$$\mathbf{P}_O \left[\xi_l^{(j)} \in \mathcal{G}(j-1, s+t), \tau_{\text{Esc}^c} > s+t \right] \leq 2k(s+t)\Delta^{R(L+1)} e^{-C_2 l / (\log n)^\beta}.$$

for some constant $C_2 > 0$. For $l = (\log n)^\beta (\log \log n)^3$ this is $o(1)$.

Combining the two previous observations, we deduce that conditional on B_{j-1} , the last exit of $\mathcal{G}(j-1, s+t)$ by the trajectory $(\mathcal{X}_{t'}^{(j)})_{t' \leq s+t}$ occurs before s , leaving at least t subsequent steps. Letting $L^{(j)}$ denote this last exit time, we can reason conditional on L_j and use Proposition 5.13 and Lemma 5.21 about shifted trajectories to deduce the claim.

The proof for the reversible case is similar, with the difference that escape probabilities are $\Theta(1)$ so we can deduce it $O(\log \log n)$ steps are enough for the k trajectories to split apart. Finally, the proof of Lemma 5.3 uses the same arguments, considering the set of paths $\Gamma(R, L, M)$ instead of Γ_{Ent} . \square

5.9 Approximation by nice paths: proof of Proposition 5.2

We now move to the second part of the proof of Theorem 5.1: the approximation of the Markov kernel by restricting to the consideration of nice paths only. For earlier works that used this argument, see for instance [19, 28].

5.9.1 Forward neighbourhood

Nice paths between x and y will have their first steps and last steps contained respectively in some quasi-tree-like neighbourhoods of x and y . We define here the forward neighbourhood of x .

Let $x \in V$, $l \geq 1$ an integer and $w_{\min} \geq 0$. From Lemma 5.1, long-range balls of radius $O(\log \log n)$ are likely to be tree-like but this is not true as the radius grow towards $\log n$. The forward graph $K(x, l, w_{\min})$ is thus designed essentially as a "spanning quasi-tree" of the ball $B_{\text{LR}}(x, l)$, obtained by exploring this ball algorithmically, pretty much as explained in Section 5.2.7, but giving priority to paths with large weights and truncating whenever cycles are encountered. This process will thus build iteratively a sequence $(K_m)_{m=0}^\tau$ of subsets of $B_{\text{LR}}(x, l)$, until it stops at a random time τ to yield $K(x, l, w_{\min}) := K_\tau$. Unless the procedure is initiated at a vertex x whose ball $B_{\text{LR}}(x, L)$ is not quasi-tree-like, K_m remains at all time quasi-tree-like. In this case for every long-range edge $e \in E_m$ there exists a unique long-range path $\xi(e)$ from x to e contained in K_m . If e has its L long range neighbourhood entirely contained in K_m , define the cumulative weight as $\hat{w}(e) := w_x(\xi(e))$, which can be computed from K_m . Because weights require the knowledge of $(L-1)$ -neighbourhoods, the exploration queue will consist in subsets E_m of long-range edges for which the whole long-range $(L-1)$ -neighbourhood is contained in K_m . Finally, a constraint of minimal weights is added during the procedure, in order to keep the number of vertices explored as $o(n)$.

Exploration of the forward neighbourhood The procedure is initiated with $K_0 := B_{\text{LR}}(x, L)$. If K_0 contains a long-range cycle, $E_0 := \emptyset$ and the procedure stops. Otherwise let E_0 be the set of long-range edges at distance 0 from x . Then for all $m \geq 0$ the $(m + 1)$ -th step goes as follows:

1. Among all long-range edges e in E_m at long-range depth at most l from x in K_m and such that $\hat{w}(e) \geq w_{\min}$, pick the edge e_{m+1} with maximal cumulative weight, using an arbitrary ordering of the vertices to break ties. If there is no such edge, the procedure stops.
2. Explore the depth- L neighbourhood of e_{m+1} : for each descendant $z \in \partial K_m$ at long-range distance $L - 1$ from e_{m+1} , sample $\eta(z)$. This exploration phase stops if a revealed edge violates the quasi-tree structure: this occurs if for some z the small-range ball $B_{\text{SR}}(\eta(z), R)$ has a non-empty intersection with K_m or one of the previously revealed balls.
3. If the previous exploration phase stopped because of intersecting small-range balls, then $E_{m+1} := E_m \setminus \{e_{m+1}\}$ and $K_{m+1} := K_m$. If it stopped because of an intersection with K_m , let Z_m be this intersection. Then E_{m+1} is obtained by deleting from E_m every long-range edge which has either a descendant or an ancestor in Z_m , as well as the edge e_{m+1} , and set $K_{m+1} := K_m$. Finally, if the exploration ended without a violation of the quasi-tree structure, add the subsequent long-range edges of e_{m+1} to E_{m+1} , whereas the newly revealed vertices are added to K_{m+1} .

When the procedure ends, the set E_τ consists by construction of edges at long-range distance $l - L$ from x , which contain no long-range cycles in their $(L - 1)$ -neighbourhood and whose weights are measurable with respect to $K(x, l, w_{\min})$.

Lemma 5.22. *Let κ_m denote the number of long-range edges revealed during the first m steps m and $\kappa(x, l, w_{\min}) := \kappa_\tau$ the total number of long-range edges revealed during the construction of $K(x, l, w_{\min})$. Suppose $B_{\text{LR}}(x, L)$ is quasi-tree-like so that $\tau \geq 1$. There exists a constant $C > 0$ such that for all $m \in [1, \tau]$*

$$\hat{w}(e_m) \leq C \frac{l \Delta^{RL}}{\kappa_m}. \quad (5.76)$$

In particular

$$\kappa(x, l, w_{\min}) \leq C \frac{l \Delta^{RL}}{w_{\min}}. \quad (5.77)$$

Proof. The set $E_{\leq m} := \bigcup_{k \leq m} E_k$ is the set of all long-range edges contained in K_m which have their L long-range neighbourhood contained in K_m . Furthermore, at time m the

procedure did not explore beyond long-range distance $(L - 1)$ from these edges. Since long-range balls of radius L contain $O(\Delta^{R(L+1)})$ vertices, one has $\kappa_m \leq C|E_{\leq m}|\Delta^{RL}$ for some constant $C > 0$.

The quasi-tree structure implies the set E can be arranged as a rooted tree, where long-range edges are linked to a same vertex if they are at long-range distance 0 from each other. The set E_m is the set of edges furthest from the root, that is edges that lead to leaves. For every $e \in E_m$, there is a unique shortest path from the root to e , and conversely every edge is on such a path. These paths correspond to long-range paths in K_m from x to e , which have long-range length at most l , so we can deduce $|E_{\leq m}| \leq l|E_m|$.

On the other hand, the sum of weights over all shortest long-range paths from the root to E_m is bounded by 1. Furthermore, the choice of a maximal weight in step 1 ensures that the weights consecutively chosen are non-increasing. Therefore every shortest path from the root to E_m has weight at least $\hat{w}(e_m)$, so we deduce $|E_m|\hat{w}(e_m) \leq 1$. All in all this shows $\kappa_m \leq Cl\Delta^{RL}/\hat{w}(e_m)$. Finally, $\hat{w}(e_m) \geq w_{\min}$ for all m , in particular when the procedure stops at $m = \tau$, hence the bound on $\kappa_\tau = \kappa(x, l, w_{\min})$. \square

The following Lemma bounds the probability to exit $K(x, l, w_{\min})$ at an edge where the quasi-tree structure was violated.

Lemma 5.23. *If $m > \tau$, let $\text{cycle}(e_{m+1})$ be the event that the exploration of the $(L + 1)$ neighbourhood of the long-range edge $e_{m+1} \in E_m$ considered at the $(m + 1)$ -th step revealed a cycle. Let $\varepsilon \in (0, 1)$ and consider the following process. If $B_{\text{LR}}(x, L)$ is not quasi-tree-like, let $W_m := 1$ for all $m \geq 0$, otherwise set $W_0 := 0$ and for all $m \geq 0$ define*

$$W_{m+1} := W_m + (\hat{w}(e_{m+1}) \wedge \varepsilon/2) \mathbb{1}_{m < \tau} \mathbb{1}_{\text{cycle}(e_{m+1})}.$$

This is the total cumulative weight of edges that violated the quasi-tree structure at step $m + 1$ and that are below $\varepsilon/2$. Suppose $l = O(\log n)$, $R, L = O(\log \log n)$ and that $w_{\min} \geq e^{(\log \log n)^3}/n$. Then for all $s = s(n)$, with high probability, for all $x \in V$,

$$W_\tau \leq W_s + \varepsilon. \tag{5.78}$$

Proof. Fix $\varepsilon \in (0, 1)$, $x \in V$ and $s = s(n)$. Suppose $B_{\text{LR}}(x, L)$ is quasi-tree-like, otherwise the result is trivial. Let $(\mathcal{F}_m)_{m \geq 0}$ be the standard filtration of the random graphs $(K_m)_m$. The choice of the edge e_{m+1} is \mathcal{F}_m -measurable. Averaging conditional on \mathcal{F}_m , the generation of the $L + 1$ long-range neighbourhood of the edge e_{m+1} requires the sampling of at most Δ^{RL} long-range edges. For the first edge, there are exactly $n - \kappa_m$ possibilities. The quasi-tree structure is violated if an edge is sampled whose endpoint

is at small-range distance at most R from a previous ball explored in the same phase or from K_m . In this case it is necessarily at small-range distance at most R from a long-range edge of K_m . Hence the conditional probability of $\text{cycle}(e_{m+1})$ is upper bounded by $\Delta^R(\Delta^{RL} + \kappa_m)/(n - \kappa_m - \Delta^{RL})$. Since $K_0 := B_{LR}(x, L)$ and the sets K_m are increasing, $\kappa_m \geq \Delta^{RL}$ thus we can bound the conditional probability

$$\mathbb{P}[\text{cycle}(e_{m+1}) \mid \mathcal{F}_m] \leq \frac{2\Delta^{R(L+1)}\kappa_m}{n - 2\kappa_m}.$$

Since

$$\mathbb{E}[W_{m+1} - W_m \mid \mathcal{F}_m] = \mathbb{1}_{m < \tau} \hat{w}(e_m) \mathbb{P}[\text{cycle}(e_{m+1}) \mid \mathcal{F}_m]$$

(5.76) implies for some constant $C > 0$

$$\begin{aligned} \mathbb{E}[W_{m+1} - W_m \mid \mathcal{F}_m] &\leq \mathbb{1}_{m < \tau} \frac{Cl\Delta^{R(2L+1)}}{n - 2\kappa_\tau}, \\ \mathbb{E}[(W_{m+1} - W_m)^2 \mid \mathcal{F}_m] &\leq \mathbb{1}_{m < \tau} \frac{Cl^2\Delta^{R(3L+1)}}{\kappa_m(n - 2\kappa_\tau)} \end{aligned}$$

Furthermore, since every iteration of the procedure reveals at least one long-range edge, $\kappa_m \geq \Delta^{RL} + m$, in particular $\tau \leq \kappa_\tau$, hence summing over m yields

$$\begin{aligned} a &:= \sum_{m=s}^{\tau-1} \mathbb{E}[W_{m+1} - W_m \mid \mathcal{F}_m] \leq C' \kappa_\tau \frac{l\Delta^{R(2L+1)}}{n - 2\kappa_\tau}, \\ b &:= \sum_{m=s}^{\tau-1} \mathbb{E}[(W_{m+1} - W_m)^2 \mid \mathcal{F}_m] \leq C' \log(\kappa_\tau) \frac{l^2\Delta^{R(3L+1)}}{n - 2\kappa_\tau} \end{aligned}$$

for some other constant $C' > 0$. Using (5.77), the assumptions made on the different parameters imply that $a = o(1)$ and $b = n^{-1+o(1)}$. Consider now the martingale M_k defined by

$$M_k := \frac{2}{\varepsilon} \left(W_k - W_s - \sum_{m=s}^{k-1} \mathbb{E}[W_{m+1} - W_m \mid \mathcal{F}_m] \right).$$

Its increments are bounded by 1 and by construction $W_\tau - W_s = \frac{\varepsilon}{2} M_\tau + a$. Since $a = o(1)$, we infer that for large enough n

$$\mathbb{P}[W_\tau - W_s > 2\varepsilon] \leq \mathbb{P}[M_\tau > 2] \leq \mathbb{P}[\exists k > 0 : M_k > 2].$$

Thus we can apply Theorem 1.6 of [74] to bound

$$\mathbb{P}[\exists k > 0 : M_k > 2] \leq e^2 \left(\frac{4b/\varepsilon}{2 + 4b/\varepsilon} \right)^{2+4b/\varepsilon} \leq (2b\varepsilon^{-2})^2.$$

Since $b = n^{-1+o(1)}$, the right hand side is $o(1/n)$. □

5.9.2 Nice paths: definition

We now define nice paths in order to prove Proposition 5.2. For the rest of this section set

$$R := C_R \log \log n, \quad C_L := C_L \log \log n, \quad M := (\log \log n)^\kappa.$$

where $C_R, C_L > 0$ and $\kappa \geq 1$ are constants chosen large enough so that the conclusion of Lemma 5.3 holds. By the second point of Lemma 5.2, there exists a constant $\alpha > 0$ such that with high probability, from any starting point the chain has \mathcal{P} quasi-tree-like neighbourhood of radius $\lfloor \alpha \log n \rfloor$ after $\log n / 2h$ steps. Fix $\epsilon \in (0, 1)$ for the rest of this section. In the sequel, we consider several constant C_0, \dots, C_5 , defined in terms of the constants $C_{LR}(\epsilon), C_h(\epsilon), C'_{LR}(\epsilon)$ of Proposition 5.1 which can in particular depend on ϵ . They are indexed in the reverse order in which they are fixed, so C_0 is chosen after C_1 , which is chosen after C_2 , etc. Let

$$t := \left\lfloor \frac{\log n}{h} + C_0 \sqrt{\log n} \right\rfloor, \quad s := \lfloor \alpha \log n \rfloor \wedge \left\lfloor \frac{\log n}{10h} \right\rfloor, \quad l_1 := d(t-s) - C_4 \sqrt{t} \quad (5.79)$$

$$w_{\min} := e^{-h(t-s) - C_h \sqrt{t}} \quad w_{\max} := e^{-ht + C_1 \sqrt{t}}.$$

Given $r \in \mathbb{N}/2$, $L \leq r \leq M$ and $l_3 \in [\ell s - C_5 \sqrt{s}, \ell s + C_5 \sqrt{s}]$, consider the following three-stage generation of the environment:

1. Explore $K = K(x_0, l_1, w_{\min})$ as explained in the previous section. Let $E := E_\tau$ be the set of long-range edge remaining in the exploration queue at the end of the procedure, and consider the set E' of boundary vertices at long-range distance l_1 from x_0 , whose image under η is yet to determine.
2. The backward neighbourhood is $B = B(r, r+s, l_3) := B_{\mathcal{P}}^-(y, r+s) \cap B_{LR}^-(y, l_3)$.
3. Finally, reveal everything else.

It will be crucial in the sequel to control the numbers N_1, N_2 of long-range edges revealed during the two first stages. By definition $N_1 := \kappa(x, l_1, w_{\min})$. Observe then that if $\epsilon' \in (0, h/10 \log \Delta)$, then $h(t-s) < (1-\epsilon') \log n$. Thus by (5.77),

$$N_1 = O(l_1 \Delta^{RL} e^{h(t-s) + C_h \sqrt{t}}) = O(\log n \Delta^{C_R C_L (\log \log n)^2} n^{1-\epsilon'} e^{C \sqrt{\log n}})$$

for some constant C . The cardinality of B is bounded by that of $B_{\mathcal{P}}^-(y, r+s)$, which is $O(\Delta^{r+s}) = O(n^{2/10})$ by choice of $r \leq M$ and s . All in all, for any $\epsilon' \in (0, h/10 \log \Delta)$,

$$N_1 = O(n^{1-\epsilon'}), \quad (5.80)$$

$$N_2 = O(n^{1/10}).$$

Let $\mathcal{F}_{r,l}$ be the σ -algebra generated by the long-range edges revealed during the two first stages. Since K is quasi-tree-like, a non-backtracking long-range path ξ from x

to E' entirely contained in K must cross a unique edge of E . Let ξ_E denote this edge and define $w_E(\xi) := w_{x,R,L}(\xi_1 \cdots \xi_E)$. This is essentially the weight $w(\xi)$, but where last steps of the path were truncated to keep a weight that is $\mathcal{F}_{r,l}$ -measurable.

In B , let F' be the set of boundary vertices which are at long-range distance exactly l from y , for which the shortest long-range path to y is unique and has a tree-like neighbourhood in $B_{\mathcal{D}}(y, r+s)$, that is $B \cap B_{LR}(z, L)$ contains no long-range cycle for all vertex z on this path. Consider now the set F of long-range edges in B that are at long-range distance $L-1$ from F' . If ξ is a non-backtracking long-range path from F' to y , let ξ_F denote the unique edge of F crossed by ξ and ξ_{F+1} the subsequent edge. Set $w_F(\xi) := w_{\xi_F^+, R, L}(\xi_{F+1} \cdots \xi_{|\xi|-L+1})$, where ξ_F^+ is the endpoint of ξ_F closest from y . Here we truncate the path at both ends: the first steps to have a $\mathcal{F}_{r,l}$ -measurable weight but also the last steps, as the trajectories considered afterwards end at y .

Definition 5.7. Given $r \in \mathbb{N}/2$, $r \leq M$ and $l \in [\ell s - C_5 \sqrt{\log n}, \ell s + C_5 \sqrt{\log n}]$ let $\mathfrak{N}_{r,l}^t(x, y)$ be the set of length t paths \mathfrak{p} between x and y such that

- (i) $\mathfrak{p} \in \Gamma(R, L, M)$
- (ii) \mathfrak{p} can be decomposed as the concatenation $\mathfrak{p} = \mathfrak{p}_1 \mathfrak{p}_2 \mathfrak{p}_3$ of three paths such that: \mathfrak{p}_1 is a path from x to E' entirely contained in $K(x, l_1)$, whose endpoint is the only vertex of E' it contains (which implies that it starts and ends with a long-range step),
- (iii) \mathfrak{p}_2 is a path between E' and F' which starts with a small-range step but ends with a long-range step such that

$$C_3 \sqrt{\log n} \leq |\mathfrak{p}_2| \leq C_2 \sqrt{\log n}.$$

for some C_2, C_3 . In addition, \mathfrak{p}_2 contains a regeneration edge outside $K(x, l_1, w_{\min})$ before step ML , and the endpoint of \mathfrak{p}_2 is the only vertex of $B(y, r+s, l)$ it contains.

- (iv) \mathfrak{p}_3 is a path of length $r+s$ from F' to y entirely contained in $B(y, r+s, l)$, which starts and ends with a small-range step and does not contains any regeneration edge in its first r steps,
- (v) $w_E(\xi(\mathfrak{p}_1))w_F(\xi(\mathfrak{p}_3)) \leq w_{\max}$.

In the sequel we consider

$$\mathcal{P}_{r,l}^t(x, y) := \sum_{\mathfrak{p} \in \mathfrak{N}_{r,l}^t(x, y)} \mathcal{P}(\mathfrak{p}).$$

The complete set of nice paths is obtained by taking the union over parameters r, l , namely

$$\mathfrak{N}^t(x, y) := \bigcup_{\substack{r \leq M \\ r \in \mathbb{N}+1/2}} \bigcup_{l=\ell_s-C_5\sqrt{s}}^{\ell_s+C_5\sqrt{s}} \mathfrak{N}_{r,l}(x, y).$$

and the total probability of nice paths by $\mathcal{P}_{\mathfrak{N}^t}^t(x, y) := \sum_{\mathbf{p} \in \mathfrak{N}^t(x, y)} \mathcal{P}(\mathbf{p})$.

Conditions (i) and (ii) allow to relate the probability of following a nice path with the weight constraint (v): thanks to the tree structure of $K(x, l_1)$, each vertex in E' has a unique ancestor edge in E . Given $e \in E$, let $E'(e)$ denote the set of vertices in E' with ancestor e and recall $\xi(e)$ is the unique non-backtracking long-range path from x to e . Similarly for $f \in F$, let $F'(f)$ be the set of vertices of F' from which the unique non-backtracking long-range path to y goes through f and $\xi(f)$ the unique non-backtracking long-range path from f^+ to y (which thus does not include f). Then for a fixed total long-range length l :

$$\begin{aligned} \sum_{\substack{\mathbf{p}: \xi(\mathbf{p}_1)_{E=e} \\ \xi(\mathbf{p}_3)_{F=f} \\ |\xi(\mathbf{p})|=l}} \mathcal{P}(\mathbf{p}) &\leq \sum_{\substack{\xi: \xi_E=e \\ \xi_F=f, \\ |\xi|=l}} \mathbf{P}_x [\xi(X_0 \cdots X_{\tau_l})_{\leq l-L+1} = \xi, (X_0 \cdots X_{\tau_l}) \in \Gamma(R, L, M)] \\ &\leq \sum_{\substack{\xi: \xi_E=e \\ \xi_F=f, \\ |\xi|=l}} w_{x,R,L}(\xi). \end{aligned}$$

by Lemma 5.4. For each ξ in the sum,

$$\begin{aligned} w_{x,R,L}(\xi) &= w_{x,R,L}(\xi_1 \cdots \xi_E) w_{\xi_E^+, R, L}(\xi_{E+1} \cdots \xi_F) w_{\xi_F^+, R, L}(\xi_{F+1} \cdots \xi_{l-L+1}) \\ &= w_E(\xi_1 \cdots \xi_E) w_{\xi_E^+, R, L}(\xi_{E+1} \cdots \xi_F) w_F(\xi_{F+1} \cdots \xi_{l-L+1}) \end{aligned}$$

Observe now that for fixed e and f the first and third factors are fixed as well and determined by $\xi(\mathbf{p}_1), \xi(\mathbf{p}_3)$, so the sum is only over the steps $\xi_{E+1} \cdots \xi_{F-1}$. Since weights sum up to 1, we can also sum over l and bound

$$\sum_{\substack{\mathbf{p}: \xi(\mathbf{p}_1)=\xi_1 \\ \xi(\mathbf{p}_3)=\xi_3}} \mathcal{P}(\mathbf{p}) \leq w_{\max} \tag{5.81}$$

thanks to the weight constraint (v). In words: the total probability of nice paths with prescribed long-range edges in E and F is upper bounded by w_{\max} .

5.9.3 Nice paths are typical

We prove here points (i)-(ii) of Proposition 5.2. The argument is the same for both points and relies on Lemmas 5.1, 5.2, 5.3 and Proposition 5.1. Let $t = \log n/h + C_0(\varepsilon)\sqrt{\log n}$. To prove (i) we can consider a fixed starting state $x \in V$ and then need

to ensure that the t steps of a trajectory are nice with quenched probability at least $1 - \varepsilon$, while for point (ii) we need to prove the last t steps of a length $s' + t$ trajectory are nice, for $s' = (\log n)^a$, $a \geq 1$ or $s' = C \log \log n$, $C > 0$ in the reversible case, this time independently of the starting state. The proof thus only differs by which statement of the aforementioned Lemmas and Proposition we use. Let us for instance focus on point (ii) in the general model. Consider $a \geq 1$ such that the statement of Lemmas 5.2, 5.3 and Proposition 5.1 hold. There are several properties to check. Since there are finitely many of them, once a property is shown to hold with probability $1 - o_{\mathbb{P}}(1)$ or $1 - \varepsilon$ we can automatically assume it is satisfied when checking the remaining properties. Write $t' := s' + t$ to simplify notation.

We already proved in Lemma 5.3 that a failure of the requirement $(X_{s'} \cdots X_{t'}) \in \Gamma(R, L, M)$ occurs with probability $o_{\mathbb{P}}(1)$. For fixed r, l the very definition of the backward neighbourhood $B(r, r + s, l)$ implies that it necessarily contains the last $r - s$ steps of the trajectory provided they have the prescribed long-range length l . Summing over r and l , the constraint that \mathfrak{p}_2 only has its last endpoint in the backward neighbourhood while \mathfrak{p}_3 is contained in it amounts to conditioning by the last regeneration time occurring before s . In particular this requires the existence of a regeneration time in the interval $[t' - s - M, t' - s]$ but this is exactly ensured by the fact that $(X_{s'} \cdots X_{t'}) \in \Gamma(R, L, M)$. Assume now this property hold and let $T_{F'}$ denote the last regeneration before time $t' - s$. The remaining obstructions to following a nice paths are:

1. the first steps of $(X_{s'} \cdots X_{t'})$ are not contained in $K = K(X_{s'}, l, w_{\min})$, which occurs if $B_{\text{LR}}(X_{s'}, L)$ is not quasi-tree-like. This occurs with probability $o_{\mathbb{P}}(1)$ by Lemma 5.2.
2. from $X_{s'}$, the chain leaves K before it reaches long-range distance l_1 , which occurs if:
 - the loop-erased trace exits K through the L long-range neighbourhood of an edge e which satisfied $\hat{w}(e) < w_{\min}$: since cumulative weights along a path are non-increasing this implies $-\log w(\xi(X_{s'} \cdots X_{t'-s})) > -\log w_{\min} = h(t-s) + C_h \sqrt{t}$ which occurs with probability less than ε by Proposition 5.1.
 - the chain crosses an edge which violated the quasi-tree structure. Recall the process considered in Lemma 5.23. As mentionned above we can suppose $B_{\text{LR}}(X_{s'}, L)$ is quasi-tree-like. Then by Proposition 5.1, the total probability of paths with long-range length L and weights above $\varepsilon/2$ is $o_{\mathbb{P}}(1)$. Thus up to this $o_{\mathbb{P}}(1)$ error the quantity W_{κ} considered in the lemma exactly counts the probability to exit K at an edge which violated the tree structure. Thanks

to the choice of w_{\min} (5.79) the lemma establishes that with high probability, $W_\kappa \leq W_{\lfloor L/2 \rfloor} + \varepsilon$ for any value of $X_{s'}$. However $W_{\lfloor L/2 \rfloor} = 0$ as $B_{\text{LR}}(X_{s'}, L)$ is quasi-tree-like.

3. the path from $X_{t'-s-T_{F'}}$ to y is not unique or its long-range length is not in the interval $[ds - C_5\sqrt{\log n}, ds + C_5\sqrt{\log n}]$. If this path is not unique, this implies in particular that the ball $B(X_{t'-s-M}, t' - s - M)$ is not quasi-tree-like. Since $s + M = O(s)$ Lemma 5.2 shows this occurs with probability $o_{\mathbb{P}}(1)$. For the long-range distance requirement, Proposition 5.1 shows that $|\xi(X_{t'-s} \dots X_{t'})| - ds| \leq C_{\text{LR}}\sqrt{s}$ with probability at least $1 - \varepsilon - o(1)$. Then note that the long-range distance traveled in the intervals $[T_{F'}, t']$ and $[t' - s, t']$ differ by at most M , hence $|\xi(X_{t'-s-T_{F'}} \dots X_{t'})| - ds| \leq C_{\text{LR}}\sqrt{s} + M \leq C_5\sqrt{s}$ for a large enough constant C_5 , using that $M = o(s)$.
4. No regeneration occurred outside outside K during the first ML steps of the intermediate path \mathfrak{p}_2 or the latter does not have length $O(\sqrt{t})$. Since the trajectory is in $\Gamma(R, L, M)$ there must be regeneration edge in less than M steps after having first reached E' . By the non-backtracking property, this regeneration edge is either outside K or at long-range distance less than L from E' , after which the chain is forced to exit K . Thus the L -th consecutive regeneration edge must be outside K . Using again that regeneration edges are spaced by at most M steps, this makes in total ML steps at most until a regeneration occurs outside $K(x, l_1)$. For the length requirement observe that the long-range length is sub-additive. Since by definition a nice trajectory decomposes as the concatenation $(X_{s'} \dots X_{t'}) = \mathfrak{p}_1 \mathfrak{p}_2 \mathfrak{p}_3$ the sub-additivity implies

$$|\xi(X_{s'} \dots X_{s'+t})| \leq |\xi(\mathfrak{p}_1)| + |\xi(\mathfrak{p}_2)| + |\xi(\mathfrak{p}_3)|.$$

The path $\xi(\mathfrak{p}_1)$ has length l_1 , whereas $\xi(\mathfrak{p}_3)$ has variable length but from the bounds on l in Definition 5.7 and (5.79) we infer that their combined length is

$$|\xi(\mathfrak{p}_1)| + |\xi(\mathfrak{p}_3)| \leq dt - (C_4 - C_5)\sqrt{t},$$

while the intermediate path obviously has length $|\xi(\mathfrak{p}_2)| \leq |\mathfrak{p}_2|$. Choose $C_4 \geq C_5 + 2C_{\text{LR}}$. Hence if $|\mathfrak{p}_2| < C_3\sqrt{t}$ with $C_3 := C_{\text{LR}}$, then $|\xi(X_{s'} \dots X_{s'+t})| < dt - C_{\text{LR}}\sqrt{t}$, which occurs with probability at most $o_{\mathbb{P}}(1)$ by Proposition 5.1. To prove the upper bound on $|\mathfrak{p}_2|$, observe $|\mathfrak{p}_1| + s'$ coincides with the first hitting time τ_{l_1} of long-range distance l_1 after step s' . From Proposition 5.1 we can deduce the existence of $C_2 > 0$ such that $\tau_{l_1} - s' \geq t - s - C_2\sqrt{t}$ with probability at least $1 - \varepsilon$. Since $|\mathfrak{p}_3| \geq s$, we deduce that

$$|\mathfrak{p}_2| = t - |\mathfrak{p}_1| - |\mathfrak{p}_3| \leq C_2\sqrt{t}.$$

5. $w_E(\xi(X_{s'} \cdots X_{\tau_{l_1}}))w_F(\xi(X_{t'-s-T_{F'}} \cdots X_{t'})) > w_{\max}$: let $\xi := \xi(X_{s'} \cdots X_t)$. When deriving (5.81) we used that

$$w_{x,R,L}(\xi) = w_E(\xi_1 \cdots \xi_E)w_{\xi_E^+,R,L}(\xi_{E+1} \cdots \xi_F)w_F(\xi_{F+1} \cdots \xi_l)$$

wher ξ_l is the last edge of ξ . Now by the non-backtracking property of nice paths $\xi(X_{s'} \cdots X_{\tau_{l_1}})$ and $\xi(X_{t'-s-T_{F'}} \cdots X_{t'})$ contain $\xi_1 \cdots \xi_E$ and $\xi_{F+1} \cdots \xi_l$ respectively so the goal is to show that $w_E(\xi_1 \cdots \xi_E)w_F(\xi_{F+1} \cdots \xi_l) > w_{\max}$ with probability at most ε . As we argued for the previous point, Proposition 5.1 implies $\tau_{l_1} - s' \geq t_1 := t - s - C_2\sqrt{t}$ with probability at least $1 - \varepsilon$. Since cumulative weights are non-increasing along a path, we deduce that

$$w_E(\xi_1 \cdots \xi_E) \leq w_{x,R,L}(\xi(X_{s'} \cdots X_{t_1})) \leq e^{-t_1 h + C_h \sqrt{t_1}}$$

with probability at least $1 - \varepsilon$. Similarly we know that $T_{F'} \leq t' - s$, hence

$$w_E(\xi_1 \cdots \xi_F) \geq w(\xi(X_{s'} \cdots X_{t'-s})) \geq e^{-(t-s)h - C_h \sqrt{t-s}}$$

with probability at least $1 - \varepsilon - o(1)$. From these two bounds, we deduce

$$w_{\xi_E^+,R,L}(\xi_{E+1} \cdots \xi_F) \geq e^{-C_h \sqrt{t-s} - C_h \sqrt{t_1} - C_h \sqrt{t}} \geq e^{-C' \sqrt{t}}$$

for some $C' = C'(\varepsilon) > 0$. Thus if $w_E(\xi_1 \cdots \xi_E)w_F(\xi_{F+1} \cdots \xi_l) > w_{\max}$ we obtain that

$$w_{x,R,L}(\xi) \geq w_{\max} e^{-C' \sqrt{t}}$$

which has probability at most ε if the constant C_1 in w_{\max} (5.79) is taken sufficiently large.

5.9.4 Concentration of nice paths

Recall $\mathcal{F}_{r,l}$ is the σ -algebra generated by the long-range edges revealed during the two first stages. For fixed r and l , we prove concentration of $\mathcal{P}_{\mathfrak{N}_{r,l}}^t(x, y)$ conditional on $\mathcal{F}_{r,l}$ using Theorem 4.1. The concentration will be strong enough for a union bound, which will yield Proposition 5.2. Conditional on $\mathcal{F}_{r,l}$ all the randomness of $\mathcal{P}_{\mathfrak{N}_{r,l}}^t(x, y)$ comes from the intermediate path \mathfrak{p}_2 . It is contained in a random environment that arises from a uniform bijection σ' between the two subsets $V_1' \subset V_1$, $V_2' \subset V_2$ that remain to be matched after the two first stages. As was observed in the previous section, the two first stages revealed a number of long-range edges which is $O(n^{1-\varepsilon'})$, so σ' is a uniform bijection between sets of $n' = n - o(n)$ elements, which can be identified with a uniform permutation of n' elements.

The following lemma proves that conditions of concentration are fulfilled. We use Corollary 4.1 which is more practical to use than Theorem 4.1.

Lemma 5.24. *For all $r \leq M, l \geq 0$, conditional on $\mathcal{F}_{r,l}$ $\mathcal{P}_{\mathfrak{N}_{r,l}}^t(x, y)$ can be realized as a multilinear function ϕ on $S_{n'}$ of degree at most*

$$d := C_2 \sqrt{\log n'}. \tag{5.82}$$

With the notations of Corollary 4.1,

$$\alpha_\phi = O(d^3 (4\Delta)^d w_{\max} \log n), \quad A_\phi = O(d^2 \Delta^d n^{2/10} w_{\max})$$

and $A_{\nabla\phi} = O(d^2 (2\Delta)^d w_{\max}).$

Proof. A pair of multi-indices (\mathbf{i}, \mathbf{j}) of size k identifies with a sequence of k potential long-range edges $(i_1, j_1), \dots, (i_k, j_k)$. Note that all properties required of nice paths are either $\mathcal{F}_{r,l}$ -measurable or can be determined from the path (this would not be the case if we had chosen for instance to take the weight of $w(\xi(\mathbf{p}_2))$ into account). Therefore it is possible to define the coefficient $a_{\mathbf{i}, \mathbf{j}}$ as the total probability of nice paths which meet exactly these long-range edges in addition to those that are contained in K or B (which do not count as random here). Nice paths do not have to cross the long-range edges but may only pass through one endpoint, due to the fact that the probability $p = p(x, \sigma(x))$ of crossing a long-range edge depends in general on σ .

The upper bound on the degree follows easily from this definition. By Point (iii) in Definition 5.7, the random part \mathbf{p}_2 of nice paths has length bounded by $d = C_2 \sqrt{\log n'}$ which thus also upper bounds the number of long-range edges met by nice paths and consequently the degree.

To prove concentration recall the notation of Proposition 4.1. Consider a set S of long-range edges of size d or less. It is connected to at most d vertices of E' and F' , which in turn correspond to at most d edges in E and F . Therefore the set of nice paths that meet the edges of S exactly can cross at most d edges of E and d edges of F , hence by (5.81) we deduce that the maximal coefficient of ϕ is bounded by $d^2 w_{\max}$. By multilinearity, the maximal coefficient is non-increasing with respect to partial differentiation (see Lemma 4.6) so this proves

$$M(\partial_{i,j}\phi, \sigma) \leq M(\phi, \sigma) \leq d^2 w_{\max}$$

for any potential long-range edge (i, j) . On the other hand, the number of monomials in ϕ can be upper bounded by the maximal number of paths between E' and F' of length d . Since F' has cardinality at most $n^{2/10}$ by (5.80)

$$N(\phi, \sigma) \leq |F'| \Delta^d = O(\Delta^d n^{2/10}).$$

Restricting to a potential long-range edge (i, j) , the latter can be part of at most $O((2\Delta)^d)$ paths of length d between E' and F' and thus as many monomials. Therefore

$$N(\partial_{i,j}(\phi, \sigma)) \leq O((2\Delta)^d).$$

We deduce that $A_\phi = O(d^2 \Delta^d n^{2/10} w_{\max})$ and $A_{\nabla\phi} = O(d^2 (2\Delta)^d w_{\max})$. The log factor in Theorem 4.1 is thus $O(\log n)$ hence

$$\alpha_\phi = O(d^3 (4\Delta)^d w_{\max} \log n).$$

□

The following Lemma will be sufficient to deduce Proposition 5.2. The proof is postponed to the next section. Recall the measure $\mathbf{Q}_u^{(L)}$ was defined in (5.19). To ease notation we will drop the exponent (L) in the sequel. Recall also that $T_1^{(G,l)}, T_1^{(\mathcal{G},l)}$ denote regeneration times with horizon L in G^* and \mathcal{G} respectively, μ is the invariant measure of the regeneration chain in the quasi-tree and ν was considered in Proposition 5.12. Below we write $\mathbf{Q}_{\nu+c} = \sum_u (\nu(u) + c(u)) \mathbf{Q}_u$.

Lemma 5.25. *There exists a measure c on V such that $\sum_{v \in V} c(v) = o_{\mathbb{P}}(1)$ and*

$$1 - o_{\mathbb{P}}(1) \leq \mathbb{E} \left[\mathcal{P}_{\mathfrak{N}_{r,l}}^t(x, y) \mid \mathcal{F}_{r,l} \right] \left(\frac{\mathbf{Q}_{\nu+c}[X_{r+s} = y, |\xi(X_0 \cdots X_{r+s})| = l, r < T_1 \leq M]}{\mathbb{E}_{\mathbf{Q}_\mu}[T_1^{(G,\infty)}]} \right)^{-1} \leq 1. \quad (5.83)$$

Proof of Proposition 5.2. Points (i)-(ii) were established proved in Section 5.9.3 so we now prove (iii) - (iv). Suppose first $x, y \in V$ and r, l are fixed. Let $\phi := \mathcal{P}_{\mathfrak{N}_{r,l}}^t(x, y)$ as in Lemma 5.24 and $z := \frac{\varepsilon}{2} \mathbb{E}[\phi] + \frac{\varepsilon}{2C_6 M n \sqrt{\log n}}$ for some $C_6 > 0$. Note that $w_{\max} \leq e^{-C\sqrt{\log n}}/n$ for some constant C which tends to $+\infty$ as the constant C_0 in the definition of t grows (5.79), while other factors of $\alpha_\phi, A_{\nabla\phi}$ are of all of order at most $e^{C'\sqrt{\log n}}$. Thus for any choice of $C = C(\varepsilon) > 0$ Lemma 5.24 shows that $\alpha_\phi, A_{\nabla\phi}$ can both be bounded by $e^{-C\sqrt{\log n}}/n$, provided the constant C_0 is sufficiently large, while $d^2(d-1)A_\phi/n = o(n^{-3/5})$. In particular we can choose C so that $A_{\nabla\phi} \leq z$. Since $\mathbb{E}[\phi] \leq 2z/\varepsilon$ and $z \geq \varepsilon/(2C_6 M n \sqrt{\log n})$, applying Corollary 4.1 yields

$$\mathbb{P}[|\phi - \mathbb{E}[\phi]| \geq z] \leq 2 \exp\left(\frac{-C'\varepsilon^2}{\alpha_\phi M n \sqrt{\log n}}\right) \quad (5.84)$$

for some $C' = C'(\varepsilon) > 0$. Up to increasing again the value of C , we can ensure that $\alpha_\phi M n \sqrt{\log n} \leq (\log n)^{-2}$, hence

$$\mathbb{P}\left[|\phi - \mathbb{E}[\phi]| > \frac{\varepsilon}{2} \mathbb{E}[\phi] + \frac{\varepsilon}{2C_6 M n \sqrt{\log n}}\right] \leq \exp(-C'\varepsilon^2 (\log n)^2).$$

This is sufficient to take a union bound over $x, y \in V$ and $r \leq M, l \in [ds - C_5 \sqrt{\log n}, ds + C_5 \sqrt{\log n}]$. Thus summing over r, l we obtain that with high probability, for all $x, y \in V$,

$$\left| \mathcal{P}_{\mathfrak{N}}^t(x, y) - \sum_{r,l} \mathbb{E} \left[\mathcal{P}_{\mathfrak{N}_{r,l}}^t(x, y) \mid \mathcal{F}_{r,l} \right] \right| \leq \frac{\varepsilon}{2} \sum_{r,l} \mathbb{E} \left[\mathcal{P}_{\mathfrak{N}_{r,l}}^t(x, y) \mid \mathcal{F}_{r,l} \right] + \frac{\varepsilon}{n}.$$

Lemma 5.25 gives an estimate of the conditional expectation which shows:

$$\begin{aligned} \mathcal{P}_{\mathfrak{N}}^t(x, y) &\leq \frac{1 + \varepsilon/2}{\mathbf{E}_{\mathbf{Q}_\mu} [T_1^{(\mathcal{G}, \infty)}]} \sum_{r=0}^M \mathbf{Q}_{\nu+c} [X_{r+s} = y, r < T_1^{(G, L)} \leq M] + \frac{\varepsilon}{n} \\ &= (1 + \varepsilon/2) A \hat{\pi}(y) + \frac{(1 + \varepsilon/2) A}{\mathbf{E}_{\mathbf{Q}_\nu} [T_1^{(G, L)}]} \sum_{r=0}^M \mathbf{Q}_c [X_{r+s} = y, r < T_1^{(G, L)} \leq M] + \frac{\varepsilon}{n}. \end{aligned} \quad (5.85)$$

with $A := \frac{\mathbf{E}_{\mathbf{Q}_\nu} [T_1^{(G, L)}]}{\mathbf{E}_{\mathbf{Q}_\mu} [T_1^{(\mathcal{G}, \infty)}]}$. Summing over $y \in V$ and $r \in [0, M]$ in the second term yields $(1 + \varepsilon/2) A \sum_{v \in V} c(v) \leq 2A\varepsilon$ with high probability. On the other hand, from Lemma 5.25 we also have the lower bound

$$\mathcal{P}_{\mathfrak{N}}^t(x, y) \geq (1 - \varepsilon/2)(1 - o_{\mathbb{P}}(1)) A \hat{\pi}(y) - \varepsilon/n,$$

which by summing over $y \in V$ shows that $A \leq 1 + 10\varepsilon$ with high probability. Plugging this in the upper bound above yields the second statement of the proposition.

Let us move to the proof of (iv). Define

$$\hat{\pi}_1(v) := \frac{1}{\mathbf{E}_{\mathbf{Q}_\nu} [T_1^{(G, L)}]} \sum_{r=0}^M \sum_{u \in S} \nu(u) \mathbf{Q}_u [X_{r+s} = v, r < T_1 \leq M, \tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} > M + s]$$

for all $v \in V$ and decompose $\hat{\pi} =: \hat{\pi}_1 + \hat{\pi}_2$.

We start bounding $\hat{\pi}_2(V)$. Summing over y in (5.85), the left hand side becomes the total probability of following a nice path, which by the first part of the proposition is at least $1 - \varepsilon$ with high probability for a typical $x \in V$. Consequently A can be lower bounded by a constant with high probability, which implies that for some constant $c > 0$, $\mathbf{E}_{\mathbf{Q}_\nu} [T_1^{(G, L)}] \geq c$ with high probability. Thus by summing over $v \in V$ we can bound

$$\hat{\pi}_2(V) \leq c^{-1} M \mathbf{Q}_\nu [\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq M + s] + o_{\mathbb{P}}(1).$$

Then notice that the quenched probability $\mathbf{P} [\tau_{\eta(u)} > \tau_L \mid X_{1/2} = u]$ can always be lower bounded by $q_n := \delta^{CL}$ for some constant $C > 0$, hence

$$\begin{aligned} \hat{\pi}_2(V) &\leq c^{-1} M \sum_{u \in V} \nu(u) \frac{\mathbf{P} [\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq M + s, \tau_{\eta(u)} > \tau_L \mid X_{1/2} = u]}{\mathbf{P} [\tau_{\eta(u)} > \tau_L, X_{1/2} = u]} + o_{\mathbb{P}}(1) \\ &\leq c^{-1} q_n^{-1} M \sum_{u \in V} \nu(u) \mathbf{P} [\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq M + s \mid X_{1/2} = u] + o_{\mathbb{P}}(1). \end{aligned}$$

As $L = O(\log \log n)$ one has $q_n^{-1} = O((\log n)^{-b})$ for some $b > 0$. Next by Markov's

property for all $u \in V$

$$\begin{aligned} \mathbf{P} \left[\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq M + s \mid X_{1/2} = u \right] &= \mathbf{P} \left[\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq M + s \mid X_0 = u, X_{1/2} = u \right] \\ &= \frac{\mathbf{P}_u \left[\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq M + s, X_{1/2} = u \right]}{p(u, \eta(u))} \\ &\leq c' \mathbf{P}_u \left[\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq M + s \right] \end{aligned}$$

for some constant $c' > 0$ as the entries of p are bounded (Assumption (H2)). For fixed $u \in V$, Lemma 5.3 shows $\mathbf{P}_u \left[\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq M + s \right] = o_{\mathbb{P}}((\log n)^{-b}) = o_{\mathbb{P}}(q_n^{-1})$ as $M + s = O(\log n)$. Letting $U \in V$ be a random variable independent from the environment, this can be rephrased as the fact that $\mathbf{P}_U \left[\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq M + s \right] = o_{\mathbb{P}}(q_n^{-1})$ conditional on U , but then this statement must also hold unconditionally. As the measure ν is deterministic we can take U of law ν , after which by taking the conditional expectation with respect to the environment we get that

$$\mathbb{E} \left[\mathbf{P}_U \left[\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq M + s \right] \mid \eta \right] = \sum_{u \in V} \nu(u) \mathbf{P}_u \left[\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} \leq M + s \right] = o_{\mathbb{P}}(q_n^{-1}).$$

All in all, this proves $\hat{\pi}_2(V) = o_{\mathbb{P}}(1)$.

To bound the ℓ^2 norm of $\hat{\pi}_1$, we can use what precedes to first bound

$$\begin{aligned} \hat{\pi}_1 &\leq c^{-1} M \max_{r \leq M} \mathbf{Q}_{\nu} \left[X_{r+s} = \cdot, \tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} > s \right] \\ &\leq c^{-1} q_n^{-1} M \max_{r \leq M} \sum_{u \in V} \nu(u) \mathbf{P} \left[X_{r+s} = \cdot, \tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} > M + s \mid X_{1/2} = u \right]. \end{aligned}$$

Consequently

$$\mathbb{E} \left[\sum_{v \in V} \hat{\pi}_1(v)^2 \right] \leq c^{-2} q_n^{-2} M^2 \max_{r \leq M} \mathbb{P} \left[\begin{array}{l} X_{r+s} = X'_{r+s}, \tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} > s, \\ \tau_{\eta(X_{1/2})} > \tau_L, \tau_{\eta(X'_{1/2})} > \tau_L \end{array} \right]$$

where X, X' are two independent chains conditional on the same environment, both started at time $1/2$ under the law ν . Here the event $\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} > s$ applies to both chains. Fix $r \leq M$. Observe that if all long-range edges crossed by X' are at distance more than R from the trajectory followed by the first chain X , then X' can reach X_{r+s} only by deviating by a small-range distance R or by crossing the long-range edge $(X'_{1/2}, \eta(X'_{1/2}))$ if $\eta(X'_{1/2}) \in B_{\text{SR}}^-(X_{1/2} \dots X_{r+s}, R)$. The event $\{\tau_{\eta(X'_{1/2})} > \tau_L\}$ proscribes this last possibility to occur before the chain has crossed a long-range distance L , so the chain must backtrack. Thus either of these two possibilities contradicts $\tau_{\text{SR}}^{(R)} \wedge \tau_{\text{NB}}^{(L)} > s$. Consequently the only possibility that the two chains meet is that the second one crosses a long-range edge at small-range distance less than R from the first chain. From (5.56) $d_{\text{SR}}(X_{1/2}, X'_{1/2}) \leq R$ with probability $O(\Delta^R/n) = o(1)$ under ν . Then assuming this does not hold, under the annealed law the environment can be generated sequentially

along the trajectories of the chains as explained in Section 5.2.7. The number of long-range edges met by X' up to time $M + s$ which fall in a R small-range neighbourhood of X is thus stochastically dominated by a binomial $\text{Bin}(M + s, \Delta^R(M + s)/n)$. As $q_n = O((\log)^b)$, $R = O(\log \log n)$, $s = O(\log n)$, $M = o(\log n)$ we deduce that

$$\mathbb{E} \left[\sum_{v \in V} \hat{\pi}_2(v) \right] \leq O \left(\frac{C(\log n)^{b'}}{n} \right)$$

for some $b' > 0$, which gives the result by the first moment argument (5.4). \square

5.9.5 Expectation of nice paths

Proof of Lemma 5.25. Let $\tau_{E'}$ be the hitting time of E' and $U_{t'}(K)$ the time spent in K before the time $t' \geq 0$. By definition, a nice path requires that $\tau_{E'} \leq t - s$ and the first part \mathfrak{p}_1 of the path is the trajectory until $\tau_{E'}$. The second part of a nice path is the trajectory until hitting F' and is assured to spend less than $m := ML$ steps in K . Therefore by strong Markov's property, one can bound

$$\begin{aligned} \sum_{\mathfrak{p} \in \mathfrak{N}_{r,l}} \mathcal{P}(\mathfrak{p}) &\leq \sum_{\substack{t_1+t_2=t-(r+s) \\ C_3\sqrt{\log n} \leq t_2 \leq C_2\sqrt{\log n}}} \sum_{u \in E'} \sum_{v \in F'} (\mathbf{P}_x [\tau_{E'} = t_1, X_{t_1} = u]) \\ &\times \mathbf{P} \left[\exists k \geq 0 : T_k^{(G,L)} = t_2, X_{t_2} = v, U_{t_2}(K) \leq m \mid X_{1/2} = u \right] \\ &\mathbf{P} \left[X_{r+s} = y, r < T_1 \leq M \mid X_{1/2} = v, \tau_{\eta(v)} > \tau_L \right]. \end{aligned} \quad (5.86)$$

In each term of this sum, the first and third factor are $\mathcal{F}_{r,l}$ -measurable, so only the second factor gets averaged when taking conditional expectation. We claim this expectation satisfies

$$\sum_{v \in F'} \left| \mathbb{P} \left[\exists k \geq 0 : T_k^{(G,L)} = t_2, X_{t_2} = v, U_{t_2}(K) \leq m \mid X_{1/2} = u, \mathcal{F}_{r,l} \right] - \frac{\nu(v)}{\mathbb{E}_{\mathbb{Q}_\mu} [T_1^{(G,\infty)}]} \right| + o_{\mathbb{P}}(1).$$

Note that $\nu/\mathbb{E}_{\mathbb{Q}_\mu} [T_1^{(G,\infty)}]$ is independent of u and t_1 , while the first factor in the sum considered above can be summed to at most 1. Also we can recognize the measure \mathbf{Q}_v (5.19) in the third factor. Therefore provided the claim holds one obtains that for some $c = (c_v)_v$ satisfying $\sum_{v \in V} c(v) = o_{\mathbb{P}}(1)$,

$$\begin{aligned} \mathbb{E} \left[\sum_{\mathfrak{p} \in \mathfrak{N}_{r,l}} \mathcal{P}(\mathfrak{p}) \mid \mathcal{F}_{r,l} \right] &\leq \sum_{v \in F'} \left(\frac{\nu(v)}{\mathbb{E}_{\mathbb{Q}_\mu} [T_1^{(G,\infty)}]} + c(v) \right) \mathbf{Q}_v [X_{r+s} = y, r < T_1 \leq M] \\ &\leq \sum_{v \in V} \frac{\nu(v) + c(v)}{\mathbb{E}_{\mathbb{Q}_\mu} [T_1^{(G,\infty)}]} \mathbf{Q}_v [X_{r+s} = y, d_{\text{LR}}(v, y) = l, r < T_1 \leq M], \end{aligned}$$

where in the second we implicitly changed the definition of c , using that $\mathbb{E}_{\mathbb{Q}_\mu} [T_1^{(G,\infty)}] = O(1)$. This proves the upper bound in the lemma. To prove the lower bound, note that

the inequality (5.86) is not sharp if the trajectory is not in $\Gamma(R, L, M)$ or there is no regeneration time in the interval $[\tau_{l_1}, ML]$. This was shown to occur with probability $o_{\mathbb{P}}(1)$ in Section 5.9.3, so the upper bound is also a lower bound up to a $o_{\mathbb{P}}(1)$ error.

Let us prove the claim. Let $u \in E', v \in F'$. The idea is to couple the chains X and \mathcal{X} on a quasi-tree to relate their regeneration times, in order to use Proposition 5.12. However the conditioning by $\mathcal{F}_{r,l}$ already revealed some long range edges, which requires in turn a similar conditioning in the quasi-tree. We argue that the only conditioning required is by $B_{\text{LR}}(O, L)$, the ball of radius L in a quasi-tree. Let u_0 be the ancestor of E at long-range distance L from u . By definition of $K(x, l_1)$, the L long-range neighbourhood $B_{\text{LR}}(u_0, L)$ contains no long-range cycle and thus is a possible realization of $B_{\text{LR}}(O, L)$ around the root of a quasi-tree. Consider a quasi-tree \mathcal{G} which has this ball as the neighbourhood of its root and is completed with the standard procedure (so without taking consideration of the long-range edges revealed in $K \cup B$). Using the coupling of Section 5.2.7, the chain X started at u can thus be coupled with the chain \mathcal{X} on \mathcal{G} , started at the vertex of $B_{\text{LR}}(O, L)$ that identifies with u . This coupling fails after X enters B or if it re-enters K by another path than the one it used. Let τ_{coup} denote this new coupling time.

From (5.80), the probability of sampling an element of either K or B under the uniform measure on V is $O(1/n^{\varepsilon'})$ for some $\varepsilon' > 0$. Consequently, using the same comparison with a binomial we used in the proof of Lemma 5.1, this implies that the chain X re-enters K or reaches B in $O(\sqrt{\log n})$ steps is $O(\log n/n^{2\varepsilon'}) = o(1)$. Note that considering regeneration times requires the knowledge of the L steps ahead but this is $O(\log \log n)$. Consequently it remains true that for $t_2 + L = O(\sqrt{\log n})$,

$$\sum_{v \in F'} \left| \mathbb{P} \left[\exists k \geq 0 : T_k^{(G,L)} = t_2, X_{t_2} = v, U_{t_2}(K) \leq m \mid X_{1/2} = u, \mathcal{F}_{r,l} \right] \right. \\ \left. - \mathbb{P} \left[\exists k \geq 0 : T_k^{(G,L)} = t_2, \iota(\mathcal{X}_{t_2}) = v, U_{t_2}(K) \leq m \mid \mathcal{X}_{1/2} = x, B_{\text{LR}}(O, L) \right] \right| = o(1)$$

where $x \in B_{\text{LR}}(O, L)$ is a vertex at long-range distance L with type u . Obviously, $\{T_k^{(G,\infty)}, k \geq 1\} \subset \{T_k^{(G,L)}, k \geq 1\}$. This inclusion may be strict however, if the chain backtracks over a long-range distance L . Lemma 5.18 shows this occurs before time $O(\sqrt{\log n})$ with probability $o(1)$. Consequently, with high probability $T_k^{(G,L)} = T_k^{(G,\infty)}$ for all regeneration times that occur before t_2 , so we can exchange these random times in the equation above. Now since $t_2 = \Theta(\sqrt{\log n})$ and $m^2 = O((\log \log n)^{2\kappa+2}) = o(t_2)$, Proposition 5.12 proves that

$$\sum_{v \in V} \left| \mathbb{P}_u \left[\exists k \geq 0 : T_k^{(G,\infty)} = t_2, \iota(\mathcal{X}_{t_2}) = v, U_{t_2}(K) \leq m \mid B_{\text{LR}}(O, L) \right] - \frac{\nu(v)}{\mathbb{E}_{\mathbb{Q}_\mu} [T_1^{(G,\infty)}]} \right| \\ \leq \mathbb{P}_u [U_{t_2}(K) > m \mid B_{\text{LR}}(O, L)] / \mathbb{E}_{\mathbb{Q}_\mu} [T_1^{(G,\infty)}] + o(1).$$

By Lemma 5.14, the first term is stretched exponential in m and thus $o(1)$. Using triangle inequality to combine the two previous bounds yields the claim.

□

Bibliography

- [1] Roy L Adler and Benjamin Weiss. Entropy, a complete metric invariant for automorphisms of the torus. *Proceedings of the National Academy of Sciences of the United States of America*, 57(6):1573, 1967. [90](#)
- [2] Elie Aidékon. Large deviations for transient random walks in random environment on a galton–watson tree. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 46(1):159–189, 2010. [doi:doi:10.1214/09-AIHP204](#). [178](#), [180](#)
- [3] Joseph S. Alameda, Caroline Bang, Zachary Brennan, David P. Herzog, Jürgen Kritschgau, and Elizabeth Sprangel. Cutoff in the Bernoulli-Laplace model with $O(n)$ swaps, 2022. [arXiv:2203.08647](#). [22](#), [33](#)
- [4] David Aldous. Random walks on finite groups and rapidly mixing markov chains. pages 243–297, 1983. URL: <https://doi.org/10.1007%2Fbfb0068322>, [doi:10.1007/bfb0068322](#). [21](#), [22](#), [33](#)
- [5] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, 1986. URL: <http://www.jstor.org/stable/2323590>. [21](#), [33](#)
- [6] David Aldous and James Allen Fill. Reversible markov chains and random walks on graphs, 2002. Unfinished monograph, recompiled 2014, available at <http://www.stat.berkeley.edu/~aldous/RWG/book.html>. [13](#), [25](#)
- [7] Noga Alon, Itai Benjamini, Eyal Lubetzky, and Sasha Sodin. Non-backtracking random walks mix faster. *Commun. Contemp. Math.*, 9(4):585–603, 2007. [doi:10.1142/S0219199707002551](#). [18](#), [30](#)
- [8] Claudio Ascì. Generating uniform random vectors. *Journal of Theoretical Probability*, 14(2):333–356, 2001. [19](#), [31](#), [67](#), [70](#), [75](#)
- [9] Claudio Ascì. Generating uniform random vectors in Z_p^k : The general case. *Journal of Theoretical Probability*, 22(3):791–809, 2009. [19](#), [31](#), [67](#), [68](#), [70](#)

- [10] Koenraad MR Audenaert. A sharp continuity estimate for the von Neumann entropy. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8127, 2007. [74](#)
- [11] Luca Avena, Hakan Gldaş, Remco van der Hofstad, and Frank den Hollander. Random walks on dynamic configuration models: A trichotomy. *Stochastic Processes and their Applications*, 129(9):3360–3375, sep 2019. URL: <https://doi.org/10.1016/j.spa.2018.09.010>, doi:10.1016/j.spa.2018.09.010. [119](#)
- [12] Andr Avez. Entropie des groupes de type fini. *C. R. Acad. Sci. Paris, Sr. A-B* 275:A1363–A1366, 1972. [14](#), [15](#), [26](#), [44](#)
- [13] Zsuzsanna Baran, Jonathan Hermon, Anela Œarkovi, and Perla Sousi. Phase transition for random walks on graphs with added weighted random matching. 2023. [arXiv:2306.13077](#). [119](#), [120](#), [121](#)
- [14] Riddhipratim Basu, Jonathan Hermon, and Yuval Peres. Characterization of cutoff for reversible markov chains. *The Annals of Probability*, 45(3):1448–1487, 2017. URL: <http://www.jstor.org/stable/44245575>. [23](#), [34](#)
- [15] Dave Bayer and Persi Diaconis. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, 2:294–313, 1992. doi:doi:10.1214/aoap/1177005705. [22](#), [33](#)
- [16] Anna Ben-Hamou. A threshold for cutoff in two-community random graphs. *The Annals of Applied Probability*, 30(4):1824 – 1846, 2020. doi:10.1214/19-AAP1544. [119](#)
- [17] Anna Ben-Hamou, Eyal Lubetzky, and Yuval Peres. Comparing mixing times on sparse random graphs. *Annales de l’Institut Henri Poincar, Probabilits et Statistiques*, 55(2):1116 – 1130, 2019. doi:10.1214/18-AIHP911. [118](#)
- [18] Anna Ben-Hamou and Yuval Peres. Cutoff for permuted Markov chains. *Annales de l’Institut Henri Poincar, Probabilits et Statistiques*, 59(1):230 – 243, 2023. doi:10.1214/22-AIHP1248. [67](#), [71](#), [119](#)
- [19] Anna Ben-Hamou and Justin Salez. Cutoff for nonbacktracking random walks on sparse random graphs. *The Annals of Probability*, 45(3):1752–1770, 2017. [115](#), [118](#), [120](#), [122](#), [207](#)
- [20] Itai Benjamini, Gady Kozma, and Nicholas Wormald. The mixing time of the giant component of a random graph. *Random Structures & Algorithms*,

- 45(3):383–407, 2014. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20539>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20539>, doi:<https://doi.org/10.1002/rsa.20539>. 115, 118
- [21] Yves Benoist and Frédéric Paulin. Systèmes dynamiques élémentaires. *Cours de Magistère ENS*, 3, 2002. URL: https://www.imo.universite-paris-saclay.fr/~paulin/notescours/cours_sysdyn.pdf. 79, 90
- [22] Nathanaël Berestycki, Eyal Lubetzky, Yuval Peres, and Allan Sly. Random walks on the random graph. *The Annals of Probability*, 46(1):456–490, 2018. doi:[10.1214/17-AOP1189](https://doi.org/10.1214/17-AOP1189). 115, 118, 121, 146, 148, 149
- [23] Kenneth Richard Berg. *On the conjugacy problem for K-systems*. PhD thesis, University of Minnesota, 1967. 90
- [24] Sébastien Blachère, Peter Haïssinsky, and Pierre Mathieu. Asymptotic entropy and Green speed for random walks on countable groups. *Ann. Probab.*, 36(3):1134–1152, 2008. doi:[10.1214/07-AOP356](https://doi.org/10.1214/07-AOP356). 44
- [25] E. Bolthausen. An estimate of the remainder in a combinatorial central limit theorem. *Probability Theory and Related Fields*, 66(3):379–386, 1984. doi:[10.1007/BF00533704](https://doi.org/10.1007/BF00533704). 101
- [26] Erwin Bolthausen and Friedrich Götze. The rate of convergence for multivariate sampling statistics. *Ann. Stat.*, 21(4):1692–1710, 1993. doi:[10.1214/aos/1176349393](https://doi.org/10.1214/aos/1176349393). 101
- [27] Charles Bordenave. A new proof of Friedman’s second eigenvalue theorem and its extension to random lifts. *Ann. Sci. Éc. Norm. Supér. (4)*, 53(6):1393–1439, 2020. doi:[10.24033/asens.2450](https://doi.org/10.24033/asens.2450). 119
- [28] Charles Bordenave, Pietro Caputo, and Justin Salez. Random walk on sparse random digraphs. *Probab. Theory Relat. Fields*, 170(3-4):933–960, 2018. doi:[10.1007/s00440-017-0796-7](https://doi.org/10.1007/s00440-017-0796-7). 115, 119, 120, 122, 200, 207
- [29] Charles Bordenave, Pietro Caputo, and Justin Salez. Cutoff at the “entropic time” for sparse Markov chains. *Probability Theory and Related Fields*, 173(1):261–292, February 2019. doi:[10.1007/s00440-018-0834-0](https://doi.org/10.1007/s00440-018-0834-0). 115, 119, 120, 122, 200
- [30] Charles Bordenave and Bastien Dubail. Markovian Linearization of Random Walks on Groups. *International Mathematics Research Notices*, 05 2022. doi:[10.1093/imrn/rnac045](https://doi.org/10.1093/imrn/rnac045). 13, 25, 37

- [31] Charles Bordenave and Hubert Lacoïn. Cutoff at the entropic time for random walks on covered expander graphs. *J. Inst. Math. Jussieu*, 21(5):1571–1616, 2022. doi:[10.1017/S1474748020000663](https://doi.org/10.1017/S1474748020000663). 47, 119, 172
- [32] Charles Bordenave, Yanqi Qiu, and Yiwei Zhang. Spectral gap of sparse bistochastic matrices with exchangeable rows. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 56(4):2971–2995, 2020. doi:[10.1214/20-AIHP1065](https://doi.org/10.1214/20-AIHP1065). 70
- [33] Stéphane Boucheron, Gábor Lugosi, and Pacal Massart. On concentration of self-bounding functions. *Electron. J. Probab.*, 14:no. 64, 1884–1899, 2009. doi:[10.1214/EJP.v14-690](https://doi.org/10.1214/EJP.v14-690). 104
- [34] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux. doi:[10.1093/acprof:oso/9780199535255.001.0001](https://doi.org/10.1093/acprof:oso/9780199535255.001.0001). 20, 31, 106, 107
- [35] Philippe Bougerol et al. *Products of random matrices with applications to Schrödinger operators*, volume 8. Springer Science & Business Media, 2012. 61, 65
- [36] Rufus Bowen. Markov partitions and minimal sets for Axiom A diffeomorphisms. *Amer. J. Math.*, 92:907–918, 1970. doi:[10.2307/2373402](https://doi.org/10.2307/2373402). 79
- [37] Rufus Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, revised edition, 2008. With a preface by David Ruelle, Edited by Jean-René Chazottes. 79, 90, 93
- [38] Pietro Caputo and Matteo Quattropiani. Mixing time of PageRank surfers on sparse random digraphs. *Random Structures & Algorithms*, 59(3):376–406, apr 2021. URL: <https://doi.org/10.1002/2FrSa.21009>, doi:[10.1002/rsa.21009](https://doi.org/10.1002/rsa.21009). 119
- [39] Pietro Caputo and Matteo Quattropiani. Mixing time trichotomy in regenerating dynamic digraphs. *Stochastic Processes and their Applications*, 137:222–251, jul 2021. URL: <https://doi.org/10.1016/j.spa.2021.03.003>, doi:[10.1016/j.spa.2021.03.003](https://doi.org/10.1016/j.spa.2021.03.003). 119
- [40] René Carmona and Jean Lacroix. *Spectral theory of random Schrödinger operators*. Springer Science & Business Media, 2012. 61, 65

- [41] Sourav Chatterjee. *Concentration inequalities with exchangeable pairs*. PhD thesis, Stanford University, 2005. [arXiv:math/0507526](https://arxiv.org/abs/math/0507526). 20, 32, 97, 101, 104, 106, 108
- [42] Sourav Chatterjee. Concentration of Haar measures, with an application to random matrices. *J. Funct. Anal.*, 245(2):379–389, 2007. [doi:10.1016/j.jfa.2007.01.003](https://doi.org/10.1016/j.jfa.2007.01.003). 20, 32, 101, 104, 108
- [43] Sourav Chatterjee. Stein’s method for concentration inequalities. *Probab. Theory Related Fields*, 138(1-2):305–321, 2007. [doi:10.1007/s00440-006-0029-y](https://doi.org/10.1007/s00440-006-0029-y). 20, 32, 97, 101, 104
- [44] Sourav Chatterjee. A short survey of Stein’s method, 2014. [arXiv:1404.1392](https://arxiv.org/abs/1404.1392). 101
- [45] Sourav Chatterjee and Persi Diaconis. Speeding up Markov chains with deterministic jumps. *Probab. Theory Related Fields*, 181(1-3):377–400, 2021. [doi:10.1007/s00440-021-01049-1](https://doi.org/10.1007/s00440-021-01049-1). 19, 30, 67, 69, 70, 71, 73, 119
- [46] Sourav Chatterjee, Persi Diaconis, and Elizabeth Meckes. Exchangeable pairs and Poisson approximation. *Probability Surveys*, 2(none):64 – 106, 2005. [doi:10.1214/154957805100000096](https://doi.org/10.1214/154957805100000096). 101
- [47] Fang Chen, László Lovász, and Igor Pak. Lifting markov chains to speed up mixing. In *Proceedings of the thirty-first annual ACM symposium on Theory of Computing*. ACM, may 1999. URL: <https://doi.org/10.1145/2F301250.301315>, [doi:10.1145/301250.301315](https://doi.org/10.1145/301250.301315). 18, 30
- [48] Guan-Yu Chen and Laurent Saloff-Coste. The cutoff phenomenon for ergodic markov processes. *Electronic Journal of Probability*, 13(3):26–78, 2008. [doi:10.1214/EJP.v13-474](https://doi.org/10.1214/EJP.v13-474). 22, 33
- [49] F. R. K. Chung, Persi Diaconis, and R. L. Graham. Random walks arising in random number generation. *Ann. Probab.*, 15(3):1148–1165, 1987. URL: <https://www.jstor.org/stable/2244046>. 19, 30, 67, 68, 69, 70, 72, 73, 78, 85, 88
- [50] Guillaume Conchon-Kerjan. Cutoff for random lifts of weighted graphs. *The Annals of Probability*, 50(1):304 – 338, 2022. [doi:10.1214/21-AOP1534](https://doi.org/10.1214/21-AOP1534). 119
- [51] Alain Connes and EJ Woods. Hyperfinite von Neumann algebras and Poisson boundaries of time dependent random walks. *Pacific journal of mathematics*, 137(2):225–243, 1989. 39

- [52] Amir Dembo, Nina Gantert, Yuval Peres, and Ofer Zeitouni. Large deviations for random walks on Galton-Watson trees: averaging and uncertainty. *Probab. Theory Related Fields*, 122(2):241–288, 2002. doi:10.1007/s004400100162. 149, 178, 179
- [53] Yves Derriennic. Marche aléatoire sur le groupe libre et frontière de Martin. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32(4):261–276, December 1975. doi:10.1007/BF00535840. 62
- [54] Yves Derriennic. Quelques applications du théoreme ergodique sous-additif. *Astérisque*, 74:183–201, 1980. 14, 15, 26, 27, 53
- [55] P. Diaconis and L. Saloff-Coste. What do we know about the metropolis algorithm ? *Journal of Computer and System Sciences*, 57(1):20–36, 1998. URL: <https://www.sciencedirect.com/science/article/pii/S0022000098915768>, doi:<https://doi.org/10.1006/jcss.1998.1576>. 18, 30
- [56] Persi Diaconis. *Group representations in probability and statistics*, volume 11 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1988. URL: <https://www.jstor.org/stable/4355560>. 75
- [57] Persi Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. USA*, 93(4):1659–1664, 1996. doi:10.1073/pnas.93.4.1659. 22, 33, 118
- [58] Persi Diaconis. Some things we’ve learned (about Markov chain Monte Carlo). *Bernoulli*, 19(4):1294 – 1305, 2013. doi:10.3150/12-BEJSP09. 18, 30
- [59] Persi Diaconis and Ron Graham. An affine walk on the hypercube. *J. Comput. Appl. Math.*, 41(1-2):215–235, 1992. Asymptotic methods in analysis and combinatorics. doi:10.1016/0377-0427(92)90251-R. 70
- [60] Persi Diaconis, Susan Holmes, and Radford M. Neal. Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.*, 10(3):726–752, 2000. doi:10.1214/aoap/1019487508. 18, 30
- [61] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981. doi:10.1007/BF00535487. 21, 22, 33, 104, 108
- [62] Persi Diaconis and Mehrdad Shahshahani. Time to Reach Stationarity in the Bernoulli–Laplace Diffusion Model. *SIAM Journal on Mathematical Anal-*

- ysis*, 18(1):208–218, 1987. [arXiv:https://doi.org/10.1137/0518016](https://doi.org/10.1137/0518016), [doi:10.1137/0518016](https://doi.org/10.1137/0518016). 22, 33
- [63] Jian Ding, Eyal Lubetzky, and Yuval Peres. Total variation cutoff in birth-and-death chains. *Probability Theory and Related Fields*, 146(1):61, November 2008. [doi:10.1007/s00440-008-0185-3](https://doi.org/10.1007/s00440-008-0185-3). 22, 33
- [64] Jian Ding, Eyal Lubetzky, and Yuval Peres. The Mixing Time Evolution of Glauber Dynamics for the Mean-Field Ising Model. *Communications in Mathematical Physics*, 289(2):725–764, July 2009. [doi:10.1007/s00220-009-0781-9](https://doi.org/10.1007/s00220-009-0781-9). 22, 33
- [65] Bastien Dubail and Laurent Massoulié. Accelerating Abelian Random Walks with Hyperbolic Dynamics. *Probability Theory and Related Fields*, 04 2022. [doi:10.1007/s00440-022-01128-x](https://doi.org/10.1007/s00440-022-01128-x). 13, 25, 67
- [66] Richard Durrett. *Essentials of stochastic processes*. Springer-Verlag Inc, Berlin; New York, 1999. 13, 25
- [67] Ken Dykema. Symmetric random walks on certain amalgamated free product groups. In *Topological and asymptotic aspects of group theory*, volume 394 of *Contemp. Math.*, pages 87–99. Amer. Math. Soc., Providence, RI, 2006. [doi:10.1090/conm/394/07436](https://doi.org/10.1090/conm/394/07436). 46
- [68] Sean Eberhard and Péter P. Varjú. Mixing time of the Chung–Diaconis–Graham random process. *Probability Theory and Related Fields*, 179(1):317–344, February 2021. [doi:10.1007/s00440-020-01009-1](https://doi.org/10.1007/s00440-020-01009-1). 70, 74, 119
- [69] David Ellis, Ehud Friedgut, and Haran Pilpel. Intersecting families of permutations. *J. Amer. Math. Soc.*, 24(3):649–682, 2011. [doi:10.1090/S0894-0347-2011-00690-5](https://doi.org/10.1090/S0894-0347-2011-00690-5). 99
- [70] Alexandros Eskenazis and Evita Nestoridi. Cutoff for the Bernoulli–Laplace urn model with $o(n)$ swaps. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 56(4):2621 – 2639, 2020. [doi:10.1214/20-AIHP1052](https://doi.org/10.1214/20-AIHP1052). 22, 33
- [71] William Feller. *An Introduction to Probability Theory and Its Applications*, Volume I. 1967. 169
- [72] Behrang Forghani. Asymptotic entropy of transformed random walks. *Ergodic Theory and Dynamical Systems*, 37(5):1480–1491, jan 2016. URL: <https://doi.org/10.1017/etds.2015.90>, [doi:10.1017/etds.2015.90](https://doi.org/10.1017/etds.2015.90). 44

- [73] N. Fountoulakis and B.A. Reed. The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Structures & Algorithms*, 33(1):68–86, 2008. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20210>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20210>, doi:<https://doi.org/10.1002/rsa.20210>. 115, 118
- [74] David A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, 3:100–118, 1975. doi:doi:10.1214/aop/1176996452. 210
- [75] Joel Friedman. A proof of alon’s second eigenvalue conjecture and related problems. *Memoirs of the American Mathematical Society*, 195(910):0–0, 2008. URL: <https://doi.org/10.1090%2Fmemo%2F0910>, doi:10.1090/memo/0910. 119
- [76] Harry Furstenberg. Noncommuting random products. *Transactions of the American Mathematical Society*, 108(3):377–428, 1963. URL: <https://doi.org/10.1090%2Fs0002-9947-1963-0163345-0>, doi:10.1090/s0002-9947-1963-0163345-0. 14, 26, 42, 60
- [77] Lorenz Gilch. Asymptotic Entropy of Random Walks on Free Products. *Electronic Journal of Probability*, 16(none):76 – 105, 2011. doi:10.1214/EJP.v16-841. 41, 53, 60
- [78] Lorenz A. Gilch. Rate of escape of random walks on free products. *J. Aust. Math. Soc.*, 83(1):31–54, 2007. doi:10.1017/S1446788700036375. 41, 53, 60
- [79] Lorenz A Gilch and François Ledrappier. Regularity of the drift and entropy of random walks on groups. *arXiv preprint arXiv:1512.03552*, 2015. 46
- [80] Sébastien Gouëzel. Analyticity of the entropy and the escape rate of random walks in hyperbolic groups. *Discrete Anal.*, pages Paper No. 7, 37, 2017. doi:10.19086/da.1639. 46
- [81] Geoffrey Grimmett and Harry Kesten. Random Electrical Networks on Complete Graphs II: Proofs. 08 2001. 149
- [82] Yves Guivarc’h. Sur la loi des grands nombres et le rayon spectral d’une marche aléatoire. In *Journées sur les marches aléatoires*, number 74 in Astérisque. Société mathématique de France, 1980. URL: http://www.numdam.org/item/AST_1980__74__47_0/. 14, 26
- [83] Uffe Haagerup and Steen Thorbjørnsen. A new application of random matrices: $\text{Ext}(C_{\text{red}}^*(F_2))$ is not a group. *Annals of Mathematics*, 162(2):711–

- 775, Sep 2005. URL: <https://doi.org/10.4007%2Fannals.2005.162.711>, doi: [10.4007/annals.2005.162.711](https://doi.org/10.4007/annals.2005.162.711). 15, 27, 38
- [84] Robert H Haring-Smith. Groups and simple languages. *Transactions of the American Mathematical Society*, 279(1):337–356, 1983. 42
- [85] Yair Hartman, Yuri Lima, and Omer Tamuz. An abramov formula for stationary spaces of discrete groups. *Ergodic Theory and Dynamical Systems*, 34(3):837–853, jan 2013. URL: <https://doi.org/10.1017%2Fetds.2012.167>, doi: [10.1017/etds.2012.167](https://doi.org/10.1017/etds.2012.167). 44
- [86] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970. 18, 30
- [87] Jimmy He. Markov chains on finite fields with deterministic jumps. *Electronic Journal of Probability*, 27, Jan 2022. URL: <https://doi.org/10.1214%2F22-ejp757>, doi: [10.1214/22-ejp757](https://doi.org/10.1214/22-ejp757). 67, 71
- [88] Jimmy He, Huy Tuan Pham, and Max Wenqiang Xu. Mixing time of fractional random walk on finite fields. *Electronic Journal of Probability*, 27, Jan 2022. URL: <https://doi.org/10.1214%2F22-ejp858>, doi: [10.1214/22-ejp858](https://doi.org/10.1214/22-ejp858). 67, 71, 72, 74
- [89] J. William Helton, Tobias Mai, and Roland Speicher. Applications of realizations (aka linearizations) to free probability. *J. Funct. Anal.*, 274(1):1–79, 2018. doi: [10.1016/j.jfa.2017.10.003](https://doi.org/10.1016/j.jfa.2017.10.003). 15, 27, 37, 41, 46
- [90] Jonathan Hermon and Sam Olesker-Taylor. Cutoff for almost all random walks on abelian groups, 2021. [arXiv:2102.02809](https://arxiv.org/abs/2102.02809). 119
- [91] Jonathan Hermon and Sam Olesker-Taylor. Cutoff for random walks on upper triangular matrices, 2021. [arXiv:1911.02974](https://arxiv.org/abs/1911.02974). 119
- [92] Jonathan Hermon, Allan Sly, and Perla Sousi. Universality of cutoff for graphs with an added random matching. *The Annals of Probability*, 50(1), Jan 2022. URL: <https://doi.org/10.1214%2F21-aop1532>, doi: [10.1214/21-aop1532](https://doi.org/10.1214/21-aop1532). 23, 34, 113, 115, 118, 119, 120, 121, 125, 127, 146, 148, 176
- [93] Jonathan Hermon, Anđela Šarković, and Perla Sousi. Cutoff for random walk on random graphs with a community structure. 2022. [arXiv:2212.04469](https://arxiv.org/abs/2212.04469). 119, 120, 121

- [94] Martin Hildebrand. Random processes of the form $x_{n+1} = a_n x_n + b_n \pmod{p}$. *Ann. Probab.*, 21(2):710–720, 04 1993. doi:10.1214/aop/1176989264. 69, 70, 71
- [95] Martin Hildebrand. Random processes of the form $X_{n+1} = a_n X_n + b_n \pmod{p}$ where b_n takes on a single value. In *Random discrete structures (Minneapolis, MN, 1993)*, volume 76 of *IMA Vol. Math. Appl.*, pages 153–174. Springer, New York, 1996. doi:10.1007/978-1-4612-0719-1_10. 69, 70, 88
- [96] Martin Hildebrand. On the Chung-Diaconis-Graham random process. *Electron. Comm. Probab.*, 11:347–356, 2006. doi:10.1214/ECP.v11-1237. 70
- [97] Martin Hildebrand. A lower bound for the Chung-Diaconis-Graham random process. *Proc. Amer. Math. Soc.*, 137(4):1479–1487, 2009. doi:10.1090/S0002-9939-08-09687-1. 70
- [98] Martin Hildebrand. On a lower bound for the Chung-Diaconis-Graham random process. *Statist. Probab. Lett.*, 152:121–125, 2019. doi:10.1016/j.spl.2019.04.020. 70
- [99] Martin Hildebrand. A multiplicatively symmetrized version of the Chung-Diaconis-Graham random process. *Journal of Theoretical Probability*, 2021. doi:10.1007/s10959-021-01088-3. 70, 71
- [100] Martin Hildebrand and Joseph McCollum. Generating random vectors in $(\mathbb{Z}/p\mathbb{Z})^d$ via an affine random process. *J. Theoret. Probab.*, 21(4):802–811, 2008. doi:10.1007/s10959-007-0135-5. 19, 31, 67, 68, 70
- [101] Wassily Hoeffding. A combinatorial central limit theorem. *The Annals of Mathematical Statistics*, 22(4):558–566, 1951. doi:doi:10.1214/aoms/1177729545. 100
- [102] V. A. Kaimanovich and A. M. Vershik. Random walks on discrete groups: Boundary and entropy. *The Annals of Probability*, 11:457–490, 1983. doi:doi:10.1214/aop/1176993497. 14, 15, 26, 27, 41, 44, 53, 60, 192
- [103] Vadim Kaimanovich. The poisson boundary of covering markov operators. *Israel Journal of Mathematics*, 89(1-3):77–134, 1995. 39, 43
- [104] Vadim A Kaimanovich. "Münchhausen trick" and amenability of self-similar groups. *International Journal of Algebra and Computation*, 15(05n06):907–937, 2005. 39, 44

- [105] Vadim A. Kaimanovich. Self-similarity and random walks. pages 45–70, 2009. URL: https://doi.org/10.1007/2F978-3-0346-0030-9_2, doi:10.1007/978-3-0346-0030-9_2. 39, 44
- [106] Kseniya Klyachko. *Random Processes of the Form $X_N + 1 = AX_N + B_N \pmod{p}$* . PhD thesis, State University of New York at Albany, 2020. 19, 31, 67, 70, 91
- [107] András Krámli and Domokos Szász. Random walks with internal degrees of freedom. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 63(1):85–95, 1983. 39
- [108] Hubert Lacoin. Mixing time and cutoff for one dimensional particle systems, 2021. [arXiv:2111.06436](https://arxiv.org/abs/2111.06436). 22, 33
- [109] Steven P Lalley. Finite range random walk on free groups and homogeneous trees. *The Annals of Probability*, pages 2087–2130, 1993. 46, 62, 64
- [110] Serge Lang. *Algebra*. Springer New York, 2002. 68, 73, 81
- [111] Michel Ledoux. *The concentration of measure phenomenon*. American Mathematical Society, Providence, 2001. 20, 31
- [112] François Ledrappier. Analyticity of the entropy for some random walks. *Groups Geom. Dyn.*, 6(2):317–333, 2012. doi:10.4171/GGD/159. 46
- [113] François Ledrappier. Regularity of the entropy for random walks on hyperbolic groups. *Ann. Probab.*, 41(5):3582–3605, 2013. doi:10.1214/12-AOP748. 46
- [114] François Ledrappier. Some asymptotic properties of random walks on free groups. *Topics in probability and Lie groups: boundary theory*, 28:117–152, 2001. 15, 27, 41, 46, 53, 55, 56, 59, 60, 61
- [115] Franz Lehner. Computing norms of free operators with matrix coefficients. *American Journal of Mathematics*, 121(3):453–486, 1999. 46, 60
- [116] David A. Levin, Malwina J. Luczak, and Yuval Peres. Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. *Probability Theory and Related Fields*, 146(1):223, December 2008. doi:10.1007/s00440-008-0189-z. 22, 33
- [117] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on “Coupling from

- the past” by James G. Propp and David B. Wilson. [doi:10.1090/mbk/107](https://doi.org/10.1090/mbk/107). 13, 17, 18, 23, 25, 29, 30, 34, 172
- [118] Elon Lindenstrauss and Péter P. Varjú. Spectral gap in the group of affine transformations over prime fields. *Ann. Fac. Sci. Toulouse Math. (6)*, 25(5):969–993, 2016. [doi:10.5802/afst.1518](https://doi.org/10.5802/afst.1518). 72
- [119] Torgny Lindvall. *Lectures on the coupling method*. Dover Publications, Inc., Mineola, NY, 2002. 167
- [120] Torgny Lindvall and L. C. G. Rogers. On coupling of random walks and renewal processes. *Journal of Applied Probability*, 33:122–126, 1996. 167
- [121] Eyal Lubetzky and Yuval Peres. Cutoff on all Ramanujan graphs. *Geometric and Functional Analysis*, 26(4):1190–1216, July 2016. [doi:10.1007/s00039-016-0382-7](https://doi.org/10.1007/s00039-016-0382-7). 119
- [122] Eyal Lubetzky and Allan Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 153(3):475 – 510, 2010. [doi:10.1215/00127094-2010-029](https://doi.org/10.1215/00127094-2010-029). 23, 34, 118
- [123] Eyal Lubetzky and Allan Sly. Cutoff for general spin systems with arbitrary boundary conditions. *Communications on Pure and Applied Mathematics*, 67(6):982–1027, 2014. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21489>, [arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.21489](https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.21489), [doi:https://doi.org/10.1002/cpa.21489](https://doi.org/10.1002/cpa.21489). 22, 33
- [124] Eyal Lubetzky and Allan Sly. Universality of cutoff for the Ising model. *The Annals of Probability*, 45(6A):3664 – 3696, 2017. [doi:10.1214/16-AOP1146](https://doi.org/10.1214/16-AOP1146). 22, 33
- [125] Alexander Lubotzky and Igor Pak. The product replacement algorithm and Kazhdan’s property (T). *J. Amer. Math. Soc.*, 14(2):347–363, 2001. [doi:10.1090/S0894-0347-00-00356-8](https://doi.org/10.1090/S0894-0347-00-00356-8). 73
- [126] Russell Lyons, Robin Pemantle, and Yuval Peres. Ergodic theory on Galton–Watson trees: speed of random walk and dimension of harmonic measure. *Ergodic Theory and Dynamical Systems*, 15(3):593–619, 1995. [doi:10.1017/S0143385700008543](https://doi.org/10.1017/S0143385700008543). 192
- [127] Russell Lyons, Robin Pemantle, and Yuval Peres. Biased random walks on Galton–Watson trees. *Probab. Theory Relat. Fields*, 106(2):249–264, 1996. [doi:10.1007/s004400050064](https://doi.org/10.1007/s004400050064). 175

- [128] Russell Lyons and Yuval Peres. *Probability on trees and networks*, volume 42 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016. doi:10.1017/9781316672815. 13, 25, 53, 116, 151, 152
- [129] Jean Mairesse. Random walks on groups and monoids with a Markovian harmonic measure. *Electron. J. Probab.*, 10:1417–1441, 2005. Id/No 43. 41, 42, 43, 46, 54, 56, 60
- [130] Jean Mairesse and Frédéric Mathéus. Random walks on free products of cyclic groups. *J. Lond. Math. Soc., II. Ser.*, 75(1):47–66, 2007. doi:10.1112/jlms/jdl006. 41, 53, 56, 60, 61
- [131] Fabio Martinelli. *Relaxation Times of Markov Chains in Statistical Mechanics and Combinatorial Structures*, pages 175–262. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. doi:10.1007/978-3-662-09444-0_4. 18, 29
- [132] P. Mathieu. Differentiating the entropy of random walks on hyperbolic groups. *Ann. Probab.*, 43(1):166–187, 2015. doi:10.1214/13-AOP901. 46
- [133] Bernard Maurey. Construction de suites symétriques. *CR Acad. Sci. Paris Sér. AB*, 288(14):A679–A681, 1979. 97
- [134] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, jun 1953. URL: <https://doi.org/10.1063%2F1.1699114>, doi:10.1063/1.1699114. 18, 30
- [135] James A. Mingo and Roland Speicher. *Free probability and random matrices*, volume 35 of *Fields Institute Monographs*. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2017. doi:10.1007/978-1-4939-6942-5. 38
- [136] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006. URL: <http://dx.doi.org/10.1561/0400000003>, doi:10.1561/0400000003. 17, 29
- [137] Arunava Mukherjea and Nicolas A. Tserpes. *Measures on topological semigroups: convolution products and random walks*. Lecture Notes in Mathematics, Vol. 547. Springer-Verlag, Berlin-New York, 1976. 86

- [138] Tatiana Nagnibeda and Wolfgang Woess. Random walks on trees with finitely many cone types. *J. Theoret. Probab.*, 15(2):383–422, 2002. doi:10.1023/A:1014810827031. 60
- [139] Radford M. Neal. *Bayesian learning for neural networks*. Springer-Verlag Inc, Berlin; New York, 1996. 18, 29
- [140] Radford M. Neal. MCMC using Hamiltonian dynamics. In *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC Handb. Mod. Stat. Methods, pages 113–162. CRC Press, Boca Raton, FL, 2011. 18, 30
- [141] Richard Neville III. *On lower bounds of the Chung-Diaconis-Graham random process*. PhD thesis, State University of New York at Albany, 2011. 70
- [142] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original. 13, 25
- [143] Narutaka Ozawa et al. An entropic proof of cutoff on ramanujan graphs. *Electronic Communications in Probability*, 25, 2020. 119
- [144] Igor Pak. What do we know about the product replacement algorithm? In *Groups and computation, III (Columbus, OH, 1999)*, volume 8 of *Ohio State Univ. Math. Res. Inst. Publ.*, pages 301–347. de Gruyter, Berlin, 2001. 73
- [145] Daniel Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.*, 20:no. 79, 32, 2015. doi:10.1214/EJP.v20-4039. 172, 173
- [146] Patrick Rebeschini and Sekhar C Tatikonda. Accelerated consensus via min-sum splitting. *Advances in Neural Information Processing Systems*, 30, 2017. 18, 30
- [147] Christian Robert and George Casella. A short history of Markov Chain Monte Carlo: Subjective recollections from incomplete data. *Statistical Science*, 26(1):102–115, 2011. doi:doi:10.1214/10-STS351. 18, 29
- [148] Christian P. Robert and George Casella. *Monte Carlo statistical methods*. Springer-Verlag Inc, Berlin; New York, 1999. 18, 29
- [149] Justin Salez. Cutoff for non-negatively curved markov chains. *Journal of the European Mathematical Society*, May 2023. URL: <https://doi.org/10.4171/2Fjems%2F1348>, doi:10.4171/jems/1348. 23, 34, 118, 119

- [150] Laurent Saloff-Coste. Random walks on finite groups. In *Probability on Discrete Structures*, pages 263–346. Springer Berlin Heidelberg, 2004. URL: https://doi.org/10.1007%2F978-3-662-09444-0_5, doi:10.1007/978-3-662-09444-0_5. 22, 33, 118
- [151] Dominik Schmid. *Mixing times and limit theorems for exclusion processes*. PhD thesis, Technische Universität München, 2021. 22, 33
- [152] Eugene Seneta. *Non-negative matrices and Markov chains*. Springer Science & Business Media, 2006. 61, 62
- [153] Nikita Sidorov. Arithmetic dynamics. In *Topics in dynamics and ergodic theory*, volume 310 of *London Math. Soc. Lecture Note Ser.*, pages 145–189. Cambridge Univ. Press, Cambridge, 2003. doi:10.1017/CB09780511546716.010. 90
- [154] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Berkeley Symposium on Mathematical Statistics and Probability*, 1972, 1972. URL: <http://projecteuclid.org/euclid.bsmsp/1200514239>. 101
- [155] Charles Stein. Approximate computation of expectations. *Lecture Notes-Monograph Series*, 7:i–164, 1986. URL: <http://www.jstor.org/stable/4355512>. 101
- [156] Michel Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications mathématiques de l’IHÉS*, 81(1):73–205, Dec 1995. URL: <https://doi.org/10.1007%2Fbf02699376>, doi:10.1007/bf02699376. 97
- [157] Terence Tao. *Expansion in finite simple groups of Lie type*, volume 164 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2015. doi:10.1090/gsm/164. 72
- [158] A. M. Vershik. Arithmetic isomorphism of hyperbolic automorphisms of a torus and of sofic shifts. *Funktsional. Anal. i Prilozhen.*, 26(3):22–27, 1992. doi:10.1007/BF01075629. 90
- [159] Wolfgang Woess. *Random walks on infinite graphs and groups*, volume 138 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2000. doi:10.1017/CB09780511470967. 42, 154

RÉSUMÉ

Cette thèse aborde trois sujets principaux liés aux marches aléatoires et aux chaînes de Markov. Le premier concerne l'introduction d'une technique appelée "linéarisation" pour simplifier l'étude des marches aléatoires dans des groupes discrets, en les réduisant à des marches aux plus proches voisins dans un espace d'états élargi. Ensuite, elle étudie un mécanisme d'accélération des dynamiques markoviennes par l'intégration d'étapes de dynamique déterministe, où l'effet accélérateur est attribué au caractère chaotique de cette dynamique. Enfin, elle explore le phénomène de cutoff en démontrant un phénomène de cutoff au temps entropique pour un modèle général de chaînes de Markov en environnement aléatoire. Un nouveau résultat de concentration de la mesure pour le groupe symétrique est développé à cet effet.

MOTS CLÉS

Temps de mélange, chaîne de Markov, cutoff, environnement aléatoire, accélération, linéarisation, marche aléatoire

ABSTRACT

This thesis covers three main topics related to random walks and Markov chains. The first involves the introduction of a technique called "linearization" to simplify the study of random walks in discrete groups, by reducing them to nearest-neighbor walks in an expanded state space. Next, it investigates a mechanism to accelerate Markovian dynamics by incorporating deterministic steps, where the acceleration effect is attributed to the chaotic nature of this dynamic. Finally, it explores the "cutoff" phenomenon, demonstrating a cutoff phenomenon at the entropic time for a general model of Markov chains in random environment. A new measure concentration result for the symmetric group is developed for this purpose.

KEYWORDS

Mixing times, Markov chains, cutoff, random environment, acceleration, linearization, random walks

