HAL
open science

Diffusion décentralisée d'information dans les systèmes distribués

Hervé Baumann

To cite this version:

Hervé Baumann. Diffusion décentralisée d'information dans les systèmes distribués. Réseaux et télécommunications [cs.NI]. Paris Diderot University, 2012. Français. NNT : . tel-01274130

HAL Id: tel-01274130

https://inria.hal.science/tel-01274130v1

Submitted on 15 Feb 2016

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.

Université Paris Diderot – Paris 7
Laboratoire d'Informatique Algorithmique : Fondements et Applications

Diffusion décentralisée d'information dans les systèmes distribués

Thèse présentée pour l'obtention du diplôme de

Docteur de l'Université Paris Diderot, spécialité Informatique

à l'École Doctorale de Sciences Mathématiques de Paris Centre

Par

Hervé Baumann

Thèse dirigée par Pierre Fraigniaud

Soutenue publiquement
le 24 septembre 2012
devant le jury constitué de :

M. Andrea Clementi	Examinateur
M. Philippe Duchon	Rapporteur
M. Pierre Fraigniaud	Directeur
Mme Anne-Marie Kermarrec	Examinatrice
M. Vlady Ravelomanana	Examinateur
M. Yann Vaxès	Rapporteur

Résumé

Cette thèse porte sur la diffusion d'informations dans les réseaux. La diffusion consiste à transmettre une information à partir d'un nœud source vers tous les autres nœuds. Elle est très présente en informatique, aussi bien pour transmettre du contenu par internet que pour le calcul distribué ou parallèle. Dans les protocoles que nous considérons, chaque nœud informé participe à la diffusion d'information : dans l'inondation, un nœud informé transmet l'information à tous ses voisins en une seule étape de temps, et dans la diffusion 1-Port chaque nœud informé communique, à chaque étape, avec un unique voisin.

Notre première contribution porte sur l'étude du pire temps de diffusion 1-Port où un nœud n'appelle pas un voisin qu'il sait être informé. Nous étudions trois variantes. Dans la première, un nœud ne rappelle simplement pas un voisin auquel il a déjà envoyé l'information ; dans la deuxième, les voisins qu'un nœud sait être informés sont tous ceux avec lesquels il a déjà échangé de l'information (dans un sens ou dans l'autre) ; dans la troisième, chaque nœud sait exactement quels sont ses voisins informés. Nous montrons principalement que dans la première variante, il est possible de calculer en temps polynomial le pire temps de diffusion et un pire schéma correspondant. Pour les deux autres variantes, nous montrons que ce calcul est NP-difficile, en fait difficilement approximable.

Notre deuxième contribution concerne l'étude de la diffusion dans des réseaux dynamiques modélisés par des graphes à évolution arête-markovienne. Nous proposons une technique de réduction permettant de ramener le calcul du temps d'inondation au calcul du diamètre d'un graphe pondéré statique tiré aléatoirement. Cette technique nous permet de calculer le temps de diffusion par inondation dans les graphes dynamiques arête-markoviens. Nous avons également calculé le temps de diffusion probabiliste 1-Port dans certains graphes à évolution arête-markovienne.

Notre dernière contribution porte sur la diffusion spatiale où tous les nœuds peuvent communiquer les uns avec les autres. Les performances des protocoles de diffusion spatiale sont mesurées en fonction de l'éloignement des destinataires par rapport à la source. Nous décrivons un protocole de diffusion spatiale probabiliste généralise, c'est-à-dire, valable pour toute métrique. Nous analysons sa vitesse de propagation dans les métriques de graphes et dans les métriques à densité uniforme.

Remerciements

Table des matières

1 Introduction **1**
1.1 Contexte . . . 1
1.2 Structure et contenu de la thèse . . . 4

2 Diffusion désordonnée **7**
2.1 Bref de l'art de la diffusion 1-Port . . . 8
2.1.1 Recherche de schémas de diffusion optimaux . . . 8
2.1.2 Le protocole de diffusion probabiliste . . . 10
2.1.3 Le protocole de diffusion quasi-aléatoire . . . 12
2.2 Introduction à la diffusion désordonnée . . . 13
2.2.1 Les modèles désordonnés . . . 15
2.2.2 État de l'art . . . 17
2.2.3 Nos résultats . . . 18
2.3 Résultats préliminaires . . . 19
2.3.1 Caractérisation des pires graphes . . . 19
2.3.2 Écarts exponentiels entre les modèles . . . 20
2.3.3 Les diodes . . . 22
2.4 Un algorithme polynomial pour SkipNone . . . 24
2.5 Inapproximabilité de SkipSenders et SkipInformed . . . 27
2.6 Extensions . . . 32
2.6.1 Le cas des graphes orientés . . . 32
2.6.2 Boîte à outils . . . 33
2.6.3 Liens entre les modèles . . . 37
2.6.4 Variations sur SkipInformed . . . 38
2.7 Conclusion . . . 39

3 Diffusion dans les graphes évolutifs **43**
3.1 Introduction . . . 44
3.1.1 État de l'art sur les graphes évolutifs . . . 44
3.1.2 Inondation dynamique . . . 47
3.1.3 Résultats antérieurs . . . 49
3.1.4 Modèle et notations . . . 50

3.1.5 Bref aperçu de nos contributions à la diffusion dans les réseaux arêtes-markoviens . . . 52
3.2 Méthodologie générale, cas stationnaire . . . 54
3.2.1 Réduction du temps d'inondation au diamètre . . . 54
3.2.2 Trois scénarios particuliers . . . 62
3.3 Application aux graphes à évolution homogène . . . 64
3.3.1 Au-dessus du seuil de connectivité . . . 67
3.3.2 En dessous du seuil de connectivité . . . 68
3.4 Méthodologie générale, cas arbitraire . . . 71
3.4.1 Majoration du temps d'inondation par un diamètre . . . 72
3.4.2 Minoration du temps d'inondation . . . 74
3.5 Applications . . . 75
3.5.1 Application aux graphes à évolution arête-markovienne homogène 76
3.5.2 Application aux graphes évolutifs avec suite de degré arbitraire 77
3.5.3 Application aux graphes évolutifs en loi de puissance . . . 82
3.6 Diffusion 1-Port aléatoire dans $\mathcal{M}_{n,p,q,g}$ . . . 84
3.6.1 Graphes orientés . . . 84
3.6.2 Graphes non orientés . . . 87
3.7 Conclusion . . . 89
3.7.1 Inondation . . . 89
3.7.2 Diffusion 1-Port . . . 91

4 Diffusion spatiale 93
4.1 Introduction . . . 93
4.1.1 Définition et état de l'art . . . 93
4.1.2 Résumé des résultats présentés dans ce chapitre . . . 96
4.2 Protocole Random revisité . . . 97
4.3 Diffusion spatiale et routage glouton . . . 98
4.3.1 Le routage . . . 98
4.3.2 Les graphes augmentés . . . 99
4.3.3 Lien entre routage glouton et diffusion spatiale . . . 102
4.4 Le protocole de diffusion Rankscale . . . 104
4.4.1 Temps de propagation de Rankscale, première approche . . 105
4.4.2 Temps de propagation de Rankscale, seconde approche . . 108
4.4.3 Cas des métriques à densité uniforme . . . 112
4.5 Conclusion . . . 115

5 Conclusion et perspectives 117
5.1 Conclusion . . . 117
5.2 Perspectives . . . 119
5.2.1 Autres mesures . . . 119
5.2.2 Autres tâches . . . 120
5.2.3 Autres contraintes . . . 121

Chapitre 1

Introduction

1.1 Contexte

Un réseau est constitué de nœuds, représentant par exemple des personnes ou des machines, capables de communiquer entre eux. Les réseaux sont représentés par des graphes dont les sommets sont les nœuds et les arêtes représentent les liens de communication. Que le réseau soit social ou informatique, la communication, c'est-à-dire la transmission d'information, est un élément clé de la structure et de l'organisation du réseau.

Le processus de communication considéré dans cette thèse est la ***diffusion***. La diffusion dans un réseau consiste à transmettre une information à partir d'un nœud source vers tous les autres nœuds. Elle est très présente en informatique, aussi bien pour transmettre du contenu par internet que pour le calcul distribué ou parallèle. En effet, le processus de diffusion a une longue histoire initiée dans les années 60 par des études visant à comprendre la structure des réseaux permettant la diffusion rapide d'information (voir [HHL88] pour un recueil). Motivées par l'émergence d'architectures multi-processeurs étroitement couplées pour les ordinateurs parallèles dans les années 80, beaucoup d'études ont abordé le problème de diffusion dans des familles de réseau spécifiques, y compris les hypercubes et graphes en papillon (voir les recueils [FL94, HKMP96]).

La diffusion est proche d'autres processus que nous mentionnons dans la thèse : la diffusion partielle, le routage et l'échange total. Lorsque la transmission doit être réalisée seulement vers un sous-ensemble du réseau, on parle de ***diffusion partielle***. Le ***routage*** est le cas particulier qui consiste à transmettre une information à partir d'un nœud source vers un unique nœud cible. Dans l'***échange total***, chaque nœud possède une information unique qui doit être diffusée à l'ensemble des nœuds du réseau.

Dans toute cette thèse, et dans tous les modèles considérés, les communications s'effectuent par étapes synchrones. Ainsi, le temps est exprimé en nombre d'étapes. Nous nous intéressons à la performance des protocoles en termes de vitesse de diffusion, pour un modèle de communication donné et, pour un protocole respectant

ce modèle, nous cherchons à déterminer le ***temps de diffusion*** de ce protocole, c'est-à-dire le nombre d'étapes nécessaires pour informer tous les nœuds du réseau.

Nous considérons principalement deux modèles de communication. Lorsque chaque nœud informe tous ses voisins en une seule étape, on parle d'***inondation***. L'inondation est très efficace en termes de temps puisque le nombre d'étapes nécessaires à une diffusion est le diamètre du graphe. En revanche, elle est très coûteuse en ressources. L'autre modèle considéré est le modèle 1-Port. Dans ce modèle, à chaque étape, chaque nœud n'est capable d'appeler qu'un seul de ses voisins. Nous étudions en particulier une restriction du modèle 1-Port, dite Push. Dans le modèle Push, l'information ne peut se transmettre que du nœud appelant vers le nœud appelé. Dans la littérature on rencontre aussi le modèle Pull où l'information ne peut se transmettre que du nœud appelé vers le nœud appelant. Remarquons qu'un même nœud peut-être appelé par plusieurs nœuds à la même étape. Dans le modèle Pull, le modèle indique que le nœud appelé v transmet l'information en une seule étape à tous les nœuds u qui l'ont appelés. Il est intéressant de remarquer que malgré cette différence, les modèles Push et Pull sont parfaitement symétriques. En effet, dans [Gia11], il est montré que d'un schéma de diffusion d'un protocole Pull tel qu'un message de source u arrive en v en T étapes on peut déduire un schéma de diffusion d'un protocole Push tel qu'un message de source v arrive en u en T étapes, et inversement. De plus dans le cas de protocoles probabilistes, pour deux nœuds u et v, le temps mis par un message de source u pour informer v dans le cas Push suit la même loi de distribution que le temps mis par un message de source v pour informer u dans le cas Pull. Dans la suite de cette thèse, sauf mention contraire, lorsque nous parlons d'un protocole de diffusion, il est de type Push.

Nous rencontrons aussi le modèle PushPull où l'information peut être transmise dans les deux sens. Ce modèle a été introduit par [KSSV00]. Il est montré que dans ce modèle, sur la clique, il est possible d'envoyer beaucoup moins de messages en préservant un temps de diffusion asymptotiquement optimal. Dans d'autres études telles que [Gia11], il est montré que, sur d'autres graphes, il est possible de concevoir des protocoles respectant le modèle PushPull strictement plus rapides que dans les modèles Push ou Pull. Néanmoins, cette thèse reste focalisée sur le modèle Push.

Dans les protocoles que nous considérons dans cette thèse, les nœuds transmettent systématiquement toute l'information qu'ils possèdent. Ainsi, il existe un unique protocole de diffusion dans le modèle de l'inondation : un nœud qui a l'information la transmet à tous ses voisins. Dans un modèle 1-Port, dans tous les protocoles de diffusion, à chaque étape, chaque nœud u choisit un correspondant v dans son voisinage et l'appelle. Plus formellement, l'algorithme générique suivant est exécuté indéfiniment par tous les nœuds, en parallèle, dans le but de diffuser de l'information :

(1) le nœud u choisit un voisin v ;

(2) u transmet l'information à v.

Les protocoles de diffusion 1-Port diffèrent en (1) par la façon dont chaque nœud sélectionne le destinataire de sa prochaine communication. En revanche, comme nous considérons que les nœuds transmettent toute l'information qu'ils possèdent, l'étape

(2) est imposée par le modèle. Dans le modèle PUSH, seul le nœud appelant peut transmettre de l'information. Nous pouvons donc considérer que seuls les nœuds détenteurs de l'information sont à l'initiative d'un appel.

Nous considérons plusieurs protocoles de diffusion 1-PORT. Pour un protocole donné, un ***schéma de diffusion*** est un ordonnancement des appels effectués par le protocole. Dans le cas d'un protocole déterministe, le schéma est déterministe et il est le résultat des choix des envois effectués par chaque nœud. Dans le cas d'un protocole probabiliste, un schéma de diffusion est une instance particulière correspondant à une exécution effective du protocole.

Les protocoles déterministes que nous considérons dans cette thèse sont des protocoles LISTBASED. Dans un protocole LISTBASED, chaque nœud u possède une liste (cyclique) de ses voisins, et le correspondant v de u est sélectionné dans l'ordre de la liste. En substance, un protocole de diffusion LISTBASED s'effectue comme suit. Pour tout nœud $u \in V$ possédant une information, à chaque étape, et simultanément :

(1) u sélectionne le voisin v suivant dans sa liste ;

(2) u transmet l'information à v.

Les listes pourront être construites de manière centralisée et calculées avant le début du processus, pour rendre la diffusion optimale. Lorsque les listes sont fournies par un adversaire qui cherche à rendre la diffusion la plus lente possible (sous certaines contraintes), nous parlerons alors de diffusion désordonnée (*messy broadcasting*) selon la terminologie de la littérature [AHK94].

Le protocole aléatoire considéré dans cette thèse est le protocole RANDOM [FPRU90]. Dans le protocole RANDOM, le correspondant v de u est sélectionné aléatoirement uniformément parmi les voisins de u. Le protocole de diffusion RANDOM s'effectue comme suit. Pour tout nœud $u \in V$ possédant une information, à chaque étape et simultanément :

(1) u sélectionne aléatoirement un voisin v,

(2) u transmet l'information à v.

Le protocole RANDOM ne nécessite que très peu d'information par nœud, et le choix de l'interlocuteur peut se faire de manière décentralisée. Le fait que les nœuds n'aient pas besoin de mémoire rend le protocole très peu sensible aux événements passés. En particulier, un nœud qui se rétablit après un accident ou une panne passagère peut redémarrer l'exécution du protocole à partir de zéro, même si toutes les informations locales ont été perdues. En plus de la robustesse, un autre attrait du protocole probabiliste est sa simplicité : il y a très peu de calcul à effectuer. En particulier, un nœud n'a pas besoin de calculer un schéma de diffusion complexe, ni même de tenir compte d'éventuelles informations présentes dans l'entête des messages.

En fait, des protocoles pratiques de diffusion récemment proposés [GKG06] sont fondés sur des mécanismes aléatoires, principalement parce que l'aléa préserve également le mécanisme d'éventuels changements de l'environnement. Aussi, des protocoles du type RANDOM ont récemment trouvé plusieurs applications dans diverses tâches

système ou réseau, telles que la diffusion partielle [BGPS06, GKG06, LEH03], la diffusion en flux continu [FGKM10], la localisation de ressources [KKD04, KK02], la gestion de bases de données distribuées [DGH+88], l'agrégation de données [KDG03], et les systèmes *publish-subscribe* [CMPC04].

Nous pouvons par contre remarquer que, dans un protocole probabiliste, chaque nœud doit être capable de tirer aléatoirement un de ses voisins, à chaque étape. Il existe toutefois un protocole qui permet de diminuer l'aléa nécessaire. Dans le protocole QuasiRandom, introduit dans [DFS08], un nœud choisit le voisin qu'il va appeler dans l'ordre défini par une liste de son voisinage, comme dans le modèle ListBased. En revanche, au lieu de commencer par informer le premier voisin de la liste, il commence à une position tirée aléatoirement. Si les ordres des listes sont donnés par un adversaire, cette modification permet de diminuer le potentiel de nuisance de cet adversaire. En fait, ce protocole se comporte aussi bien, voire mieux que le protocole Random dans de nombreuses classes de graphes. Si, au contraire, l'ordre des listes cycliques est aussi donné aléatoirement, le protocole QuasiRandom revient à rajouter de la mémoire au protocole Random. En effet, la variante du protocole Random où un nœud ne rappelle un voisin que s'il a déjà appelé tous les autres se comporte comme un protocole QuasiRandom dont les listes sont données dans un ordre aléatoire.

1.2 Structure et contenu de la thèse

Après un bref état de l'art de la diffusion 1-Port, dans le chapitre 2, nous nous intéressons en premier lieu au pire cas de la diffusion 1-Port probabiliste du protocole ListBased, appelé diffusion *désordonnée*. Nous étudions la diffusion désordonnée selon trois hypothèses sur l'information possédée par chaque nœud à propos de leurs voisins, tels qu'introduits dans [HL98] :

- SkipNone est la version basique où les nœuds ne disposent d'aucune information supplémentaire à propos de leurs voisins. Ainsi, chaque nœud informé suit le protocole ListBased tel que décrit précédemment, et transmet l'information à tous ses voisins dans l'ordre défini par la liste.
- Dans la variante SkipSenders, un nœud retient quels nœuds lui ont envoyé l'information. Ainsi, dans l'exécution du protocole, il passe les nœuds desquels il a reçu l'information, et transmet cette information au prochain nœud de la liste avec lequel il n'a encore rien échangé.
- La variante SkipInformed suppose l'existence d'un mécanisme sous-jacent permettant à chaque nœud de savoir en permanence lesquels de ses voisins ont reçu l'information. Ainsi, dans l'exécution du protocole, il passe la transmission aux nœuds informés pour transmettre l'information au prochain nœud non informé de la liste.

Dans un graphe modélisant un réseau, un pire cas de diffusion ListBased, pour l'une des trois variantes, est donné par l'ensemble des listes ordonnées de voisinage. Connaître le pire temps de diffusion permet, d'une part, de connaître une majoration

du temps de diffusion de n'importe quel protocole probabiliste utilisant les mêmes informations. D'autre part, il permet de connaître la capacité de nuisance d'un adversaire susceptible de choisir ou modifier les listes. Dans un premier temps, nous montrons qu'il existe des différences très importantes en termes de temps de diffusion entre les trois variantes. Nous nous intéressons d'abord à l'existence de *diodes*, c'est-à-dire de gadgets pour lesquels l'information circule rapidement dans un sens et lentement dans l'autre. Nous donnons des diodes pour les cas SkipNone et SkipInformed, et montrons qu'il n'est pas possible d'en trouver dans le cas SkipSenders. Nous nous intéressons ensuite aux graphes ayant les pires temps de diffusion. Nous avons montré qu'il n'est pas possible de caractériser simplement les pires graphes pour SkipNone, et nous avons caractérisé complètement ceux pour SkipInformed. Pour SkipSenders, ce problème de caractérisation demeure ouvert. En fait, le temps maximal, sur l'ensemble des graphes, de la diffusion désordonnée n'est pas connu dans la variante SkipSenders. Nos résultats principaux du chapitre 2 concernent la complexité du calcul du pire temps de diffusion désordonnée. Nous montrons qu'il est possible de calculer en temps polynomial le pire temps de diffusion d'un graphe dans le cas SkipNone. Dans les deux autres cas, ce calcul est NP-difficile. Pour SkipSenders, ce pire temps est $(\frac{1}{2}+\epsilon)$-inapproximable, pour tout $\epsilon > 0$, même dans le cas de graphes bipartis. Dans le cas SkipInformed, le pire temps est $((\frac{1}{n})^{\frac{1}{2}-\epsilon})$-inapproximable, pour tout $\epsilon > 0$. Ce dernier résultat d'inapproximabilité utilise l'existence de diodes. Nous complétons finalement le chapitre par la mise en évidence de résultats pouvant être utiles pour l'étude de variantes du problème de diffusion désordonnée ou pour une étude approfondie de ce problème.

Dans le chapitre 3, nous étudions la diffusion dans un réseau *dynamique* modélisé par un graphe à *évolution arête-markovienne*. Le modèle de graphes à évolution arête-markovienne stipule que les arêtes suivent un processus markovien de vie et de mort : une arête e absente du graphe au temps t apparaît au temps $t+1$ avec probabilité p_e, appelée taux de naissance, et une arête e présente disparaît avec probabilité q_e, appelée taux de mort, indépendamment des autres arêtes. L'évolution est dite homogène si les taux de naissance et de mort sont les mêmes pour toutes les arêtes, respectivement notés p et q. Nous parlons de cas stationnaire lorsque le graphe initial est tiré selon la distribution stationnaire du processus markovien. Dans un premier temps, nous considérons l'inondation dans ces réseaux. Nous avons donné une technique générale de réduction permettant, dans le cas stationnaire, de ramener le temps d'inondation au calcul du diamètre d'un graphe pondéré aléatoire mais statique. Le résultat principale du chapitre 3, outre la technique de réduction, est son application aux réseaux à évolution arête-markovienne homogène permettant d'améliorer des résultats de [CMPS11]. Soit n le nombre de nœuds et $\widehat{p} = \frac{p}{p+q}$. Nous montrons alors que le temps d'inondation est $\Theta(\frac{\ln n}{np})$ si $\widehat{p}$ est inférieur à $\frac{1}{n}$, $\Theta(\frac{\ln n}{np} + \frac{\ln n}{\ln(n\widehat{p})})$ si $\widehat{p}$ est entre $\frac{1}{n}$ et $\frac{\ln n}{n}$, et $\Theta(\frac{\ln n}{\ln(n\widehat{p})})$ si $\widehat{p}$ est au-dessus de $\frac{\ln n}{n}$. Nous montrons également qu'il est possible, dans certains cas, de diminuer de manière importante le nombre d'étapes consécutives où les nœuds vont transmettre l'information, sans pour autant augmenter le temps d'inondation de plus d'une constante multiplicative. Nous donnons ensuite

une technique permettant de majorer et de minorer le temps d'inondation pour toute distribution initiale, en se ramenant au cas précédent. Cette technique est appliquée au cas de l'évolution arête-markovienne homogène afin de compléter les résultats de [CMM+10]. Nous montrons que pour toute valeur de p et q, quelque soit le graphe initial, le temps d'inondation est au plus $O(\frac{\log n}{\log(1+np)})$. Nous appliquons également cette technique à deux cas hétérogènes. Dans le premier cas, les taux de naissances sont définis comme dans les graphes aléatoires de [CL02], à l'aide d'une séquence de degrés arbitraire. Le temps d'inondation dépend alors des séquences considérées. Dans le second cas, les graphes évolutifs sont en loi de puissance. L'inondation se fait alors en $O(\log n)$ étapes, même lorsqu'aucun graphe n'est connexe. Finalement, nous considérons la diffusion 1-Port aléatoire dans les graphes arête-markoviens. Nous donnons une majoration sur le temps de diffusion dans certains cas. En particulier, si $p + q \geq 1$, alors le temps de diffusion est $O\left(\frac{1}{1-(1-p)^{n-1}} \log n\right)$.

Dans le chapitre 4, nous étudions la diffusion *spatiale* introduite dans [KKD04]. Dans la diffusion spatiale, les communications peuvent se faire entre tous les nœuds, disposés dans un espace métrique. Le but n'est toutefois pas uniquement d'informer l'ensemble des nœuds rapidement, mais d'informer chaque nœud en un nombre d'étapes dépendant de sa proximité à la source, indépendamment de la taille du réseau. Dans ce cadre, nous exprimons le nombre d'étapes nécessaires pour informer un nœud en fonction des caractéristiques structurelles de la plus petite boule centrée en la source contenant ce nœud. Le premier résultat principal du chapitre 4 est une méthode permettant de transposer certains résultats concernant le routage dans les graphes augmentés en des résultats sur la diffusion spatiale. Le second résultat principal du chapitre consiste en la description d'un protocole général, pour les métriques de graphes éventuellement infinis. Ce protocole permet d'informer tous les nœuds dans une boule B de taille b, et de degré maximal Δ_B, en un nombre d'étapes sous-linéaire en b, avec terme additif Δ_B. Le protocole informe également tous les nœuds de B en un nombre d'étapes sous-polynomial en b, avec un facteur multiplicatif Δ_B. Ce protocole est ensuite adapté aux métriques quelconques, et nous montrons que, dans les métriques à densité uniforme telles que définies dans [KKD04], il est légèrement plus efficace que celui donné dans [KKD04].

Enfin, nous concluons cette thèse par le chapitre 5 où nous proposons également quelques perspectives de recherche.

Chapitre 2

Diffusion désordonnée

Ce chapitre reprend et complète l'article *The worst case behavior of randomized gossip protocols* écrit avec Pierre Fraigniaud, Hovhannes Harutyunyan et Rémi De Verclos, et présenté à *9th annual conference on Theory and Applications of Models of Computation (TAMC 2012)* [BFHdV12].

Nous considérons ici le protocole de diffusion d'information ListBased où chaque nœud est muni d'une liste cyclique de tous ses voisins. Étant donnée une source disposant d'une information, le problème de la diffusion optimale consiste à déterminer un schéma de diffusion 1-Port transmettant l'information à tous les nœuds du réseau en un nombre minimal d'étapes. Le problème de la diffusion ***désordonnée*** est son opposé : il consiste à calculer le temps maximal nécessaire à une information pour se propager dans le réseau. Notons, en effet, qu'un schéma de diffusion optimale ou un pire schéma sont deux cas possibles de la diffusion probabiliste. Précisément, les temps de diffusion dans ces schémas ListBased sont respectivement une borne inférieure et une borne supérieure à la diffusion du protocole QuasiRandom introduit par [DFS08] où un nœud choisit une position aléatoire dans une liste cyclique, et dès lors appelle ses voisins dans l'ordre de la liste. Motivés par des applications potentielles du protocole de diffusion QuasiRandom à la diffusion en flux continu (*live streaming*), nous nous sommes intéressés à son comportement dans le pire cas tel qu'introduit par [AHK94]. Les résultats principaux de ce chapitre sont la conception d'un algorithme polynomial qui calcule un pire schéma de diffusion dans la variante SkipNone, où chaque nœud informé suit le protocole ListBased, c'est-à-dire transmet l'information à ses voisins dans l'ordre défini par la liste. Dans la variante SkipSenders, où un nœud passe ses voisins qui lui ont envoyé l'information, le problème est NP-difficile et $(\frac{1}{2}+\epsilon)$-inapproximable pour tout $\epsilon > 0$. Et dans la variante SkipInformed, où un nœud informé passe la transmission aux voisins déjà informés, le problème est NP-difficile et $(n^{\frac{1}{2}+\epsilon})$-inapproximable, pour tout $\epsilon > 0$. Nous donnons également des graphes pour lesquels il y a un seuil exponentiel entre les temps de diffusion dans deux variantes. Un autre résultat important de ce chapitre est l'existence de diodes pour les variantes SkipNone et SkipInformed. Les diodes sont des graphes pour lesquels, en considérant deux nœuds pendants, la diffusion est

rapide (logarithmique en la taille du graphe) de l'un vers l'autre même dans le pire cas, tandis que, dans l'autre sens, le pire temps de diffusion est lent (linéaire en la taille du graphe). Pour la variante SkipSenders, nous montrons qu'il n'existe pas de diode.

2.1 Bref de l'art de la diffusion 1-Port

Dans cette section, nous rappelons les principaux résultats de la littérature concernant la complexité du calcul d'un schéma de diffusion optimal. Nous listons ensuite les principaux résultats connus pour les deux protocoles de diffusion probabilistes : Random et QuasiRandom. Rappelons que sauf mention contraire, le modèle de communication est Push.

Remarque. Avant de rentrer dans les détails des différents protocoles, nous pouvons remarquer que dans le modèle 1-Port Push, chaque nœud informe au plus un de ses voisins à chaque étape. Il faut alors nécessairement au moins $\lceil \log n \rceil$ [1] étapes pour que les n nœuds du réseau soient informés. Par ailleurs, puisque l'information ne s'échange qu'entre voisin, une autre minoration triviale du temps de diffusion est le diamètre du graphe (à une constante près), ou, plus exactement, l'excentricité de la source.

2.1.1 Recherche de schémas de diffusion optimaux

Le problème de la diffusion consiste à calculer $b(G, s)$, le nombre minimal d'étapes pour informer tous les nœuds d'un graphe G à partir d'une source s, et à déterminer le schéma optimal associé. C'est un problème NP-difficile [GJ79]. La complexité est néanmoins polynomiale dans les arbres, mais reste NP-difficile même dans le cas de classes de graphes pour lesquels de nombreux problèmes NP-difficiles sont polynomiaux : les graphes planaires bipartis, les graphes triangulés, les sous-graphes de la grille [JM95], etc. Dans le cas où l'on autorise plusieurs sources à avoir la même information au début de la diffusion, savoir si b étapes sont suffisantes pour informer tous les nœuds reste un problème NP-difficile sur ces classes, même pour des petites valeurs de b. Savoir si $b = 2$ étapes sont suffisantes est un problème NP-difficile dans les graphes scindés (*split-graphs*), dans la grille complète [GJ79] (dans certains cas, même si le degré maximum est 3) ou dans les graphes 3-réguliers [Mid93]. Ces résultats améliorent ceux de [SCH81] qui étaient les premiers à montrer que le problème reste difficile dès $b \leq 4$. Dans [JRS95], il est montré que le problème de la diffusion est aussi NP-complet dans le cas des graphes planaires de degré borné. Mais surtout, [JRS95] montre que : sauf si $P = NP$, il n'existe pas d'algorithme d'approximation polynomial avec précision $\frac{1}{8}$, c'est-à-dire donnant un résultat entre $\frac{8}{9}$ et $\frac{9}{8}$ de l'optimal. Les meilleurs résultats d'inapproximabilité connus actuellement proviennent de [EK05] : sauf si $NP \subseteq DTIME(n^{O(\log n)})$, dans le cas G non orienté,

1. Tous les log sont en base 2.

il n'y a pas de $(3-\epsilon)$-approximation de calcul de $b(G,s)$, pour tout $\epsilon > 0$. Dans le cas G orienté, il n'y a pas de $\Omega(\sqrt{\log n})$-approximation. Il existe par contre plusieurs résultats d'approximabilité, dont, en particulier, des études donnant des algorithmes d'approximation poly-logarithmiques. Dans [Rav94], l'auteur introduit le *maintien* (*poise*) $P(G)$ d'un graphe G, comme étant le maximum, pris sur tous les arbres couvrants de G, de la somme du degré maximum et de la profondeur de l'arbre. Il montre que, pour toute source s, $\Omega(P(G)) \leq b(G,s) \leq O(P(G)\frac{\log n}{\log\log n})$ et, à l'aide d'une approximation à un facteur multiplicatif $\log n$ près (plus un terme additif $O(\log^2 n)$) du calcul de $P(G)$, il obtient un algorithme polynomial qui calcule une approximation de $b(G,s)$ à un facteur multiplicatif $O(\frac{\log^2 n}{\log\log n})$ près, et un schéma de diffusion correspondant. À l'aide d'une analyse similaire mais plus directe, les auteurs de [BNGNS00] donnent un algorithme d'approximation avec un facteur multiplicatif $O(\log n)$. On trouve le même résultat avec un algorithme de meilleure complexité dans [EK05]. Ces deux dernières études améliorent à la fois les résultats de [Rav94] mentionnés précédemment, ceux de [KP95] qui donnaient une approximation avec un facteur multiplicatif $O(\log n)$ mais un terme additif $\sqrt{n}$ supplémentaire, et ceux donnant une approximation en $O(\log n)$ sur des classes de graphes particuliers tels que les graphes triangulés. La meilleure majoration d'approximation connue à ce jour, obtenue dans [EK03], est $O(\frac{\log n}{\log\log n})$. Cette majoration améliore à la fois les résultats généraux précédents et plusieurs résultats dans des classes de graphes particuliers : [KP95] donnait une $O(\log n)$ approximation pour les graphes triangulés, et [MRS+98] donnait une $O(\frac{\log n}{\log\log n})$ approximation pour les graphes de largeur arborescente bornée.

Il est intéressant de relever la remarque faite dans [Rav94] à propos du temps d'échange total. Rappelons que, dans l'échange total, tous les nœuds du réseau ont une information unique. Si l'on suppose qu'à chaque envoi un nœud peut envoyer toutes les informations en sa possession, en une seule étape, alors le facteur d'approximation du problème de l'échange total est au plus deux fois celui de la diffusion. En effet, à partir d'un schéma de diffusion qui met t étapes on peut facilement construire un schéma d'échange total qui met $2t$ étapes : soit s la source de la diffusion, pendant les t premières étapes le protocole d'échange total exécute le schéma de diffusion "inversé" afin que s récupère toutes les informations, puis il exécute ensuite le schéma normalement pour qu'en t étapes supplémentaires tous les nœuds aient toutes les informations. Inversement, un schéma d'échange total est trivialement un schéma de diffusion. Une remarque similaire est faite dans [KCV92], elle montre la proximité du problème de diffusion et celui de l'échange total sous l'hypothèse qu'à chaque envoi un nœud peut envoyer toutes les informations en sa possession, en une seule étape. En effet, si G' est le graphe où G est dupliqué et où les deux copies de G sont reliées par une arête entre les deux copies du nœud s alors le temps de diffusion dans G d'un message de source s est t si et seulement si le temps d'échange total dans G' est $2t+1$ dans le cas *full-duplex* où l'information peut aller dans les deux sens entre le nœud appelant et le nœud appelé, et $2t+2$ dans le cas *half-duplex* où l'information ne peut aller que du nœud appelant vers le nœud appelé. De plus, de l'un quelconque

des deux schémas de diffusion on déduit l'autre très facilement.

2.1.2 Le protocole de diffusion probabiliste

Nous donnons ici les résultats les plus importants concernant le protocole de diffusion probabiliste Random. En procédant à l'analyse d'un algorithme aléatoire de calcul de plus court chemin, les auteurs de [FG85] introduisent le problème de l'*appel téléphonique* : une ville contient n personnes, dont l'une connaît une rumeur. À la première étape, elle la raconte à une personne choisie au hasard dans la ville ; puis à chaque étape, chaque personne qui connaît la rumeur la transmet à quelqu'un d'autre, choisi indépendamment au hasard dans la ville. Soit S_n le nombre d'étapes avant que toute la ville connaisse la rumeur. Les auteurs de [FG85] montrent, d'une part, que $\frac{S_n}{\log n} \to 1 + \ln 2$ en probabilité lorsque $n \to +\infty$ et, d'autre part, donne aussi un résultat avec forte probabilité [2]. Plus précisément, ils montrent que si $\gamma > 0$ alors, $\forall \epsilon > 0$,

$$\mathbb{P}\left(S_n > (1+\epsilon)\alpha(\gamma)\log n\right) = o(n^{-\gamma})$$

où $\alpha(\gamma) = 1 + (\gamma+1)\ln 2$. De plus, la constante $\alpha(\gamma)$ est la meilleure possible au sens où si $0 \leq \beta < \alpha(\gamma)$ alors $\mathbb{P}\left(S_n > \beta \log_2 n\right) \gg n^{-\gamma}$. Ce protocole de diffusion 1-Port où à chaque étape, chaque nœud informé u sélectionne le destinataire v aléatoirement uniformément parmi tous les nœuds de V et lui envoie l'information est un cas particulier du protocole Random où le réseau de communication représenté par G est une clique, il est connu pour bien s'exécuter en pratique [GKG06].

D'autres auteurs se sont intéressés par la suite au calcul du temps de diffusion du protocole Random dans la clique, par exemple, dans [Pit87] l'auteur montre que $S_n = (1+\ln 2)\log_2 n + O(1)$ en probabilité. Plus formellement, il montre que si $\omega(n) \to \infty$, même arbitrairement lentement, lorsque $n \to \infty$, alors

$$\mathbb{P}\left(|S_n - \log n - \ln n\,| \leq \omega(n)\right) \to 1.$$

La majoration $(1+\ln 2)\log n + O(1)$ donnée, en probabilité, par [Pit87] améliore la majoration $(1+\ln 2)\log n + o(\log n)$ donnée par [FG85]. Par contre, le résultat de [FG85] reste très intéressant car il permet de maîtriser la garantie probabiliste.

Dans [FPRU90], les auteurs se sont intéressés à d'autres graphes en particulier le cas général, l'hypercube et les graphes aléatoires. Ils montrent que pour un réseau quelconque G de n sommets, soit $T(G)$ le temps tel qu'après $T(G)$ étapes du protocole Random, tous les sommets de G ont reçu l'information avec probabilité $1 - \frac{1}{n}$. Alors :

- $\log_2 n \leq T(G) \leq 12n\log n$, pour tout graphe G.

2. Dans la littérature les notations "asymptotiquement presque sûrement" (a.p.s.) et "avec forte probabilité" (a.f.p.) ne désignent pas toujours la même chose. Dans cette thèse, nous considérons que les événements $\mathcal{E}_n$ surviennent

- asymptotiquement presque sûrement (a.p.s.) si $\mathbb{P}[\mathcal{E}_n] \to 1$ lorsque $n \to \infty$, c'est-à-dire, $\mathbb{P}[\mathcal{E}_n] = 1 - o(1)$,
- avec forte probabilité (a.f.p.) si $\mathbb{P}[\mathcal{E}_n] \geq 1 - O(\frac{1}{n^\alpha})$ avec $\alpha > 0$.

- Il y a des réseaux, par exemple l'étoile, tels que $T(G) = \Omega(n \log n)$.
- Il y a des réseaux, par exemple la clique, tels que $T(G) = O(\log n)$.

La preuve de ce théorème peut être trouvée dans le chapitre 4 où elle est réutilisée dans le contexte de la diffusion spatiale. L'idée de la preuve est de constater qu'en considérant un plus court chemin P entre la source et un nœud quelconque, chaque nœud est voisin d'au plus 3 nœuds de P. Cette remarque implique un temps moyen pour informer ce nœud d'au plus $3n$ étapes. Le facteur $4 \log n$ sert ensuite à garantir qu'avec probabilité au moins $1 - \frac{1}{n}$ tous les nœuds sont informés. À l'aide d'une remarque faite dans la preuve de la proposition 1 de [DFS08], on peut également montrer qu'après $8n \log n$ étapes, tous les nœuds sont informés avec probabilité $1 - \frac{1}{n}$.

Dans [FPRU90], il est aussi montré que $T(G) = O(\deg(G)(\mathrm{diam}(G) + \log n))$, ce qui devient $T(G) = O(\mathrm{diam}(G) + \log n)$ dans le cas où G est de degré borné. Pour l'hypercube k-dimensionnel (ayant $n = 2^k$ sommets) H_k, $T(H_k) = \Theta(k) = \Theta(\log n)$. Pour presque tout graphe aléatoire d'Erdős-Rényi $G \in \mathcal{G}_{n,p}$

- si $p \leq \frac{\ln n - \omega(n)}{n}$ avec $\omega(n) \to \infty$ alors $T(G) = \infty$
- si $p \leq \frac{\ln n + \omega(n)}{n}$ avec $\omega(n) \to \infty$ et $\omega(n) = O(\log \log n)$ alors $T(G) = \Theta(\log^2 n)$
- si $p \geq \frac{(1+\epsilon) \ln n}{n}$ avec $\epsilon > 0$ fixé alors $T(G) = \Theta(\log n)$

Diverses études ont été menées pour évaluer le temps de diffusion du protocole Random pour d'autres types de graphes, comme, par exemple, les graphes en loi de puissance, souvent utilisés pour modéliser les réseaux sociaux. Sur le modèle en loi de puissance défini dans [CL03] à l'aide d'une suite de degrés moyens, il est montré dans [FPS12] que le modèle de communication PushPull permet un temps de diffusion de $\Theta(\log \log n)$ étapes, lorsque l'exposant $\beta \in]2, 3[$. Par contre, il faut $\Omega(\log n)$ étapes si $\beta > 3$. Le cas général où le temps de diffusion est exprimé en fonction de la conductance

$$\phi = \min_{S \subseteq V, \mathrm{vol}(S) \leq |E|} \frac{\mathrm{cut}(S, V - S)}{\mathrm{vol}(S)},$$

où $\mathrm{vol}(A)$ est le nombre d'arêtes adjacentes à au moins un nœud de A et $\mathrm{cut}(A, B)$ le nombre d'arêtes adjacentes à un nœud de A et à un nœud de B, a été étudié dans un premier temps par [CLP10b, CLP10a]. Les meilleurs résultats connus à ce jour sont donnés par [Gia11] qui montre que le temps de diffusion du protocole Random est $O(\phi^{-1} \log n)$ dans le modèle de communication PushPull et qu'il existe des graphes tels que le temps de diffusion est $\omega(\phi^{-1} \log n)$. En outre, [Gia11] montre que dans le cas Push ou le cas Pull, il existe au moins une source pour laquelle le temps de diffusion est $O(\phi^{-1} \log n)$.

Dans [ES09], les auteurs ont montré que si les transmissions échouent avec une probabilité $1 - p$ alors le temps de diffusion est multiplié par un facteur au plus $\frac{6}{p}$. Ils donnent aussi une borne logarithmique sur le temps de la diffusion aléatoire dans le cas des graphes réguliers. Ceci leur permet de montrer, à l'aide du résultat de [Pit87], que parmi les graphes réguliers, la clique est le graphe où la diffusion aléatoire se fait le plus rapidement. Ils améliorent également la constante de la majoration du temps de diffusion protocole Random sur un graphe quelconque donné dans [FPRU90] :

ils montrent en effet que pour tout graphe, tous les nœuds sont informés en au plus $(1+o(1))n\ln n$ étapes, avec probabilité au moins $1-\frac{1}{n}$. Ils fournissent enfin la borne $O(n\frac{\Delta}{\delta})$ où δ est le degré minimum du graphe, avec probabilité au moins $1-\frac{1}{n}$.

2.1.3 Le protocole de diffusion quasi-aléatoire

Une autre approche légèrement différente de la diffusion aléatoire est de considérer qu'au lieu d'envoyer le message à n'importe quel voisin, chaque nœud possède une liste de ses voisins et sélectionne à chaque étape le suivant dans cette liste. Les résultats varient en fonction de la manière dont ces listes sont construites. Si les listes sont construites aléatoirement cela revient alors pour chaque nœud à sélectionner le destinataire uniformément parmi les voisins à qui il n'a pas déjà envoyé l'information. Par contre, lorsque les listes sont choisies par un adversaire, il peut y avoir de lourdes conséquences sur la vitesse de diffusion. En effet, dans le cas de la clique, si les listes sont toutes identiques, il faut autant d'étapes que de nœuds pour qu'ils soient tous informés.

Dans [DFS08], les auteurs se sont intéressés au cas où la sélection du destinataire commence dans chaque liste à un rang choisi aléatoirement uniformément et recommence au début de la liste une fois arrivé à la fin. C'est le modèle QuasiRandom. Ce modèle a l'avantage de diminuer fortement le besoin de l'aléa. Il permet également de ne pas envoyer un message au même voisin et permet aussi d'éviter le cas pathologique dans le choix des listes. En fait, les auteurs ont montré que ce protocole se comporte bien même lorsque les listes sont choisies par un adversaire. Ils ont montré que la diffusion quasi-aléatoire dans une clique, dans un graphe aléatoire d'Erdős-Rényi $\mathcal{G}_{n,p}$ (dans le cas $p \geq \frac{(1+\epsilon)\ln n}{n}$ et $p = \frac{\ln n + f(n)}{n}$ si $f(n) \to \infty$ et $f(n) = O(\log\log n)$), ainsi que dans un hypercube, se fait aussi, a.p.s., en $\Theta(\log n)$ étapes comme dans la diffusion aléatoire. Dans le cas des k-arbres, le temps est même $\Theta(\frac{k\log n}{\log k})$ au lieu de $\Theta(k\log n)$. Les bornes générales semblables à celles données dans [FPRU90] sont même plus simples à calculer dans le cas quasi-aléatoire puisque le terme $\log n$ n'est plus nécessaire pour obtenir la forte probabilité, dans la mesure où l'on a la garantie qu'un nœud u aura contacté tous ses voisins exactement une fois en $\deg(u)$ étapes (où $\deg(u)$ désigne le degré de u). Ainsi, dans le modèle quasi-aléatoire, avec probabilité 1, les auteurs de [DFS08] montrent aussi que le temps de diffusion d'un graphe quelconque à n sommets est borné supérieurement par $\deg(G)\mathrm{diam}(G)$ et par $2n-3$. Dans [DFS09], les auteurs améliorent leurs résultats précédents en montrant qu'avec forte probabilité le temps de diffusion est $O(\log n)$ dans un graphe vérifiant une certaine propriété d'expansion qui est vérifiée asymptotiquement presque sûrement par les graphes aléatoires connectés (épars et denses) ainsi que les graphes aléatoires d-réguliers. De plus, ils montrent que dans le cas de pannes aléatoires de probabilité f, si le protocole prévoit de renvoyer le message au même destinataire au lieu de passer au suivant tant qu'il n'y a pas eu d'acquittement, alors l'exécution n'est ralentie que d'un facteur $O(\frac{1}{1-f})$.

Une analyse expérimentale faite dans [DFKS11] montre non seulement que

QuasiRandom se comporte mieux que Random sur les cliques, les hypercubes, les graphes réguliers aléatoires et les graphes aléatoires, mais qu'en plus l'écart à la moyenne est moindre. Notons tout de même que, dans le cas des graphes aléatoires, puisque l'ordre des listes est choisi aléatoirement dans cette étude, il n'est pas étonnant que QuasiRandom se comporte mieux que Random puisqu'il correspond à une amélioration de Random où le destinataire est choisi par un nœud expéditeur uniquement parmi ses voisins qu'il n'a pas déjà contactés.

D'autres auteurs se sont intéressés à ce protocole de diffusion quasi-aléatoire. Dans [ADHP09, FH09], les auteurs ont montré que la diffusion d'information quasi-aléatoire est aussi rapide dans la clique que la diffusion aléatoire donnée par [Pit87]. Formellement si S_n est le nombre d'étapes nécessaires pour que tous les nœuds soient informés, il est montré dans [ADHP09] qu'a.p.s. $|S_n - (\log n + \ln n)| = o(\log n)$, quel que soit le choix des listes, ce résultat est amélioré dans [FH09] : a.p.s.

$$\log n + \ln n - 4 \ln \ln n \leq S_n \leq \log n + \ln n + \omega(n)$$

où $\omega(n)$ est une fonction tendant vers l'infini arbitrairement lentement.

2.2 Introduction à la diffusion désordonnée

Le problème "opposé" au problème de recherche de la diffusion optimale consiste à chercher un pire schéma de diffusion. Ce problème s'appelle le problème de la ***diffusion désordonnée***. Il a été introduit par [AHK94].

L'étude de la diffusion désordonnée est particulièrement importante dans les contextes où les résultats sur la diffusion (quasi)-aléatoire sont insuffisants pour garantir une bonne performance. Premièrement, lorsqu'on considère des réseaux arbitraires pour lesquels le comportement du protocole aléatoire n'est pas connu avec précision. En effet, le comportement du protocole aléatoire est bien connu dans le modèle de graphe aléatoire $\mathcal{G}_{n,p}$ mais il ne reflète pas nécessairement la structure des réseaux du monde réel. Des bornes supérieures sur le nombre d'étapes nécessaires à une information pour se diffuser à tous les nœuds sont connues. Cependant, bien que ces bornes soient souvent exactes, elles sont générales et ne reflètent pas nécessairement le comportement de la diffusion d'information pour chaque réseau. C'est le cas même en se restreignant à l'analyse des réseaux ayant certaines propriétés telles que d'expansion-arêtes bornée [Sau07], d'expansion-sommets bornée [SS11], ou de conductance bornée [CLP10b, CLP10a, Gia11] (voir aussi [CHS10]). Deuxièmement, les applications telles que le partage de fichiers ou la diffusion continue de données nécessitent l'envoi d'un nombre important de "morceaux" d'information appelés *paquets*, *blocs* ou *frames* selon le contexte. Dans ces conditions, les résultats en moyenne, ou même avec forte probabilité, peuvent ne pas garantir qu'aucun paquet (ou même une fraction constante de ces paquets) ne connaîtra de retard important, ce qui conduit à augmenter la perte de paquets et empêche la bonne réception des fichiers. En effet, "avec forte probabilité" se réfère à une certaine garantie statistique exprimée en fonction de la taille n du réseau. Ainsi, si le nombre de paquets est grand par

rapport à n, alors cette garantie peut ne pas être suffisante pour empêcher plusieurs "mauvais événements" de se produire. Troisièmement, les applications comme la diffusion en flux continu, audio ou vidéo, nécessitent de contrôler la *gigue*, c'est-à-dire la différence de temps entre la réception de deux paquets consécutifs. Encore une fois, les résultats en moyenne, voir a.f.p., pourraient ne pas permettre au concepteur de bien calibrer la taille des mémoires tampons des applications assurant qu'une gigue importante n'empêchera pas une bonne réception du flux audio ou vidéo.

Il est intéressant de remarquer qu'imposer une plus forte probabilité $1 - O(\frac{1}{n^\alpha})$ de succès pour chaque paquet, c'est-à-dire augmenter α, ne fournit pas nécessairement une solution aux problèmes énoncés ci-dessus. En effet, pour ce faire, nous devons supposer que l'analyse du protocole nous permet d'établir des compromis explicites entre la probabilité de succès et de la borne sur le nombre d'étapes, pour le réseau spécifique ou la classe spécifique de réseaux auxquels nous sommes confrontés. Par ailleurs, il peut arriver que même des compromis explicites ne soient pas assez précis, dans le sens où l'augmentation de la probabilité de succès au-dessus du seuil souhaité peut entraîner une borne sur le nombre d'étapes dépassant en fait le pire des comportements du protocole aléatoire. Par exemple, il est démontré dans [DFS08] que le protocole QuasiRandom diffuse une information à tous les nœuds d'un hypercube en au plus $1542 \cdot \log n$ étapes, avec probabilité au moins $1 - \frac{1}{n}$. Cette majoration du nombre d'étapes est, pour toute valeur raisonnable du nombre de nœuds, supérieure à la majoration triviale $\log^2 n$ du pire comportement du protocole QuasiRandom dans un hypercube [3]. Afin de répondre à ces problèmes, ce chapitre aborde la question de déterminer le pire cas de la diffusion probabiliste.

Plus précisément, étant donné un réseau G, nous nous sommes intéressés au calcul du *pire* comportement du protocole QuasiRandom dans G. C'est principalement parce que la version générale de diffusion aléatoire d'information n'empêche pas un paquet de faire des aller-retours entre deux nœuds adjacents pendant un nombre arbitrairement grand d'étapes que nous effectuons notre étude dans le cadre de la diffusion d'information QuasiRandom. Dans ce contexte, le comportement au pire cas du protocole est donné par le choix des pires listes cycliques de voisins imposées aux nœuds, combiné avec un pire choix des positions de départ dans ces listes. Par la suite, nous supposerons simplement que les listes sont telles que la position de départ est toujours le début de la liste. Ceci revient alors à étudier le pire comportement du protocole ListBased, c'est d'ailleurs cette vision que nous utilisons dans nos preuves. Ceci reviendrait également à étudier le protocole Random modifié de telle sorte qu'un nœud ne rappelle pas un nœud qu'il a déjà appelé.

Le pire temps de diffusion est aussi une borne supérieure absolue sur le délai maximum entre deux paquets consécutifs diffusés dans le réseau en utilisant un protocole ListBased ou un protocole QuasiRandom. Bien entendu, cette borne

3. Bien que la principale préoccupation de [DFS08] n'est pas nécessairement l'optimisation de la constante devant le facteur $\log n$, nous soulignons que diminuer cette constante même par plusieurs ordres de grandeur aboutirait encore à une majoration supérieure à $\log^2 n$ pour toutes valeurs raisonnables de n.

reflète uniquement les retards causés par le protocole lui-même, et ignore d'autres raisons qui peuvent causer des retards, comme la congestion du trafic, et qui sortent du cadre de travail cette thèse.

Il est facile de trouver des réseaux pour lesquels le pire temps de diffusion est $\Omega(n)$ (par exemple les réseaux complets et les étoiles) et d'autres réseaux pour lesquels le pire temps de diffusion est $O(\log n)$ (par exemple les arbres binaires complets). Toutefois, l'étude du pire temps de diffusion pour des réseaux arbitraires n'est pas aussi facile (voir Section 2.2.1). L'analyse devient encore plus difficile lorsque nous avons affaire à une optimisation naturelle du protocole ListBased. On trouve en particulier deux autres variantes dans la littérature [AHK94, HL98, HH02, HHL11], correspondantes à trois contraintes de communications différentes. Par exemple, à chaque étape, chaque nœud peut passer l'envoi d'un paquet aux voisins dont il avait reçu le même paquet lors des étapes précédentes, y compris le voisin à partir duquel il l'a reçu pour la première fois. Nous pouvons également supposer l'existence d'un mécanisme sous-jacent permettant aux nœuds de connaître perpétuellement l'ensemble de ses voisins déjà informés.

2.2.1 Les modèles désordonnés

Soit $G = (V, E)$ un graphe simple et connexe de n nœuds, et soit $s \in V$ une source. Nous nous intéressons au calcul du pire temps mis par une information apparaissant en s pour atteindre tous les nœuds de G dans le protocole ListBased. Les communications s'effectuent par étapes synchrones. À chaque étape, chaque nœud informé envoie l'information à l'un de ses voisins. Tous les schémas de diffusion peuvent être décrits en assignant une liste ordonnée L_u à chaque nœud $u \in V$, listant ses voisins dans G dans un certain ordre. À la première réception de l'information, par exemple à l'étape r, le nœud u propage l'information en l'envoyant à son i-ème voisin dans la liste L_u à l'étape $r + i$, $1 \leq i \leq \deg(u)$ où $\deg(u)$ désigne le degré de u dans G. Notez que u peut aussi bien envoyer l'information à un voisin qui l'a déjà reçue d'un autre nœud. En fait, il pourrait même envoyer l'information au nœud duquel il a obtenu l'information. Pour tenir compte de ce phénomène, le modèle dispose de trois variantes différentes [4].

- La variante SkipNone ou sn (passer aucun voisin). C'est la version basique du protocole ListBased telle que décrite précédemment : chaque nœud u propage l'information à tous ses voisins dans l'ordre défini par la liste L_u. Soit $\mathcal{L} = \{L_u, u \in V\}$ et soit $\varrho(G, \mathcal{L}, s, t)$ le nombre d'étapes que met une information initialisée en s pour atteindre t quand l'information est propagée dans G selon le protocole défini par les listes de $\mathcal{L}$. Définissons alors $t_{\text{SN}}(G, s, t) = \max_{\mathcal{L}} \varrho(G, \mathcal{L}, s, t)$ et $t_{\text{SN}}(G) = \max_{s,t \in V} t_{\text{SN}}(G, s, t)$.
- La variante SkipSenders ou ss (passer les voisins expéditeurs). Lorsqu'un nœud u effectue la diffusion d'information selon L_u, cette variante suppose que

4. Les trois variantes SkipInformed, SkipSenders et SkipNone sont notés respectivement M_1, M_2 et M_3 dans [AHK94].

u saute l'envoi de l'information aux nœuds desquels il a reçu l'information. Plus précisément, supposons que u reçoive l'information à l'étape r. Pour $i \geq 1$, soit S_i le sous-ensemble des voisins de u ayant envoyé l'information à une étape dans $[r, r+i-1]$. À l'étape $r+i$, le nœud u envoie l'information au voisin suivant dans la liste $L_u \setminus S_i$. Définissons alors $t_{\text{SS}}(G, s, t)$ et $t_{\text{SS}}(G)$ comme ci-dessus.

- La variante SkipInformed ou si (passer les voisins informés). Cette variante suppose l'existence d'un mécanisme sous-jacent qui permet à chaque nœud de savoir lequel de ses voisins a reçu l'information, à n'importe quel moment. Plus précisément, supposons que u reçoive l'information au temps r. Pour $i \geq 1$, soit I_i le sous-ensemble des voisins de u au courant de l'information avant l'étape $r+i$. À l'étape $r+i$, le nœud u envoie l'information au voisin suivant dans la liste $L_u \setminus I_i$. Définissons alors $t_{\text{SI}}(G, s, t)$ et $t_{\text{SI}}(G)$ comme ci-dessus.

La figure 2.1 donne un exemple d'un pire comportement de ces trois variantes sur un même graphe.

Faisons deux remarques simples mais importantes de [AHK94]. Par définition, puisqu'un schéma de diffusion de SkipInformed est un schéma de diffusion légitime de SkipSenders, et que, de même, un schéma de diffusion de SkipSenders est un schéma de diffusion légitime de SkipNone, pour chaque réseau G, et chaque paire source-cible $s, t \in V(G)$, nous avons

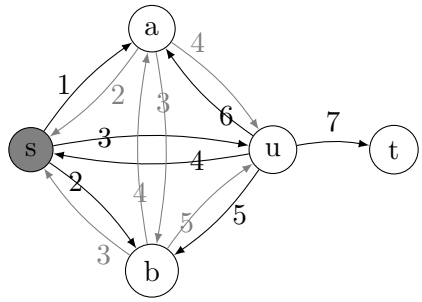
$$t_{\text{SI}}(G, s, t) \leq t_{\text{SS}}(G, s, t) \leq t_{\text{SN}}(G, s, t)$$

et

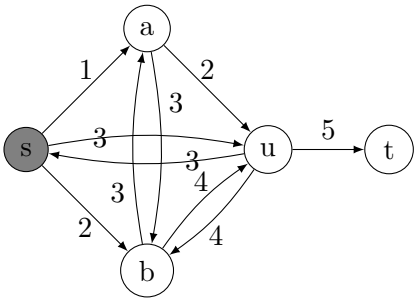
$$t_{\text{SI}}(G) \leq t_{\text{SS}}(G) \leq t_{\text{SN}}(G).$$

Dans un arbre T, on a toujours $t_{\text{SI}}(T, s, t) = t_{\text{SS}}(T, s, t)$ pour tout s et t. En effet, si on considère que l'arbre est enraciné en s, un nœud reçoit toujours l'information de son père, et il est le seul à le transmettre à ses fils.

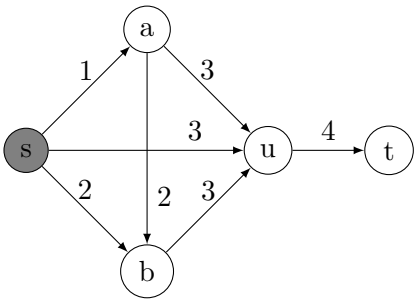
Remarquons qu'un protocole SkipSenders est réaliste dans le cas de la diffusion d'une information unique. En revanche, un protocole SkipNone est sans doute plus réaliste dans le cas d'une multitude d'informations. En effet, le premier modèle exige que chaque nœud maintienne la liste des voisins desquels il a déjà reçu tous les messages, tandis que le second modèle ignore les réceptions multiples du même message. La dernière variante, SkipInformed, est la plus efficace des trois variantes car elle évite qu'un nœud reçoive le même message encore et encore lors d'étapes différentes. Néanmoins, il existe des graphes G de n nœuds vérifiant $t_{\text{SI}}(G) = n - 1$. (Nous avons évidemment $t_{\text{SI}}(G) \leq n - 1$ pour chaque graphe de n nœuds puisqu'au moins un nouveau nœud est informé à chaque étape dans la variante SkipInformed).



(a) Exemple pour SkipNone



(b) Exemple pour SkipSenders



(c) Exemple pour SkipInformed

Figure 2.1 – Exemples de pire schémas de diffusion.

2.2.2 État de l'art

De l'étude des pire temps de diffusion [AHK94] nous retiendrons pour tout graphe G, soit $v_0, \ldots, v_k$ un chemin dans G nous avons :

- $t_{\text{SN}}(v_0, v_k) \leq \sum_{i=0}^{k-1} \deg(v_i)$
- $t_{\text{SI}}(v_0, v_k) \leq t_{\text{SS}}(v_0, v_k) \leq \sum_{i=0}^{k-1} (\deg(v_i) - 1)$

Ces inégalités étant des égalités dans le cas des arbres, on en déduit facilement un algorithme polynomial pour calculer ces trois temps dans les arbres. Nous complétons alors la séquence d'inégalités précédente :

$$\Omega(\log n) \leq t_{\text{SI}}(G, s, t) \leq t_{\text{SS}}(G, s, t) \leq t_{\text{SN}}(G, s, t) \leq O(n).$$

La majoration $O(n)$ découle du fait que la somme des degrés des nœuds le long d'un chemin le plus court de la source à n'importe quel nœud ne peut pas dépasser $3n$ [FPRU90]. La minoration $\Omega(\log n)$ découle du fait que, à chaque étape, un nœud peut informer au plus un nœud non informé, ce qui implique que le nombre de nœuds informés peut au plus doubler à chaque étape.

Des topologies de réseaux spécifiques ont été considérées dans [HL98]. Les auteurs ont calculé les temps maximaux de diffusion dans les cliques K_n où le nombre maximal d'étapes est $n-1$ dans les trois modèles considérés, dans les anneaux C_n où le nombre maximal d'étapes est $n-1$ dans le modèle SkipNone et $\lceil \frac{n}{2} \rceil$ dans les deux variantes SkipSenders et SkipInformed, sur les chemins P_n où le pire temps est $2n-3$ dans le modèle SkipNone et $n-1$ dans les deux autres variantes. D'autres graphes ont été également étudiés, nous retiendrons ici surtout le cas des hypercubes H_d de dimension d ($n = 2^d$) pour lesquels [HL98] montre que $t_{\text{SN}}(H_d) = \frac{d(d+1)}{2}$ et $t_{\text{SS}}(H_d) = \frac{d(d-1)}{2} + 1$. En revanche, $t_{\text{SI}}(H_d)$ n'est pas connu précisément. D'après [HL98], on sait seulement que $\frac{3n}{2} \leq t_{\text{SI}}(H_d) \leq t_{\text{SS}}(H_d)$.

Les résultats précédents ont été complétés par [HH02]. Dans le graphe biparti complet $K_{m,n}$, nous avons $t_{\text{SN}}(K_{m,n}) = \max\{m, n\} + 1$ et $t_{\text{SI}}(K_{m,n}) = t_{\text{SS}}(K_{m,n}) = \max\{m, n\}$. Également, $t_{\text{SI}}(H_d) \geq \frac{7d-5}{4}$ mais on ne sait toujours pas si $t_{\text{SI}}(H_d) = O(d)$, si $t_{\text{SI}}(H_d) = \Omega(d^2)$, ou si ce temps se situe entre ces deux bornes. Ceci illustre la difficulté du protocole SkipInformed.

Plus récemment, dans [HHL11], les auteurs se sont intéressés au calcul exact de bornes supérieures. Ils montrent que, pour tout réseau G, on a $t_{\text{SN}}(G) \leq 2n - 3$ [5]. L'égalité est vérifiée pour les chemins et ce sont les seuls arbres à vérifier l'égalité. [HHL11] montre aussi que $t_{\text{SI}}(G) \leq n - 1$, l'égalité étant vérifiée pour les chemins ou les cliques. Par contre une borne supérieure exacte pour t_{SS} n'est actuellement pas connue. Si nous nous restreignons à nouveau aux arbres, puisque $t_{\text{SI}}(T) = t_{\text{SS}}(T)$, pour tout arbre T, nous obtennons plus facilement la borne $n - 1$ pour les deux cas, et elle est atteinte uniquement pour les chenilles.

2.2.3 Nos résultats

Dans un premier temps nous complétons les résultats de [HHL11], puis nous exhibons diverses différences entre les trois protocoles, dont des graphes pour lesquels le comportement des protocoles diffèrent grandement. Ces graphes seront particulièrement utiles en tant que gadgets par la suite. Nous exhibons de nombreux écarts exponentiels entre t_{SN}, t_{SS} et t_{SI}. Précisément, nous montrons qu'il existe des réseaux G et des paires source-cible $s, t \in V(G)$ tels que $t_{\text{SN}}(G, s, t) = \Omega(n)$ tandis que $t_{\text{SS}}(G, s, t) = O(\log n)$. De même, nous montrons qu'il existe des réseaux G et des paires source-cible $s, t \in V(G)$ tels que $t_{\text{SS}}(G, s, t) = \Omega(n)$ tandis que $t_{\text{SI}}(G, s, t) = O(\log n)$. Ainsi, de petites variations dans la mise en œuvre du protocole de diffusion ListBased peuvent avoir un impact considérable sur le comportement

5. Un résultat de [DFS08] dans le cadre de l'étude du protocole QuasiRandom, permet aussi d'obtenir ce résultat avec une preuve différente.

au pire cas du protocole.

Le premier résultat principal de ce chapitre est la conception d'un algorithme en temps $O(m + n \log n)$ qui, pour tout réseau G de n nœuds et m arêtes, et n'importe quelle paire de nœuds (s, t) dans G, calcule $t_{\text{SN}}(G, s, t)$. Plus précisément, pour n'importe quelle cible fixée t, notre algorithme calcule un ensemble de listes $\mathcal{L} = \{L_u, u \in V\}$ tel que, pour n'importe quelle source s, la diffusion de s à t selon cet ensemble $\mathcal{L}$ prend $t_{\text{SN}}(G, s, t)$ étapes. Ainsi, notre algorithme donne directement un algorithme en temps $O(n(m + n \log n))$ qui, étant donné un réseau G de n nœuds et m arêtes, calcule $t_{\text{SN}}(G)$. Par conséquent, le calcul d'un pire protocole ListBased est polynomial dans sa version de base SkipNone.

La situation est radicalement différente lorsqu'on considère une variante du protocole dans lequel les nœuds sont au courant de la situation de leurs voisins, c'est-à-dire, savent à tout moment si oui ou non ils ont déjà reçu l'information. En effet, le second résultat principal de ce chapitre établit que, à moins que P = NP, $t_{\text{SI}}(G, s, t)$ ne peut être approximé en temps polynomial avec un facteur multiplicatif $(\frac{1}{n})^{\frac{1}{2}+\epsilon}$ près, pour tout $\epsilon > 0$. En tant que variante de ce dernier résultat, nous montrons aussi que, à moins que P = NP, il n'y a pas de schéma d'approximation polynomial (*PTAS*) permettant d'approximer $t_{\text{SI}}(G, s, t)$. En fait, le temps $t_{\text{SS}}(G, s, t)$ ne peut pas être approximé à $\frac{1}{2} + \epsilon$ près, pour tout $\epsilon > 0$.

Les résultats de cette étude peuvent être étendus aux graphes orientés. Dans le contexte de graphes orientés, le protocole SkipNone reste résoluble en temps polynomial, et le protocole SkipInformed reste inapproximable. En fait, il est possible de montrer que, à la fois pour le protocole SkipSenders et le protocole SkipInformed, les pires performances du protocole ListBased ne peuvent être approximées en temps polynomial à un facteur multiplicatif $(\frac{1}{n})^{1+\epsilon}$ près, pour tout $\epsilon > 0$.

Finalement nous donnerons une collection de résultats liant différentes variantes (multisources, multidestinaires, cas orienté, etc.) ainsi que des gadgets potentiellement utiles pour une étude ultérieure.

2.3 Résultats préliminaires

2.3.1 Caractérisation des pires graphes

Dans [HHL11] les auteurs se sont intéressés à l'étude des pires graphes pour les trois modèles, nous complétons ici leurs résultats. Le résultat suivant caractérise les graphes les plus "lents" dans la variante SkipInformed.

Propriété 2.3.1. *Pour tout graphe G de n nœuds, $t_{\text{SI}}(G) = n - 1$ si et seulement si G est un graphe d'intervalles.*

Démonstration. Nous savons [Ola91] que $G = (V, E)$ est un graphe d'intervalles si et seulement s'il y a un ordre total $\prec$ sur V tel que si $u \prec v \prec w$ et $\{u, w\} \in E$ alors $\{u, v\} \in E$. Ainsi, soit G un graphe d'intervalles et soit $u_1 \prec u_2 \prec \cdots \prec u_n$

un ordre sur les nœuds de G vérifiant la précédente caractérisation de G comme graphe d'intervalles. Le schéma de diffusion suivant met exactement $n-1$ étapes pour diffuser une information à partir du nœud u_1. (N'oublions pas que les graphes sont supposés connexes). Pour $i = 1$ à $n-1$, le nœud u_i et tous ses voisins informés envoient le message à u_{i+1}, tandis que tous les autres nœuds informés restent inactifs (parce que, comme nous le verrons, ils n'ont plus de voisins non informés). Ce schéma de diffusion garantit que l'ensemble des nœuds informés après l'étape i est exactement $\{u_1, \ldots, u_{i+1}\}$, $i = 1$ à $n-1$. Il est conforme au modèle SkipInformed, parce que si le nœud u_j, $j < i$, n'est pas un voisin de u_i, alors u_j n'a pas de voisin dans $\{u_i, \ldots, u_n\}$. Ainsi, à l'étape i, soit les nœuds informés sont dans le voisinage de u_i, soit ils sont inactifs (parce que tous leurs voisins sont dans $\{u_1, \ldots, u_{i-1}\}$ et par conséquent ils sont déjà informés). Par conséquent, si G est un graphe d'intervalles, alors $t_{\text{SI}}(G) = n-1$.

Réciproquement, soit G un graphe avec $t_{\text{SI}}(G) = n-1$. Alors, il existe un schéma de diffusion dans le modèle SkipInformed qui garantit qu'à chaque étape i, $1 \leq i \leq n-1$, exactement un nouveau nœud se trouve informé. Ainsi, l'étape au cours de laquelle un nœud reçoit l'information fournit un ordre total sur les nœuds $u_1 \prec \ldots \prec u_n$ où u_{i+1} est le nœud informé à l'étape i, et u_1 est la source. Soit i, j, k tels que $i < j < k$ et $\{u_i, u_k\} \in E$. À l'étape j, u_i a au moins un voisin non informé (parce u_k n'est pas encore informé). Donc, à cette étape, u_i envoie l'information à un nœud non informé. Ce nœud doit être u_j, simplement parce que c'est le seul nœud qui a été nouvellement informé à l'étape j. Ainsi $\{u_i, u_j\} \in E$. Par conséquent G est un graphe d'intervalles, d'après la caractérisation de [Ola91]. □

Notez que la caractérisation des graphes "les plus lents" pour le modèle SkipNone s'avère être bien difficile. Celle donnée dans [HHL11] est d'ailleurs incomplète. En particulier, ils ne peuvent pas être caractérisés par une liste finie de mineurs, ou de sous-graphes exclus. En effet, pour tout graphe H, il y a un graphe de pire cas G contenant H comme sous-graphe. Il est obtenu en reliant les nœuds de H et ceux d'un graphe complet K, de taille au moins $|V(H)|$ par un graphe biparti complet, puis avec un nœud universel u relié à tous les nœuds dans H, tous les nœuds de K et à un nœud supplémentaire t. Avec s dans le graphe complet K, nous avons $t_{\text{SN}}(G, s, t) = 2n - 3$ où $n = |V(H)| + |V(K)| + 2$.

Pour le modèle SkipSenders c'est encore plus difficile, en effet la valeur maximale n'est pas exactement connue. Dans [HHL11], le pire exemple connu donne un temps de diffusion désordonnée $2n - \lceil \log n \rceil - 2$ mais il n'y a pas la garantie que cette borne soit exacte. Et nous savons seulement qu'on ne peut pas trouver plus grand que le pire temps $2n - 3$ du modèle SkipNone.

2.3.2 Écarts exponentiels entre les modèles

Afin d'illustrer les différences entre les trois variantes du modèle, nous soulignons l'existence d'écarts exponentiels entre les trois paramètres t_{SN}, t_{SS} et t_{SI}. La propriété ci-dessous traite de la variante SkipNone comparée à la variante de SkipSenders.

Propriété 2.3.2. *Il existe une famille infinie de graphes $G = (V, E)$ de taille $n = |V|$ croissante, et $s, t \in V$ tel que $t_{\text{SN}}(G, s, t) = \Omega(n)$ tandis que $t_{\text{SS}}(G, s, t) = O(\log n)$.*

Démonstration. Soit G le graphe à $n = 2^{k+1} + 1$ nœuds composé d'un arbre binaire complet T de profondeur k enraciné en s, un simple nœud t, et un nœud universel u (c'est-à-dire u est relié à t et à tous les nœuds dans t). Voir la figure 2.2a.

Dans le modèle SkipNone, le schéma de diffusion suivant prend $n + 2$ étapes pour transmettre le message de source s à t dans T. À la réception du message, un nœud dans T appelle d'abord son père en T, puis il appelle ses deux fils dans T dans un ordre arbitraire, et enfin il appelle le nœud universel u. Lors de la réception du message, u appelle tous les nœuds de l'arbre dans un ordre arbitraire, avant de finalement appeler t. Dans ce protocole, le nœud u reçoit le message de s à l'étape 3. Puis, il faut $2^{k+1} - 1$ étapes à u pour appeler tous les autres nœuds de l'arbre binaire avant d'appeler t. Enfin, le nœud t se trouve informé à l'étape $3 + 2^{k+1} - 1 + 1 = n + 2$.

D'autre part, dans le modèle SkipSenders, chaque protocole prend au plus $3k + 2$ étapes. En effet, si un nœud interne v de l'arbre T est informé à l'étape r, alors toutes ses arêtes incidentes sont consommées à l'étape $r + 3$, parce que soit $v = s$ et $\deg(v) = 3$, soit $v \neq s$ et $\deg(v) = 4$ où l'une de ses arêtes incidentes est celle à travers laquelle v a reçu le message. Si une feuille v de l'arbre T est informée à l'étape r, alors toutes ses arêtes incidentes sont consommées à l'étape $r + 1$, car $\deg(v) = 2$. Ainsi, après $3k + 1$ étapes, toutes les arêtes incidentes à chaque nœud de l'arbre ont été consommées. Ainsi, après au plus $3k + 2$ étapes, le nœud u a appelé t. □

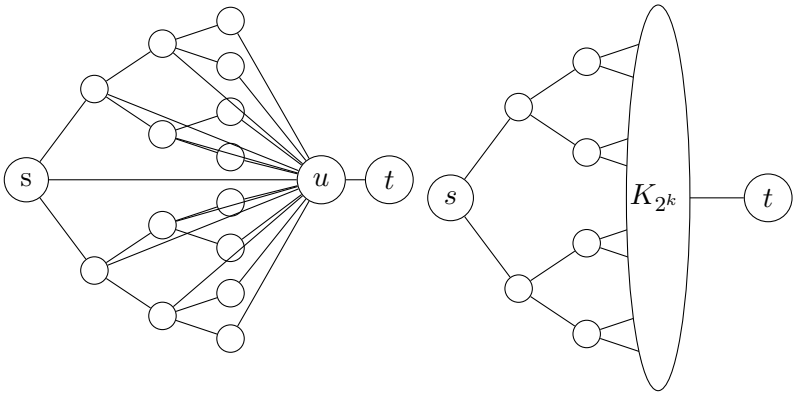
Le résultat suivant traite de la variante SkipSenders comparée à la variante SkipInformed

Propriété 2.3.3. *Il existe une famille infinie de graphe $G = (V, E)$ de taille n croissante, et $s, t \in V$ tel que $t_{\text{SS}}(G, s, t) = \Omega(n)$ tandis que $t_{\text{SI}}(G, s, t) = O(\log n)$.*

Démonstration. Soit G le graphe à $n = 2^{k+1}$ nœuds composé d'un arbre binaire complet T de profondeur k enraciné en s, avec toutes ses feuilles connectées en clique de 2^k nœuds, plus un nœud t relié à l'une des feuilles de T. Voir la figure 2.2b.

Dans le modèle SkipSenders, même si chaque nœud sauf t est initialement informé, il existe un protocole nécessitant au moins $\frac{n}{2}$ étapes supplémentaires pour informer t. Le protocole fonctionne comme suit. Soit $l_0, \ldots, l_{\frac{n}{2}-1}$ les feuilles de T. À l'étape $r = 1, \ldots, \frac{n}{2}$, les deux feuilles l_i et l_j, s'appellent l'une et l'autre si et seulement si $i = j \oplus r$, où $\oplus$ désigne le OU exclusif bit-à-bit. Les $\frac{n}{2}$ étapes correspondant à ce protocole forment $\frac{n}{2}$ couplages disjoints où, pour chaque correspondance, les deux nœuds couplés u et v s'appellent l'un l'autre. Par conséquent, $t_{\text{SS}}(G, s, t) \geq \frac{n}{2}$.

D'autre part, dans le modèle SkipInformed, tous les nœuds de t sont au courant d'une information initiée en s après au plus $2k$ étapes, pour n'importe quel protocole de diffusion (ListBased). Ainsi $t_{\text{SI}}(G, s, t) \leq 2k + 1 = 2 \log n - 1$. □



(a) Entre SkipNone et SkipSenders (b) Entre SkipNone et SkipSenders

Figure 2.2 – Écarts exponentiels

Remarque. En utilisant le même graphe que dans la Propriété 2.3.2, on peut prouver qu'il existe un graphe G tel que $t_{\text{SN}}(G) = \Omega(n)$ tandis que $t_{\text{SS}}(G) = O(\log n)$. En revanche, il n'est pas évident si un tel écart exponentiel existe pour t_{SS} par rapport à t_{SI}. Le mieux que nous sachions, c'est l'existence d'un graphe G tel que $t_{\text{SS}}(G) = \Omega(n)$ tandis que $t_{\text{SI}}(G) = O(\sqrt{n})$. Par exemple, on pourra prendre le produit cartésien d'une clique et d'un chemin, tous deux de taille $\sqrt{n}$.

2.3.3 Les diodes

Dans ce chapitre, nous définissons un type particulier de structures, appelées ***diodes***, qui nous seront utiles pour nos résultats d'inapproximabilité.

Définition. *Un graphe G est une (t_1, t_2)-diode du modèle* $\text{M} \in \{\text{SN}, \text{SS}, \text{SI}\}$ *s'il vérifie la propriété suivante. Il existe deux nœuds distincts v_{in} et v_{out}, chacun de degré 1, respectivement appelés* entrée *et* sortie*, tels que*

$$t_{\text{M}}(G, v_{in}, v_{out}) \leq t_1 \text{ et } t_{\text{M}}(G, v_{out}, v_{in}) \geq t_2.$$

Lorsque l'on considère une famille infinie de (t_1, t_2)-diodes, les $(\Theta(\log n), \Theta(n))$-diodes de taille n sont simplement appelées diodes.

L'hypothèse que v_{in} et v_{out} soient des sommets pendants est importante pour pouvoir ensuite réutiliser une diode dans un graphe plus grand tout en préservant les propriétés de la diode.

Propriété 2.3.4. *Pour* SkipNone *et* SkipInformed *il existe une famille infinie de diodes.*

Démonstration. Pour SkipNone, il suffit de reprendre la famille de graphe de la Proposition 2.3.2 où u n'est en fait universel que pour les feuilles, $v_{in} = t$ et v_{out}

est un nouveau sommet pendant accroché à s. Pour SkipInformed, il suffit de reprendre la famille de graphe de la Proposition 2.3.3 où v_{in} est un nouveau sommet pendant accroché à s et $v_{out} = t$. □

Propriété 2.3.5. *Il n'existe pas de diode dans le modèle* SkipSenders.

En fait cette propriété est un corollaire du résultat suivant qui est plus fort et indique formellement la propriété de symétrie essentielle satisfaite par le modèle SkipSenders.

Théorème 2.3.6. *Soit G un graphe connecté et soit s, t deux nœuds de G de degré* 1. *Alors $t_{\text{SS}}(G, s, t) = t_{\text{SS}}(G, t, s)$.*

Démonstration. Soit $G = (V, E)$ un graphe connexe et soient $s, t \in V$, tous deux de degré 1. Nous montrons que s'il existe un ensemble de listes $\mathcal{L}$ assurant que l'information circule de s vers t en exactement $r = t_{\text{SS}}(G, \mathcal{L}, s, t)$ étapes, alors il existe un ensemble de listes $\mathcal{L}'$ assurant que l'information circule de t vers s en au moins r étapes. Les listes de $\mathcal{L}'$ sont construites à partir des listes de $\mathcal{L}$ comme suit. Notons $\tau(u)$ l'étape au cours de laquelle $u \in V$ se trouve informé dans $\mathcal{L}$. Puis désignons par u_0 un des (éventuellement nombreux) nœuds qui envoient le message à u à l'étape $\tau(u)$. Selon $L_u \in \mathcal{L}$, le nœud u transmet cette information à ses voisins au cours des étapes subséquentes à l'étape $\tau(u)$. Soit u_i, $1 \leq i \leq k$ avec $k < \deg(u)$, le voisin de u ayant reçu l'information de u à l'étape $\tau(u) + i$. Nous allons maintenant décrire le comportement de u dans $\mathcal{L}'$. Supposons que le nœud u obtienne l'information à l'étape $\tau'(u)$ dans $\mathcal{L}'$. (Nous avons $\tau(t) = 0$). Alors u transmet cette information à ses voisins, en commençant par ceux dans $\{u_0, u_1, \ldots, u_k\}$, dans l'ordre inverse. Plus précisément, à l'étape $\tau'(u) + j$, $j \geq 1$, u transmet l'information au nœud u_i, $0 \leq i \leq k$ ayant le plus grand indice et de telle sorte que l'arête $\{u, u_i\}$ n'a pas encore été utilisée (soit pour transmettre de u vers u_i ou de u_i vers u). Lorsque toutes les arêtes $\{u, u_i\}$, $0 \leq i \leq k$, ont été traversées par l'information, le nœud u poursuit la transmission de l'information à ses voisins dans un ordre arbitraire, en suivant la règle de SkipSenders, jusqu'à ce que toute la liste de ses voisins soit épuisée. Ceci termine la description de $\mathcal{L}'$.

Pour prouver que $\mathcal{L}'$ satisfait la propriété désirée, notons $\tau(e)$ (respectivement $\tau'(e)$) l'étape où l'arête $e \in E$ est utilisée dans $\mathcal{L}$ (respectivement dans $\mathcal{L}'$). Notez que chaque arête est utilisée dans au moins une direction et qu'aucune arête ne peut être utilisée dans les deux directions opposées à des étapes différentes (par définition du modèle SkipSenders). Nous montrons que, pour chaque arête $e \in E$, nous avons

$$\tau(e) + \tau'(e) \geq r + 1 \ . \tag{2.1}$$

Par exemple, cette inégalité est vraie pour l'unique arête e_t incidente à t par définition $\tau(e_t) = r$ et puisque nécessairement $\tau'(e_t) = 1$. Dans le but d'obtenir une contradiction, supposons qu'il existe une arête e qui satisfait $\tau(e) + \tau'(e) < r + 1$. Parmi toutes ces arêtes, soit $e = \{u, v\}$ celle pour laquelle $\tau'(e)$ est minimal. Sans

perte de généralité, nous supposons que u envoie l'information à v à cette étape $\tau'(e)$ dans $\mathcal{L}'$. Nous établissons la contradiction en montrant qu'il existe un autre voisin v' de u tel que, pour $e' = \{u, v'\}$, nous avons $\tau(e') + \tau'(e') < r + 1$ avec $\tau'(e') < \tau'(e)$.

Pour prouver l'existence du nœud v', nous considérons le statut de v parmi les voisins de u. En utilisant les mêmes notations que ci-dessus, soit u_0 un nœud informant u à l'étape $\tau(u)$, et soit u_i, $1 \leq i \leq k$ avec $k < \deg(u)$, le voisin de u recevant l'information de u à l'étape $\tau(u) + i$ dans $\mathcal{L}$. Pour $i = 0, \ldots, k$, soit $e_i = \{u, u_i\}$. Si $v \notin \{u_0, u_1, \ldots u_k\}$, alors $\tau'(e) > \tau'(e_0)$ puisque, par la construction de $\mathcal{L}'$, u appelle $u_0, u_1, \ldots, u_k$ avant d'appeler d'autres voisins. Par ailleurs, $\tau(e) \geq \tau(e_0)$ par définition du nœud u_0. Ainsi, e_0 satisfait $\tau(e_0) + \tau'(e_0) < r + 1$ et ainsi $v' = u_0$ donne la contradiction.

Supposons maintenant que $v \in \{u_0, u_1, \ldots u_k\}$. Si $v = u_i$ avec $i \neq k$, alors, puisque $\tau(e_i) + \tau'(e_i) < r + 1$ et $\tau'(e_i) \geq \tau'(e_{i+1}) + 1$, nous obtenons que

$$\tau(e_{i+1}) + \tau'(e_{i+1}) = \tau(e_i) + 1 + \tau'(e_{i+1}) \leq \tau(e_i) + 1 + \tau'(e_i) - 1 < r + 1 \ .$$

Ainsi $v' = u_{i+1}$ donne la contradiction.

Enfin, si $v = u_k$, alors $\tau'(e_k) = \tau'(u) + 1$. Ceci vient du fait que, d'une part, nous avons supposé que u est un expéditeur par e et, d'autre part, e_k est par construction de $\mathcal{L}'$ la première arête utilisée par u. Alors, soit v'' et $e'' = \{u, v''\}$ tels que $\tau'(e'') = \tau'(u)$. Nous savons que $\tau(e'') \leq \tau(u) + k$. Ainsi

$$\tau(e'') + \tau'(e'') \leq \tau(u) + k + \tau'(u) = \tau(e_k) + \tau'(e_k) - 1 < r \ .$$

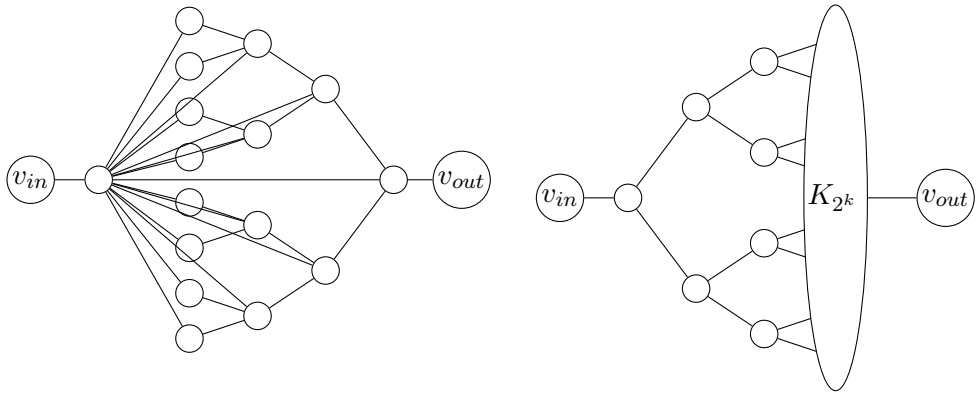
Puisque $\tau'(e'') < \tau'(e_k)$, nous obtenons que $v' = v''$ donne la contradiction. Par conséquent, l'inégalité (2.1) est vraie. Elle nous permet de compléter la preuve en considérant l'unique arête e_s incidente à s. Nous avons $\tau'(e_s) = \tau'(s)$ et $\tau(e_s) = 1$. Puisque $\tau(e_s) + \tau'(e_s) \geq r + 1$, nous obtenons que l'information circule de t à s en au moins r étapes. □

2.4 Un algorithme polynomial pour SkipNone

Cette section est dédiée au premier résultat principal de ce chapitre. Nous montrons que, pour SkipNone, il existe un algorithme polynomial qui trouve un pire schéma de diffusion.

Théorème 2.4.1. *Il existe un algorithme s'exécutant en temps $O(m + n \log n)$ sur les graphes de n nœuds et de m arêtes qui, pour tout graphe $G = (V, E)$, et toute cible $t \in V$, calcule les n valeurs $t_{\text{SN}}(G, s, t)$, $s \in V$.*

Démonstration. Nous décrivons un algorithme permettant d'atteindre la performance indiquée dans l'assertion du théorème. Cet algorithme est détaillé dans l'algorithme 1. À chaque nœud $v \in V$ est associé trois variables : un entier positif $value(v)$ un booléen



(a) SkipNone (b) SkipInformed

Figure 2.3 – Diodes

$tag(v)$, et une variable auxiliaire $maxlab(v) \in \{1, \ldots, \deg(v)\}$. Nous allons démontrer que lorsque l'algorithme termine, nous avons $t_{\text{SN}}(G, s, t) = value(s)$ pour chaque $s \in V$. Le booléen $tag(v)$ indique si la valeur $value(v)$ est figée ($tag(v) = 1$) ou si elle peut encore être mise à jour ($tag(v) = 0$). La variable auxiliaire $maxlab(v)$ mémorise l'*étiquette* maximale qui peut actuellement être assignée aux arêtes incidentes à v au cours de l'exécution de l'algorithme. En effet, à chaque arête $\{u, v\} \in E$ sont associées deux étiquettes $label(u, v)$ et $label(v, u)$. Nous allons montrer qu'une fois l'algorithme terminé, pour chaque nœud u, nous avons toutes les étiquettes $label(u, v)$ deux à deux distinctes pour tous les nœuds v dans le voisinage $N(u)$ de u, avec des valeurs comprises entre 1 et $\deg(u)$. L'étiquette $label(u, v)$ représente le "poids" de l'arête $\{u, v\}$ dans le sens de u vers v, défini comme délai par rapport à la réception du message par u après lequel u transmet ce message à v.

Premièrement, observons que par le paramétrage de $label(u, v)$ dans la boucle *while* de l'algorithme 1, nous savons que $1 \leq label(u, v) \neq label(u, v') \leq \deg(u)$ pour toute paire de voisins v, v' de u. En utilisant ce fait, nous obtenons directement que $t_{\text{SN}}(G, s, t) \geq value(s)$ pour tout s. Afin de voir pourquoi ceci est vrai, nous considérons le schéma de diffusion où chaque nœud u recevant le message au moment τ transmet le message à $v \in N(u)$ au temps $\tau + label(u, v)$, où $N(u)$ désigne l'ensemble des voisins de u. Cette diffusion est compatible avec le modèle SkipNone. Par conséquent, $t_{\text{SN}}(G, s, t) \geq \text{dist}_D(s, t)$ où la distance est calculée dans le graphe orienté pondéré D obtenue en remplaçant chaque arête $\{u, v\}$ par deux arcs (u, v) et (v, u), avec des poids respectifs $label(u, v)$ et $label(v, u)$. À présent, par construction, nous obtenons que $\text{dist}_D(s, t) = value(s)$ puisque l'assignation de $value(s)$ effectuée par l'algorithme 1 est réalisée d'une manière similaire à l'algorithme de Dijkstra, appliqué pour le calcul des plus courts chemins vers une unique destination dans les graphes orientés avec des poids positifs. Dans notre algorithme, les poids des arêtes sont les étiquettes des arêtes, qui sont déterminées à la volée lors de l'exécution de l'algorithme.

```
Entrée : G = (V, E), et t ∈ V;
Sortie : t_SN(G, s, t) pour chaque s ∈ V;

value(t) ← 0;
tag(t) ← 0;
forall the u ∈ V \ {t} do
    value(u) ← ∞;
    tag(u) ← 0;
    maxlab(u) ← deg(u)
end
while ∃u | tag(u) = 0 do
    v ← argmin{value(u) | tag(u) = 0};
    forall the u ∈ N(v) do
        label(u, v) ← maxlab(u);
        maxlab(u) ← maxlab(u) − 1;
        value(u) ← min{value(u), label(u, v) + value(v)}
    end
    tag(v) ← 1;
end
Sortie : value(s) pour chaque s ∈ V.
```

Algorithme 1: Algorithme de calcul de t_{SN}

Ainsi, $t_{\text{SN}}(G, s, t) \geq value(s)$.

Nous allons maintenant montrer que $t_{\text{SN}}(G, s, t) \leq value(s)$ pour tout s. Notez que la valeur d'un nœud ne peut que diminuer lors de l'exécution de l'algorithme, et qu'une fois un nœud marqué 1 ($tag(u) = 1$), sa valeur reste inchangée jusqu'à la fin de l'algorithme. De plus, la propriété suivante est vérifiée :

($\star$) une fois qu'un nœud est marqué 1, sa valeur est bornée supérieurement par la valeur finale de tous les nœuds encore marqués 0.

En effet, au moment où un nœud v est marqué 1, il a la valeur minimum $x = value(v)$ parmi tous les nœuds actuellement marqués 0. Considérons u avec $tag(u) = 0$. Par le choix de v, nous avons $value(u) \geq x$. Si $value(u)$ est éventuellement modifiée, la nouvelle valeur est de la forme $label(u, w) + value(w)$ où $tag(w) = 0$. Puisque $tag(w) = 0$, nous avons $value(w) \geq x$, et donc la nouvelle $value(u)$ est au moins $x + 1$.

Supposons qu'il existe un nœud s tel que $value(s) < t_{\text{SN}}(G, s, t)$ et montrons la contradiction. Parmi ces nœuds, soit v le premier nœud marqué 1. Au moment où v est marqué 1, une partie de ses voisins $w_1, \ldots, w_{k_0}$ sont marqués 0 tandis que d'autres $w_{k_0+1}, \ldots, w_{k_0+k_1}$ sont marqués 1, avec $k_0 + k_1 = \deg(v)$. Si v n'a pas de voisins marqués 1, alors soit $v = t$ et $value(t) = t_{\text{SN}}(G, t, t) = 0$, soit $value(v) = \infty$. Dans les deux cas nous avons $value(v) \geq t_{\text{SN}}(G, v, t)$. Ainsi, nous avons $k_1 > 0$. Supposons, sans perte de généralité, que les nœuds w_i sont indexés de telle sorte que pour $1 \leq i < k_0 + k_1$, w_i est (finalement) marqué 1 après que w_{i+1} le soit. Par conséquent,

pour tout i, $1 \leq i \leq k_0 + k_1$, par l'affectation des étiquettes en décroissant de $\deg(v)$ à 1, nous avons :

($\star\star$) si w_i est marqué 1, alors $label(v, w_i) = i$,

et par conséquence directe de ($\star$), nous obtenons

($\star\star\star$) si w_{i+1} est aussi marqué 1, alors $value(w_i) \geq value(w_{i+1})$.

Nous sommes maintenant prêts à établir une contradiction. Soit Σ l'ensemble des permutations de $1, \ldots, \deg(v)$. Nous avons

$$t_{\text{SN}}(G, v, t) \leq \max_{\pi \in \Sigma} \min_{1 \leq i \leq k_0 + k_1} (\pi(i) + t_{\text{SN}}(G, w_i, t))$$

où $\pi(i)$ est le moment où v envoie l'information vers w_i. Nous obtenons donc :

$$\begin{aligned}
t_{\text{SN}}(G, v, t) &\leq \max_{\pi \in \Sigma} \min_{1 \leq i \leq k_1} (\pi(k_0 + i) + t_{\text{SN}}(G, w_{k_0+i}, t)) \\
&\leq \max_{\pi \in \Sigma} \min_{1 \leq i \leq k_1} (\pi(k_0 + i) + value(w_{k_0+i})) \text{ parce que } v \text{ est le premier} \\
&\qquad \text{nœud marqué 1 parmi ceux vérifiant } value(v) < t_{\text{SN}}(G, v, t) \\
&= \min_{1 \leq i \leq k_1} (k_0 + i + value(w_{k_0+i})) \text{ par } (\star\star\star) \\
&= \min_{1 \leq i \leq k_1} (label(v, w_{k_0+i}) + value(w_{k_0+i})) \text{ par } (\star\star) \\
&= value(v)
\end{aligned}$$

Ainsi, nous obtenons $t_{\text{SN}}(G, v, t) \leq value(v)$, une contradiction.

Enfin, le temps d'exécution de l'algorithme 1 est $O(m + n \log n)$, tout comme le temps d'exécution de l'algorithme de Dijkstra. (Ici, nous supposons que les nœuds dont la marque est nulle sont stockés dans une file d'attente avec priorité implémentée par un tas de Fibonacci). □

Corollaire 2.4.2. *Il existe un algorithme s'exécutant en temps $O(n(m + n \log n))$ sur un graphe à n nœuds et m arêtes qui, pour tout graphe G, calcule $t_{\text{SN}}(G)$.*

2.5 Inapproximabilité de SKIPSENDERS et SKIPINFORMED

Dans cette section, nous montrons que, par opposition à la variante SKIPNONE du protocole LISTBASED, les deux autres variantes du modèle sont difficiles à approximer. Étant donné un problème de maximisation $\mathcal{P}$, un algorithme (en temps polynomial) $\mathcal{A}$ est considéré comme un algorithme ρ-approximant (algorithme d'approximation avec facteur ρ) de $\mathcal{P}$ si pour toute instance x de $\mathcal{P}$, nous avons

$$\rho \cdot \text{OPT}(x) \leq \mathcal{A}(x) \leq \text{OPT}(x).$$

Tout d'abord, nous montrons que la performance au pire cas du protocole LISTBASED dans la variante SKIPINFORMED est essentiellement inapproximable.

Théorème 2.5.1. *À moins que* $\mathrm{P} = \mathrm{NP}$, $t_{\mathrm{SI}}(G, s, t)$ *ne peut pas être approximé en temps polynomial à un facteur* $(\frac{1}{n})^{\frac{1}{2}-\epsilon}$ *multiplicatif près, pour tout* $\epsilon > 0$.

Démonstration. La preuve est par réduction à partir de 3-SAT. En fait, pour plus de commodité, nous faisons une réduction de la négation d'E3-SAT (exactement 3-SAT), notée E3-DNF-UNSAT. Soit Φ une formule E3-DNF, c'est-à-dire une disjonction de clauses où chaque clause est une conjonction d'exactement trois littéraux. Φ est dans E3-DNF-UNSAT si et seulement s'il existe une assignation de ses variables telle que Φ est fausse.

Nous utilisons le gadget suivant, appelé diode D. La diode D est un graphe de taille $2\nu + 1$, pour tout ν une puissance de 2, avec deux nœuds identifiés appelés *entrée* et *sortie*, tous deux de degré 1. Nous l'obtenons en prenant un arbre binaire complet T avec ν feuilles en (1) connectant toutes les feuilles de T comme un graphe complet K_ν, (2) ajoutant un nœud d'entrée v_{in} connecté à la racine de T, et (3) ajoutant un nœud de sortie v_{out} connecté à une feuille arbitraire de T. Il est aisé de vérifier que $t_{\mathrm{SI}}(D, v_{in}, v_{out}) = \Theta(\log \nu)$ et que $t_{\mathrm{SI}}(D, v_{out}, v_{in}) = \nu + \Theta(\log \nu)$. Ainsi D est bien une diode.

Soit Φ une formule E3-DNF formule avec p variables $x_1, \ldots, x_p$ et q clauses $C_1, \ldots, C_q$. Étant donnée Φ, nous construisons un graphe G avec deux nœuds identifiés s and t, tels que $t_{\mathrm{SI}}(G, s, t)$ est polynomial si Φ est dans E3-DNF-UNSAT et logarithmique sinon. Une représentation graphique de G est donnée à la figure 2.4. À chaque variable x_i correspond un nœud x_i dans G appelé nœud-variable. Le nœud source s est la racine d'un arbre binaire complet A de profondeur $\lceil \log p \rceil$, et dont chaque nœud-variable est une des feuilles. Soit $\nu \geq \max\{p, q\}$. Chaque nœud-variable x_i est l'entrée de 2ν diodes différentes, chacune ayant $2\nu + 1$ nœuds. Deux nœuds x_i^0 et x_i^1, appelés respectivement *nœud-faux* et *nœud-vrai*, sont associés à chaque variable x_i, $1 \leq i \leq p$. (Ils correspondront aux deux assignations possibles de x_i, vrai ou faux). Les nœuds x_i^0 et x_i^1 sont les racines d'arbres binaires complets T_i^0 et T_i^1 respectivement, chacun ayant ν feuilles. Les 2ν feuilles de T_i^0 et de T_i^1 sont les 2ν sorties des 2ν diodes correspondant à x_i. Les nœuds x_i^0 et x_i^1 sont aussi les racines des arbres binaires complets S_i^0 et S_i^1 respectivement, chacun de profondeur $\lceil \log q \rceil$. Les q premières feuilles de l'arbre S_i^α sont notées $x_{i,1}^\alpha, \ldots, x_{i,q}^\alpha$, pour $\alpha \in \{0, 1\}$, où $x_{i,l}^\alpha$ correspond à la variable x_i et la clause C_l, $1 \leq i \leq p$ et $1 \leq l \leq q$. Supposons que le littéral x_i apparait dans C_l. S'il apparait positivement (respectivement négativement) alors $x_{i,l}^1$ (respectivement $x_{i,l}^0$) est la racine d'un arbre binaire complet $R_{i,l}^1$ (respectivement $R_{i,l}^0$) de ν feuilles et dont les feuilles sont connectées en clique. Ces feuilles sont appelées *dupliqués* de $x_{i,l}^1$ (respectivement de $x_{i,l}^0$). À chaque clause C_l correspond un nœud C_l de G, appelé nœud-clause. Réécrivons la clause de façon à ce qu'un littéral x apparaissant positivement dans une clause C est écrit x^1, tandis qu'un littéral x apparaissant négativement dans une clause C est écrit x^0, c'est-à-dire que x devient x^1 et $\bar{x}$ devient x^0. Nous connectons le nœud-clause $C_l = x_i^{\alpha_i} \wedge x_j^{\alpha_j} \wedge x_k^{\alpha_k}$ à tous les dupliqués de $x_{i,l}^{\alpha_i}$, $x_{j,l}^{\alpha_j}$ et $x_{i,k}^{\alpha_k}$ respectivement. Finalement, le nœud cible t est la racine d'un arbre binaire complet B de profondeur $\lceil \log q \rceil$ dont les feuilles sont les q

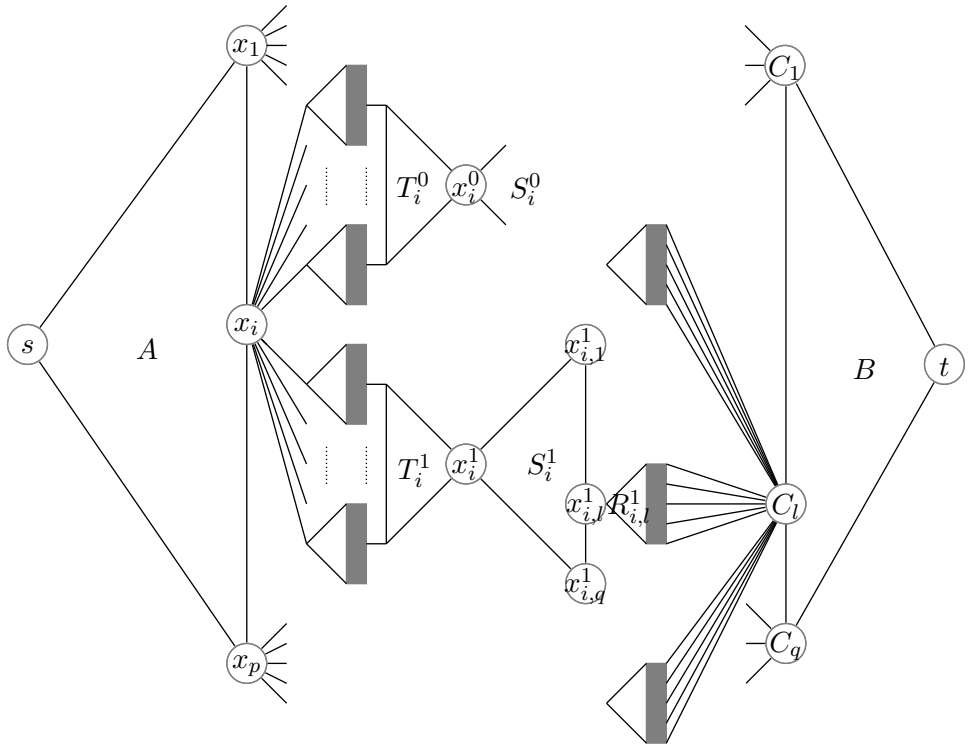


FIGURE 2.4 – Représentation graphique du gadget utilisé dans la preuve du théorème 2.5.1

nœuds-clauses . Le graphe G a $n = 4p\nu^2 + 4p\nu + 6q\nu + O(pq)$ nœuds.

Maintenant, nous allons montrer que $t_{\text{SI}}(G, s, t)$ change radicalement selon que la formule Φ soit insatisfiable ou non. Notez qu'une fois qu'un nœud intérieur de n'importe lequel des arbres binaires complets dans G est informé, tous les nœuds de cet arbre sont informés en un nombre d'étapes logarithmique en la taille de l'arbre, donc en $O(\log \nu)$ étapes.

Tout d'abord, nous montrons que si Φ est insatisfiable alors $t_{\text{SI}}(G, s, t) \geq \nu$. Le fait que Φ soit insatisfiable signifie qu'il y a une assignation $x_i^{\alpha_i}$, $\alpha_i \in \{0, 1\}$, des p variables de sorte que Φ soit fausse. Fixons une telle assignation. Une fois que le nœud-variable x_i est informé, il commence à transmettre le message aux ν entrées des ν différentes diodes menant vers $x_i^{\alpha_i}$, ce qui prend ν étapes. (Puisque $t_{\text{SI}}(D, v_{out}, v_{in}) > \nu$ pour une diode D, ces ν entrées ne reçoivent pas le message avant qu'il ne soit reçu en provenance x_i). Puis, $x_i^{\alpha_i}$ se trouve informé à l'étape $\Theta(\log \nu)$. Cependant, le nœud $x_i^{1-\alpha_i}$ n'est informé que ν étapes plus tard. Ce délai est le même pour leurs dupliqués respectifs. Puisque Φ est faux, chaque clause est fausse. Ainsi, dans toute clause C_l, d'une part, si tous les littéraux sont faux alors le nœud-clause C_l est informé après au moins ν étapes. D'autre part, si au moins un littéral est vrai, alors C_l est informé en $\Theta(\log \nu)$ étapes. Prenons les ν dupliqués du nœud $x_{i,l}^{\alpha_i}$ avec

$x_i^{\alpha_i}$ faux dans la clause C_l pour l'assignation. Ces doublons forment une clique. Il est possible d'informer un seul dupliqué par étape. En effet, jusqu'à l'étape ν, le nœud-clause C_l et les dupliqués informés envoient le message au même dupliqué non informé. En raison du retard de ν étapes causé par x_i lors de l'envoi du message aux ν diodes, ces dupliqués ne seront pas informés par l'intermédiaire de l'arbre $R_{i,l}^{\alpha_i}$ avant l'étape ν. Par conséquent, le nœud-clause C_l peut attendre jusqu'à l'étape (au moins) ν avant d'envoyer le message à un nœud intérieur de l'arbre B enraciné en t. Dans l'ensemble, t ne reçoit pas le message avant ν étapes.

Ensuite, nous montrons que si Φ n'est pas insatisfiable alors $t_{\mathrm{SI}}(G, s, t) = O(\log \nu)$. Le fait que Φ n'est pas insatisfiable signifie que, pour chaque assignation des variables, Φ est vrai, c'est-à-dire qu'il y a une clause dont les trois littéraux sont vrais. Fixons un schéma de diffusion de s à t dans G, noté $\mathcal{S}$. Par construction de G, chaque nœud-variable se trouve informé en $O(\log \nu)$ étapes de $\mathcal{S}$. Ainsi, nous ne se soucions que du premier nœud informé par chaque nœud-variable dans $\mathcal{S}$. Si ce nœud est la racine de l'arbre de la diode conduisant à son nœud-vrai, alors la variable correspondante est affectée à vrai, sinon elle est affectée à faux. De cette façon, $\mathcal{S}$ obtient une affectation des p variables. Puisque Φ n'est pas insatisfiable, il y a une clause C_r telle que ses trois littéraux sont vrais. Les trois nœuds-vrais ou nœuds-faux qui font que C_r soit vrai sont informés en $O(\log \nu)$ étapes. La même chose vaut pour tous les dupliqués correspondants à ces trois nœuds. Par conséquent, après $O(\log \nu)$ étapes, C_r doit envoyer le message dans l'arbre B enraciné en t. Ainsi t est informé en $O(\log \nu)$ étapes.

Nous en déduisons que, sauf si $\mathrm{P} = \mathrm{NP}$, $t_{\mathrm{SI}}(G, s, t)$ ne peut être approximé en temps polynomial à un facteur multiplicatif $(\frac{1}{\nu})^{1-\epsilon}$ près, pour tout $\epsilon > 0$. Le théorème en résulte en choisissant ν arbitrairement grand, polynomialement en p et q. □

Corollaire 2.5.2. *À moins que* $\mathrm{P} = \mathrm{NP}$, $t_{\mathrm{SI}}(G)$ *ne peut pas être approximé en temps polynomial à un facteur multiplicatif* $(\frac{1}{n})^{\frac{1}{3}-\epsilon}$ *près, pour tout* $\epsilon > 0$.

Démonstration. Nous utilisons ν fois la même construction que dans la preuve du théorème 2.5.1 mise bout-à-bout en anneau. Il suffit de remarquer que, quel que soit le premier nœud informé dans la construction de la preuve du théorème 2.5.1, t est informé en $O(\nu)$ étapes, puis de remarquer que $t_{\mathrm{SI}}(G, t, s) = \Theta(\nu)$ à cause des cliques de taille ν connectées aux nœuds-clauses. □

La construction du théoreme 2.5.1 peut aussi être utilisée pour prouver que, à moins que $\mathrm{P} = \mathrm{NP}$, $t_{\mathrm{SS}}(G, s, t)$ ne peut être approximé en temps polynomial à un facteur multiplicatif $\frac{2}{3} + \epsilon$ près, pour tout $\epsilon > 0$. En effet, avec le même graphe G, la même source et la même cible que dans la preuve, nous pouvons montrer que si Φ est insatisfiable alors $t_{\mathrm{SS}}(G, s, t) \geq 3\nu$, tandis que si Φ est insatisfiable alors $t_{\mathrm{SS}}(G, s, t) = 2\nu + \Theta(\log \nu)$. La suite améliore cette borne.

Théorème 2.5.3. *À moins que* $\mathrm{P} = \mathrm{NP}$, $t_{\mathrm{SS}}(G, s, t)$ *ne peut pas être approximé en temps polynomial à un facteur multiplicatif* $\frac{1}{2} + \epsilon$ *près, pour tout* $\epsilon > 0$, *même si* G *est un graphe biparti.*

Démonstration. Nous construisons un graphe G, de nœud source s et de nœud cible t, en modifiant la construction de la preuve du théorème 2.5.1 comme suit. Tout d'abord, nous remplaçons les diodes par des chemins de longueur $\frac{\nu}{2}$. La seconde transformation concerne les connexions entre les $x^1_{i,l}$ (respectivement $x^0_{i,l}$) et leurs dupliqués. Si x_i apparaît positivement (respectivement négativement) dans C_l, alors $x^1_{i,l}$ (respectivement $x^0_{i,l}$) est la racine d'un arbre binaire complet avec ν feuilles. Chacune de ces feuilles est une extrémité d'un chemin de longueur $\frac{\nu}{2}$. Les autres extrémités de ces chemins sont les nouveaux dupliqués de $x^1_{i,l}$ (respectivement $x^0_{i,l}$). Comme dans la démonstration du théorème 2.5.1 ces dupliqués sont connectés aux nœuds-clauses correspondants. Le graphe ainsi construit est biparti. En effet, nous montrons que l'algorithme qui 2-colorie le graphe G de manière gloutonne, en parcourant G en largeur à partir de s, termine sans erreur. Il commence en 2-coloriant l'arbre A. Comme tous les nœuds-variables sont au même niveau de l'arbre ils ont la même couleur. Chacun de ces nœuds-variables est relié à un arbre binaire complet par des chemins de longueur $\frac{\nu}{2}$. Colorions alors gloutonnement ces chemins. Leurs longueurs étant identiques, les feuilles des arbres ont la même couleur. Comme les feuilles ont la même couleur et sont au même niveau, ces arbres se colorient sans conflit. Ces arbres étant tous identiques, leurs racines qui sont les nœuds-vrais et le nœuds-faux ont a nouveau la même couleur. De même, tous les dupliqués ont la même couleur et les nœuds-clause l'autre couleur. Finalement, nous 2-colorions l'arbre B sans erreur puisque toutes ses feuilles sont au même niveau et ont la même couleur. Nous montrons que si Φ est insatisfiable alors $t_{\text{SS}}(G, s, t) \geq 2\nu$, sinon $t_{\text{SS}}(G, s, t) \leq \nu + \Theta(\log \nu)$. Afin de voir pourquoi nous avons ce résultat, nous observerons simplement ce qui suit.

(1) Une fois que le nœud-variable x_i est informé, l'un des deux nœuds x^0_i ou x^1_i est informé en $\frac{\nu}{2} + \Theta(\log \nu)$ étapes. (Ce nœud est celui qui est la racine de l'arbre relié au premier chemin choisi par x_i). Notez que le message prend au moins $\nu - O(\log \nu)$ étapes pour qu'un voisin de x_i soit informé par un autre nœud que x_i. Ainsi x_i passe ν étapes à envoyer le message aux ν chemins qui mènent vers le même nœud x^α_i. Le nœud $x^{1-\alpha}_i$ reçoit ainsi l'information avec un retard de ν étapes.

(2) Si une clause est vraie, alors au temps $\nu + \Theta(\log \nu)$ le nœud-clause correspondant n'a pas d'autres choix que d'envoyer le message à l'arbre B enraciné en t. En effet, comme observé en (1), le message prend au moins $\nu - O(\log \nu)$ étapes pour revenir au nœud-clause, ce nœud peut donc passer environ ν étapes avant d'envoyer le message dans l'arbre B enraciné en t.

Mis à part les deux modifications ci-dessus, chacune augmentant le temps de diffusion de $\frac{\nu}{2}$ étapes, la preuve découle des mêmes arguments que dans la démonstration du théorème 2.5.1. □

La même construction donne le même résultat pour le modèle SKIPINFORMED.

Corollaire 2.5.4. *À moins que* $\text{P} = \text{NP}$, $t_{\text{SI}}(G, s, t)$ *ne peut pas être approximé en temps polynomial à un facteur multiplicatif* $\frac{1}{2} + \epsilon$ *près, pour tout* $\epsilon > 0$, *même si* G *est un graphe biparti.*

Corollaire 2.5.5. *À moins que* $\mathrm{P} = \mathrm{NP}$, $t_{\mathrm{SS}}(G)$ *ne peut pas être approximé en temps polynomial à un facteur multiplicatif* $\frac{1}{2} + \epsilon$ *près, pour tout* $\epsilon > 0$.

Démonstration. Nous utilisons la même construction que dans la preuve du théorème 2.5.3 mise bout-à-bout autant de fois que nécessaire pour que le temps de diffusion dans les deux copies à l'extrémité de la chaîne soit négligeable (par exemple n^2) et nous utilisons le théorème 2.3.6 pour justifier que les temps sont les mêmes, quel que soit le sens de parcours dans la chaîne considérée. □

2.6 Extensions

Dans cette section, après une discussion sur le cas orienté, nous donnons quelques gadgets utiles pour lier les résultats connus aux variantes multisources et multidestinaires. Ensuite nous donnons un lien entre les modèles SkipSenders et SkipInformed. Et nous finissons par un dernier résultat concernant SkipInformed, il permet de relier le problème où la source et la cible sont fixés, au problème pour la pire source et la pire cible.

2.6.1 Le cas des graphes orientés

Les résultats de notre étude peuvent être étendus aux graphes orientés. Dans le contexte de graphes orientés, le modèle SkipNone reste praticable (c'est-à-dire résoluble en temps polynomial) : l'algorithme et la preuve qui l'accompagne fonctionnent exactement de la même manière. En fait l'algorithme donné dans le cas non orienté, agit dessus comme si c'était un graphe orienté sauf que les arcs sont toujours dans les deux sens.

Les modèles SkipSenders et SkipInformed restent difficiles à approximer. En fait, il est possible de montrer que, à la fois pour les modèles SkipSenders et SkipInformed, les pires performances de cas du protocole ListBased ne peuvent être approximées en temps polynomial à un facteur multiplicatif $(\frac{1}{n})^{1-\epsilon}$, pour tout $\epsilon > 0$. En effet, dans le cas de graphes orientés, il est facile de construire des diodes, le gadget de base utilisé dans la démonstration du théorème 2.5.1, tandis que les diodes n'existent pas dans les graphes non orientés pour le modèle SkipSenders (cf. le théorême 2.3.6). Nous obtenons donc un bien meilleur facteur d'inapproximabilité, même dans le modèle SkipInformed, où une diode existe déjà dans le cas non orienté, car dans le cas orienté on peut construire des diodes de taille constante, en prenant un chemin orienté de taille 2.

Théorème 2.6.1. *À moins que* $\mathrm{P} = \mathrm{NP}$, *dans le cas orienté* $t_{\mathrm{SS}}(G, s, t)$ *et* $t_{\mathrm{SI}}(G, s, t)$ *ne peuvent pas être approximés en temps polynomial à un facteur* $(\frac{1}{n})^{1-\epsilon}$ *multiplicatif près, pour tout* $\epsilon > 0$, *même si* G *est fortement connexe ou si* G *est acyclique.*

Démonstration. Nous construisons un graphe G, de nœud source s et de nœud cible t, en modifiant la construction de la preuve du théorème 2.5.1 tel que si Φ est

insatisfiable alors $t_{\mathrm{SS}}(G, s, t)$ et $t_{\mathrm{SI}}(G, s, t)$ sont de l'ordre de $\Theta(n)$, et de l'ordre de $\Theta(\log \nu)$, sinon.

La source s est la racine d'un arbre dont les nœuds-variables sont les racines, par contre l'arbre est orienté de la racine vers les feuilles (c'est-à-dire de s vers les nœuds-variables). Ensuite, chaque nœud-variable est relié à son nœud-vrai et son nœud-faux par ν chemins orientés, qui remplacement les diodes. Puis nous suivons la même construction que dans le théorème 2.5.1 jusqu'aux dupliqués, sauf que nous orientons les arbres S_i^α de i^α vers les feuilles, et les arbres $R_{i,l}^\alpha$ de $x_{i,l}^\alpha$ vers les feuilles qui sont les dupliquées. Enfin chaque dupliqué est relié au nœud-clause correspondant par un arc double, c'est-à-dire un arc du dupliqué vers le nœud-clause et un arc du nœud-clause vers le dupliqué. Ici, grâce à l'orientation des arcs des arbres, nous n'avons pas besoin de relier les dupliquées en clique pour ralentir le processus, puisque nous avons la garantie que si un dupliqué est informé par le nœud-clause, il ne peut informer aucun autre nœud. Enfin nous finissons comme dans la construction du théorème 2.5.1 à l'aide d'un arbre, cette fois-ci orienté des feuilles, les nœuds-clauses, vers la racine t. Le graphe G a $n = (2p + 6q)\nu + O(pq)$ nœuds. Un analyse similaire montre alors que la preuve du théorème 2.5.1 donne que si Φ est insatisfiable alors $t_{\mathrm{SI}}(G, s, t) \geq \nu$ et si Φ n'est pas insatisfiable alors $t_{\mathrm{SI}}(G, s, t) = O(\log \nu)$

Pour avoir le résultat dans le cas fortement connexe, il suffit de rajouter un arc de t vers s. Si au contraire on souhaite que G soit acyclique, nous remarquons que les seuls cycles (si on ne rajoute pas l'arc de t vers s) sont les arcs doubles entre un dupliqué et un nœud clause. Il suffit alors de remplacer ces arcs par le gadget de la figure 2.5 pour permettre au dupliqué d'informer le nœud-clause, et au nœud-clause de perdre les ν étapes dans le cas échéant. □

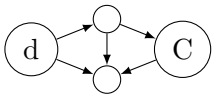


FIGURE 2.5 – Gadget remplaçant les arcs doubles entre un dupliqué d et un nœud-clause C.

2.6.2 Boîte à outils

Nous abordons ici des résultats secondaires utiles pour une analyse future de ces modèles de diffusion désordonnée ou de leurs variantes. Dans le problème de la diffusion optimale, il est possible de montrer qu'on peut considérer sans perte de généralité, que dans le schéma de diffusion optimal, toutes les communications sont utiles, c'est-à-dire que les nœuds informés n'appellent que des nœuds non informés et que chaque nœud non informé ne reçoit le message que d'un seul autre nœud. De plus, on peut supposer que toutes les étapes d'envois d'un nœud sont consécutives. En effet si un nœud fait une pause dans ses envois alors qu'il peut encore envoyer des messages utiles, la diffusion se fait au moins aussi vite s'il les fait plus tôt, sans

faire de pause. Dans le cas de la vision ListBased, on peut toujours se ramener à des listes telles que la première partie correspond à des messages utiles et la seconde à des messages inutiles, l'une ou l'autre des sous-listes pouvant être de taille nulle. Quelles sont les hypothèses du même genre pouvant permettre la simplification de l'analyse de la diffusion désordonnée ?

Lemme 2.6.2. *Dans les modèles* SkipSenders *et* SkipInformed*, il existe une liste $\mathcal{L}$ donnant un pire temps de diffusion tel que les voisins passés par u sont à la fin de la liste L_u.*

Démonstration. Si un envoi potentiel de u vers un voisin v ne se fait pas suivant la règle du modèle SkipSenders ou SkipInformed, c'est-à-dire que u passe v, alors l'envoi ne pourra se faire à aucune autre étape ultérieure. Puisque le nœud u passe ce voisin, on peut donc le placer à la toute fin de la liste sans changer le schéma effectif des envois. La liste $L_u = \{v_1, \ldots, v_i, \ldots, v_{\deg(u)}\}$ est donc telle que si u est informé à l'étape $\tau(u)$ alors il existe un temps k tel que u communique avec v_i à l'étape $\tau(u) + i$ si et seulement si $i \leq k$. □

Lemme 2.6.3. *Il existe toujours un pire schéma de diffusion dans le modèle* SkipInformed *dans lequel, à chaque étape, l'ensemble des sommets informés est minimal pour l'inclusion.*

Démonstration. Si à une étape τ dans une pire diffusion $\mathcal{L}_1$ l'ensemble I_τ est informé et qu'il existe une diffusion $\mathcal{L}_2$ n'informant qu'un sous-ensemble $J_\tau \subseteq I_\tau$, alors il existe une pire diffusion $\mathcal{L}_3$ où uniquement J_τ est informé au temps τ. Ce protocole $\mathcal{L}_3$ se comporte comme $\mathcal{L}_1$ jusqu'au temps τ puis imite $\mathcal{L}_1$. Nous montrons par récurrence que, pour tout $t \geq \tau$, l'ensemble des nœuds informés K_t satisfait $K_t \subseteq I_t$, puisqu'au temps t, si $u \in K_t \subseteq I_t$ appelle v dans $\mathcal{L}_1$, alors il peut aussi l'appeler dans $\mathcal{L}_3$. □

Lemme 2.6.4. *Il existe toujours un pire schéma de diffusion dans le modèle* SkipSenders *dans lequel, à chaque étape, l'ensemble des arêtes utilisées est minimal pour l'inclusion.*

Démonstration. La preuve est similaire à celle précédente sauf que nous nous intéressons aux arêtes. □

Ce lemme n'implique pas qu'il existe toujours un pire schéma de diffusion dans le modèle SkipSenders dans lequel, à chaque étape, l'ensemble des sommets informés est minimal pour l'inclusion. Pour s'en convaincre, nous regarderons l'analyse du modèle SkipSenders faite dans [HHL11] sur la clique à laquelle est ajouté un sommet pendant. Le schéma donné met plus de temps que celui qui imposerait à tous les nœuds de la clique d'appeler simultanément le même nœud. En effet, le dernier nœud informé dans la clique aura été appelé par tous ses voisins sauf le sommet pendant qui est le seul nœud non informé restant. Ce dernier sera alors informé à l'étape $n - 1$.

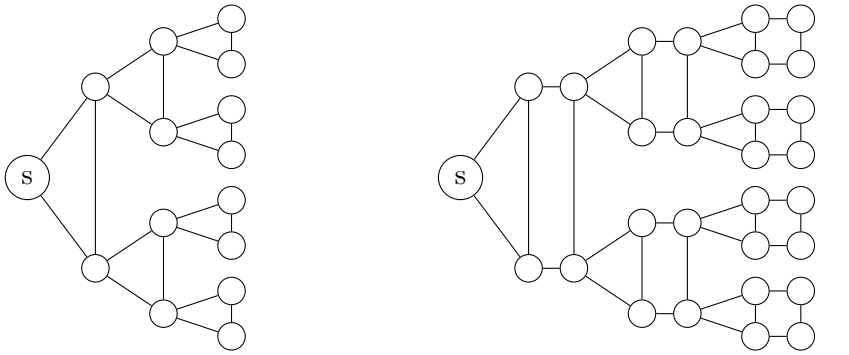
Lemme 2.6.5. *Il existe toujours un pire schéma de diffusion dans le modèle* SkipNone *dans lequel, à chaque étape, l'ensemble des arêtes utilisées est minimal pour l'inclusion. En particulier, il existe toujours un pire schéma de diffusion dans le modèle* SkipNone *dans lequel un nœud commence par appeler ceux qui lui ont envoyé l'information, puis informe des nœuds voisins déjà informés si possible avant d'informer un nouveau nœud.*

Démonstration. C'est ce qui est effectué par l'algorithme 1. L'idée est que, puisque les arêtes peuvent toujours être utilisées une fois pour chaque sens, il suffit de ne pas informer de nouveaux nœuds tant que possible. □

Appelons ***mobile en triangle*** un arbre binaire complet auquel on rajoute une arête entre deux nœuds frères (fils d'un même nœud), voir la figure 2.6a. Si la source d'un mobile en triangle est la racine alors, dans le modèle SkipInformed, à l'étape $2k$, tous les nœuds de profondeur au plus k sont informés. Considérons le schéma où un nœud appelle un de ses fils à la première étape impaire suivant immédiatement celle où il est informé, et où à l'étape (paire) suivante, ce nœud et son fils informé appellent tous les deux le second fils. Aux étapes $2k$, seuls les nœuds de profondeur au plus k sont informés. Nous pouvons donc considérer que c'est le schéma qui est suivi par un pire protocole dans un mobile en triangle. Nous obtenons la même chose dans le modèle SkipSenders avec un ***mobile en maison*** (voir la figure 2.6b) qui donne une synchronisation : aux étapes $4k$ toutes les arêtes entre des nœuds à distance au plus $2k$ de s ont été utilisées. Ce graphe donne, d'ailleurs, aussi une synchronisation dans le modèle SkipInformed. Ainsi, à l'aide de ces gadgets appelés ***synchronisateurs***, nous obtenons une synchronisation qui permet d'une part de transformer facilement un problème avec plusieurs sources en un problème avec une seule source, et d'autre part, en utilisant ces gadgets à la place des arbres, nous pourrions alors calculer bien plus simplement le temps de diffusion exact dans les démonstrations d'inapproximabilité, mais nous perdons alors le fait que le graphe de la preuve de 2.5.3 est biparti.

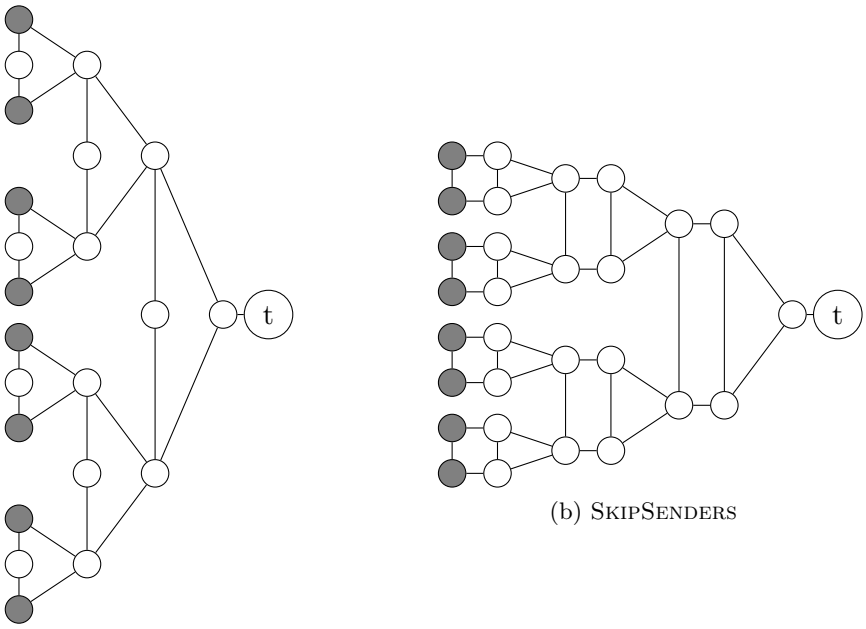
De manière similaire, nous pouvons nous intéresser au cas où il y a plusieurs cibles. Dans un premier temps, nous supposerons que chacune des cibles est de degré 1. Construisons, comme précédemment, un gadget appelé ***entonnoir*** tel que l'étape à laquelle une cible t est informée ne dépend que de la première étape à laquelle n'importe lequel des nœuds appelés ***points de connexion*** (les extrémités gauches en gris dans le figure 2.7) est informé, mais ne dépend pas du nombre de points de connexion informés à cette étape ou à des étapes ultérieures.

Si les cibles ne sont pas de degré 1, alors nous pouvons nous y ramener en "éclatant" chaque cible t, c'est-à-dire en remplaçant t par autant de copies que son degré et où chacune des copies est connectée à un seul de ses voisins. Chacune des copies d'un des destinataires est alors accordée à un entonnoir. Nous nous ramenons ainsi à un cas où toutes les cibles sont de degré 1. On fera attention de prendre des entonnoirs de même taille, quitte à ne pas utiliser tous les points de connexion, afin



(a) SkipInformed (b) SkipSenders (et SkipInformed)

Figure 2.6 – Synchronisateurs de source s



(b) SkipSenders

(a) SkipInformed

Figure 2.7 – Entonnoirs pour une cible t, avec points de connexion en gris.

que la différence, en nombre d'étapes, induite par cette transformation soit la même pour toutes les cibles. Nous terminons comme précédemment en reliant toutes ces cibles de degrés 1 à un nouvel entonnoir dont l'extrémité est la nouvelle cible.

2.6.3 Liens entre les modèles

Théorème 2.6.6. *Soit G' le graphe obtenu à partir d'un graphe G en rajoutant un sommet au milieu de chaque arête. Alors pour toute paire (s,t), nous avons $t_{\text{SS}}(s,t,G) \leq t_{\text{SI}}(s,t,G') \leq 2t_{\text{SS}}(s,t,G)$*

En effet, intuitivement, communiquer dans le graphe G' dans le modèle SkipInformed revient à communiquer dans G avec des communications qui prennent deux étapes dans G, et où deux nœuds voisins ne peuvent s'appeler l'un l'autre qu'exactement à la même étape.

Démonstration. Pour simplifier les notations, $(G, \text{SS}, \mathcal{L})$ désigne le schéma de diffusion dans le graphe G avec le modèle SkipSenders et l'ensemble de listes $\mathcal{L}$. $(G', \text{SI}, \mathcal{L}')$ désigne de manière similaire le schéma de diffusion dans le graphe G' avec le modèle SkipInformed et l'ensemble de listes $\mathcal{L}'$. Notons uv le nœud rajouté au milieu de l'arête $\{u,v\}$.

Soit (s,t) fixé. Soit $\mathcal{L}$ un ensemble de listes correspondant à un pire schéma de diffusion dans G pour le modèle SkipSenders, tel que pour chaque nœud, les voisins de ce nœud qui sont "passés" (c'est-à-dire ceux que le modèle empêche d'appeler) se retrouvent à la fin, cf. le lemme 2.6.2. $\mathcal{L}'$ est construit à partir de $\mathcal{L}$. Donnons des listes quasiment identiques aux nœuds correspondant dans G et G' : chaque voisin v dans la liste $L_u \in \mathcal{L}$ est remplacé par le voisin uv dans la liste $L'_u \in \mathcal{L}'$. Affectons une liste quelconque aux nœuds rajoutés dans G'. Puisque ces derniers sont de degrés deux, une fois qu'ils ont reçu l'information d'un voisin, ils ne peuvent ensuite l'envoyer qu'à leur autre voisin.

Notons $\tau(u)$ l'étape au cours de laquelle u se trouve informé dans $(G, \text{SS}, \mathcal{L})$ et $\tau(u,v)$ l'étape au cours de laquelle l'arête $\{u,v\}$ est utilisée dans $(G, \text{SS}, \mathcal{L})$. Définissons de même τ' dans $(G', \text{SI}, \mathcal{L}')$. Puisque toutes les arêtes ne sont pas nécessairement utilisées dans le modèle SkipInformed, τ' peut valoir l'infini. Remarquons que $\tau'(uv) = \min\{\tau'(u,uv), \tau'(v,uv)\}$.

Nous montrons que pour toute paire (u,v), nous avons $\tau(u,v) \leq \tau'(uv)$. Cela implique que pour tout u, nous avons $\tau(u) \leq \tau'(u)$. Par l'absurde, supposons que u envoie le message à uv à un temps $\tau'(uv) < \tau(u,v)$, et que $\tau'(uv)$ est minimal pour cette inégalité. Si uv est le premier voisin dans la liste L'_u, cela signifie que $\tau'(u) < \tau(u)$. Il existe donc un voisin w qui a envoyé le message à un temps $\tau'(wu)$, où $\tau'(wu)$ est antérieur à $\tau'(uv)$ et vérifie aussi l'inégalité $\tau'(wu) = \tau'(u) - 1 < \tau(u) - 1 \leq \tau(w,u)$. Ceci contredit la minimalité. Si uv n'est pas le premier voisin dans la liste L'_u, alors appelons w le voisin de u précédent uv dans la liste. Nous avons alors que $\tau'(uw) < \tau'(uv)$. En effet, ou bien u appelle uw avant uv, ou bien uw a déjà été appelé (par w). Donc

$$\tau'(uw) < \tau'(uv) < \tau(u,v) = \tau(u,w) - 1.$$

Nous avons donc $\tau'(uw) < \tau(u,w)$, ce qui contredit aussi la minimalité. Ainsi, pour toute paire (u,v), nous avons bien $\tau(u,v) \leq \tau'(uv)$.

La preuve de l'autre inégalité est symétrique. Partons d'un ensemble de listes $\mathcal{L}'$ correspondant à un pire schéma de diffusion dans G' pour le modèle SkipInformed, tel que, pour chaque nœud, les voisins de ce nœud qui sont passés par celui-ci se retrouvent à la fin de sa liste, cf. le lemme 2.6.2. $\mathcal{L}$ est construit à partir de $\mathcal{L}'$ comme suit : si u est dans G' et G alors un voisin uv dans la liste L'_u est remplacé par le voisin v dans la liste L_u. Par ailleurs, les listes des nœuds uv qui sont uniquement dans G' sont supprimées. Les temps τ et τ' sont définis comme précédemment.

La diffusion dans $(G', \text{SI}, \mathcal{L}')$ est plus rapide que le même protocole où les nœuds qui correspondent à des nœuds de G n'exécutent leur algorithme qu'aux temps pairs. Dans cette dernière variante, les nœuds de G' correspondant aux nœuds de G agissent aux temps pairs et les nœuds correspondant aux (milieu des) arêtes de G agissent aux étapes impaires. Ainsi, soit u et v envoient simultanément leur message à uv, soit un seul d'entre eux enverra à uv. C'est exactement ce qui se passe dans le modèle SkipSenders, où deux voisins s'appellent simultanément ou un seul peut utiliser l'arête (à condition de considérer que les envois à travers une arête mettent 2 étapes au lieu d'une). Ainsi, en ne considérant que les temps pairs 2τ de la variante sur G', il se passe exactement la même chose que dans SkipSenders au temps τ. D'où $t_{\text{SI}}(s, t, G') \leq 2t_{\text{SS}}(s, t, G)$. □

Remarque. Le nombre d'arêtes double dans la transformation du théorème 2.6.6, par contre le nombre de nœuds peut passer de n à $\Theta(n^2)$.

À cause de la symétrie dans le modèle SkipSenders il y a peu d'espoir qu'une transformation locale permette d'exprimer, de manière similaire, un problème avec SkipInformed en une approximation d'un problème avec SkipSenders. En effet transformer une diode de SkipInformed ne donne pas une diode de SkipSenders. De même l'algorithme polynomial pour le modèle SkipNone ne donne, a priori, aucune piste pour un algorithme d'approximation du modèle SkipSenders bien que ces deux modèles soient très ressemblants.

2.6.4 Variations sur SkipInformed

Si nous connaissons $t_{\text{SI}}(G, s, t)$, pour toute paire (s, t), nous pouvons alors évidemment calculer $t_{\text{SI}}(G)$ en prenant le maximum sur toutes les paires, au prix d'une complexité multipliée par n^2. Il est possible de faire mieux, c'est-à-dire en faisant un seul appel à un oracle donnant t_{SI}, en rajoutant un nombre constant de nœuds à G.

Théorème 2.6.7. *Soit G un graphe connexe quelconque, et G' obtenu à partir de G en ajoutant un sommet universel u et un sommet v pendant à u. Nous avons $t_{\text{SI}}(G', u, v) = t_{\text{SI}}(G') = 2 + t_{\text{SI}}(G)$.*

Démonstration. Soit un schéma de diffusion qui prend $t_{\text{SI}}(G)$ étapes sur G. Construisons un schéma qui prend $t_{\text{SI}}(G) + 2$ étapes dans G' en démarrant en u. Soit s la source du premier schéma de diffusion. À la première étape du nouveau schéma, u informe s. Les nœuds de G gardent le même schéma de diffusion, qui est alors exécuté avec un décalage dans le temps d'une étape. Les nœuds de G ont les mêmes

listes dans G et dans G', sauf u qui est ajouté à la fin de chaque liste. À chaque étape, dans G', le sommet universel u appelle un nœud qui est nouvellement informé à cette étape par un nœud de G, c'est-à-dire qui serait informé une étape plus tôt dans G. Finalement, à la dernière étape, c'est-à-dire à l'étape $t_{\text{SI}}(G) + 2$, u appelle v. Nous obtenons donc que $2 + t_{\text{SI}}(G) \leq \max\{t_{\text{SI}}(G', u, t), t \in V(G)\} \leq t_{\text{SI}}(G')$.

Montrons maintenant que $t_{\text{SI}}(G') \leq 2 + t_{\text{SI}}(G)$. Soit $\mathcal{L}'$ un pire schéma de diffusion dans G' de source s. Si $s = v$ alors, après la première étape, les nœuds informés sont u et v, ce qui reviendrait à démarrer avec le source u qui appelle v à la première étape. Nous pouvons donc supposer, sans perte de généralité, que $s \neq v$. Si $s = u$ alors, par hypothèse de minimalité du lemme 2.6.3, nous pouvons aussi considérer que u n'appelle v que lorsqu'aucun autre nœud n'est informé (c'est-à-dire à l'étape 1, quand seul u est informé, ou quand tous les autres nœuds sont informés). Sinon, sans perte de généralité, nous pouvons supposer que dans ce schéma, à l'étape où le nœud universel u est informé, il l'est simultanément par tous les nœuds informés. Et, comme précédemment, nous pouvons supposer que u n'appelle v qu'à la dernière étape. Le schéma $\mathcal{L}$ déterminé à partir de $\mathcal{L}'$ où u a été supprimé des listes des nœuds de G est un schéma de diffusion qui prend $t_{\text{SI}}(G') - 2$ étapes dans G. Dans le cas où $s = u$ et s appelle v à l'étape 1, le premier nœud de G est informé à l'étape 2 et le dernier à l'étape $t_{\text{SI}}(G')$ le schéma $\mathcal{L}$ mettrait donc au moins $t_{\text{SI}}(G') - 2$ étapes. De même dans le cas où $s = u$ et s appelle v à l'étape $t_{\text{SI}}(G')$ le premier nœud de G est informé à l'étape 1 et le dernier à l'étape $t_{\text{SI}}(G') - 1$ (sinon u n'appellerait pas u à l'étape $t_{\text{SI}}(G')$) le schéma $\mathcal{L}$ mettrait donc au moins $t_{\text{SI}}(G') - 2$ étapes. Dans le cas où $s \in G$, par l'hypothèse simplificatrice, tout $G \cup u$ est informé exactement à l'étape $t_{\text{SI}}(G') - 1$. Il est donc possible de prendre $t_{\text{SI}}(G') - 1$ étapes pour informer $G \cup u$, dont une étape où tous les nœuds de G informés appellent u. En supprimant cette étape, nous voyons que $\mathcal{L}$ prendrait donc $t_{\text{SI}}(G') - 2$ étapes pour informer tout le monde dans G. □

Nous ne connaissons pas de transformation aussi simple que celle du théorème 2.6.7 qui permette de déterminer $t_{\text{SI}}(G, s, t)$ à partir de $t_{\text{SI}}(G')$ où G' est construit à partir de G en ajoutant un nombre constant de nouveaux nœuds. Cependant, si t est un sommet pendant, nous pouvons construire G' en accrochant un chemin de taille n à t, ce qui implique que l'extrémité du chemin est soit la source soit le dernier nœud d'un pire schéma. Et nous terminons en attachant s à l'entrée d'une diode suffisamment grande pour que s soit le premier nœud de G informé dans une pire diffusion. Si t n'est pas un sommet pendant, nous pouvons utiliser la technique qui consiste à "éclater" t en $\deg(t)$ copies, chacune étant reliée à un point de connexion d'un entonnoir.

2.7 Conclusion

Nous avons montré d'importantes différences entre les modèles SkipNone, SkipSenders et SkipInformed. En particulier, nous avons montré des graphes dont les

temps de diffusions sont très différents selon les modèles considérés. De plus, nous avons montré l'existence de diodes pour les modèles SkipNone, et SkipInformed. Par contre, pour SkipSenders les diodes n'existent pas. En effet, dans ce dernier modèle, nous avons montré que si l'information circule à une certaine vitesse de s à t, alors elle circule à la même vitesse de t à s, si s et t sont des sommets pendants.

Puis, nous avons donné un algorithme en temps $O(m + n \log n)$ qui, pour tout réseau G de n nœuds et m arêtes, et n'importe quelle cible t de G, calcule $t_{\text{SN}}(G, s, t)$ pour toutes sources s. Nous avons donc un algorithme en temps $O(n(m+n \log n))$ qui calcule $t_{\text{SN}}(G)$. Nous avons alors montré que le calcul de $t_{\text{SI}}(G, s, t)$ est un problème NP-difficile, qui ne peut pas être approximé en temps polynomial avec un facteur multiplicatif $(\frac{1}{n})^{\frac{1}{2}-\epsilon}$ près, pour tout $\epsilon > 0$, à moins que P = NP. De même, $t_{\text{SI}}(G)$ ne peut pas être approximé en temps polynomial avec un facteur multiplicatif $(\frac{1}{n})^{\frac{1}{3}-\epsilon}$. Dans le cas de SkipSenders, le calcul de $t_{\text{SS}}(G, s, t)$ ou de $t_{\text{SS}}(G)$ est NP-difficile et ne peut pas être approximé en temps polynomial avec un facteur multiplicatif $\frac{1}{2} + \epsilon$, à moins que P = NP.

Il est intéressant de remarquer qu'il y a une raison fondamentale pour laquelle la construction prévue pour le modèle SkipInformed dans la démonstration du théorème 2.5.1 ne fournit pas de meilleures bornes d'inapproximabilité dans le modèle SkipSenders. C'est essentiellement à cause de l'absence de diodes dans le modèle SkipSenders. En effet, les diodes jouent un rôle important dans la démonstration du théorème 2.5.1

Ainsi, l'intrigante propriété de symétrie satisfaite par le modèle SkipSenders dans les graphes non orientés vue précédemment laisse entrevoir la possibilité de déduire un algorithme en temps polynomial pour l'approximation de $t_{\text{SS}}(G, s, t)$ à une constante près. En dehors de la curiosité intellectuelle, un tel algorithme, s'il existe, serait particulièrement utile dans la pratique, pour des raisons déjà mentionnées. En particulier, est-il possible de concevoir un algorithme en temps polynomial pour le modèle SkipSenders, avec un facteur d'approximation de $\frac{1}{2}$?

Ces résultats sont importants sous deux aspects différents. En effet, selon le point de vue, nos résultats peuvent être considérés comme de bonnes ou de mauvaises nouvelles. D'une part, ce sont de bonnes nouvelles parce que l'existence d'un algorithme en temps polynomial pour le modèle SkipNone garantit que le concepteur peut calibrer son application en fonction du délai maximal subi par un paquet. D'autre part, le même résultat signifie que si le protocole de diffusion n'utilise pas l'aléa et utilise uniquement des listes données pour distribuer les paquets aux voisins à tour de rôle, alors un adversaire peut facilement calculer des listes qui donneront la pire performance pour la diffusion ListBased. En contrepartie, notre résultat d'inapproximabilité pour le modèle SkipInformed montre qu'un adversaire déterministe aura en général de sérieuses difficultés à calculer des listes de pire cas. Ceci n'est qu'une relative bonne nouvelle pour le concepteur, puisque le modèle SkipInformed nécessite des échanges perpétuels de messages de signalisation entre voisins, qui peuvent être difficiles à mettre en œuvre. Par contre, supprimer les entrées correspondantes aux voisins à partir desquels un nœud a reçu un paquet de

sa liste ne nécessite pas de moyens de communication supplémentaires. Ainsi, utiliser le modèle SkipSenders peut être une bonne solution pour le concepteur, s'il se révèle difficile à approximer (comme, par exemple, dans les réseaux orientés). Jusqu'à présent, notre borne supérieure $\frac{1}{2} + \epsilon$ ne garantit pas que l'adversaire ne puisse pas générer efficacement des listes pour lesquelles le protocole ListBased (déterministe) aurait de mauvais résultats.

Chapitre 3

Diffusion dans les graphes évolutifs

Depuis le début du document, nous avons considéré uniquement des graphes statiques (déterministes ou aléatoires). Or, plusieurs technologies émergentes sont intrinsèquement dynamiques, comme, par exemple, les réseaux sans fil ad-hoc, les réseaux de capteurs, les réseaux mobiles ou les réseaux pair-à-pair. Les protocoles de diffusion sans mémoire supportent un haut niveau de dynamisme de leur environnement [GKG06]. Il est toutefois important d'évaluer précisément l'impact du dynamisme sur l'efficacité de ces protocoles. On sait en effet que le dynamisme dans un réseau peut avoir un impact énorme dans certaines circonstances. Cet impact peut être positif lorsque l'évolution du réseau est ergodique (par exemple, lorsqu'on mesure la bande passante globale d'un réseau radio ad-hoc [GT02]) ou lorsque le dynamisme représente la mobilité de nœuds [CPS09]. Il peut néanmoins également être négatif lorsque l'évolution du réseau est arbitraire (par exemple, lorsqu'on mesure le temps de couverture des marches aléatoires [AKL08]) ou simplement lorsque le dynamisme sert à modéliser des pannes dans un réseau.

Dans ce chapitre nous nous sommes intéressés à un modèle particulier de graphes évolutifs dit ***à évolution arête-Markovienne***. Dans ce modèle de graphes évolutifs, nous avons étudié l'inondation et la diffusion 1-Port. Une partie de ce travail à été fait en collaboration avec Pierluigi Crescenzi et Pierre Fraigniaud, et a été publié dans Distributed Computing sous le titre *Parsimonious Flooding in Dynamic Graphs* [BCF11], une version préliminaire étant parue dans les actes de *28th Symposium on Principles of Distributed Computing* (PODC 2009).

Dans un premier temps nous effectuons un survol des résultats principaux de la littérature sur les graphes évolutifs. Nous étudions en particulier le modèle de graphes à évolution arête-markovienne où les arêtes suivent un processus markovien de vie et de mort : un arête e absente du graphe au temps t apparait avec probabilité p_e et une arête présente disparaît avec probabilité q_e, indépendamment des autres arêtes. Nous définissons ensuite formellement l'inondation et surtout l'inondation *parcimonieuse* dans un graphe dynamique. Nous donnons alors notre premier résultat principal,

qui est une technique générale permettant de réduire le temps d'inondation dans un graphe à évolution arête-markovienne au calcul du diamètre d'un graphe pondéré aléatoirement, lorsque la distribution initiale est la distribution stationnaire. Nous appliquons cette technique aux graphes à évolution arête-markovienne homogène. Nous donnons le temps d'inondation pour toutes les valeurs de p et q. Nous complétons et améliorons ainsi des résultats de [CMPS11]. Ensuite, nous donnons une seconde technique permettant de majorer le temps d'inondation pour toute distribution initiale arbitraire. Cette technique est appliquée au calcul du temps de diffusion dans les graphes à évolution homogène, ainsi qu'à deux cas à évolution hétérogène : dans le premier les taux de naissances sont définie en fonction du degré espéré du donné et ces taux sont celles des graphes en loi de puissance. Finalement, nous donnons un résultat sur le temps de diffusion 1-Port dans les graphes à évolution homogène pour lesquels nous majorons le temps de diffusion dans le cas $p + q \geq 1$.

3.1 Introduction

Dans cette section, nous présentons tout d'abord les principales études sur les graphes évolutifs et en particulier les résultats sur l'inondation dynamique. Ensuite, nous définissons formellement le modèle des graphes à évolution arête-markovienne et nous résumons nos principaux résultats de ce chapitre.

3.1.1 État de l'art sur les graphes évolutifs

Une première approche des graphes dynamiques est de considérer un graphe statique dans lequel les liens ont des pannes avec une certaine probabilité, voir par exemple [AAG87]. On peut aussi trouver une définition de graphes dynamiques dans [HG97] ainsi que des approches pour les étudier, mais nous retiendrons particulièrement le concept de graphe *évolutif*. De ce que nous en savons, il a été formellement introduit dans [Fer02], où est considéré le calcul des plus courts chemins.

Définition (Ferreira [Fer02])**.**
Un **graphe évolutif** *est une suite* $(G_t)_{t\geq 0}$ *de sous-graphes d'un même graphe* G*, appelé* **graphe de base***.*

La description des graphes évolutifs est assez générale pour permettre de modéliser de nombreuses sortes de dynamismes. Ils incluent les graphes évolutifs "adversaires", dans lesquels un adversaire choisit, à chaque étape, le graphe G_t, par exemple, afin de ralentir le processus de diffusion. Ils incluent également les graphes à évolution aléatoire, dans lesquels, à chaque étape, le graphe G_t est choisi aléatoirement selon une distribution de probabilité sur une famille de graphe donnée. On peut aussi supposer que le graphe G_t considéré au temps t est un graphe aléatoire d'Erdős-Rényi tiré dans $\mathcal{G}_{n,p}$. Soit un graphe G fixé. Ne garder que certaines des arêtes de G à chaque étape, avec une probabilité p, revient à prendre $G \cap \mathcal{G}_{n,p}$. Ceci modélise alors, le graphe statique G avec pannes dont la probabilité est $1 - p$.

Une certain dynamisme peut permettre d'accélérer la diffusion de l'information (par rapport à un graphe statique ayant la même densité). L'étude des graphes évolutifs reste toutefois complexe. Par exemple, le problème du calcul des composantes connexes, facile dans le cas statique, devient NP-difficile dans le cas des graphes évolutifs [BF03], même dans le cas où le graphe de base est le graphe des nœuds à distance au plus deux dans une grille [JL04]. (Le problème est par contre toujours polynomial si le graphe de base est un arbre.) Plusieurs recherches ont récemment été effectuées sur le modèle des graphes évolutifs afin de mesurer l'impact de l'évolution aléatoire du réseau sur les performances des algorithmes. Par exemple, il a été démontré que même pour une probabilité p inférieure au seuil de connectivité de $\mathcal{G}_{n,p}$, la radiodiffusion (inondation avec interférence modélisant la diffusion dans les réseaux radio) s'effectue efficacement (voir [CMPS09]). Plus précisément, le temps de diffusion est logarithmique si p est connu, et $O(\log^2 n)$ sinon. Par contre, si c'est un adversaire qui impose chacun des graphes (avec l'hypothèse qu'il y a au moins un lien d'un sommet informé vers un sommet non informé), alors le temps devient $\Theta(\frac{n^2}{\log n})$. Il n'est que $\Theta(n)$ dans le pire cas statique [KM98, CR06].

Dans le cas statique, le temps de couverture d'une marche aléatoire, est entre $\Omega(n \log n)$, atteint pour la clique, et $O(n^3)$, atteint pour le graphe "sucette" [Fei95b, Fei95a]. Dans [AKL08] il est démontré que le temps de couverture d'une marche aléatoire reste polynomiale quelque soit $p > 0$, mais elle peut être exponentielle si la séquence de graphes est donnée par un adversaire, même si tous les graphes de la séquence sont connectés.

Il n'est pas étonnant qu'un adversaire puissant gène fortement le processus considéré (par exemple la diffusion ou le parcours), c'est pourquoi on peut souhaiter limiter son pouvoir. Par exemple, comme dans les cas mentionnés précédemment, on peut considérer que tous les graphes soient connectés. Une manière de limiter la puissance d'un adversaire peut être de définir l'*attrition* (*churn*), c'est-à-dire le nombre de liens qu'il est autorisé à modifier à chaque étape [KSW05a, KSW05b, APRU12]. Une autre manière de mesurer la puissance d'un adversaire est de définir la *connexité par intervalle* [KLO10], c'est-à-dire un temps T pour lequel, sur tout intervalle de longueur T, l'intersection des graphes de cet intervalle de temps est connexe.

En ce qui concerne la modélisation des réseaux réels, le processus aléatoire $(G_0, G_1, G_2, \dots)$ où $G_t \in \mathcal{G}_{n,p}$ souffre de l'absence de dépendance temporelle. Le réseau évolue mais sa structure à l'instant t est indépendante de la structure au temps $t' < t$. Ceci ne reflète pas précisément ce qui est observé dans de nombreux contextes tels que les réseaux sans fil (la connexion entre deux nœuds est fortement corrélé à l'existence antérieure de cette connexion) et les réseaux P2P (un échange d'informations entre deux participants est fortement corrélée à l'existence de précédents échanges) [SMZ03]. Un modèle plus évolué capturant la dépendance temporelle a été étudié récemment dans [CMM$^+$10, CMPS11] dans le cadre de l'inondation. Ce modèle est appelé processus ***arête-markovien*** ou graphe a évolution arête-markovienne, et est noté $\mathcal{M}_{n,p,q}$. Il génère une séquence aléatoire de graphes $(G_0, G_1, G_2, \dots)$, sur le même ensemble de nœuds $[n]$ définie par un taux de naissance p et un taux de mort q,

d'après la règle suivante. Soit un graphe G_0 pouvant être choisi arbitrairement. Pour $t > 0$, une arête absente $e \notin E(G_{t-1})$ apparaît dans $E(G_t)$ avec probabilité p, tandis qu'une arête présente $e \in E(G_{t-1})$ disparaît de $E(G_t)$ avec probabilité q. On parle alors de graphe à évolution arête-markovienne. Lorsque G_0 est un graphe aléatoire d'Erdős-Rényi dans $\mathcal{G}_{n,\widehat{p}}$ où $\widehat{p} = \frac{p}{p+q}$ on parle alors de cas *stationnaire* car la distribution stationnaire du processus satisfait $G_t \in \mathcal{G}_{n,\widehat{p}}$.

Remarquons que le cas d'une suite de graphes aléatoires indépendants dans $\mathcal{G}_{n,p}$ est un cas particulier du processus arête-markovien où $p + q = 1$. En effet pour ces valeurs de p et q, une arête présente au temps t demeurera au temps $t + 1$ avec probabilité $1-q = p$ et une arête absente au temps t apparaîtra au temps $t+1$ avec la même probabilité p. Ainsi, indépendamment du passé, une arête est présente au temps $t+1$ avec une probabilité p. Par conséquent, le modèle d'évolution arête-markovienne généralise la suite de graphes $\mathcal{G}_{n,p}$ de deux manières : premièrement en introduisant d'autres types de probabilités pour la présence d'une arête, et, deuxièmement, en introduisant des dépendances temporelles.

Le modèle de graphe à évolution arête-markovienne peut être facilement étendu au cas où les arêtes n'ont pas toutes les mêmes taux de naissance et de mort. En fait, bien que les graphes aléatoires d'Erdős-Rényi fournissent un modèle intéressant, ils sont différents des graphes rencontrés en pratique. L'une des différences les plus connues est le fait que les graphes rencontrés en pratique sont éparses quoique certains nœuds puissent avoir des degrés très grands. Les sommets d'un graphe aléatoire d'Erdős-Rényi ont approximativement le même degré. Le modèle de graphe dynamique à évolution arête-markovienne, dont les taux de naissance et mort sont identiques pour toutes les arêtes, souffre des mêmes défauts. Pour cette raison, nous pouvons nous inspirer, dans l'attribution des taux de naissance et de mort, d'un modèle de graphe aléatoire plus général qui a été introduit afin de pouvoir produire des graphes aléatoires avec une séquence de degrés espérés donnée (pour avoir une définition formelle de ce modèle, voir [CL06a]). Dans ce modèle, étant donnée une séquence $\mathbf{w} = w_1, \ldots, w_n$ de nombres positifs, le graphe $G_{\mathbf{w}}$ est un graphe aléatoire avec n nœuds dans lequel chaque arête $\{i, j\}$ est présente avec probabilité $p_{i,j} = \frac{w_i w_j}{\sum_{k=1}^{n} w_k}$ (indépendamment des autres arêtes). Il est alors facile de voir que le degré espéré d'un nœud i est w_i. Ainsi, si on impose que la séquence $\mathbf{w}$ satisfasse une loi de puissance, alors $G_{\mathbf{w}}$ est un graphe en loi de puissance, tandis que si on impose que $w_i = pn$, alors $G_{\mathbf{w}}$ est le graphe aléatoire d'Erdős-Rényi $\mathcal{G}_{n,p}$. Ce modèle de graphe aléatoire plus général a aussi été beaucoup étudié, et plusieurs de ses propriétés sur la connectivité ont déjà été établies (telles que l'apparition d'une composante géante).

Par la suite, si les taux de naissance et de mort sont les mêmes pour toutes les arêtes on parlera d'évolution arête-markovienne ***homogène***. Sinon, on parlera d'évolution arête-markovienne ***hétérogène***.

Dans [CMPS11] est aussi considéré une autre variante : les graphes à évolution géométrie-markovienne. Ce modèle capture ayant à la fois une forme de dépendance temporelle et de dépendance spatiale. Les nœuds se déplacent et deux nœuds sont dit voisins si leur distance est inférieure au rayon de transmission. Dans [CMPS11], le

rayon de transmission y est relativement petit par rapport à la vitesse de déplacement des nœuds. Ceci peut être vu comme un moyen d'économiser de l'énergie car le déplacement des nœuds participe involontairement à la propagation de l'information et permet de compenser un rayon de transmission plus petit et donc moins gourmand en énergie. Depuis d'autres recherches sur des modèles visant à modéliser la mobilité ont été effectuées, voir par exemple [CPS09, CS11].

3.1.2 Inondation dynamique

L'inondation est le protocole de diffusion où un nœud non informé devient informé lorsqu'il a un voisin informé. (Cette formulation est valable aussi bien dans le cas statique que dynamique.) Du point de vue des nœuds informés, dans le cas statique, nous pouvons considérer que juste après avoir été informé, un nœud informe tous ses voisins en même temps. Dans le cas où le réseau est dynamique, puisque le voisinage des nœuds change, un nœud informé transmet l'information à tous ses voisins après chaque modification du réseau. Puisque les communications et l'évolution du réseau se font à des étapes synchrones, à chaque étape, chaque nœud qui a reçu le message original lors d'une étape précédente transmet ce message à tous ses voisins.

L'inondation est un mécanisme essentiel pour la diffusion de l'information dans des contextes où la topologie du réseau est très dynamique et inconnue tels que les réseaux P2P, les réseaux mobiles, ou tout autre réseau sensible aux défaillances ; et plusieurs variantes d'inondation visant à limiter la consommation de bande passante ont été proposées [CL07, Gnu, LCC+02].

Dans le cas statique, le calcul du temps nécessaire à l'inondation pour informer tous les nœuds revient à calculer le diamètre du graphe, ou plus exactement l'excentricité de la source. Ce n'est pas aussi simple dans le cas dynamique. Au contraire, le temps d'inondation donne une mesure du graphe dynamique, qui est le pendant dynamique du diamètre. Le temps d'inondation pourrait donc être appelé *diamètre dynamique*. Ce diamètre dynamique donne une borne inférieure au temps d'exécution de tout protocole de diffusion où les communications se font entre voisins.

Dans leurs articles complémentaires [CMM+10, CMPS11], les auteurs ont analysé le protocole d'inondation dans les graphes à dynamique arête-markovienne, c'est-à-dire, dans une suite de graphes générés par le processus arête-markovien décrit précédemment. L'inondation dans les graphes dynamiques est le mécanisme de base dans lequel chaque nœud recevant une information à l'étape t transmet cette information à tous ses voisins, à toutes les étapes à venir $t' > t$. À travers ces deux études, les auteurs ont fourni plusieurs bornes sur le temps d'inondation dans les graphes à dynamique arête-markovienne. En particulier, ils ont prouvé qu'il y a une large classe de graphes dynamiques dans lesquels le temps d'inondation ne dépend (asymptotiquement) pas du taux de mort q.

Inondation dynamique parcimonieuse. Malgré l'intérêt de l'inondation en termes de simplicité et de robustesse, ce protocole souffre d'un inconvénient majeur

dans les graphes dynamiques : il exige que chaque nœud, après réception d'un message, le transmette au cours de *toutes* les étapes de temps à venir. Il en résulte un gaspillage de ressources aussi bien en termes de bande passante que de temps de calcul des nœuds. Bien sûr, si nous savons que l'inondation se termine en T étapes, nous pouvons alors restreindre à tous les nœuds de n'être actif que pour ce laps de temps. Néanmoins, T est généralement croissant avec la taille n du réseau, et les nœuds sont donc actifs T étapes, ce qui entraîne encore un important gaspillage de ressources. Notre objectif est de forcer l'inondation à s'effectuer avec plus de parcimonie, en limitant le nombre d'étapes au cours desquelles chaque nœud est actif dans le transfert d'un message à ses voisins, tout en permettant au message d'atteindre éventuellement tous les nœuds dans un délai court. L'inondation parcimonieuse permet d'économiser de la bande passante, des ressources de calcul et donc de l'énergie, ce dernier paramètre étant connu pour être crucial pour les réseaux ad-hoc et les réseaux de capteurs.

Soit un entier positif k, le protocole d'inondation (sur les graphes dynamiques) est dit ***k-actif***, si chaque nœud envoie le message uniquement pendant les k étapes de temps immédiatement après l'étape à laquelle le nœud a reçu ce message pour la première fois. Par exemple, le protocole d'inondation 1-actif est le protocole d'inondation standard pour les réseaux statiques : un message n'est transmis qu'une seule fois, à l'étape immédiatement après sa réception. Toutefois, dans des réseaux dynamiques, le protocole d'inondation peut avoir à être actif pour $k > 1$ étapes afin que le message atteigne tous les nœuds. Plus le paramètre k est petit, moins le protocole consomme de ressources.

Notre objectif est de déterminer le plus petit k tel que le protocole d'inondation k-actif se termine correctement, c'est-à-dire que le message est parvenu à tous les nœuds. Définissons le *seuil de terminaison* pour le protocole d'inondation comme le plus petit entier k tel que pour toute source $s \in [n]$, le protocole d'inondation k-actif termine correctement asymptotiquement presque sûrement. Évidement si l'inondation se termine en T étapes alors le seuil de terminaison de l'inondation est au plus T. En fait, nous montrerons que dans le cas de l'évolution arête-markovienne hétérogène, pour un large spectre de paramètres p et q, ce seuil est seulement $o(T)$, et même souvent simplement une constante. De plus nous montrerons aussi qu'être actif un nombre d'étapes égal au seuil de terminaison (à une constante près) est suffisant pour que l'inondation se termine correctement en un temps *optimal*, c'est-à-dire, en asymptotiquement le même nombre d'étapes que lorsque les nœuds restent actifs indéfiniment.

Remarque. Le cas de la diffusion dans les réseaux dynamiques est à rapprocher des modèles utilisés par la communauté des épidémiologistes afin d'analyser la propagation d'une infection au sein d'une population. Parmi ces modèles, les plus connus sont SIR (pour *Susceptible-Infective-Removed*, c'est-à-dire Susceptible-Infectée-Retiré) et SIS (*Susceptible-Infective-Susceptible*, c'est-à-dire Susceptible-Infectée-Retiré), ils visent à représenter la manière dont un virus se propage par la réaction de chaque personne en présence du virus. Dans le premier cas, une personne peut être soit saine (et susceptible d'être infectée), soit infectée, soit retirée du graphe (parce

qu'elle est décédée ou parce qu'elle a été immunisé), dans le second cas une personne soignée peut être infectée à nouveau. Un nœud infecte ses voisins avec une certaine probabilité avant d'être retiré ou à nouveau sain selon le modèle. Un protocole de diffusion 1-actif est similaire au modèle SIR, un nœud étant sain s'il est non informé, infecté s'il vient de recevoir l'information et peut donc la transmettre pendant une étape, et retiré ensuite. Le cas perpétuellement actif serait alors similaire au cas SI où les nœuds restent indéfiniment infectés et retentent d'infecter leurs voisins. Un voisin de u étant dans notre modèle un nœud pour lequel l'arête le joignant à u a une probabilité non nulle d'apparaître.

Des phénomènes de seuil ont aussi été identifiés dans le modèle SIR. Notamment, il a été démontré que l'épidémie SIR ne contamine qu'un nombre constant de nœuds lorsque la probabilité d'infection est $p = \frac{c}{n}$ avec $c < 1$, mais contamine une fraction constante de nœuds si $c > 1$ (voir [DGM08]). Comme une partie de nos résultats, cette dernière étude est étroitement liée aux graphes aléatoires d'Erdős-Rényi et en particulier à la taille de la composante géante.

3.1.3 Résultats antérieurs

Il y a de nombreux articles concernant les protocoles de diffusion dans les réseaux statiques (par exemple voir les études [FL94, HHL88, HKMP96]). La diffusion dans les graphes aléatoires $\mathcal{G}_{n,p}$ a été étudiée dans [FPRU90, FG85, Pit87] et l'échange total dans des métriques particulières a été étudiée dans [KKD04, KK02]. Dans chacun de ces cas, il n'y a pas d'évolution de la structure du réseau au cours du temps.

Plusieurs travaux se sont intéressés à la propagation de l'information dans les réseaux sans fil (par exemple, [CLF+07, SD12]). La dépendance temporelle est généralement soit ignorée, soit surmontée en considérant des intervalles temps suffisamment grands pour que ces dépendances soient négligeables. En fait, lorsque nous avons travaillé sur l'article [BCF11], les seuls travaux sur la diffusion d'information considérant une dépendance temporelle sur la présence des liens étaient ceux de [CMM+10, CMPS11]. Dans le premier article, le modèle considéré est celui sur lequel porte nos résultats, les graphes à dynamique arête-markovienne homogène. Dans le second, les auteurs considèrent une généralisation à d'autres types de dynamiques markoviennes dont les graphes à évolution géométrie-markovienne. Ils ont donné des bornes générales sur le temps d'inondation dans les graphes à dynamique markovienne homogène satisfaisant certaines propriétés. Plus spécifiquement, dans [CMM+10] il est montré que, quel que soit le graphe initial G_0 (c'est-à-dire pas nécessairement $G_0 \in \mathcal{G}_{n,\widehat{p}}$) et pour tout taux de naissance et de mort $0 < p, q < 1$, en au plus $O(\frac{\ln n}{\ln(1+np)})$ étapes, a.f.p., tous les nœuds sont informés. Si G_0 est un stable ($E(G_0) = \emptyset$), il est montré que, pour tout taux $0 < p, q < 1$, le temps d'inondation est au moins $\Omega(\frac{\ln n}{np})$. De plus si nous supposons que $p \geq c\frac{\ln n}{n}$ for $c > 1$ alors le temps d'inondation est au moins $\Omega(\frac{\ln n}{\ln(1+np)})$. Le cas où $G_0 \in \mathcal{G}_{n,\widehat{p}}$ a été étudié dans [CMPS11] où il est montré que, pour $\widehat{p} \geq c\frac{\ln n}{n}$ avec c assez grand, le temps

d'inondation est au plus

$$O\left(\frac{\ln n}{\ln n\widehat{p}} + \ln\ln n\widehat{p}\right)$$

et au moins

$$\Omega\left(\frac{\ln n}{\ln n\widehat{p}}\right).$$

3.1.4 Modèle et notations

Rappelons d'abord la définition des graphes à dynamique arête-markovienne, introduits dans [CMM+10, CMPS11]. Par la suite, n désignera un entier positif, et p et q désigneront deux valeurs réelles telles que $0 < p < 1$ et $0 < q < 1$, et g une distribution de graphe. Un processus arête-markovien $\mathcal{M}_{n,p,q,g}$ de distribution initiale g, de taux de naissance p et de mort q produit une suite de graphes $(G_0, G_1, G_2, \ldots)$ sur le même ensemble de nœuds $[n] = \{1, \ldots, n\}$ tels que G_0 est tiré selon la loi g, et pour tout $t > 0$, G_t est obtenu à partir de G_{t-1} en suivant le processus Markovien de paramètres p et q suivants :

- si $e \notin E(G_{t-1})$ alors $e \in E(G_t)$ avec probabilité p et $e \notin E(G_t)$ avec probabilité $1 - p$;
- si $e \in E(G_{t-1})$ alors $e \notin E(G_t)$ avec probabilité q et $e \in E(G_t)$ avec probabilité $1 - q$.

Par abus de terminologie, nous dirons qu'un nœud appartient à $\mathcal{M}_{n,p,q,g}$ dans le sens où il appartient à tous les graphes G_t, $t \geq 0$. Pour une suite de graphes $S = (G_0, G_1, G_2, \ldots)$, nous allons écrire $S \in \mathcal{M}_{n,p,q,g}$ pour signifier que l'espace de probabilité de S est celui de toutes les suites générées par $\mathcal{M}_{n,p,q,g}$.

Posons $\widehat{p} = \frac{p}{p+q}$. Il est aisé de montrer que, quel que soit le graphe initial, un graphe à évolution arête-Markovienne converge vers le graphe aléatoire d'Erdős-Rényi $\mathcal{G}_{n,\hat{p}}$ (voir [ER59, Gil59]), c'est-à-dire, le graphe où chaque arête est présente avec probabilité $\widehat{p}$ (indépendamment des autres arêtes).

Dans le cas particulier où le graphe initiale $G_0 \in \mathcal{G}_{n,\widehat{p}}$, nous parlons de cas ***stationnaire*** et le processus est noté $\mathcal{M}_{n,p,q}$ Comme $G_0 \in \mathcal{G}_{n,\widehat{p}}$, nous savons que $G_t \in \mathcal{G}_{n,\widehat{p}}$ pour tout $t \geq 0$. Par contre, la distribution de G_t sachant $G_{t'}$ pour $t' \leq t-1$ est plus complexe : pour $p + q \neq 1$ le graphe G_t dépend du graphe G_{t-1}. Ce n'est que lorsque $p + q = 1$ que G_{t+1} ne dépend pas de $G_{t'}$, quelque soit $t' \leq t$.

Nous nous intéressons dans un premier temps au protocole d'inondation dans le cas des graphes à évolution arête-markovienne tels qu'analysés dans [CMM+10] et [CMPS11]. Soit s la source et soit I_t l'ensemble de tous les nœuds qui détiennent le message à l'instant t. Ainsi $I_0 = \{s\}$. Pour $t \geq 0$, I_{t+1} est défini à partir de I_t par $I_{t+1} = I_t \cup N_t$ où N_t est l'ensemble de tous les nœuds de G_t voisins d'au moins un nœud dans I_t, ceux qui sont nouvellement informés. Si $p > 0$ alors, l'inondation atteint finalement toujours tous les nœuds, puisque, pour tout nœud $u \in [n]$, la probabilité qu'aucune arête n'est présente entre u et s au cours des étapes t est $(1 - \widehat{p})(1 - p)^{t-1}$, ce qui tend vers 0 lorsque t tend vers ∞. Notons que le nœud u

peut bien sûr recevoir le message bien avant que l'arête $\{s, u\}$ n'est pas présente, par des nœuds intermédiaires et des chemins plus longs.

Étant donnés s, n, p, q, g, la variable aléatoire $T_s(\mathcal{M}_{n,p,q,g})$ est le temps d'inondation à partir de s dans $\mathcal{M}_{n,p,q,g}$, définie par

$$T_s(\mathcal{M}_{n,p,q,g}) = \min\{t \geq 0 \mid I_t = [n]\}.$$

Lors qu'aucune confusion sera possible, $T_s(\mathcal{M}_{n,p,q,g})$ sera notée $T_{s,g}$. Si la distribution g n'est pas précisée, alors, par défaut, c'est la distribution stationnaire. Ainsi, dans le cas stationnaire, la variable aléatoire égale au temps d'inondation dans $\mathcal{M}_{n,p,q}$ est notée T_s.

Le protocole d'inondation décrit plus haut suppose que tous les nœuds sont perpétuellement actifs. C'est-à-dire qu'à la réception du message original, chaque nœud va transmettre ce message à tous ses voisins durant toutes les étapes de temps à venir. Nous nous sommes intéressés à délimiter l'intervalle de temps durant lequel un nœud est actif tout en permettant de terminer l'inondation. Soit k un entier positif, un protocole d'inondation k-actif est un protocole d'inondation dans lequel chaque nœud transmet le message uniquement lors des k étapes de temps immédiatement après l'étape à laquelle il a reçu le message pour la première fois. C'est-à-dire, soit $I_t^{(k)}$ l'ensemble de tous les nœuds qui possèdent le message à l'instant t de l'exécution du protocole d'inondation k-actif, nous avons $I_{t+1}^{(k)} = I_t^{(k)} \cup N_t^{(k)}$ où $N_t^{(k)}$ est l'ensemble des nœuds qui sont voisins dans G_t d'au moins un nœud dans $I_t^{(k)} \setminus I_{t-k}^{(k)}$. Par conséquent, si $k = \infty$, le protocole d'inondation k-actifs coïncide avec le protocole standard d'inondation analysé dans [CMM+10] et dans [CMPS11], tandis que si $k = 1$, le protocole d'inondation k-actif correspond au cas dans lequel chaque nœud transmet le message qu'une seule fois (à l'étape suivant immédiatement la réception du message), comme c'est le cas pour les inondations dans un réseau statique.

Nous nous sommes intéressés aux protocoles d'inondation k-actif où les nœuds sont actifs suffisamment longtemps pour garantir que le message atteigne tous les nœuds. Ainsi, pour le protocole d'inondation k-actif, nous définissons la variable aléatoire

$$T_{s,g}^{(k)} = \min\{t \geq 0 \mid I_t^{(k)} = [n]\}.$$

et nous nous intéressons au paramètre suivant :

Définition. *Le* **seuil de terminaison** *du protocole d'inondation dans $\mathcal{M}_{n,p,q,g}$ est le plus petit entier k tel que $T_{s,g}^{(k)} < \infty$ presque sûrement*[1]*, pour tout $s \in [n]$.*

Une généralisation naturelle du modèle $\mathcal{M}_{n,p,q,g}$ est de considérer qu'à chaque arête e est associée une probabilité d'apparition p_e et de disparition q_e. Le modèle

1. Formellement la valeur du seuil de terminaison dépend de la précision donnée à l'événement $T_{s,g}^{(k)} < \infty$ et par conséquent varie selon la fonction utilisée pour le terme $o(1)$ dans la définition de "presque sûrement".

sera alors noté $\mathcal{M}_{n,\mathbf{p},\mathbf{q},g}$ avec $\mathbf{p} = \{p_e, e \in E\}$ et $\mathbf{q} = \{q_e, e \in E\}$. Dans ce cas aussi il y a un cas stationnaire, noté $\mathcal{M}_{n,\mathbf{p},\mathbf{q}}$, le graphe aléatoire stationnaire correspondant est celui où chaque arête e est tirée selon la probabilité $\widehat{p_e} = \frac{p_e}{p_e+q_e}$.

Dans les sections à venir, nous allons, tout d'abord, donner une technique générale permettant de simplifier l'analyse de la diffusion dans les graphes à évolution markovienne dans le cas stationnaire que nous appliquerons au cas stationnaire $\mathcal{M}_{n,p,q}$. Nous donnerons alors une variante moins forte pour le cas général que nous appliquerons alors à $\mathcal{M}_{n,p,q,g}$ ainsi qu'à un autre cas particulier de $\mathcal{M}_{n,\mathbf{p},\mathbf{q},g}$, inspirée de la généralisation des graphes aléatoires où le degré moyen de chaque nœud est imposé [CL02].

3.1.5 Bref aperçu de nos contributions à la diffusion dans les réseaux arêtes-markoviens

Dans un premier temps, nous nous intéressons à une généralisation du modèle de graphe à évolution arête-markovienne introduit dans [CMM+10] et nous donnons une réduction générale pour ramener le calcul temps de l'inondation dans ces graphes évolutifs, dans le cas stationnaire, à un calcul d'un diamètre de graphe pondéré aléatoirement qui est appelé le graphe représentatif. Cette réduction est accompagnée d'une technique facilitant le calcul d'une majoration de ce diamètre. La réduction et la technique, forment une méthode générale permettant de calculer une majoration du temps de diffusion. La méthode générale est ensuite appliquée au cas hétérogène afin d'améliorer et compléter une partie des résultats de [CMM+10, CMPS11]. Dans ce cas, nous donnons des bornes exactes sur le temps d'inondation (lorsqu'il n'y a aucune contrainte sur l'activité des nœuds) pour toutes les valeurs de $p, q \in (0, 1)$ lorsque $G_0 \in \mathcal{G}_{n,\widehat{p}}$. Ces bornes sont résumées dans la table 3.1. Dans le cas stationnaire, pour $\widehat{p} \geq \frac{c \ln n}{n}$ avec $c > 1$, l'inondation prend a.p.s. $\Theta(\frac{\ln n}{\ln(n\widehat{p})})$ étapes. Pour $0 < \widehat{p} \leq \frac{c}{n}$ avec $c > 0$, l'inondation prend a.p.s. $\Theta(\frac{\ln n}{np})$ étapes. Entre ces deux valeurs, la situation est plus complexe et dépend alors des valeurs relatives de $\widehat{p}$ et p. Si $np \geq \ln n\widehat{p}$ alors le temps d'inondation est $\Theta(\frac{\ln n}{\ln(n\widehat{p})})$, tandis que si $np \leq \ln n\widehat{p}$ alors le temps d'inondation est $\Theta(\frac{\ln n}{np})$. Parallèlement au calcul de ces bornes sur le temps d'inondation, nous avons établi des bornes asymptotiquement exactes sur le seuil de terminaison. Si $\widehat{p} \geq \frac{c \ln n}{n}$ avec $c > 1$, alors ce paramètre vaut 1. C'est-à-dire qu'une seule étape d'activité est suffisante pour assurer (a.p.s.) que le message atteigne tous les nœuds. Si $\widehat{p} \leq \frac{c \ln n}{n}$ avec $c < 1$, alors le seuil de terminaison est $\Theta(\frac{\ln n}{np})$. (Notons que la condition dépend de $\widehat{p}$ tandis que la valeur du seuil de terminaison dépend de p).

Il est intéressant de remarquer que pour tout $p, q \in (0, 1)$, nous avons également montré que si l'activité k vaut ce seuil de terminaison, alors a.p.s., un protocole d'inondation $O(k)$-actif termine, à une constante près, aussi vite que l'inondation sans contrainte sur l'activité. En d'autres termes, le seuil de terminaison de l'inondation est du même ordre de grandeur que le seuil d'activité nécessaire afin que le protocole termine en un temps asymptotiquement optimal. En particulier, si $\widehat{p} \geq \frac{c \ln n}{n}$ avec $c > 1$, une seule étape d'activité pour chaque nœud est suffisante pour que l'inondation

se termine en temps optimal. Pareillement, pour $\frac{1}{n} \ll \widehat{p} \leq \frac{c \ln n}{n}$ avec $c < 1$, et $np \geq \ln n\widehat{p}$, le seuil de terminaison $\Theta(\frac{\ln n}{np})$ est significativement plus petit que le temps optimal de l'inondation $\Theta(\frac{\ln n}{\ln(n\widehat{p})})$; pourtant, que chaque nœud soit actif pendant $O(\frac{\ln n}{np})$ étapes suffit pour que l'inondation se termine en temps optimal asymptotiquement $O(\frac{\ln n}{\ln(n\widehat{p})})$. Pour les cas restants, le seuil de terminaison et le temps optimal coïncident à une constante près.

Nous complétons ensuite nos résultats par l'étude du cas de la distribution initial arbitraire. En nous ramenant à un cas particulier des résultats précédents où le graphe initiale est tiré selon la distribution stationnaire. Plus précisément, nous nous ramenons au cas où pour chaque arête e, $q_e = p_e - 1$, c'est-à-dire au cas où tous les graphes de la séquence arête-markovienne sont indépendants. Nous donnons une méthodologie générale pour majorer le temps d'inondation dans un graphe à évolution arête-markovienne par le temps d'inondation où graphes sont indépendants, et donc par le diamètre de son graphe représentatif.

Nous appliquerons à nouveau cette seconde méthodologie générale au cas homogène. En particulier, dans le cas homogène, nous montrons que, quelle que soit la distribution initiale g (pas nécessairement $\mathcal{G}_{n,\widehat{p}}$), et pour tout taux de naissance et de mort $0 < p, q < 1$, le temps d'inondation k-active est presque sûrement $O\left(\frac{\ln n}{\ln(1+np)}\right)$ lorsque k est suffisamment grand. En étudiant les cas pour lesquels cette borne est exacte nous montrons que contrairement à ce que nous pouvions observer lors de l'étude du cas $g = \mathcal{G}_{n,\widehat{p}}$, le taux de mort q joue un rôle dans le comportement des graphes dynamiques à évolution markovienne. Nous appliquerons également cette seconde méthodologie générale à deux cas hétérogènes. Dans le premier cas hétérogène considéré, les graphes évolutifs sont régis par une séquence de distribution de degré $\mathbf{w}$ donnée arbitrairement. Nous donnons diverses bornes qui dépendent de certaines propriétés spécifiques à la séquence des degrés. Par exemple, si la séquence est "spécialement admissible", alors le temps d'inondation est borné par $O\left(\log n(\frac{1}{\log(1+\tilde{d})} + \frac{1}{w_{\min}})\right)$, où $w_{\min}$ désigne la plus petite valeur dans la distribution des degrés et $\tilde{d}$ est le degré moyen de deuxième ordre. Dans le second cas hétérogène considéré,lorsque la séquence de degré est loi de puissance, nous montrons que l'inondation se termine en temps presque sûrement (a.p.s.) logarithmique en le nombre de nœud. Remarquons que, tandis que dans le cas des graphes en loi de puissance évoluant aléatoirement, si $\beta > 2$ la connexion à la composante géante n'augmente pas significativement le temps d'inondation, et si $\beta < 2$ la plupart du temps est passé dans la phase de connexion.

Finalement, nous étudions la diffusion 1-Port dans les graphes à évolution arête markovienne homogène. Nous montrons que si $p + q \geq 1$ alors le temps de diffusion du protocole Random est au plus $O\left(\frac{1}{1-(1-p)^{n-1}} \log n\right)$, avec forte probabilité.

3.2 Méthodologie générale dans le cas de la distribution initiale stationnaire

3.2.1 Réduction du temps d'inondation au diamètre

Une des principales difficultés dans l'analyse des graphes dynamiques est de gérer les dépendances temporelles. Dans cette section, nous introduisons un outil qui se révélera très utile pour réduire le problème du calcul du temps de réalisation d'un protocole d'inondation dans des graphes à évolution arête-markovienne, dans le cas stationnaire, à un problème du calcul du diamètre d'une famille de graphes pondérés aléatoirement. L'avantage de cet outil est double. Premièrement, l'analyse de l'inondation se réduit à un calcul de diamètre (les deux paramètres qui coïncident dans le cas statique, mais sont a priori différents dans le cadre dynamique). Deuxièmement, l'analyse de suites de graphes aléatoires avec dépendance temporelle est réduite à l'analyse d'un graphe aléatoire unique.

Nous réduisons l'analyse de l'inondation dans $\mathcal{M}_{n,\mathbf{p},\mathbf{q}}$ vers le calcul du diamètre d'un graphe *pondéré* aléatoirement défini comme suit : soient a et b deux nombres réels tels que $0 < a < 1$ et $0 < b < 1$. Soit $Z_{a,b}$ une variable aléatoire à valeur dans $\mathbb{N}$ et définie par

$$\mathbb{P}\left(Z_{a,b} = z\right) = \begin{cases} a & \text{si } z = 1, \\ (1-a)(1-b)^{z-2}b & \text{si } z > 1. \end{cases}$$

Si a et b valent tous les deux 0, alors nous considérons que $Z_{a,b} = \infty$. Les arêtes de poids infini sont celles qui ne sont pas présentes.

Étant données deux variables aléatoires X et X', notons $X \sim X'$ le fait que X et X' sont identiquement distribuées (c'est-à-dire, $\mathbb{P}(X = x) = \mathbb{P}(X' = x)$ pour tout x). Une clique de n nœuds G appartient à la famille des graphes pondérés aléatoirement $\mathcal{G}_{n,\mathbf{a},\mathbf{b}}$ (où $\mathbf{a} = \{a_e, e\}$ et $\mathbf{b} = \{b_e, e\}$) lorsqu'à chacune de ses $\binom{n}{2}$ possibles arêtes e est attribué un poids (entier et positif) aléatoire de valeur Z_e, où $Z_e \sim Z_{a_e,b_e}$ indépendamment des autres arêtes. Afin de prendre en compte l'activité des protocoles d'inondation nous définissons la famille des graphes pondérés aléatoirement *k-bornés* $\mathcal{G}^{(k)}_{n,\mathbf{a},\mathbf{b}}$ comme suit : l'ensemble des nœuds est $[n]$, une arête e est présente si et seulement si $Z_e \leq k$, et lorsqu'une arête est présente elle obtient le poids Z_e.

Par définition, $\mathcal{G}_{n,\mathbf{a},\mathbf{b}} = \mathcal{G}^{(\infty)}_{n,\mathbf{a},\mathbf{b}}$. De plus, chaque arête e de $\mathcal{G}^{(k)}_{n,\mathbf{a},\mathbf{b}}$ est présente avec probabilité

$$\varrho_e = \sum_{z=1}^{k} \mathbb{P}\left(Z_e = z\right),$$

indépendamment des autres arêtes. Ainsi, en ignorant le poids des arêtes, $\mathcal{G}^{(k)}_{n,\mathbf{a},\mathbf{b}}$ devient $\mathcal{G}_{n,\boldsymbol{\varrho}}$. En particulier, par définition $\mathcal{G}^{(1)}_{n,\mathbf{a},\mathbf{b}} = \mathcal{G}_{n,\mathbf{a}}$ quelque soit $\mathbf{b}$. Par conséquent, par plusieurs aspects, les graphes pondérés aléatoirement k-bornés généralisent les graphes aléatoires.

Dans le cas où toutes les arêtes ont les mêmes probabilités de naissance et de mort, nous notons simplement $\mathcal{G}_{n,a,b}$ et $\mathcal{G}^{(k)}_{n,a,b}$ la famille de graphes pondérés aléatoirement. $\mathcal{G}_{n,\varrho}$ et $\mathcal{G}_{n,a}$ sont alors des familles de graphes aléatoires d'Erdős-Rényi classiques avec probabilité respectivement ϱ et a.

La distance entre deux nœuds d'un graphe pondéré est le poids minimum d'un chemin reliant ces nœuds dans le graphe, où le poids d'un chemin est la somme des poids de ses arêtes. L'excentricité d'un nœud est la plus grande distance entre ce nœud et un autre nœud dans le graphe. Pour tout $s \in [n]$, soit $X^{(k)}_s$ la variable aléatoire égale à l'excentricité de s dans le graphe pondéré aléatoirement k-borné $\mathcal{G}^{(k)}_{n,\widehat{\mathbf{p}},\mathbf{p}}$. Le lemme suivant relie la variable aléatoire $X^{(k)}_s$ au temps d'inondation $T^{(k)}_s$ dans le processus arête-markovien, et stipule que ces deux variables ont la même distribution de probabilité.

Lemme 3.2.1 (Lemme de réduction)**.**
Quelque soient $n \geq 1$, $\mathbf{p}$ et $\mathbf{q}$ tels que $0 < p_e, q_e < 1$, $k \in \mathbb{N}^+ \cup \{\infty\}$, et $s \in [n]$ nous savons que $T^{(k)}_s \sim X^{(k)}_s$, à savoir

$$\mathbb{P}(T^{(k)}_s = x) = \mathbb{P}(X^{(k)}_s = x), \qquad \forall x \geq 0,$$

où l'espace de probabilité du terme de gauche de l'égalité est $\mathcal{M}_{n,\mathbf{p},\mathbf{q}}$ et celui du terme de droite est $\mathcal{G}^{(k)}_{n,\widehat{\mathbf{p}},\mathbf{p}}$.

Avant d'entrer dans la preuve du lemme 3.2.1, donnons l'intuition de ce résultat. Prenons un nœud u qui reçoit le message à l'instant t, et un nœud $v \neq u$. Soit $e = \{u, v\}$. Intéressons-nous à la probabilité pour e d'être présente à un temps ultérieur à t. Il y a deux scénarios possibles :

- Un des scénarios est que v ait reçu le message à l'instant $t' \leq t$, dans ce cas, d'un point de vue de l'inondation, nous n'avons aucun intérêt à considérer e puisqu'au temps t le message a déjà atteint ses deux extrémités. De même, du point de vue de l'excentricité, l'arête e ne joue plus aucun rôle. En effet, une fois que le temps auquel le message a atteint ses deux extrémités est fixé, c'est l'excentricité de s dans $\mathcal{G}^{(k)}_{n,\widehat{\mathbf{p}},\mathbf{p}}$ sachant que u est à distance t de s et que v est à la distance t' de s qui nous intéresse. Ainsi, le poids de e n'est pas pertinent dans ce calcul.
- L'autre scénario possible est que v n'a pas encore reçu le message à l'instant t. Dans ce cas nous affirmons que la probabilité que l'arête e soit présente au temps $t+1$ est $\widehat{p_e}$. En effet, les arêtes sont mutuellement indépendantes. Ainsi, le fait que d'autres arêtes ont été présentes dans le passé (en particulier celles par lesquelles le message transité de s à u) n'a aucun impact sur la présence ou l'absence de e. Maintenant, en admettant que la probabilité pour l'arête e d'être présente au temps $t+1$ est $\widehat{p_e}$, la probabilité pour cette arête d'être présente au temps $t+t'$ pour la première fois après l'instant t est, pour $t \geq 2$, égal à $(1-\widehat{p_e})(1-p_e)^{t'-2}p_e$. C'est précisément le poids de e dans $\mathcal{G}^{(k)}_{n,\widehat{\mathbf{p}},\mathbf{p}}$.

Plus formellement, à partir d'une source $s \in [n]$ fixée, de l'activité $k \in \mathbb{N} \cup \{\infty\}$ et de $S = (G_0, G_1, ...)$ une suite de graphes, nous définissons un graphe pondéré H appelé graphe k-représentatif de S. Dans la preuve du Lemme de réduction, il sera montré que l'excentricité de s dans le graphe k-représentatif de S est égale au temps d'inondation du protocole k-actif dans S, à partir de la source s, et que la probabilité d'un graphe H d'être le graphe représentatif de S réalisation arbitraire d'un processus arête-markovien $\mathcal{M}_{n,\mathbf{p},\mathbf{q}}$ est la même que d'être sélectionné dans $\mathcal{G}_{n,\widehat{\mathbf{p}},\mathbf{p}}$ Ainsi, la famille de graphes pondérés aléatoirement $\mathcal{G}_{n,\widehat{\mathbf{p}},\mathbf{p}}$ est appelée ***famille représentative*** de $\mathcal{M}_{n,\mathbf{p},\mathbf{q}}$

Nous donnons à présent la construction formelle du graphe k-représentatif de la suite S. Pour un nœud $u \in [n]$, $t(u)$ désigne l'étape de temps auquel u reçoit le message pour la première fois, c'est-à-dire :

$$t(u) = \min\{t \mid u \notin \cup_{i=0}^{t-1} I_i^{(k)} \wedge u \in I_t^{(k)}\}.$$

Rappelons que $I_t^{(k)}$ désigne l'ensemble des nœuds informés lors des t premières étapes du protocole d'inondation k-actif, quelque soit $t \geq 0$, avec $I_0^{(k)} = \{s\}$. Notez qu'il peut arriver que $t(u) = \infty$. Pour chaque arête $e = \{u, v\}$ où $u, v \in [n]$ et $u \neq v$, nous définissons

$$t(e) = \begin{cases} \min\{t(u), t(v)\} & \text{si } t(u) < \infty \text{ ou } t(v) < \infty, \\ \perp & \text{sinon.} \end{cases}$$

Alors, pour chaque arête e, soit

$$t_{\text{after}}(e) = \min\{t \mid t > t(e) \wedge e \in E(G_t)\}.$$

Ainsi, si $t(e) \neq \perp$, $t_{\text{after}}(e)$ désigne la première étape à laquelle l'arête e apparaît après qu'une de ses extrémités ait été informée pour la première fois. Sinon, $t_{\text{after}}(e)$ désigne la première étape à laquelle l'arête e apparaît.

Définition. *Le graphe pondéré H est appelé le* **graphe k-représentatif** *de la suite S est défini comme suit. $V(H) = [n]$ et pour chaque paire de nœuds distincts $u, v \in [n]$:*

$$e = \{u, v\} \in E(H) \iff t_{after}(e) - t(e) \leq k$$

De plus si $e \in E(H)$, alors le poids de e est égal à

$$w(e) = t_{after}(e) - t(e).$$

La valeur $\perp$ dans la définition ci-dessus correspond aux arêtes dont les extrémités n'ont pas reçu le message, et le choix d'imposer que, dans ce cas, $t_{\text{after}}(e)$ désigne la première étape à laquelle l'arête e apparaît permet essentiellement de simplifier la preuve. Cette valeur n'intervient pas dans le calcul des distances dans H puisque dans ce cas les deux extrémités sont déconnectées de s, c'est-à-dire dans une autre composante connexe.

La preuve ci-dessous formalise ces arguments.

Démonstration. (du Lemme de réduction)

Par souci de simplicité, nous omettons de noter l'indice de l'arête auquel s'applique la probabilité. Ceci revient à donner uniquement la preuve dans le cas $\mathcal{M}_{n,p,q}$, c'est-à-dire que toutes les arêtes ont les mêmes probabilités de naissance ou de mort. Nous notons ϕ l'application qui à S renvoie son graphe k-représentatif H. Pour chaque $u \in V(H)$, soit $d(u)$ la distance pondérée de s à u dans H. L'importante propriété suivante est vérifiée par ϕ :

***Affirmation* 3.2.2.** *Soit $T_s(S)$ le temps de d'inondation du protocole k-actif de source s dans S et soit $\mathrm{ecc}_H(s)$ l'excentricité de s dans $H = \phi(S)$. Pour chaque $u \in [n]$, nous avons l'égalité $d(u) = t(u)$ et par conséquent $T_s(S) = \mathrm{ecc}_H(s)$.*

Preuve. Soit $u \in [n]$. Montrons d'abord que $t(u) \leq d(u)$. Si u n'est pas connecté à s dans H, alors $d(u) = \infty$ et donc $t(u) \leq d(u)$. Supposons alors que u est connecté à s dans H. Soit $u_0, u_1, u_2, \ldots, u_{\ell-1}, u_\ell$ un plus court chemin dans H de $s = u_0$ à $u = u_\ell$. En raisonnant par l'absurde, supposons qu'il existe une arête e de ce chemin pour laquelle $t(e) = \perp$ et montrons que nous arrivons alors à une contradiction. Soit alors $i > 0$ le plus petit index tel que $t(\{u_i, u_{i+1}\}) = \perp$. Le nœud u_i n'a pas reçu le message dans S, bien que u_{i-1} l'ait reçu. Puisque l'inondation est k-actif, ceci implique que $t_{\text{after}}(\{u_{i-1}, u_i\}) - t(\{u_{i-1}, u_i\}) > k$, et par conséquent $\{u_{i-1}, u_i\} \notin E(H)$, la contradiction voulue. Ainsi toutes les arêtes du chemin vérifient $t(\{u_i, u_{i+1}\}) \geq 0$. Par conséquent, pour tout $i = 0, \ldots, \ell-1$, nous savons que $t(u_{i+1}) \leq t(u_i) + w(\{u_i, u_{i+1}\})$. Le poids d'un chemin est alors égal à

$$\begin{aligned} d(u) &= \sum_{i=0}^{\ell-1} w(\{u_i, u_{i+1}\}) \\ &\geq \sum_{i=0}^{\ell-1} (t(u_{i+1}) - t(u_i)) \\ &= t(u_\ell) - t(u_0) \\ &= t(u) - t(s) \\ &= t(u). \end{aligned}$$

Montrons à présent que $t(u) \geq d(u)$. Si $t(u) = \infty$, alors l'inégalité est évidente. Supposons donc que $t(u) < \infty$. Il existe une suite de nœuds $u_0, u_1, u_2, \ldots, u_{\ell-1}, u_\ell$ tels que $s = u_0$, $u = u_\ell$ et que pour tout i où $1 \leq i \leq \ell$, u_i a été informé par u_{i-1}. Ceci implique que pour tout i où $1 \leq i \leq \ell$, l'arête $e_i = \{u_{i-1}, u_i\}$ appartient à $E(H)$ puisque le protocole d'inondation est k-actif. De plus

$$w(e_i) = t_{\text{after}}(e_i) - t(e_i) = t_{\text{after}}(e_i) - t(u_{i-1}) = t(u_i) - t(u_{i-1}).$$

Ainsi, il existe un chemin de s à u dont le poids $\sum_{i=1}^{\ell} w(e_i)$ est égal à $t(u) - t(s) = t(u)$. Donc, pour tout nœud u, la distance $d(u)$ de s à u dans H est au plus $t(u)$. Ce qui complète la preuve de l'affirmation 3.2.2. $\diamond$

D'après l'affirmation 3.2.2, le temps d'inondation du protocole k-actif, de source s dans S peut être obtenu en calculant l'excentricité de s dans le graphe $H = \phi(S)$. Ce graphe H est clairement un graphe aléatoire (qui dépend de la réalisation S du processus arête-markovien $\mathcal{M}_{n,p,q}$). D'autre part, H peut aussi être vu comme une réalisation de $\mathcal{G}^{(k)}_{n,\widehat{p},p}$, simplement parce que le poids d'une arête est entre 1 et k. Ainsi, en particulier, ϕ peut être vu comme une application de $\mathcal{M}_{n,p,q}$ dans $\mathcal{G}^{(k)}_{n,\widehat{p},p}$. Soit $\sim$ la relation d'équivalence dans $\mathcal{M}_{n,p,q}$ définie par

$$S \sim S' \iff \phi(S) = \phi(S').$$

Par cette définition, ϕ devient une injection de $\mathcal{M}_{n,p,q}/\sim$ vers $\mathcal{G}^{(k)}_{n,\widehat{p},p}$. En fait, ϕ est aussi une surjection. Pour montrer cette affirmation prenons H une réalisation de $\mathcal{G}^{(k)}_{n,\widehat{p},p}$. Pour toute arête $e = \{u, v\}$, $u, v \in [n]$, $u \neq v$, soit

$$d(e) = \begin{cases} \min\{d(u), d(v)\} & \text{si } d(u) < \infty \text{ ou } d(v) < \infty\,; \\ \perp & \text{sinon.} \end{cases}$$

Définissons $S = (G_0, G_1, \dots)$ comme la réalisation de $\mathcal{M}_{n,p,q}$ dans laquelle, pour chaque arête e et pour chaque temps $t \in [0, kn)$, nous avons que $e \in E(G_t)$ si et seulement si $e \in E(H)$ et $t = d(e) + w(e)$. (Pour $t \geq kn$, G_t est arbitraire). Par construction $\phi(S) = H$. Donc,

$$\begin{array}{rccc} \phi : & \mathcal{M}_{n,p,q}/\sim & \to & \mathcal{G}^{(k)}_{n,\widehat{p},p} \\ & S & \mapsto & H \end{array}$$

est une injection et une surjection. Nous montrons maintenant que ϕ est un morphisme qui préserve les probabilités :

***Affirmation* 3.2.3.** *Pour tout $H \in \mathcal{G}^{(k)}_{n,\widehat{p},p}$, $\mathbb{P}\left(\phi^{-1}(H)\right) = \mathbb{P}(H)$, où l'espace de probabilité du terme de gauche de l'égalité est $\mathcal{M}_{n,p,q}$, tandis que celui du terme de droite est $\mathcal{G}^{(k)}_{n,\widehat{p},p}$.*

Preuve. Soit $H \in \mathcal{G}^{(k)}_{n,\widehat{p},p}$. Par définition de ϕ, l'ensemble $\phi^{-1}(H)$ est égal à l'ensemble de toutes les suites $S = (G_0, G_1, \dots) \in \mathcal{M}_{n,p,q}$ telles que, pour chaque arête e, nous avons :

- si $e \in E(H)$ alors $\left(\bigwedge_{i=1}^{w(e)-1} \left(e \notin E(G_{t(e)+i})\right)\right) \wedge \left(e \in E(G_{t(e)+w(e)})\right)$;
- si $e \notin E(H)$ alors $\bigwedge_{i=1}^{k} \left(e \notin E(G_{t(e)+i})\right)$.

Définissons l'événement correspondant avec $d(e)$ qui remplace $t(e)$ dans la formule ci-dessus :

$$\begin{aligned} \mathcal{E}(e) \;=\; & \left[(e \in E(H)) \wedge \left(\bigwedge_{i=1}^{w(e)-1} \left(e \notin E(G_{d(e)+i})\right)\right) \wedge \left(e \in E(G_{d(e)+w(e)})\right)\right] \\ & \vee \left[(e \notin E(H)) \wedge \left(\bigwedge_{i=1}^{k} \left(e \notin E(G_{d(e)+i})\right)\right)\right]. \end{aligned}$$

Grâce à l'affirmation 3.2.2, nous savons que si $S \in \phi^{-1}(H)$, alors $t(u) = d(u)$ pour tout nœud $u \in [n]$, et donc que pour toute arête e, $d(e) = t(e)$. Soit $\mathcal{F}(e)$ désignant l'événement correspondant conjoint avec $\mathcal{E}(e)$, c'est-à-dire,

$$\mathcal{F}(e) \equiv \Big(\mathcal{E}(e) \wedge (t(e) = d(e))\Big).$$

Alors,

$$\mathbb{P}\left(\phi^{-1}(H)\right) = \mathbb{P}\left(\bigwedge_{e} \mathcal{F}(e)\right).$$

Trions les événements $\mathcal{E}(e)$ dans l'ordre croissant de la distance de e par rapport à s dans H. Donc, pour $e = \{u, v\}$, soit $d_{min}(e) = \min\{d(u), d(v)\}$. Notons que lorsque $d_{min}(e) < \infty$, alors $d_{min}(e) = d(e)$. Ainsi, nous obtenons une liste des événements $\mathcal{E}(e_1), \ldots, \mathcal{E}(e_m)$ avec $m = \binom{n}{2}$, où $0 \leq d_{min}(e_i) \leq d_{min}(e_{i+1})$ pour $i = 1, \ldots, m-1$. Nous montrons par induction sur z que

$$\mathbb{P}\left(\bigwedge_{i=1}^{z} \mathcal{F}(e_i)\right) = \prod_{i=1}^{z} \mathcal{P}(e_i)$$

où pour $e \in E(H)$,

$$\mathcal{P}(e) = \begin{cases} \widehat{p} & \text{si } w(e) = 1, \\ (1-\widehat{p})(1-p)^{w(e)-2}p & \text{sinon} \end{cases}$$

et pour $e \notin E(H)$,

$$\mathcal{P}(e) = (1-\widehat{p})(1-p)^{k-1}.$$

Cette propriété est triviale pour $z = 0$. Supposons qu'elle soit vraie pour $z \geq 0$. Par l'hypothèse d'induction, nous avons

$$\mathbb{P}\left(\bigwedge_{i=1}^{z+1} \mathcal{F}(e_i)\right) = \mathbb{P}\left(\mathcal{F}(e_{z+1}) \mid \bigwedge_{i=1}^{z} \mathcal{F}(e_i)\right) \cdot \prod_{i=1}^{z} \mathcal{P}(e_i).$$

Nous montrons à présent que la probabilité du membre de gauche du produit est égal à $\mathcal{P}(e_{z+1})$. Pour cela nous réécrivons cette expression ainsi :

$$\mathbb{P}\left(\mathcal{E}(e_{z+1}) \mid (t(e_{z+1}) = d(e_{z+1})) \wedge \bigwedge_{i=1}^{z} \mathcal{F}(e_i)\right) \cdot \mathbb{P}\left(t(e_{z+1}) = d(e_{z+1}) \mid \bigwedge_{i=1}^{z} \mathcal{F}(e_i)\right)$$

Nous calculons séparément les deux termes du produit ci-dessus. Concentrons-nous, dans un premier temps, sur le premier événement conditionnel. Par l'indépendance entre les arêtes du processus arête-markovien considéré, l'événement $\mathcal{E}(e_{z+1})$ est indépendant de l'événement $\bigwedge_{i=1}^{z} \mathcal{E}(e_i)$ car ce sont des arêtes différentes et parce que les $d(e_i)$ sont des paramètres fixés (il n'y a donc pas de problème de dépendance temporelle, uniquement une présence ou absence indépendante des arêtes).

Nous affirmons que l'événement $\mathcal{E}(e_{z+1})$ est aussi indépendant de l'événement $\bigwedge_{i=1}^{z+1}(t(e_i) = d(e_i))$. L'intuition de cette affirmation est que le comportement d'une arête après un temps t donné est indépendant des événements liés à la réception du message par les nœuds aux temps $t' < t$. Et le comportement d'une arête dont les extrémités n'ont pas reçu le message est indépendant de la réception du message par des nœuds à des temps antérieurs ou de la non-réception du message par d'autres nœuds. Plus formellement, intéressons-nous aux événements de la forme "$t(e) = d(e)$" pour $e = \{u, v\}$, $u \neq v$. Pour cela notons $\mathcal{T}(e, t)$ l'événement "$t(e) = t$". Nous avons

$$\begin{aligned}
&\mathcal{T}'(e,t) = \\
&\quad \Big(\big(\{s,u\} \in G_t \wedge t \leq k\big) \vee \big(\{s,v\} \in G_t \wedge t \leq k\big)\Big) \vee \\
&\quad \bigvee_{\ell=1}^{t-1} \bigvee_{\mathbf{w} \in ([n]\setminus\{u,v\})^{\ell}} \bigvee_{\substack{\mathbf{t} \in \{1,\ldots,k\}^{\ell+1}, \\ \sum_{r=1}^{\ell+1} t_r = t}} \left[\begin{array}{l} \Big(\{s, w_1\} \in E(G_{t_1})\Big) \\ \wedge \left(\bigwedge_{j=1}^{\ell-1} \left(\{w_j, w_{j+1}\} \in G_{\sum_{r=1}^{j+1} t_r}\right)\right) \\ \wedge \left(\{w_\ell, u\} \in G_{\sum_{r=1}^{\ell+1} t_r} \vee \{w_\ell, v\} \in G_{\sum_{r=1}^{\ell+1} t_r}\right) \end{array} \right]
\end{aligned}$$

Finalement, pour $t = \infty$,

$$\mathcal{T}(e, \infty) = \bigwedge_{t \geq 0} \overline{\mathcal{T}(e,t)}.$$

Ainsi, l'événement $\mathcal{T}(e, t)$ peut s'exprimer comme une formule booléenne ne contenant que des expressions du type "$u = s$", "$v = s$" ou "$e' \in E(G_{t'})$", où $e' \neq e$ et $t' \leq t$. Plus important, en définissant $t_{min}(e') = \min\{t(w), t(w')\}$ pour chaque arête $e' = \{w, w'\}$, nous obtenons que les événements dans $\mathcal{T}(e, t)$ ne font pas référence à des arêtes e' vérifiant $t_{min}(e') < t$.

Ceci a pour conséquence que pour $e = e_{z+1}$ et $t = d_{min}(e_{z+1})$, nous obtenons que l'événement $\bigwedge_{i=1}^{z+1}(t(e_i) = d(e_i))$ ne dépend que des arêtes e' telles que $t_{min}(e') < d_{min}(e_{z+1})$. (Ceci vient du fait que $d_{min}(e_i) \leq d_{min}(e_{z+1})$ pour tout $i = 1, \ldots, z$). L'événement $\bigwedge_{i=1}^{z+1}(t(e_i) = d(e_i))$ ne dépend donc pas de $\mathcal{E}(e_{z+1})$ puisque ce dernier événement décrit le comportement de e_{z+1} à des temps au moins $d_{min}(e_{z+1})$. Ainsi,

$$\mathbb{P}\left(\mathcal{E}(e_{z+1}) \mid (t(e_{z+1}) = d(e_{z+1})) \wedge \bigwedge_{i=1}^{z} \mathcal{F}(e_i)\right) = \mathbb{P}\left(\mathcal{E}(e_{z+1})\right).$$

À présent, il nous reste plus qu'à effectuer le calcul de

$$\mathbb{P}\left(t(e_{z+1}) = d(e_{z+1}) \mid \bigwedge_{i=1}^{z} \mathcal{F}(e_i)\right).$$

Soit $H' = \cup_{1 \leq i \leq z} e_i$ le graphe (pondéré) induit par les arêtes e_i, $1 \leq i \leq z$. Lorsque l'événement $\bigwedge_{i=1}^{z} \mathcal{F}(e_i)$ se produit, le poids de ces arêtes e_i dans H' est le même que leurs poids $w(e_i)$ dans H (en donnant un poids infini aux arêtes absentes de H). Ainsi, pour chaque nœud u de H', les deux affirmations suivantes sont vraies :

1. la distance $d'(u)$ de u à s dans H' est la même que la distance $d(u)$ de u à s dans H
2. le temps $t'(u)$ auquel le nœud u reçoit le message de s dans le protocole d'inondation restreint aux arêtes de H' est le même que le temps $t'(u)$ auquel le nœud u reçoit le message de s dans S.

Par conséquent, $t'(u) = d'(u)$ pour chaque nœud de H'. D'autre part, d'après notre analyse de l'événement $\mathcal{T}$, nous obtenons que l'événement "$t(e_{z+1}) = d(e_{z+1})$" ne dépend que des arêtes e telles que $t(e) < d(e_{z+1})$, donc uniquement des arêtes e_i pour $1 \leq i \leq z$. En d'autres termes, l'événement "$t(e_{z+1}) = d(e_{z+1})$" est entièrement déterminé par H'. Ainsi, au moins une des extrémités de e_{z+1} appartient à H' et nous avons $t(e_{z+1}) = t'(e_{z+1}) = d'(e_{z+1}) = d(e_{z+1})$. Et par conséquent,

$$\mathbb{P}\left(t(e_{z+1}) = d(e_{z+1}) \mid \bigwedge_{i=1}^{z} \mathcal{F}(e_i)\right) = 1.$$

Ainsi

$$\mathbb{P}\left(\mathcal{F}(e_{z+1}) \mid \bigwedge_{i=1}^{z} \mathcal{F}(e_i)\right) = \mathbb{P}\left(\mathcal{E}(e_{z+1})\right).$$

Puisque $\mathbb{P}\left(\mathcal{E}(e_{z+1})\right) = \mathcal{P}(e_{z+1})$, nous obtenons finalement que

$$\mathbb{P}\left(\bigwedge_{i=1}^{z+1} \mathcal{F}(e_i)\right) = \prod_{i=1}^{z+1} \mathcal{P}(e_i)$$

ce qui complète la preuve de l'induction.

Pour $z = m$, nous obtenons

$$\mathbb{P}\left(\phi^{-1}(H)\right) = \mathbb{P}\left(\bigwedge_{i=1}^{m} \mathcal{F}(e_i)\right) = \prod_{i=1}^{m} \mathcal{P}(e_i).$$

Cette probabilité est précisément la probabilité de $H \in \mathcal{G}^{(k)}_{n,\widehat{p},p}$. Ceci complète la preuve du fait. $\diamond$

Nous avons maintenant tous les outils nécessaires pour démontrer le lemme. Soit $x \geq 0$,

$$\mathcal{S} = \{S \in \mathcal{M}_{n,p,q} \mid T^{(k)}_s(S) = x\},$$

et

$$\mathcal{H} = \{H \in \mathcal{G}^{(k)}_{n,\widehat{p},p} \mid \mathrm{ecc}_H(s) = x\},$$

tels que

$$\mathbb{P}(T^{(k)}_s = x) = \sum_{S \in \mathcal{S}} \mathbb{P}(S) \text{ et } \mathbb{P}(X^{(k)}_s = x) = \sum_{H \in \mathcal{H}} \mathbb{P}(H).$$

D'après l'affirmation 3.2.2, l'application ϕ transforme $\mathcal{S}$ en $\mathcal{H}$. De plus, puisque cette application $\phi : \mathcal{S}/\sim \to \mathcal{H}$ est à la fois injective et surjective donc bijective, nous obtenons

$$\mathbb{P}(T_s^{(k)} = x) = \sum_{H \in \mathcal{H}} \mathbb{P}(\phi^{-1}(H)).$$

Ainsi , par l'affirmation 3.2.3,

$$\mathbb{P}(T_s^{(k)} = x) = \sum_{H \in \mathcal{H}} \mathbb{P}(H),$$

qui donne

$$\mathbb{P}(T_s^{(k)} = x) = \mathbb{P}(X_s^{(k)} = x),$$

ce qui complète la preuve du lemme. □

Nous complétons cette section en soulignant un corollaire direct du lemme de réduction. Soit $\mathrm{diam}(\mathcal{G}_{n,a,b}^{(k)})$ la variable aléatoire égale au diamètre du graphe pondérée aléatoirement $\mathcal{G}_{n,a,b}^{(k)}$, le corollaire ci-dessous découle directement de la preuve en remarquant le simple fait que le diamètre d'un graphe est au moins l'excentricité de chacun de ses nœuds et au plus deux fois cette excentricité.

Corollaire 3.2.4. *Pour tout $n \geq 1$, tout $0 < p, q < 1$, tout $k \in \mathbb{N} \cup \{\infty\}$ et tout $s \in [n]$, nous avons*

$$\mathbb{P}\Big(\mathrm{diam}(\mathcal{G}_{n,\widehat{\mathbf{p}},\mathbf{p}}^{(k)}) \leq x\Big) \leq \mathbb{P}\Big(T_s^{(k)} \leq x\Big) \leq \mathbb{P}\Big(\mathrm{diam}(\mathcal{G}_{n,\widehat{\mathbf{p}},\mathbf{p}}^{(k)}) \leq 2x\Big), \qquad \forall x \geq 0.$$

Démonstration. Soit $x \geq 0$. Le lemme de réduction indique que $\mathbb{P}(T_s^{(k)} = x) = \mathbb{P}(X_s^{(k)} = x)$. Nous avons :

$$\mathrm{diam}(\mathcal{G}_{n,\widehat{\mathbf{p}},\mathbf{p}}^{(k)}) \leq x \;\Rightarrow\; X_s^{(k)} \leq x \;\Rightarrow\; T_s^{(k)} \leq x$$

et

$$T_s^{(k)} \leq x \;\Rightarrow\; X_s^{(k)} \leq x \;\Rightarrow\; \mathrm{diam}(\mathcal{G}_{n,\widehat{\mathbf{p}},\mathbf{p}}^{(k)}) \leq 2x.$$

Nous en déduisons donc le corollaire. □

Le lemme de réduction et le corollaire ci-dessus sont les principaux outils que nous utilisons pour calculer le seuil de terminaison de l'inondation. Ceci est fait dans la section suivante.

3.2.2 Trois scénarios particuliers

Nous donnons ici une méthodologie générale utilisée pour calculer une majoration du diamètre du graphe représentatif et, d'après le lemme de réduction, donc du temps d'inondation.

Soit un graphe pondéré, et soit C une composante connexe de ce graphe, nous notons $\mathrm{diam}(C)$ le diamètre pondéré dans C, et $\mathrm{diam}_1(C)$ le diamètre de C en ignorant les poids (c'est-à-dire, lorsque tous les poids sont à 1). Fixons C une composante connexe de $H^{(k)}$, défini comme le graphe pondéré aléatoirement obtenu à partir du graphe pondéré représentatif H d'une famille de graphes aléatoires $\mathcal{G}$ de n sommets. Soit $i \notin C$. Nous définissons $\mathrm{connect}(i, C)$ comme la variable aléatoire égale au plus petit h tel qu'il y ait une arête de i à C dans $H^{(h)}$, et définissons

$$\mathrm{connect}(C) = \max_{i \notin C} \mathrm{connect}(i, C) \ .$$

Le lemme de réduction établit que le temps d'inondation dans une suite de $\mathcal{S}$ de graphes aléatoires $(G_t)_{t \geq 0}$ dans $\mathcal{G}$ est égale à $\mathrm{diam}(H^{(\infty)})$. Le lemme suivant fournit une autre formulation, qui est plus facile à manipuler lorsqu'on s'intéresse aux cas spécifiques des graphes aléatoires.

Lemme 3.2.5. *$\mathcal{S} = (G_t)_{t \geq 0}$ une suite de graphes aléatoires obtenue à partir de $\mathcal{M}_{n,\mathbf{p},\mathbf{q},g}$. Quelque soit $k \geq 1$ et quelque soit une composante connexe C de $H^{(k)}$, le temps d'inondation dans $\mathcal{S}$ est au plus* $\mathrm{diam}(C) + 2\,\mathrm{connect}(C)$*, donc, en particulier, au plus* $k\,\mathrm{diam}_1(C) + 2\,\mathrm{connect}(C)$.

Démonstration. Fixons $k \geq 1$. D'après le lemme de réduction, il est suffisant de montrer que $\mathrm{diam}(H^{(\infty)})$ est majoré par l'expression dans l'énoncé du lemme. Considérons un nœud i en dehors de C. (S'il n'y a pas de tels nœuds alors évidemment $\mathrm{diam}(H^{(\infty)}) \leq \mathrm{diam}(C)$). Soit $h = \mathrm{connect}(C)$. Par définition, il y a un chemin de i vers tous les autres nœuds j de $H^{(h)}$, de poids total au plus $\mathrm{diam}(C) + 2h$. Donc, $\mathrm{diam}(H^{(\infty)}) \leq \mathrm{diam}(C) + 2\,\mathrm{connect}(C)$, qui donne le résultat désiré. De plus, puisque le poids de chaque arête de C est au plus k, nous obtenons que $\mathrm{diam}(C) \leq k\,\mathrm{diam}_1(C)$. □

Nous pouvons identifier trois scénarios d'applications du lemme 3.2.5, qui fournissent une méthodologie pour l'analyse de l'inondation dans les réseaux dynamiques.

Scénario 1 : $H^{(1)}$ est connecté. Alors, le temps d'inondation est majoré par le diamètre de $H^{(1)}$, c'est-à-dire le diamètre d'un graphe aléatoire G dans $\mathcal{G}$.

Scénario 2 : $H^{(1)}$ n'est pas connecté mais a une composante géante C. Alors le temps d'inondation est majoré par le diamètre de la composante géante plus le temps requis par les nœuds hors de la composante géante pour finalement avoir une arête les connectant à la composante géante. Le premier est le diamètre de la composante géante d'un graphe aléatoire G de $\mathcal{G}$. Le dernier nécessite uniquement un calcul individuel, pour chaque nœud i hors de la composante, impliquant seulement les $p_{i,j}$ pour $j \in C$.

Scénario 3 : $H^{(1)}$ n'a pas de composante géante (voire figure 3.1). Alors, le temps d'inondation peut être majoré en cherchant, dans un premier temps, la plus petite valeur k telle que $H^{(k)}$ ait une composante géante C. Une fois ceci atteint, le diamètre de C est majoré par k fois le diamètre (non pondéré) de

la composante géante dans un modèle de graphe aléatoire $\mathcal{G}'$ relié à k et $\mathcal{G}$. Attention, contrairement à $H^{(1)}$, le graphe $H^{(k)}$ n'est pas nécessairement un graphe aléatoire de la famille de graphes aléatoires originale $\mathcal{G}$. C'est par contre toujours le cas dans $\mathcal{M}_{n,p,q}$. Finalement, nous majorons le temps d'inondation en sommant la majoration sur le diamètre de C à la valeur de connect(C) calculé comme dans le deuxième scénario.

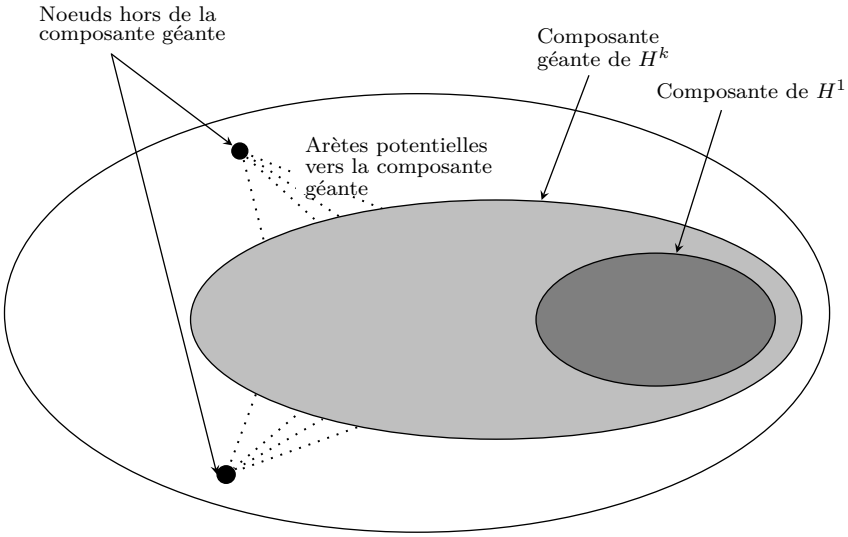


FIGURE 3.1 – Illustration des trois scénarios.
La figure est une représentation abstraite du graphe pondéré représentatif H, avec la plus grande, ou plus grosse, composante $C^{(1)}$ de $H^{(1)}$ dépeinte en gris foncé, et la plus grande, ou la plus grosse, composante $C^{(k)}$ de $H^{(k)}$ dépeinte en gris clair. $C^{(1)}$ n'est pas une composante géante, tandis que $C^{(k)}$ est géante (c'est-à-dire, contient une fraction du nombre total de nœuds ou du nombre total d'arêtes) pour k assez grand. L'inondation prend au plus le diamètre de $C^{(k)}$, plus le temps pour un nœud externe, dépeint en noir, pour avoir une arête le connectant à $C^{(k)}$.

Le lemme 3.2.5 est suffisamment flexible pour être utilisé pour tous les types de composantes "géantes", incluant ceux qui sont géantes, non pas en termes de taille (c'est-à-dire le nombre de nœud) mais en termes de *volume* (c'est-à-dire le nombre d'arêtes). En fait, il peut être utilisé avec n'importe quelle sous-graphe connexe C, et ce lemme 3.2.5 peut même être utilisé récursivement sur C.

3.3 Application aux graphes à évolution arête-markovienne homogène

Dans cette section, nous calculons le seuil de terminaison du protocole inondation dans les graphes à évolution arête-markovienne $\mathcal{M}_{n,p,q}$, pour l'ensemble du spectre

	$0 < \widehat{p} \leq \frac{c}{n}, c > 0$	$\frac{1}{n} \ll \widehat{p} \leq \frac{c \ln n}{n}, c < 1$		$\widehat{p} \geq \frac{c \ln n}{n}, c > 1$
		$np \leq \ln n\widehat{p}$	$np \geq \ln n\widehat{p}$	
Graphe de $\mathcal{G}_{n,\widehat{p}} = \mathcal{G}^{(1)}_{n,\widehat{p}}$				
Temps d'inondation	$\Theta(\frac{\ln n}{np})$	$\Theta(\frac{\ln n}{np})$	$\Theta(\frac{\ln n}{\ln(n\widehat{p})})$	$\Theta(\frac{\ln n}{\ln(n\widehat{p})})$
Seuil de terminaison	$\Theta(\frac{\ln n}{np})$	$\Theta(\frac{\ln n}{np})$		1

TABLE 3.1 – Récapitulatif des résultats pour $\mathcal{M}_{n,p,q}$

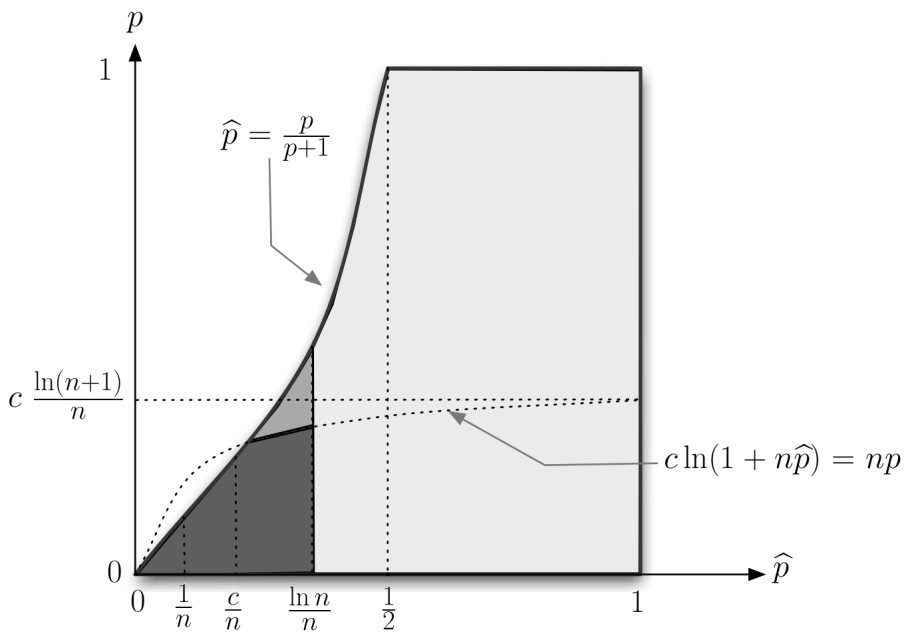


FIGURE 3.2 – Représentation graphique de nos résultats pour $\mathcal{M}_{n,p,q}$.

de $p, q \in (0, 1)$. Nous montrons que le temps d'inondation reste du même ordre de grandeur que l'optimal lorsque l'inondation est restreinte au cas k-actif, si k est (à une constante multiplicative près) égal au seuil de terminaison. Ces résultats sont résumés dans la table 3.1.

Nous distinguons les trois régimes standards : au-dessus du seuil de connectivité de $\mathcal{G}_{n,\widehat{p}}$, en dessous de son seuil de connectivité en présence d'une composante géante, et en l'absence de composante géante.

Intuition de nos résultats pour $\mathcal{M}_{n,p,q}$.
Dans la figure 3.2 nous fournissons une représentation graphique de nos résultats pour $\mathcal{M}_{n,p,q}$.

La zone en gris clair correspond à des combinaisons de paramètres p et $\widehat{p}$ pour lesquels 1 étape d'activité garantit un temps d'inondation optimale. Dans cette zone, $\widehat{p} \geq c\frac{\ln n}{n}$ avec $c > 1$. Dans ce cas, le graphe aléatoire $\mathcal{G}_{n,\widehat{p}}$ est, asymptotiquement presque sûrement, connexe. Ainsi, intuitivement, le temps de l'inondation est juste le temps de traverser $\mathcal{G}_{n,\widehat{p}}$ avec une activité de 1. Comme le diamètre de $\mathcal{G}_{n,\widehat{p}}$ est $\Theta(\frac{\ln n}{\ln(n\widehat{p})})$, asymptotiquement presque sûrement, nous obtenons le résultat affiché dans la table 3.1.

La zone en gris moyen est celle où le seuil d'accessibilité est nettement plus petit que le temps d'inondation optimale. Elle correspond au cas où $\frac{1}{n} \ll \widehat{p} \leq \frac{c \ln n}{n}$ avec $c < 1$. Dans ce cas, $\mathcal{G}_{n,\widehat{p}}$ est, asymptotiquement presque sûrement, déconnecté mais a une composante géante d'un diamètre $\Theta(\frac{\ln n}{\ln(n\widehat{p})})$. Considérons un nœud en dehors de la composante géante. Le temps moyen pour produire une arête reliant ce nœud à un nœud de la composante géante est approximativement $\frac{1}{np}$. Ainsi, le nombre d'étapes jusqu'à ce que tous les nœuds en dehors de la composante géante aient une arête reliant celle-ci est approximativement $\frac{\ln n}{np}$. Par conséquent, les nœuds doivent être actifs pendant au moins $\frac{\ln n}{np}$ étapes. Le temps d'inondation est alors la somme du temps de connexion à la composante géante $\frac{\ln n}{n}$ et du diamètre de la composante géante $\frac{\ln n}{\ln(n\widehat{p})}$. La table 3.1 distingue deux sous-cas, en fonction de la valeur relative de np et $\ln(n\widehat{p})$. Sur la figure 3.2, la ligne pointillée $np = \ln(n\widehat{p})$ permet de séparer la zone de gris moyen de la zone gris foncé.

Enfin, la zone en gris foncé est celle où le seuil d'accessibilité est du même ordre de grandeur que le temps d'inondation optimale. Elle correspond au cas où $\widehat{p} < \frac{1}{n}$. Dans ce cas, $\mathcal{G}_{n,\widehat{p}}$ est une union de petites composantes. Intuitivement, obtenir une composante géante nécessite d'attendre $t = \frac{c}{np}$ étapes. En fait, on peut intuitivement redimensionner le temps d'un facteur $t = \frac{c}{np}$. Ce redimensionnement nous ramène à une situation similaire au cas précédent, avec une composante géante d'un diamètre d'ordre $\ln n$. Par ce redimensionnement, nous obtenons un temps total de $t \ln n = \frac{c \ln n}{np}$ étapes nécessaires pour traverser la composante géante. Ce temps est égal au temps requis pour connecter les nœuds en dehors de la composante géante à celle-ci. Ainsi le temps d'inondation est $\frac{c \ln n}{np}$.

Avant d'entrer dans les détails de la preuve, donnons une explication intuitive des valeurs du seuil de terminaison dérivée dans les différents régimes. Supposons que le protocole est actif pendant k étapes et effectuons une estimation approximative du nombre d'appels effectués au total. Chaque nœud exécute environ $n\widehat{p}$ appels à la première étape, et, pendant les $k-1$ autres étapes, il réalise environ np autres appels. Ainsi, le nombre total d'appels effectués par un nœud est environ $n(\widehat{p}+(k-1)p)$. Ainsi, le nombre total d'appels est d'environ $n^2(\widehat{p} + (k-1)p)$. À partir de là, le problème du collectionneur de coupons donne l'intuition que, pour tous les nœuds reçoivent le message, le nombre d'appels doit être $\Theta(n \ln n)$. Par conséquent, nous obtenons que $\widehat{p} + (k-1)p$ doit être $\Theta(n \ln n)$. Ainsi, pour $\widehat{p} \gg \frac{\ln n}{n}$, une seule étape d'activité est suffisante. Par contre, pour $\widehat{p} \ll \frac{\ln n}{n}$, une activité de $\Theta(\frac{\ln n}{np})$ est nécessaire pour que l'information atteigne tous les nœuds. Dans la suite de cette section, nous prouvons

$\mathcal{G}_{n,\widehat{p}} = \mathcal{G}^{(1)}_{n,\widehat{p}}$	k tel que $\mathcal{G}^{(k)}_{n,\widehat{p}}$ ait une composante géante C		$\mathrm{diam}_1(C)$		$\mathrm{connect}(C)$
	1	$*$	$O(\frac{\log n}{\log n\widehat{p}})$	$+$	0
	1	$*$	$O(\frac{\log n}{n\widehat{p}} + \frac{\log n}{\log n\widehat{p}})$	$+$	$O(\frac{\log n}{np})$
	$\Theta(\frac{1}{np})$	$*$	$O(\log n)$	$+$	$O(\frac{\log n}{np})$

TABLE 3.2 – Calcul des bornes supérieures pour $\mathcal{M}_{n,p,q}$

que les bornes fournies par cette explication intuitive sont réellement correctes. Pour les bornes supérieures, nous utilisons le lemme 3.2.5, tel que résumé dans la table 3.2

3.3.1 Au-dessus du seuil de connectivité

Si $\widehat{p} \geq c\,\frac{\ln n}{n}$, où $c > 1$, alors, a.p.s., chaque graphe d'une suite générée par $\mathcal{M}_{n,p,q}$ est connecté. Nous sommes alors dans le scénario 1. On peut donc s'attendre à ce que l'inondation dans $\mathcal{M}_{n,p,q}$ termine au moins aussi vite que dans $\mathcal{G}_{n,\widehat{p}}$. C'est ce qui a été établi dans [CMPS11] où il a été montré que si $\widehat{p} \geq c\,\frac{\ln n}{n}$ pour une constante c assez grande alors, a.f.p., $T_s = O(\frac{\ln n}{\ln(n\widehat{p})} + \ln\ln(n\widehat{p}))$ et $T_s = \Omega(\frac{\ln n}{\ln(n\widehat{p})})$. Le lemme de réduction de la section précédente nous permet d'étendre ces résultats en montrant que, a.p.s., $T_s = \Theta(\frac{\ln n}{\ln(n\widehat{p})})$ lorsque $n\widehat{p} - \ln n \to \infty$. D'autre part, le lemme de réduction nous permet aussi de montrer que si $n\widehat{p} - \ln n \to \infty$ alors, a.p.s., seulement *une* étape d'activité par nœud est suffisante pour que l'inondation termine en $O(\frac{\ln n}{\ln(n\widehat{p})})$ étapes. Donc, le seuil de terminaison doit l'inondation dans les graphes à évolution arête-markovienne avec $n\widehat{p} - \ln n \to \infty$ est juste de 1 et cette unique étape d'activité (par nœud) est suffisante à ce que l'inondation soit optimale, c'est-à-dire, asymptotiquement aussi rapide (à une constante près) que lorsque l'activité est non bornée.

Théorème 3.3.1. *Pour tout taux de naissance p et de mort q tels que $n\widehat{p} - \ln n \to \infty$, le temps d'inondation du protocole 1-actif dans $\mathcal{M}_{n,p,q}$ pour une source $s \in [n]$ est, a.p.s., $O(\frac{\ln n}{\ln(n\widehat{p})})$. De plus ce temps de d'inondation est asymptotiquement égal au temps d'inondation optimal.*

Démonstration. Pour la majoration, nous appliquons le lemme 3.2.5, avec $k = 1$. Ici, $\mathcal{G}^{(1)}_{n,\widehat{p},p} = \mathcal{G}_{n,\widehat{p}}$. En outre, si $n\widehat{p} - \ln n \to \infty$ alors $\mathcal{G}_{n,\widehat{p}}$ est a.p.s. connecté [Bol01] donc la composante géante C est l'intégralité du graphe, et tout le graphe $\mathrm{connect}(C) = 0$.

En fait, il a été démontré (voir [CL01]) que le diamètre[2] de $\mathcal{G}_{n,\widehat{p}}$ est, a.p.s., $O(\frac{\ln n}{\ln(n\widehat{p})})$. Par conséquent, a.p.s., $T_s^{(1)} \leq O(\frac{\ln n}{\ln(n\widehat{p})})$.

Comme mentionné précédemment dans ce chapitre, le fait que le protocole d'inondation (perpétuellement actif) ne termine pas avant $\Omega(\frac{\ln n}{\ln(n\widehat{p})})$ étapes a été montré dans [CMPS11] pour $\widehat{p} \geq c\,\frac{\ln n}{n}$ avec c une constante suffisamment grande. Le fait que cette borne inférieure reste vraie pour toutes les valeurs possibles de $\widehat{p}$ tel que $n\widehat{p} - \ln n \to \infty$ sera démontré plus tard dans la preuve du théorème 3.3.3. □

Remarque. La plupart de nos résultats sont donnés avec la garantie "asymptotiquement presque sûrement" (a.p.s.). La raison est que, dans la littérature, les résultats concernant les graphes aléatoires sont aussi donnés avec la garantie "asymptotiquement presque sûrement". Cependant, par le lemme de réduction, tout résultat, concernant la connectivité ou le diamètre des graphes aléatoires, exprimé avec forte probabilité se transpose directement à nos résultats avec la même garantie probabiliste.

3.3.2 En dessous du seuil de connectivité

Nous considérons à présent le cas où $\mathcal{G}_{n,\widehat{p}}$ n'est, asymptotiquement presque sûrement, pas connecté $\widehat{p} \leq c\,\frac{\ln n}{n}$ avec $c < 1$. (En particulier, p et $\widehat{p}$ tendent tous les deux vers zéro lorsque n tend vers l'infini). Commençons par établir une condition nécessaire pour que le protocole d'inondation k-actif termine, asymptotiquement presque sûrement.

Lemme 3.3.2. *Pour tout taux de naissance p et de mort q tels que $p \to 0$ et $\widehat{p} \to 0$ lorsque $n \to \infty$, le seuil de terminaison pour une source $s \in [n]$ dans $\mathcal{M}_{n,p,q}$ est, a.p.s., $\Omega(\frac{\ln n - n\widehat{p}}{np})$.*

Démonstration. Nous calculons ici une borne inférieure sur le poids $w(e)$ d'une arête de $\mathcal{G}^{(k)}_{n,\widehat{p},p}$ pour que l'inondation termine. Par définition de $\mathcal{G}^{(k)}_{n,\widehat{p},p}$, nous avons que $\mathbb{P}\left(w(e) \leq k\right) = 1 - (1-\widehat{p})(1-p)^{k-1}$. Afin que l'inondation termine a.p.s., le lemme de réduction affirme que cette probabilité doit être supérieure à $\frac{\ln n}{n}$ car sinon le graphe aléatoire induit par les arêtes de poids au plus k serait a.p.s. déconnecté. Par conséquent, $k \geq 1 + \frac{\ln(1-\frac{\ln n}{n}) - \ln(1-\widehat{p})}{\ln(1-p)} = 1 + \frac{\ln n - n\widehat{p}}{np}(1+o(1))$. □

Théorème 3.3.3. *Considérons des taux de naissance p et de mort q tels que $\frac{c}{n} \ll \widehat{p} \leq \frac{c' \ln n}{n}$ avec $c' < 1 < c$. Alors le seuil de terminaison à partir d'une source $s \in [n]$ dans $\mathcal{M}_{n,p,q}$ est, a.p.s., égal à $\Theta(\frac{\ln n}{np})$. Par ailleurs, le temps d'inondation optimal à partir d'une source $s \in [n]$ dans $\mathcal{M}_{n,p,q}$ est, a.p.s., égal à*

$$\Theta\left(\frac{\ln n}{np} + \frac{\ln n}{\ln n\widehat{p}}\right).$$

2. Le diamètre d'un graphe déconnecté est le plus grand diamètre (au sens usuel) de ses composantes connexes.

De plus, le protocole d'inondation k-actif à partir d'une source $s \in [n]$ dans $\mathcal{M}_{n,p,q}$ se termine a.p.s. en temps optimal dès que k est supérieur, à une constante près, au seuil de terminaison.

Démonstration. Par le lemme 3.3.2 et en utilisant le fait que le temps des inondations optimales est au moins égal au seuil de terminaison, nous obtenons que le temps de l'inondation est moins $\Omega(\frac{\ln n}{np})$ lorsque $\widehat{p} \leq \frac{c \ln n}{n}$ avec $c < 1$. Montrons aussi que le temps d'inondation est au moins $\Omega(\frac{\ln n}{\ln(n\widehat{p})})$. (Ce dernier résultat sera vrai pour tout $\widehat{p} \geq \frac{1}{n}$, et donc vrai dans le contexte du théorème 3.3.1). Rappelons que I_t désigne l'ensemble des nœuds qui ont connaissance du message à l'étape de temps t pendant l'exécution du protocole d'inondation. Nous supposons qu'au lieu d'une seule source s, c'est un ensemble I_0 de sources qui démarrent le processus d'inondation avec le même message, avec $|I_0| = \beta \ln n$ et $\beta > 1$. L'inondation à partir de I_0 se termine au moins aussi vite que l'inondation à partir d'une unique source. Soit $\tilde{p} = \max\{p, \widehat{p}\}$. Puisque $\widehat{p} \geq \frac{1}{n}$, nous avons $\tilde{p}n \geq 1$. Pour $U \subseteq [n]$, soit $\Gamma_t(U) = \{e = \{u, v\} \in E(G_t) \mid u \in U, v \notin U\}$. Nous avons alors

$$\mathbb{P}\left(|I_{t+1}| > (1 + e\tilde{p}n)|I_t|\right) \leq \mathbb{P}\left(|\Gamma_t(I_t)| > e\tilde{p}n|I_t|\right).$$

À présent, une arête reliant $u \in I_t$ à $v \notin I_t$ est présente dans G_t avec probabilité p si $u \in I_{t-1}$, et avec probabilité $\widehat{p}$ si $u \notin I_{t-1}$. En fait, $u \in I_{t-1}$ et $v \notin I_t$ impliquent que l'arête était absente dans G_{t-1}. Et si $u \notin I_{t-1}$, alors la présence d'une arête $\{u, v\}$ dans G_t est indépendante du passé, comme nous l'avons vu dans la preuve du lemme de réduction, cet événement survient alors dans G_t avec probabilité $\widehat{p}$. Toutes ces arêtes de $\Gamma_t(I_t)$ sont mutuellement indépendantes. Par conséquent, si $\mathrm{Bin}(a, b)$ désigne une variable aléatoire binomiale de paramètres a et b (i.e, la somme de a tirages de Bernoulli indépendants de probabilité b), nous obtenons

$$\mathbb{P}\left(|\Gamma_t(I_t)| > e\tilde{p}nx \mid |I_t| = x\right) \leq \mathbb{P}\left(\mathrm{Bin}((n-x)x, \tilde{p}) > e\tilde{p}nx \mid |I_t| = x\right)$$

pour tout $x \geq |I_0|$. Par la borne de Chernoff, nous en déduisons que

$$\mathbb{P}\left(\mathrm{Bin}(nx, \tilde{p}) > e\tilde{p}nx \mid |I_t| = x\right) \leq e^{-\tilde{p}nx}.$$

Puisque $\tilde{p} \geq \frac{1}{n}$ et $x \geq |I_0| \geq \beta \ln n$, nous obtenons que

$$\mathbb{P}\left(|I_{t+1}| > (1 + e\tilde{p}n)x \mid |I_t| = x\right) \leq n^{-\beta}.$$

Donc

$$\mathbb{P}\left(|I_{t+1}| > (1 + e\tilde{p}n)|I_t|\right) \leq n^{-\beta}.$$

Par l'inégalité de Boole, nous en déduisons

$$\mathbb{P}\left(|I_t| > (1 + e\tilde{p}n)^t|I_0|\right) \leq \sum_{i=0}^{t-1} \mathbb{P}\left(|I_{i+1}| > (1 + e\tilde{p}n)|I_i|\right) \leq \frac{t}{n^{\beta}}.$$

Pour $t < \frac{\ln n - \ln(\beta \ln n)}{\ln(1+e\tilde{p}n)}$, nous avons $(1+e\tilde{p}n)^t |I_0| > n$, et donc

$$\mathbb{P}\left(|I_{t+1}| > n\right) \leq \frac{\ln n - \ln |I_0|}{\ln(1+e\tilde{p}n)} n^{-\beta} \leq \frac{\ln n}{\ln(1+e)} n^{-\beta} = o(\frac{1}{n}).$$

À partir de ceci, nous pouvons conclure que lorsque $\widehat{p} \geq \frac{1}{n}$, le temps d'inondation, a.f.p., est au moins $\frac{\ln n - \ln |I_0|}{\ln(1+e\tilde{p}n)}$, ce qui est au moins $\Omega(\frac{\ln n}{\ln(\widehat{p}n)})$. Ainsi, indépendamment de p, si $\frac{1}{n} \leq \widehat{p} \leq \frac{c \ln n}{n}$ avec $c < 1$, alors, a.p.s.,

$$T_s \geq \Omega\left(\frac{\ln n}{np} + \frac{\ln n}{\ln(n\widehat{p})}\right).$$

Pour calculer une borne supérieure sur le temps d'inondation nous appliquons le lemme 3.2.5 dans le cas du scénario 2, c'est-à-dire que nous nous concentrons sur $\mathrm{diam}(\mathcal{G}^{(h)}_{n,\widehat{p},p})$. Puisque $\widehat{p} \geq \frac{c}{n}$ pour une constante $c > 1$, $\mathcal{G}_{n,\widehat{p}}$ contient a.p.s. une composante géante de taille au moins $g(c)n$, où $g(c) \in (0,1)$ est l'unique racine positive de $1 - x = e^{-cx}$ (voir par exemple, [AS00] ou [JLR00]). De plus, le diamètre de cette composante géante est, a.p.s., au plus $O(\frac{\ln n}{n\widehat{p}} + \frac{\ln n}{\ln(n\widehat{p})})$ (voir [CL01]). Ainsi, le sous graphe $\mathcal{G}^{(h)}_{n,\widehat{p},p}$ induit par toutes les arêtes de poids 1 contient, a.p.s., une composante géante C de taille au moins $g(c)n$ et de diamètre au plus $O(\frac{\ln n}{n\widehat{p}} + \frac{\ln n}{\ln(n\widehat{p})})$. Pour chaque nœud $u \notin C$, nous estimons maintenant $\mathrm{connect}(u, C)$, c'est-à-dire, le poids d'une arête reliant ce nœud à C. Soit $\mathcal{E}_u$ l'événement "le nœud $u \notin C$ n'a pas d'arête vers C dans $\mathcal{G}^{(h)}_{n,\widehat{p},p}$". Nous avons

$$\mathbb{P}\left(\mathcal{E}_u\right) \leq (1-p)^{(h-1)\cdot g(c)\cdot n}.$$

Soit $\mathcal{E} = \cup_u \mathcal{E}_u$. Par l'inégalité de Boole,

$$\mathbb{P}\left(\mathcal{E}\right) \leq (1 - g(c))n(1-p)^{(h-1)\cdot g(c)\cdot n}.$$

Considérons l'activité suivante :

$$h = 1 + \frac{2 \ln n}{g(c)np}.$$

Avec ce paramétrage de l'activité, nous obtenons que $\mathbb{P}\left(\mathcal{E}\right) \leq \frac{1}{n}$ et donc que le diamètre de $\mathcal{G}^{(h)}_{n,\widehat{p},p}$ est a.p.s. au plus $O(\frac{\ln n}{np} + \frac{\ln n}{\ln(n\widehat{p})} + \frac{\ln n}{n\widehat{p}})$. Puisque $\widehat{p} \geq \frac{p}{2}$, nous obtenons que, a.p.s.,

$$T_s^{(h)} \leq O\left(\frac{\ln n}{np} + \frac{\ln n}{\ln(n\widehat{p})}\right).$$

La preuve se termine en remarquant que le lemme 3.3.2 indique qu'une activité de $1 + \frac{2 \ln n}{g(c)np}$ est du même ordre de grandeur que le seuil de terminaison. □

Théorème 3.3.4. *Considérons des taux de naissance p et de mort q tels que $0 < \widehat{p} \leq \frac{c}{n}$ pour une constante $c > 0$. Alors, a.p.s., le seuil de terminaison et le temps d'inondation optimal à partir d'une source $s \in [n]$ dans$\mathcal{M}_{n,p,q}$ sont tous deux égaux à $\Theta(\frac{\ln n}{np})$.*

Démonstration. Nous considérons deux cas en fonction de la présence probable d'une composante géante dans $\mathcal{G}_{n,\widehat{p}}$ ou non. Supposons d'abord que $\widehat{p} = \frac{c}{n}$ pour une constante $c > 1$, il y a alors, a.p.s., une composante géante dans $\mathcal{G}_{n,\widehat{p}}$. Nous sommes alors aussi dans le scénario 2. Ainsi, avec les mêmes arguments que dans la preuve du théorème 3.3.3, nous obtenons que, a.p.s., $T_s \geq \Omega(\frac{\ln n}{np} + \frac{\ln n}{\ln(n\widehat{p})})$ et $T_s^{(k)} \leq O(\frac{\ln n}{np} + \frac{\ln n}{\ln(n\widehat{p})})$ pour l'activité $k = 1 + \frac{2\ln n}{g(c)np}$.

Supposons à présent que $\widehat{p} < \frac{1}{n}$, c'est-à-dire, $\mathcal{G}_{n,\widehat{p}}$ n'a probablement pas de composante géante. Ce qui correspond au scénario 2. Nous nous ramenons au cas précédent en effectuant un nombre suffisant d'étapes d'inondation afin d'augmenter la probabilité qu'une arête soit présente. Précisément, soit $d > 1$ une constante arbitraire, et soit $\gamma = \frac{d}{n}$. Calculons la plus petite valeur de k telle que pour toute arête e, $\mathbb{P}(w(e) \leq k) \geq \gamma$. Nous obtenons

$$\mathbb{P}(w(e) \leq k) = 1 - (1 - \widehat{p})(1 - p)^{k-1}$$

et donc

$$\begin{aligned} k &\geq 1 + \frac{\ln(1-\gamma) - \ln(1-\widehat{p})}{\ln(1-p)} \\ &= 1 + \frac{d - n\widehat{p}}{np}(1 + o(1)) \\ &= \Theta\left(\frac{1}{np}\right). \end{aligned}$$

Pour un tel $k = \Theta(\frac{1}{np})$, nous assurons que $\mathcal{G}^{(k)}_{n,\widehat{p},p}$ a une composante géante C de taille au moins $g(d)n$ et de diamètre majoré par $O(k\frac{\ln n}{\ln n\gamma}) \leq O(k \ln n)$. C'est-à-dire, après $O(\frac{\ln n}{np})$ étapes de temps, nous pouvons appliquer le résultat du cas précédent, et, après connect$(C) \leq \frac{\ln n}{np}$ étapes supplémentaires de, l'inondation se termine. La borne inférieure $\Omega(\frac{\ln n}{np})$ est donnée par le lemme 3.3.2. □

3.4 Une méthodologie générale dans le cas de la distribution initiale arbitraire

Dans les sections précédentes, nous nous sommes intéressés au cas stationnaire, où le graphe initial G_0 est tiré dans $\mathcal{G}_{n,\widehat{\mathbf{p}}}$. Dans cette dernière section, nous allons considérer un cas plus général, en supposant que G_0 n'est pas nécessairement un graphe aléatoire de la distribution stationnaire, mais peut être tiré selon une distribution

arbitraire g. En particulier, nous pourrions très bien imposer que le graphe G_0 soit le n-stable, le graphe de n nœuds avec aucune arête. Nous rappelons que le processus arête-markovien, où G_0 est pris selon une probabilité initiale de distribution g sur une famille de graphes est notée $\mathcal{M}_{n,\mathbf{p},\mathbf{q},g}$, et $\mathcal{M}_{n,p,q,g}$ lorsque toutes arêtes ont les mêmes probabilités. Le temps de d'inondation du protocole k-actif à partir de la source s dans $\mathcal{M}_{n,\mathbf{p},\mathbf{q},g}$ est alors noté $T_{s,g}^{(k)}$.

3.4.1 Majoration du temps d'inondation par un diamètre

Dans le cas où $g = \mathcal{G}_{n,\widehat{p}}$, le lemme de réduction nous a permis de relier le temps d'inondation $T_s^{(k)}$ dans $\mathcal{M}_{n,p,q}$ à l'excentricité $X_s^{(k)}$ de s dans le graphe pondéré aléatoirement $\mathcal{G}_{n,\widehat{p},p}^{(k)}$. Dans le cas de la distribution arbitraire g une relation similaire bien que moins forte peut être établie. Soit $Y_s^{(k)}$ une variable aléatoire égale à l'excentricité de s dans le graphe pondéré aléatoirement k-borné $\mathcal{G}_{n,0,\mathbf{p}}^{(k)}$. Notons que $Y_s^{(k)}$ se réfère à $\mathcal{G}_{n,0,\mathbf{p}}^{(k)}$, tandis que $X_s^{(k)}$ (utilisé dans le lemme de réduction) se réfère à $\mathcal{G}_{n,\widehat{\mathbf{p}},\mathbf{p}}^{(k)}$.

Lemme 3.4.1. *Pour tout $n \geq 1$, tout $0 < p, q < 1$, tout $k \in \mathbb{N}^+ \cup \{\infty\}$ et tout $s \in [n]$, $T_{s,g}^{(k)}$ est stochastiquement dominé par $Y_s^{(k)}$, c'est-à-dire*

$$\mathbb{P}(T_{s,g}^{(k)} \geq x) \;\leq\; \mathbb{P}(Y_s^{(k)} \geq x), \qquad \forall x \geq 0,$$

où l'espace probabilité du terme de gauche de l'égalité est $\mathcal{M}_{n,p,q,g}$ et celui du terme de droite est $\mathcal{G}_{n,0,\mathbf{p}}^{(k)}$.

Démonstration. Afin de relier $T_{s,g}^{(k)}$ et $Y_s^{(k)}$, nous utilisons une troisième variable aléatoire. Pour définir cette dernière, nous utilisons un processus arête-markovien légèrement différent. Pour toutes les arêtes possibles e, considérons deux suites infinies de bits $\mathbf{P}_e$ et $\mathbf{Q}_e$, où chaque bit est le résultat d'un tirage de Bernoulli indépendant de probabilité p_e dans la suite $\mathbf{P}_e$, et de probabilité q_e dans la suite $\mathbf{Q}_e$. On peut voir le processus arête-markovien comme défini par la collection des $\frac{n(n-1)}{2}$ paires $(\mathbf{P}_e, \mathbf{Q}_e)$ de suites de Bernoulli mutuellement indépendantes, une pour chaque arête e, comme suit. Initialement G_0 est tiré selon g, ensuite pour chaque arête e,

- si $e \in G_t$ alors : $e \in G_{t+1}$ si et seulement si $\mathbf{Q}_e[t] = 0$;
- si $e \notin G_t$ alors : $e \in G_{t+1}$ si et seulement si $\mathbf{P}_e[t] = 1$;

Soit $\mathbf{P}[t]$ et $\mathbf{Q}[t]$ les vecteurs $(\mathbf{Q}_e[t])_e$ et $(\mathbf{Q}_e[t])_e$, respectivement, et soit $\mathbf{P}$ et $\mathbf{Q}$ les suites $(\mathbf{P}[t])_{t\geq 0}$ et $(\mathbf{Q}[t])_{t\geq 0}$, respectivement. Ainsi, à partir de maintenant nous verrons une suite de graphes $(G_t)_{t\geq 0}$ définie par le processus arête-markovien comme une suite de graphes $(G_t)_{t\geq 0}$ définie par le triplet $(G_0, \mathbf{P}, \mathbf{Q})$.

Nous pouvons à présent décrire la troisième variable aléatoire utilisée pour relier $T_{s,g}^{(k)}$ et $Y_s^{(k)}$. Considérons un triplet $H = (G_0, \mathbf{P}, \mathbf{Q})$ et définissons $S_{s,g}^{(k)}(H)$ comme le temps d'inondation du protocole k-actif dans la suite de graphes correspondant à H où, dans l'exécution du protocole, lorsque un nœud u reçoit un message pour la

première fois au temps noté t toutes les arêtes incidentes à u sont retirées du graphe au temps $t+1$. Après le temps $t+1$, le comportement de ces arêtes redevient "normal", c'est-à-dire qu'il suit les règles spécifiées par $\mathbf{P}$ et $\mathbf{Q}$. Plus précisément, considérons une arête e qui a été retirée au temps $t+1$, si $e \notin G_{t+1}$ alors le comportement de e sera le même à partir de l'étape $t+1$, dans les deux processus correspondant aux variables aléatoires $S_{s,g}^{(k)}$ et $T_{s,g}^{(k)}$; si $e \in G_{t+1}$, alors le comportement de e sera opposé dans $S_{s,g}^{(k)}$ et $T_{s,g}^{(k)}$ jusqu'à un moment $t'+1$ où t' est le plus petit entier tel que $t' \geq t+1$ et $\mathbf{P}_e[t'] \neq \mathbf{Q}_e[t']$. Après l'étape $t'+1$ le comportement de e restera perpétuellement identique dans les deux processus.

Notons que $S_{s,g}^{(k)}(H)$ est défini uniquement comme une fonction de H, et donc les espaces de probabilité pour $S_{s,g}^{(k)}$ et pour $T_{s,g}^{(k)}$ sont les mêmes. En fait, $S_{s,g}^{(k)}(H)$ correspond aux performances du protocole d'inondation k-actif qui est "localement" retardé à chaque nœud.

Nous affirmons alors que pour tout $x \geq 0$, $\mathbb{P}(T_{s,g}^{(k)} \geq x) \leq \mathbb{P}(S_{s,g}^{(k)} \geq x)$. Pour établir cette affirmation, nous montrons en fait un résultat plus fort : soit $H = (G_0, \mathbf{P}, \mathbf{Q})$ une suite de graphes, nous montrons que $T_{s,g}^{(k)}(H) \leq S_{s,g}^{(k)}(H)$. C'est-à-dire que nous montrons que si le protocole d'inondation k-actif est "localement" retardé à chaque nœud, alors il est aussi "globalement" retardé. Pour tout nœud u, soit t_u l'étape de temps à laquelle le nœud u reçoit le message pour la première fois dans le protocole d'inondation k-actif dans H, et soit t'_u l'étape de temps à laquelle le nœud u reçoit le message pour la première fois dans le protocole d'inondation k-actif localement retardé dans H. Supposons qu'il existe un nœud u tel que $t'_u < t_u$ dans le but d'obtenir une contradiction. Prenons u l'un de ces nœuds avec le plus petit t'_u. Soit v un nœud qui a envoyé le message à u à cette étape t'_u dans le protocole d'inondation k-actif localement retardé. Nous avons que $t'_v < t'_u$. Ainsi, par la minimalité de t'_u, nous avons $t_v \leq t'_v$. Ainsi

$$t_v \leq t'_v < t'_u < t_u.$$

Par la définition du protocole d'inondation k-actif localement retardé, l'arête $e = \{u, v\}$ n'est pas présente dans le graphe à l'instant t, $t'_v + 1 \leq t < t'_u$, pendant l'exécution du protocole d'inondation k-actif localement retardé. Cependant, elle était présente dans le graphe à l'instant t'_u. Ainsi, $\mathbf{P}_e[t'_u - 1] = 1$. Alors, pendant l'intervalle de temps $t_v + 1 \leq t < t_u$, e n'était pas présente dans le graphe lors de l'exécution du protocole d'inondation k-actif. En particulier, e n'était pas présent au temps $t'_u - 1$. Le fait que $\mathbf{P}_e[t'_u - 1] = 1$ implique donc que e doit être présent au temps t'_u, en contradiction avec le fait que u reçoit le message de v au temps $t_u > t'_u$. Ceci achève la preuve de l'affirmation.

À partir de l'affirmation, nous pouvons facilement compléter la démonstration du lemme. En effet, considérons la preuve du lemme de réduction où la probabilité pour chaque arête du graphe pondéré aléatoirement d'avoir un poids de 1 est nulle, au lieu d'être $\widehat{p}$. Cela correspond au comportement du protocole d'inondation localement retardé. Par les mêmes arguments que ceux utilisés dans la preuve du lemme de réduction, nous obtenons que, pour tout $x \geq 0$, $\mathbb{P}(S_{s,g}^{(k)} \geq x) = \mathbb{P}(Y_s^{(k)} \geq x)$. Par

conséquent, pour tout $x \geq 0$, $\mathbb{P}(T_{s,g}^{(k)} \geq x) \leq \mathbb{P}(Y_s^{(k)} \geq x)$, comme souhaité. □

Rappelons que $Y_s^{(k)}$ désigne la variable aléatoire égale à l'excentricité de s dans le graphe pondéré aléatoirement k-borné $\mathcal{G}_{n,0,\mathbf{p}}^{(k)}$. À présent, soit $Z_s^{(k)}$ la variable aléatoire égale à l'excentricité de s dans le graphe pondéré aléatoirement $(k-1)$-borné $\mathcal{G}_{n,\mathbf{p},\mathbf{p}}^{(k-1)}$.

Corollaire 3.4.2. *Pour tout $n \geq 1$, tout $0 < p, q < 1$, tout $k \in \mathbb{N}^+ \cup \{\infty\}$ et tout $s \in [n]$, $T_{s,g}^{(k)}$ est stochastiquement dominé par $Y_s^{(k)}$, c'est-à-dire*

$$\mathbb{P}(T_{s,g}^{(k)} \geq x) \ \leq \ \mathbb{P}(Y_s^{(k)} \geq x), \qquad \forall x \geq 0,$$

où l'espace probabilité du terme de gauche de l'égalité est $\mathcal{M}_{n,p,q,g}$ et celui du terme de droite est $\mathcal{G}_{n,\mathbf{p},\mathbf{p}}^{(k)}$.

Démonstration. Pour tout $x \geq 0$, nous avons

$$\mathbb{P}(Y_s^{(k)} \leq x) \leq 2 \cdot \mathbb{P}(Z_s^{(k)} \leq x). \tag{3.1}$$

En effet, soit ϕ une bijection de $\mathcal{G}_{n,\mathbf{p},\mathbf{p}}^{(k-1)}$ vers $\mathcal{G}_{n,0,\mathbf{p}}^{(k)}$ consistant à additionner 1 aux poids de chaque arête. ϕ est un morphisme qui préserve les probabilités, c'est-à-dire, $\mathbb{P}(G) = \mathbb{P}(\phi(G))$ pour tout $G \in \mathcal{G}_{n,\mathbf{p},\mathbf{p}}^{(k-1)}$. Comme tous les poids des arêtes de $G \in \mathcal{G}_{n,\mathbf{p},\mathbf{p}}^{(k-1)}$ sont strictement positifs (et dans $\mathbb{N}$), leur ajouter 1 les multiplie par au plus 2. Ainsi, l'excentricité de $\phi(G)$ est au plus le double de celle de G.

En combinant l'équation (3.1) avec le lemme 3.4.1, nous obtenons que le temps de l'inondation du protocole k-actif dans les graphes à évolution arête-markovienne est, à une constante près, asymptotiquement le même que le diamètre $Z^{(k)}$ □

Remarque. Puisque $\mathcal{G}_{n,\mathbf{p},\mathbf{p}}$ est exactement la famille k -représentative d'une suite aléatoire de graphes aléatoires tirés indépendamment dans $\mathcal{G}_{n,\mathbf{p}}$, étudier les suites aléatoires permet également d'obtenir une majoration dans cas général.

3.4.2 Minoration du temps d'inondation

Une preuve similaire à celle du lemme 3.4.1 donne le résultat suivant, qui permet de se ramener au cas des graphes qui se densifient, c'est-à-dire dont aucune arête ne disparaît, pour toute arête e, $q_e = 0$.

Lemme 3.4.3. *Pour tout $n \geq 1$, tout $0 < p, q < 1$, tout $k \in \mathbb{N}^+ \cup \{\infty\}$ et tout $s \in [n]$, $T_{s,g}^{(k)}$ est stochastiquement dominé par $Y_s^{(k)}$, c'est-à-dire*

$$\mathbb{P}\left(T_s^{(k)}(\mathcal{M}_{\mathbf{p},\mathbf{q},g} \geq x)\right) \ \leq \ \mathbb{P}\left(T_s^{(k)}(\mathcal{M}_{\mathbf{p},\mathbf{0},g}) \geq x\right), \qquad \forall x \geq 0,$$

où l'espace probabilité du terme de gauche de l'égalité est $\mathcal{M}_{n,p,q,g}$ et celui du terme de droite est $\mathcal{G}_{n,0,\mathbf{p}}^{(k)}$.

L'étude de l'inondation dans un graphe qui se densifie donne donc une minoration de l'inondation dans un graphe évolutif.

Dans le cas où $g = \emptyset$, la famille de graphes pondérés aléatoirement $\mathcal{G}_{n,0,\mathbf{p}}$ sert aussi à représenter l'évolution d'un graphe évolution arête-markovienne quand $q = 0$ et que le graphe initial est vide. Le poids d'une arête du graphe tiré dans $\mathcal{G}_{n,0,\mathbf{p}}$ ne correspond,alors, plus au temps mis pour apparaître une fois que l'une de ses extrémités est informée, mais simplement l'étape à laquelle elle apparaît dans le processus.

Remarque. On pourrait aussi bien tirer le graphe initial aléatoirement dans $\mathcal{G}_{n,\mathbf{p}}$ et utiliser la famille $\mathcal{G}_{n,\mathbf{p},\mathbf{p}}$. Ceci revient au même, à un décalage dans le temps prêt d'une seule étape.

Le résultat suivant donne une méthode de calcul du temps d'inondation dans $\mathcal{M}_{\mathbf{p},0,g}$.

Théorème 3.4.4. *Si avec une certaine garantie de probabilité, pour un graphe aléatoire H de $\mathcal{G}_{n,0,\mathbf{p}}$ vérifie* $\mathrm{diam}_1(H^{(k)}) \leq k$ *où $H^{(k)}$ est le sous-graphe de H contenant les arêtes de poids $\leq k$, alors, avec la même garantie de probabilité, le temps d'inondation dans $\mathcal{M}_{n,\mathbf{p},0,g}$ est $O(k)$. De plus, si k est le plus petit vérifiant la propriété, avec cette garantie, et que $g = \emptyset$, alors le temps est $\Theta(k)$.*

Démonstration. Soit H tiré dans $\mathcal{G}_{n,0,\mathbf{p}}$, soit k tel que l'excentricité ne tenant pas compte des poids des arêtes (c'est-à-dire, en nombre d'arêtes et non en poids d'un chemin), noté ecc_1, de s dans $H^{(k)}$ vérifie $\mathrm{ecc}_{1,H^{(k)}}(s) \leq k$ alors $2k$ étapes sont suffisantes pour que l'inondation termine. En effet, même si l'inondation commençait à l'étape k, alors il ne faudrait que k étapes supplémentaires pour atteindre tous les nœuds dans la mesure où après la k-ième étape, l'excentricité de s est au plus k.

La borne est exacte si k est le plus petit tel que $\mathrm{diam}_1(H)^{(k)} \leq k$ et que $g = \emptyset$. Si $g = \emptyset$, $\mathcal{G}_{n,0,\mathbf{p}}$ est la classe représentative de $\mathcal{M}_{n,\mathbf{p},0,g}$. Le diamètre du premier suit donc la même loi que le temps d'inondation du second. Et si $\mathrm{diam}_1(H^{(k)}) > k$ alors $\mathrm{ecc}_{H^{(k)}}(s) > \frac{k}{2}$ et $\mathrm{ecc}_{H^{(\frac{k}{2})}}(s) > \frac{k}{2}$ donc $frack2$ étapes ne sont pas suffisantes pour informer tout le monde. □

Ce théorème généralise, au cas $\mathcal{M}_{n,\mathbf{p},\mathbf{q},g}$, une technique qui a été employée dans l'étude [CMM+10], dans le cadre de $\mathcal{M}_{n,p,q,g}$ pour donner une borne inférieure. Il est montré qu'avec forte probabilité que si $g = \emptyset$ et $q = 0$:

- si $p \leq \frac{\log\log n}{n}$ le temps d'inondation est $\Theta(\frac{\log n}{np})$.
- si $\frac{\log\log n}{n} \leq p \leq \frac{\log^\alpha n}{n}$ le temps d'inondation est $\Theta(\frac{\log n}{\log\log n})$.
- si $\frac{\log^\alpha n}{n} \leq p$ le temps d'inondation est $\Theta(\frac{\log n}{\log(np)})$.

3.5 Applications

Dans cette section nous appliquons le résultat principal de la section précédente à trois modèles de graphes à évolution arête-markovienne. Comme le résultat du

lemme 3.4.1 permet de se ramener à une suite de graphes indépendants, c'est-à-dire au cas où pour toute arête e, $p_e + q_e = 1$. Nous nous intéressons donc uniquement à la définition des taux de naissance.

Dans un premier temps nous considérons le cas homogène, où toutes les arêtes apparaissent avec probabilité p. Ensuite, nous nous intéressons au cas où la probabilité p_e dépend du degré espéré des deux extrémités de e. Et finalement, nous étudions le cas particulier du précédent où les degrés suivent une loi de puissance.

3.5.1 Application aux graphes à évolution arête-markovienne homogène

En utilisant le lemme 3.4.1 nous pouvons montrer une borne supérieure générale sur le temps d'inondation du protocole k-actif, quand G_0 est tiré selon une distribution de graphe g arbitraire. Le théorème ce-dessous considère une activité k suffisamment grande. Il y a en effet des distributions g qui empêchent l'inondation k-active de terminer pour k trop petit. Un cas extrême est quand G_0 est un stable, dans ce cas le protocole d'inondation 1-actif ne termine évidemment pas, simplement parce que la source n'a pas d'arête incidente pour transmettre le message à l'étape 0.

Théorème 3.5.1. *Pour toute distribution initiale g, quelques soient les taux de naissance p et de mort q tels que $0 < p, q < 1$, le temps d'inondation k-active dans les graphes à évolution arête-markovienne $\mathcal{M}_{n,p,q,g}$ est,a.p.s., $O(\frac{\ln n}{\ln(1+np)})$ lorsque $k \geq 2$ si $np - \ln n \to \infty$, ou $k = \Omega(\frac{\ln n}{np})$ sinon.*

Démonstration. Nous appliquons directement le corollaire 3.4.2 aux résultats que nous avions déjà. Si $np - \ln n \to \infty$, cette dernière valeur est, a.p.s., $O(\frac{\ln n}{\ln(np)})$ pour $k \geq 2$ (voir la preuve du théorème 3.3.1). Sinon, si $\frac{c}{n} \leq p$ pour $c > 1$ alors $Z^{(k)}$ est, a.p.s., $O(\frac{\ln n}{np} + \frac{\ln n}{\ln(np)}) = O(\frac{\ln n}{\ln(1+np)})$ pour $k = \Omega(\frac{\ln n}{np})$ (voir la preuve du théorème 3.3.3). Finalement, si $0 < p \leq \frac{c}{n}$ pour une constante $c > 0$, alors $Z^{(k)}$ est, a.p.s., $O(\frac{\ln n}{np})$ pour $k = \Omega(\frac{\ln n}{np})$ (voir la preuve du théorème 3.3.4). □

Nous complétons cette section par une brève discussion sur l'exactitude de la borne du théorème 3.5.1. Cette borne est exacte dans le sens où pour tout taux de naissance $p \in (0, 1]$, il existe une distribution de probabilité g et un taux de mort q telle que la borne $O(\frac{\ln n}{\ln(1+np)})$ soit exacte. Il suffit de considérer le cas $g = \mathcal{G}_{n,p}$, alors pour $q = 1 - p$ nous avons que $\widehat{p} = p$, pour lequel la table 3.1 montre l'exactitude de la borne.

Un autre cas intéressant est lorsque $g = \{G_0\}$ où G_0 est le n-stable (c'est-à-dire, un graphe de n nœuds sans arêtes). Dans ce cas [CMM+10] a montré que, pour tout $0 < p, q < 1$, le temps d'inondation est, a.f.p., au moins $\Omega(\frac{\ln n}{np})$. Pour $p \leq \frac{1}{n}$, cette borne est équivalente à $\Omega(\frac{\ln n}{\ln(1+np)})$. Pour $p \geq c\frac{\ln n}{n}$ avec $c > 1$, [CMM+10] a aussi montré que le temps d'inondation est au moins $\Omega(\frac{\ln n}{\ln(1+np)})$. Dans l'intervalle $\frac{1}{n} \leq p \leq \frac{\ln n}{n}$, [CMM+10] a donné des bornes pour le cas particulier $q = 0$. Le temps d'inondation est, a.f.p., $\Theta(\frac{\ln n}{np})$ si $\frac{1}{n} \leq p \leq \frac{\ln \ln n}{n}$ et est, a.f.p., $\Theta(\frac{\ln n}{\ln \ln n})$ si

<table>
<tr><td></td><td>$0 < p \leq \frac{1}{n}$,</td><td colspan="2">$\frac{1}{n} \leq p \leq \frac{\log n}{n}$,
$p \leq \frac{\log\log n}{n}$ | $\frac{\log\log n}{n} < p$</td><td>$p \geq \frac{c \log n}{n}$, $c > 1$</td></tr>
<tr><td rowspan="2">Temps d'inondation</td><td rowspan="2">$\Theta(\frac{\log n}{\log(1+np)})$
$= \Theta(\frac{\log n}{np})$</td><td colspan="2">$O(\frac{\log n}{\log(1+np)})$
atteint dans le cas $q = 1 - p$.</td><td rowspan="2">$\Theta(\frac{\log n}{\log(1+np)})$</td></tr>
<tr><td>$\Omega(\frac{\log n}{np})$</td><td>$\Omega(\frac{\log n}{\log\log n})$
atteint dans le cas $q = 0$.</td></tr>
</table>

TABLE 3.3 – Bornes pour $\mathcal{M}_{n,p,q,g}$ avec $g = \emptyset$, [CMM+10] et cette section.

$\frac{\ln\ln n}{n} \leq p \leq \frac{\ln n}{n}$. Ces deux bornes ne coïncident pas avec la borne supérieure du théorème 3.5.1. En fait, comme mentionné dans [CMM+10], le taux de mort q peut jouer un rôle (asymptotiquement) significatif lorsque $\frac{1}{n} \leq p \leq \frac{\ln n}{n}$. Nous montrons que c'est effectivement le cas en démontrant que la borne du théorème 3.5.1 est atteinte dans cet intervalle pour au moins $q = 1 - p$. En effet, pour $q = 1 - p$, la suite $(G_t)_{t\geq 1}$ est composée de graphes aléatoires indépendants dans $\mathcal{G}_{n,p}$. Puisque G_0 est un stable, inonder dans $(G_t)_{t\geq 0}$ requiert exactement une étape de plus que l'inondation dans $(G_t)_{t\geq 1}$, c'est-à-dire $1 + \Theta(\frac{\ln n}{\ln(1+np)})$. Ces bornes sont rappelées dans la table 3.3.

3.5.2 Application aux graphes évolutifs avec suite de degré arbitraire

Nous considérons ici la généralisation des graphes aléatoires d'Erdős-Rényi introduite dans [CL02] et nous étudions le temps d'inondation dans une suite de graphes tirés aléatoirement indépendamment dans cette famille de graphes aléatoires.

Étant donnée une liste de n réels strictement positifs $\mathbf{w} = (w_1, \ldots, w_n)$ avec $\max_{1\leq i\leq n} w_i^2 < \sum_{i=1}^n w_i$, cette partie traite du modèle de graphes aléatoires $\mathcal{G}_{\mathbf{w}}$ tels que définis dans [CL02]. La probabilité de présence d'une arête $e_{i,j}$ entre les nœuds i et j (pas nécessairement distincts) est fixée à

$$p_{i,j} = \frac{w_i w_j}{\sum_{k=1}^n w_k} \ . \tag{3.2}$$

Il est facile de voir qu'avec de tels paramètres, le degré espéré d_i d'un nœud i dans $\mathcal{G}_{\mathbf{w}}$ est précisément le "poids" w_i du nœud i. De plus le degré espéré moyen du graphe G dans $\mathcal{G}_{\mathbf{w}}$ est

$$d = \frac{1}{n} \sum_{i=1}^n w_i$$

et le degré moyen de deuxième ordre est

$$\tilde{d} = \frac{\sum_{i=1}^n w_i^2}{\sum_{i=1}^n w_i} \ .$$

L'analyse de l'inondation trouve tout son intérêt quand chacun des graphes G_t dans $\mathcal{G}$ a peu de chances d'être connecté. Dans un tel cadre, la façon dont l'information se diffuse est alors entièrement gouvernée par la dynamique des graphes dans la suite, où de nouvelles arêtes apportent de nouvelles opportunités à l'information de se propager alors que la disparition d'arêtes pourrait engendrer des délais dans la propagation du message à travers le réseau (dynamique). Ainsi cette étude se concentre sur le cas où le degré moyen d est constant. Il doit être comparé au cas où $p = \frac{c}{n}$ pour une constante c, dans le modèle $\mathcal{G}_{n,p}$.

Deux paramètres importants vont impacter l'efficacité de l'inondation dans les réseaux dynamiques. Le premier est la *taille* (c'est-à-dire, le nombre de nœuds) de la plus grande composante connexe, la seconde est le *volume* (c'est-à-dire, approximativement le nombre d'arêtes) de la plus "grosse" composante connexe. Plus spécifiquement, en suivant les notations de [CL02], la taille d'un ensemble de sommet S est notée $|S|$, et son volume est noté

$$\mathrm{Vol}(S) = \sum_{i \in S} w_i \ .$$

Ainsi le volume de S est la somme des degrés espérés des nœuds de S et $\mathrm{Vol}(G) = \mathrm{Vol}(V(G)) = nd$.

Par la suite, sans perte de généralité, nous supposons que $w_1 \leq w_2 \leq \cdots \leq w_n$.

Dans cette section, nous portons notre attention sur le cas général d'une suite de graphes mutuellement indépendants de $\mathcal{G}_{\mathbf{w}}$, telle que définie au début de la sous-section, avec $\mathbf{w} = (w_1, \ldots, w_n)$ pour une suite $\mathbf{w}$ arbitraire.

Nous commençons par un cas particulier qui s'applique uniquement lorsque la suite $\mathbf{w}$ satisfait certaines propriétés c'est-à-dire lorsqu'elle est *spécialement admissible*. Cette contrainte demande que $d > 1 + \epsilon$ pour une constante ϵ strictement positive et l'existence d'un ensemble de nœuds U satisfaisant des contraintes spécifiques sur son volume (voir [CL03] pour plus de détails [3]).

Théorème 3.5.2. *Supposons que* $\mathbf{w}$ *soit une suite spécialement admissible. Soit* $\mathcal{S} = (G_t)_{t \geq 0}$ *une suite de graphes aléatoires mutuellement indépendants,* $G_t \in \mathcal{G}_{\mathbf{w}}$, $t \geq 0$. *Alors, a.p.s., l'inondation dans* $\mathcal{S}$ *prend au plus* $O\Big(\log n(\frac{1}{\log \widetilde{d}} + \frac{1}{w_1})\Big)$ *étapes.*

Démonstration. D'après [CL02, CL06b], pour $d > 1$, un graphe de $\mathcal{G}_{\mathbf{w}}$ a, a.p.s., une composante géante C de volume au moins $g\ \mathrm{Vol}(G)$ où g ne dépend que de d. Nous appliquons alors la méthodologie résumée dans le scénario 2 de la sous-section 3.2.2. Il est prouvé dans [CL03] que, puisque $\mathbf{w}$ est spécialement admissible, le diamètre de C est $\Theta(\frac{\log n}{\log \widetilde{d}})$.

3. Nous ne mentionnons pas la condition sur $\tilde{d}$ établie dans [CL03], puisqu'elle n'est utilisée que pour la minoration.

Soit $\mathcal{E}_i$ l'événement "le nœud $i \notin C$ n'a pas d'arêtes vers C dans $H^{(h)}$".

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_i) &= \prod_{j\in C}(1-p_{ij})^{h-1} \\
&\leq \prod_{j\in C} e^{-(h-1)p_{ij}} \\
&= e^{-(h-1)\sum_{j\in C} p_{ij}} \\
&= e^{-(h-1)\sum_{j\in C} \frac{w_i w_j}{\mathrm{Vol}(H^{(h-1)})}} \\
&\leq e^{-(h-1)w_i \sum_{j\in C} \frac{w_j}{\mathrm{Vol}(H^{(h-1)})}} \\
&\leq e^{-(h-1)w_i \frac{\mathrm{Vol}(C)}{\mathrm{Vol}(H^{(h-1)})}} \\
&\leq e^{-(h-1)w_i g} .
\end{aligned}$$

Comme les poids des nœuds sont rangés par ordre croissant, nous obtenons : $\mathbb{P}(\mathcal{E}_i) \leq e^{-(h-1)w_1 g}$. Par conséquent, en choisissant $h = 1 + \frac{2\log n}{w_1 g}$, nous obtenons que $\mathbb{P}(\mathcal{E}_i) \leq \frac{1}{n^2}$. Puis en appliquant l'inégalité de Boole, nous déduisons que $\mathrm{connect}(C) = O(\frac{\log n}{w_1})$, avec probabilité au moins $1 - \frac{1}{n}$. Nous concluons en appliquant Lemme 3.2.5. □

Le théorème suivant ne fait aucune hypothèse sur $\mathbf{w}$. Soit Λ l'ensemble des fonctions décroissantes $\lambda : \mathbb{N} \to \mathbb{N}$, satisfaisant $1 \leq \lambda(n) \leq (1-\epsilon)n$ pour n assez grand et pour une constante $\epsilon > 0$, et satisfaisant $\lambda(n) \to \infty$ quand $n \to \infty$.

Théorème 3.5.3. *Soit $\mathcal{S} = (G_t)_{t\geq 0}$ une suite de graphes aléatoires mutuellement indépendants, $G_t \in \mathcal{G}_{\mathbf{w}}$, $t \geq 0$. Alors, a.p.s., l'inondation dans $\mathcal{S}$ prend au plus $O\Big(\min_{\lambda\in\Lambda}\{\frac{\log\lambda(n)}{w_1} + \frac{\log n}{\log(1+w_{\lambda(n)}^2/d)}\}\Big)$ étapes.*

Démonstration. Soient $\mathbf{w} = (w_1, \ldots, w_n)$ et $\lambda \in \Lambda$. Rappelons que, dans $\mathcal{G}_{\mathbf{w}}$, une arête est présente entre i et j avec probabilité

$$p_{i,j} = \frac{w_i w_j}{\sum_k w_k} = \frac{w_i w_j}{\mathrm{Vol}(G)}.$$

Pour obtenir le résultat établi dans le théorème, nous utilisons le scénario 3 comme décrit dans la sous-section 3.2.2. Pour ce faire, nous allons construire une composante géante C obtenue en modifiant l'espace de probabilité initiale $\mathcal{G}_{\mathbf{w}}$. Soient $w' = w_{\lambda(n)}$ et $p' = \frac{w'^2}{nd}$. Définissons alors le modèle de graphe aléatoire $\mathcal{G}'$ où une arête est présente entre i et j avec probabilité

$$p'_{i,j} = \begin{cases} p_{i,j} & \text{si } i < \lambda(n) \text{ ou } j < \lambda(n) \\ p' & \text{sinon.} \end{cases}$$

D'où, pour chaque, $t \geq 0$, tout tirage donné de $\frac{n(n-1)}{2}$ valeurs mutuellement indépendantes $r_{i,j} \in [0,1]$, $i \neq j$, choisies aléatoirement uniformément résulte en un graphe

G_t de $\mathcal{G}_{\mathbf{w}}$ (l'arête $e_{i,j}$ est présente si $r_{i,j} \leq p_{i,j} = \frac{w_i w_j}{\sum_k w_k}$) et un graphe G'_t de $\mathcal{G}'$ (l'arête $e_{i,j}$ est présente si $r_{i,j} \leq p'_{i,j}$). Par définition de $\mathcal{G}'$, puisque $p'_{i,j} \leq p_{i,j}$ pour tout i, j, nous avons que G'_t est une sous-graphe de G_t. Par conséquent, pour tout $t \geq 0$, si $I_\tau \in \{1, \ldots, n\}$ est l'ensemble des nœuds informés dans $\mathcal{S} = (G_t)_{t\geq 0}$ au temps τ, et $I'_\tau \in \{1, \ldots, n\}$ est l'ensemble des nœuds informés dans $\mathcal{S}' = (G'_t)_{t\geq 0}$ au temps τ, nous avons

$$I'_\tau \subseteq I_\tau \Rightarrow I'_{\tau+1} \subseteq I_{\tau+1} \ .$$

Si le temps d'inondation dans $\mathcal{S}'$ est $f(\mathcal{S}')$, alors le temps d'inondation dans $\mathcal{S}$ est au plus $f(\mathcal{S}')$. Par conséquent, le temps d'inondation dans $\mathcal{G}_{\mathbf{w}}$ est stochastiquement dominé par le temps d'inondation dans $\mathcal{G}'$. Nous nous concentrons donc maintenant uniquement sur $\mathcal{G}'$.

Soit C l'ensemble des $n' = n - \lambda(n) + 1$ nœuds $i \geq \lambda(n)$. Chacun des sous-graphes aléatoires, sur l'ensemble de sommets dans C, pour un graphe aléatoire de $\mathcal{G}'$, est un graphe aléatoire d'Erdős-Rényi de $\mathcal{G}_{n',p'}$. Soit H'_C le graphe pondéré représentatif de $\mathcal{G}_{n',p'}$. Puisque $\lambda(n) \leq (1-\epsilon)n$, nous avons $n' \to \infty$ quand $n \to \infty$. Par conséquent, de la section précédente, nous savons que le diamètre de H'_C est au plus $O(\frac{\log n'}{\log(1+n'p')})$ où la notation en grand-O est pour $n \to \infty$. À présent, soit H' le graphe pondéré représentatif de $\mathcal{G}'$. Le diamètre de C dans H' est le diamètre de H'_C, et ainsi est au plus $O(\frac{\log n'}{\log(1+n'p')})$. Donc,

$$\mathrm{diam}(C) = O\Big(\frac{\log(n - \lambda(n) + 1)}{\log(1 + (n - \lambda(n) + 1)\frac{w'^2}{nd})}\Big)$$

Pour calculer le temps d'inondation, il reste juste à majorer connect(C). Soit $k \geq 1$ et soit $\mathcal{E}_i$ l'événement "le nœud $i \notin C$ n'a pas d'arête vers C dans $H'^{(k)}$". Avec les mêmes arguments que dans la preuve du théorème 3.5.2, nous obtenons :

$$\mathbb{P}(\mathcal{E}_i) = \prod_{j\in C}(1 - p_{ij})^k \leq e^{-k\sum_{j\in C}\frac{w_i w_j}{\mathrm{Vol}(G)}} \leq e^{-kw_i\frac{\mathrm{Vol}(C)}{\mathrm{Vol}(G)}} \ .$$

Nous avons $\mathrm{Vol}(G) = nd$, et puisque C contient les n' nœuds de plus grand poids, nous avons $\mathrm{Vol}(C) \geq n'd$. Puisque nous avons aussi $\mathrm{Vol}(C) \geq n'w'$, nous obtenons finalement

$$\mathbb{P}(\mathcal{E}_i) \leq e^{-kw_1\frac{n-\lambda(n)+1}{nd}\max\{d,w'\}}$$

Par conséquent, en choisissant

$$k = \frac{2nd\ \log\lambda(n)}{w_1\ (n - \lambda(n) + 1)\ \max\{d, w'\}}$$

nous obtenons que $\mathbb{P}(\mathcal{E}_i) \leq \frac{1}{\lambda(n)^2}$. En appliquant l'inégalité de Boole, sur les $\lambda(n) - 1$ nœuds les plus "faibles", nous en déduisons que

$$\mathrm{connect}(C) = O\Big(\frac{nd\ \log\lambda(n)}{w_1\ (n - \lambda(n) + 1)\ \max\{d, w'\}}\Big)$$

avec probabilité au moins $1 - O(\frac{1}{\lambda(n)})$.

Par le lemme 3.2.5, nous pouvons donc conclure que, a.p.s., le temps d'inondation dans $\mathcal{S}$ est majoré par

$$O\Big(\frac{\log(n-\lambda(n)+1)}{\log(1+(n-\lambda(n)+1)\frac{w'^2}{nd})} + \frac{nd\ \log\lambda(n)}{w_1\ (n-\lambda(n)+1)\ \max\{d,w'\}}\Big)\ .$$

Puisque $\lambda \in \Lambda$, nous obtenons que $n - \lambda(n) \geq \epsilon n$. Ainsi, a.p.s., le temps d'inondation dans $\mathcal{S}$ est majoré par

$$O\Big(\frac{\log n}{\log(1+\frac{w'^2}{d})} + \frac{\log\lambda(n)}{w_1}\Big)$$

comme voulu. □

Citons deux applications intéressantes du théorème 3.5.3. La première suppose de prendre le nœud λ dont le degré espéré w_λ est approximativement égal au degré espéré moyen d. La seconde suppose de prendre le nœud λ dont le degré espéré au carré w_λ^2 est approximativement égal au degré moyen de deuxième ordre $\frac{1}{n}\sum_i w_i^2 = d\tilde{d}$. Remarquons que le résultat obtenu dans ce dernier cas a la même forme que celle du théorème 3.5.2.

Corollaire 3.5.4. *S'il existe $\lambda \in \Lambda$ tel que $w_{\lambda(n)} = \Theta(d)$, alors l'inondation prend, a.p.s., un nombre d'étapes au plus*

$$O\Big(\frac{\log\lambda(n)}{w_1} + \frac{\log n}{\log(1+d)}\Big)\ .$$

S'il existe $\lambda \in \Lambda$ tel que $w_{\lambda(n)}^2 = \Theta(d\tilde{d})$, alors l'inondation prend, a.p.s., un nombre d'étapes au plus

$$O\Big(\frac{\log\lambda(n)}{w_1} + \frac{\log n}{\log(1+\tilde{d})}\Big)\ .$$

Remarque. La démonstration du théorème 3.5.3 n'utilise ni les résultats connus sur l'existence d'une composante géante dans $\mathcal{G}_{\mathbf{w}}$, ni ceux sur le diamètre d'une telle composante. En effet, en l'absence d'information a priori sur la suite de degré espéré $\mathbf{w} = (w_1, \ldots, w_n)$, ces résultats peuvent difficilement être exploités en utilisant notre méthodologie pour déduire de meilleures bornes que celles indiquées dans le théorème 3.5.3. Afin de voir pourquoi, rappelons que, pour $d > 1$, un graphe de $\mathcal{G}_{\mathbf{w}}$ a, a.p.s., une composante géante en volume (voir [CL02, CL06b]). Pour $d \leq 1$, nous pouvons appliquer le scénario 3 de la sous-section 3.2.2, en augmentant la probabilité de présence de chaque arête, comme suit. Par définition, il y a une arête entre les nœuds i et j dans $H^{(k)}$ avec probabilité

$$1 - (1-p_{ij})^k = 1 - \Big(1 - \frac{w_i w_j}{nd}\Big)^k\ .$$

En fait, par construction, nous avons $H^{(k)} \in \mathcal{G}_{\mathbf{w}^{(k)}}$ avec $\mathbf{w}^{(k)} = (w_1^{(k)}, \ldots, w_n^{(k)})$ et

$$w_i^{(k)} = \sum_j \left[1 - \left(1 - \frac{w_i w_j}{nd}\right)^k\right] .$$

Nous pouvons alors calculer k tel que

$$\sum_i w_i^{(k)} \geq 2n \ ,$$

pour lequel nous déduisons que le degré espéré moyen $d^{(k)}$ de $\mathbf{w}^{(k)} = (w_1^{(k)}, \ldots, w_n^{(k)})$ satisfaisant $d^{(k)} > 2$. C'est-à-dire, $H^{(k)}$ a a.p.s. une composante géante C de volume $\mathrm{Vol}(C) \geq g_2 \cdot \mathrm{Vol}(G)$ pour une constante g_2, où l'indice 2 se réfère au fait que $d^{(k)} \geq 2$. Continuer le calcul de l'inondation par notre méthode fait cependant face à deux obstacles. Le premier, [CL03] a montré que le diamètre de C est $\Theta(\frac{\log n}{\log \widetilde{d^{(k)}}})$ où $\tilde{d}^{(k)}$ est le degré espéré moyen de deuxième ordre de la suite $\mathbf{w}^{(k)}$, mais ce résultat n'est valable que sous l'hypothèse que la suite $\mathbf{w}^{(k)}$ est *spécialement admissible*, ce qui peut ou non être le cas, en fonction de la suite $\mathbf{w}$ initiale. Deuxièmement, même si le diamètre de C peut être exprimé par une formule close, le temps pour se connecter à C est minoré par le temps mis pour se connecter par le nœud le plus "faible". Plus précisément, C contient une fraction ϵ des nœuds. Pour calculer une majoration sur le temps nécessaire aux $(1-\epsilon)n$ nœuds hors de C pour se connecter à C, l'utilisation de l'inégalité de Boole introduit un facteur multiplicatif intrinsèque $\ln(1-\epsilon)n = O(\log n)$, ce qui résulte en un temps d'inondation de $O(\frac{\log n}{w_1})$, qui n'est pas meilleur que la majoration du théorème 3.5.3.

3.5.3 Application aux graphes évolutifs en loi de puissance

Dans cette section, nous appliquons la méthodologie décrite précédemment aux cas où les suites sont mutuellement indépendantes de graphes aléatoires avec une distribution de degrés en loi de puissance, c'est-à-dire les graphes dont le nombre y de nœuds de degrés x est distribué selon $\frac{1}{x^\beta}$ pour un $\beta > 0$.

Il y a plusieurs modèles visant à décrire des familles manipulables de graphes aléatoires en loi de puissance (voir, par exemple, [ACL01, CL03, Lu01]). Le modèle de [ACL01], appelé modèle à duplication (*duplicate model*), permet de montrer que pour $\beta < 1$, un graphe aléatoire tiré selon le modèle à duplication est a.p.s. connecté. Malheureusement, ce n'est pas suffisant pour nous permettre d'appliquer le scénario 1 décrit précédemment. En fait, de ce que nous en savons, le diamètre de ces graphes aléatoires (connectés) n'est pas connu. Nous avons donc considéré le modèle de graphe aléatoire en loi de puissance $\mathcal{G}_{\alpha,\beta}$ introduit et analysé dans [Lu01]. Ce modèle possède deux paramètres, α et β, qu'on tâchera de ne pas confondre avec ceux de $\mathcal{G}_{n,p}$. Lorsque la distribution des degrés est tracée avec une échelle logarithmique, α est l'ordonnée à l'origine (*Y-intercept*), tandis que $-\beta$ est la pente (*slope*). Plus précisément, un graphe de $\mathcal{G}_{\alpha,\beta}$ a une distribution de degré telle que le nombre

moyen de nœuds de degrés k, est égal à $\left\lfloor \frac{e^\alpha}{k^\beta} \right\rfloor$. Dans ce modèle, le degré maximum est $\Delta = \left\lfloor e^{\frac{\alpha}{\beta}} \right\rfloor$, et le nombre de nœuds est $n = \sum_{k=1}^{\Delta} \left\lfloor \frac{e^\alpha}{k^\beta} \right\rfloor$. Nous avons alors

$$n \simeq \begin{cases} \zeta(\beta)\, e^\alpha & \text{si } \beta > 1 \\ \alpha\, e^\alpha & \text{si } \beta = 1 \\ c\, e^{\frac{\alpha}{\beta}} & \text{si } 0 < \beta < 1 \end{cases}$$

où $\zeta(x) = \sum_{k=1}^{\infty} \frac{1}{k^x}$ est la fonction zêta de Riemann, et $\frac{\beta}{1-\beta} \leq c \leq \frac{1}{1-\beta}$. Pour tirer un graphe aléatoire dans $\mathcal{G}_{\alpha,\beta}$, on fixe une suite $\mathbf{w} = (w_1, \ldots, w_n)$ vérifiant que pour chaque $k = 1, \ldots, \Delta$,

$$|\{i : k \leq w_i \leq k+1\}| = \left\lfloor \frac{e^\alpha}{k^\beta} \right\rfloor,$$

et on applique l'égalité de l'équation 3.2 pour tirer les arêtes. Le lecteur est invité à se référer à [Lu01] pour plus de détails sur cette construction. Nous montrons le résultat suivant :

Théorème 3.5.5. *Soit $\mathcal{S} = (G_t)_{t \geq 0}$ une suite de graphes aléatoires mutuellement indépendants, $G_t \in \mathcal{G}_{\alpha,\beta}$. Presque sûrement, l'inondation dans $\mathcal{S}$ nécessite au plus $O(\log n)$ étapes.*

Démonstration. Quelque soit $\beta > 0$, un graphe dans $\mathcal{G}_{\alpha,\beta}$ a a.p.s. une composante géante en termes de volume [Lu01], c'est-à-dire une composante connexe C dont le nombre d'arêtes, $\mathrm{Vol}(C)$, est une fraction constante g du nombre total d'arêtes du graphe. Ainsi, nous pouvons appliquer le Scenario 2 décrit dans la section précédente. En particulier, nous nous intéressons à $H^{(1)} \in \mathcal{G}_{\alpha,\beta}$, et notons C sa composante géante.

Soit $\mathcal{E}_i$ l'événement "le nœud $i \notin C$ n'a pas d'arêtes vers C dans $H^{(h)}$".

Par les mêmes arguments que dans la preuve du théorème 3.5.2, en choisissant $h = 1 + \frac{2 \log n}{g}$, nous obtenons que $\mathbb{P}(\mathcal{E}_i) \leq \frac{1}{n^2}$. En appliquant l'inégalité de Boole, nous en déduisons que $\mathrm{connect}(C) = O(\log n)$, avec probabilité au moins $1 - \frac{1}{n}$.

À présent, comme nous savons par [Lu01] que le diamètre de la composante géante C est a.p.s. au plus $O(1)$ pour $0 < \beta < 2$, et au plus $O(\log n)$ pour $\beta \geq 2$, en appliquant le lemme 3.2.5, nous en déduisons que a.p.s., l'inondation dans $\mathcal{S}$ prend au plus $O(\log n)$ étapes. □

Notons que le diamètre de la composante géante C utilisée dans la preuve précédemment, est a.p.s. $\Theta(\log n)$ pour $\beta > 2$. Par conséquent, la connexion à la composante géante ne retarde pas l'inondation de manière significative par rapport au temps mis pour traverser la composante géante. Au contraire, pour $0 < \beta < 2$, le diamètre de la composante géante est $O(1)$. Par conséquent, le coût en étapes supplémentaires, pour atteindre les nœuds hors de la composante géante, retarde significativement l'inondation en comparaison du temps mis pour atteindre tous les nœuds de la composante géante.

3.6 Diffusion 1-Port aléatoire dans $\mathcal{M}_{n,p,q,g}$

Dans cette section, nous étudions le temps de diffusion du protocole 1-Port aléatoire Random dans un réseau dynamique modélisé par un graphe à évolution arête markovienne homogène $\mathcal{M}_{n,p,q,g}$. En fait, par souci de simplicité nous allons tout d'abord considérer le cas orienté. En effet, si l'indépendance entre les arêtes était suffisante pour avoir une indépendance spatiale pour le processus d'inondation ce n'est plus exactement le cas de la diffusion. Dans le cas de l'inondation, si deux nœuds u et v sont informés, nous pouvons alors complètement ignorer l'état de l'arête $\{u, v\}$, par contre dans le cas de la diffusion probabiliste, l'événement u appelle v et l'événement v appelle u sont corrélés. Ce n'est plus le cas si nous considérons des graphes orientés. Par la suite nous verrons comment se ramener au cas non orienté.

3.6.1 Graphes orientés

Nous considérons dans un premier temps le cas où $p + q = 1$, que nous utiliserons ensuite pour démontrer une borne sur le temps de diffusion dans le cas $p + q > 1$. Malheureusement, la technique utilisée ne s'adapte pas correctement au cas $p + q < 1$. Nous donnerons alors les raisons fondamentales.

Théorème 3.6.1. *Si $p + q \geq 1$, alors le temps de diffusion dans la version orienté de $\mathcal{M}_{n,p,q,g}$ est au plus $O\left(\frac{1}{1-(1-p)^{n-1}} \log n\right)$, avec forte probabilité.*

Démonstration. Lorsque $p + q = 1$, alors $p = \hat{p}$, et $(G_1, G_2, \ldots)$ est une séquence indépendante de graphes aléatoires orientés tirés dans $\mathcal{D}_{n,p}$ (défini comme $\mathcal{G}_{n,p}$). Sans perte de généralité, nous pouvons supposer que $G_0 \in \mathcal{D}_{n,p}$. En effet, même si aucun message n'est envoyé à la première étape (par exemple si la distribution initiale est celle qui force G_0 à être un stable) alors le processus de diffusion n'est décalé dans le temps que d'une seule étape.

Dans le cas de la séquence de graphe aléatoire orienté considéré ici, à chaque étape, chaque nœud informé u n'a aucun voisin avec probabilité $1 - \rho = (1 - p)^{n-1}$. Et avec probabilité $\rho = 1 - (1 - p)^{n-1}$ il a au moins un voisin, et u envoie le message à l'un des voisins choisi uniformément. Alors, par symétrie, à chaque étape chaque nœud $v \neq u$ reçoit le message de u avec la même probabilité $\frac{1-(1-p)^{n-1}}{n-1}$.

Le processus de diffusion 1-Port aléatoire dans une séquence indépendante de graphes aléatoires orientés est alors équivalente à la diffusion 1-Port aléatoire dans une clique (non orientée) avec une sélection uniforme et une probabilité d'erreur dans les transmissions de $1 - \rho = (1 - p)^{n-1}$. Le temps de diffusion dans ce cas est alors majoré par $O\left(\frac{1}{1-(1-p)^{n-1}} \log n\right)$ a.f.p..

Dans le cas $p > 1 - q$, une arête absente a une meilleure probabilité de naître qu'une arête présente de rester. Nous considérons donc deux étapes consécutives, pour utiliser ensuite les mêmes arguments que dans le cas précédent. Nous montrons que sur deux étapes consécutives, alors chaque nœud est sélectionné par u avec une

probabilité au moins $\frac{1-(1-p)^{n-1}}{n-1}$, ce qui nous donne la même majoration asymptotique que dans le cas $p+q=1$.

Nous considérons les deux étapes consécutives $2t$ et $2t+1$. Soit les événements suivants

- $\mathcal{V}_{pair}$: "v est sélectionné par u à l'étape $2t$"
- $\mathcal{V}_{impair}$: "v est sélectionné par u à l'étape $2t+1$"
- G : "$G_{2t} = G$"

Nous voulons montrer que, quel que soit le graphe à l'étape G_t, nous avons $\mathbb{P}(\mathcal{V}_{pair}) \geq \frac{1-(1-p)^{n-1}}{n-1}$ ou $\mathbb{P}(\mathcal{V}_{impair}) \geq \frac{1-(1-p)^{n-1}}{n-1}$.

***Affirmation* 3.6.2.** *Si u et v sont voisins dans G alors $\mathbb{P}(\mathcal{V}_{pair} \mid G) \geq \frac{1-(1-p)^{n-1}}{n-1}$.*

Preuve. Dans le cas où u et v sont voisins dans G, $\mathbb{P}(\mathcal{V}_{impair} \mid G) \geq \frac{1}{n-1} \geq \frac{1-(1-p)^{n-1}}{n-1}$.
$\diamond$

***Affirmation* 3.6.3.** *Si u et v ne sont pas voisins dans G alors $\mathbb{P}(\mathcal{V}_{impair} \mid \bar{G}) \geq \frac{1-(1-p)^{n-1}}{n-1}$*

Preuve. Par hypothèse, u et v ne sont pas voisins dans G. À G fixé, nous savons déjà que si $p = 1-q$, alors $\mathbb{P}(\mathcal{V}_{impair} \mid G) = \frac{1-(1-p)^{n-1}}{n-1}$. Nous montrons ici que $\mathbb{P}(\mathcal{V}_{impair} \mid G)$ est croissante en q (à p fixé). Notons N le nombre de voisins et M le nombre de non voisins de u au temps $2t$, $N+M=n-1$. Par hypothèse sur G nous avons $M>0$.

$$\mathbb{P}(\mathcal{V}_{impair} \mid G) = p \sum_{k=0}^{M-1} \sum_{l=0}^{N} \frac{1}{k+l+1} \binom{M-1}{k} p^k (1-p)^{M-1-k} \binom{B}{l} (1-q)^l q^{N-l}$$

Montrons que chaque terme $t_k(q) = \sum_{l=0}^{N} \frac{1}{k+l+1} \binom{B}{l} (1-q)^l q^{N-l}$ est croissant en q.

$$\begin{aligned}
&\frac{dt_k(q)}{dq} \\
&= \sum_{l=0}^{N-1} \frac{1}{k+l+1} \binom{B}{l} (1-q)^{l-1} q^{N-l-1} - \sum_{l=1}^{N} \frac{1}{k+l+1} \binom{B}{l} (1-q)^{l-1} q^{N-l} \\
&= \sum_{l=0}^{N-1} \frac{1}{k+l+1} \binom{B}{l} (1-q)^{l-1} q^{N-l-1} - \sum_{l=0}^{N-1} \frac{1}{k+l+2} \binom{B}{l+1} (1-q)^{l} q^{N-l-1} \\
&= \sum_{l=0}^{N-1} (1-q)^l q^{N-l-1} \left(\frac{(N-l)\binom{B}{l}}{k+l+1} - \frac{l\binom{B}{l+1}}{k+l+2} \right)
\end{aligned}$$

Remarquons que

$$\frac{(N-l)\binom{B}{l}}{k+l+1} - \frac{l\binom{B}{l+1}}{k+l+2} = \frac{(N-l)\binom{B}{l}}{k+l+1} - \frac{l(N-l)\binom{B}{l}}{(l+1)(k+l+2)}$$

a le même signe que

$$(l+1)(k+l+2) - l(k+l+1).$$

Or, $(l+1)(k+l+2) - l(k+l+1)$ est positif, donc $\frac{dt_k(q)}{dq}$ et $\frac{\partial \mathbb{P}(\mathcal{V}_{impair} \mid G)}{\partial q}$ sont aussi positifs. Nous en déduisons que

$$\mathbb{P}(\mathcal{V}_{impair} \mid G) = p \sum_{k=0}^{M-1} \sum_{l=0}^{N} \frac{1}{k+l+1} \binom{M-1}{k} p^k (1-p)^{M-1-k} \binom{B}{l} (1-q)^l q^{N-l}.$$

Une autre manière de le voir est de considérer deux évolutions arc-markoviennes couplées, l'une avec probabilité de mort $q = 1-p$ et l'autre avec probabilité de mort $q' > 1-q$, à partir du graphe G_{2t}, qui donnent respectivement les graphes G_{2t+1} et G'_{2t+1}. Ces deux évolutions sont couplées de la façon suivante :

- un arc absent de G_{2t} apparait à la fois dans G_{2t+1} et G'_{2t+1} avec probabilité p.
- un arc présente dans G_{2t} disparait à la fois dans G_{2t+1} et G'_{2t+1} avec probabilité q et disparaît uniquement de G'_{2t+1} avec probabilité $q' - q$.

Les arcs présents dans G_{2t+1} et absent de G'_{2t+1} sont appelés des arcs *fantômes*. Soit v un nœud non voisin de u dans G_{2t}, v reçoit l'information de u au temps $2t+1$ avec une plus forte probabilité dans G'_{2t+1} que dans G_{2t+1}. En effet, soit v n'est voisin de u ni dans G_{2t+1} ni dans G'_{2t+1}, soit il est voisin dans les deux. Dans ce derniers cas, u a moins de voisins, autres que v, dans G'_{2t+1}.

◇

Ainsi, puisque si $p+q \geq 1$ alors pendant deux étapes consécutives, chaque nœud non informé v reçoit le message d'un nœud informé u avec probabilité au moins $\frac{1-(1-p)^{n-1}}{n-1}$, le temps de diffusion dans la version orientée de $\mathcal{M}_{n,p,q}$ est dominé stochastiquement par le double du temps d'inondation dans la clique dont les transmissions ne réussissent qu'avec probabilité $\rho = 1-(1-p)^{n-1}$. Ce qui nous donne alors la majoration $O\left(\frac{1}{1-(1-p)^{n-1}} \log n\right)$. □

Remarque. On remarque aisément que la majoration est exacte lorsque $p = 1-q$. Et si $np = \Omega(1)$, alors $1-(1-p)^{n-1} = \Omega(1)$ et $\frac{1}{1-(1-p)^{n-1}} \log n = O(\log n)$. La borne est donc aussi exacte dans ce cas. De plus, si $np = o(1)$, alors $1-(1-p)^{n-1}$ et $\log(1+np)$ sont tous les deux $\Theta(np)$. La majoration est alors asymptotiquement la même que celle trouvée pour l'inondation pour une distribution du graphe initial quelconque.

Dans le cas $p+q < 1$ une analyse similaire donnerait $O\left(\frac{1-q}{p(1-q^{n-1})} \log n\right)$. En effet, le pire cas pour que v soit sélectionné par u à l'étape $t+1$ est lorsqu'au temps t, tous les nœuds sauf v sont voisins de u. Dans ce pire cas u sélectionne v avec probabilité $p \sum_{k=0}^{n-2} \frac{1}{k+1} \binom{n-1}{k} (1-q)^k q^{n-2-k} = \frac{p}{1-q} \frac{1-(q)^{n-1}}{n-1}$. Si ce calcul est adapté quand p est du même ordre de grandeur que $1-q$, il ne l'est pas pour p et q petits, en effet, nous perdons alors un facteur n. Pour s'en convaincre, nous pouvons considérer le cas extrême où $q = 0$, G_0 est le n-stable et $p = o(\frac{1}{n})$. Nous savons qu'il faudra $\Theta(\frac{\log n}{np})$

étapes pour que le graphe soit connecté et que pendant les $\Theta(\log n)$ étapes suivantes le nombre d'arêtes aura au plus doublé. Intuitivement la diffusion est alors celle dans un graphe aléatoire. Une autre manière de s'en convaincre est de prendre q petit et p encore plus petit $p = o(\frac{q}{n^2})$, de telle manière à ce qu'il y ait tellement peu d'arêtes dans les graphes que l'inondation et la diffusion reviennent au même.

Une raison intuitive pour le fait que la technique employée n'est pas efficace est qu'elle revient à considérer la probabilité d'utiliser une arête qu'elle soit présente ou non à un instant donné. Si le dynamisme est important, cette hypothèse n'est pas très restrictive. Par contre, si le graphe évolue peu, il faut alors tenir compte des arêtes qui survivent et qui n'ont pas encore été utilisées, si on souhaite avoir une bonne majoration.

3.6.2 Graphes non orientés

Nous allons montrer que les bornes asymptotiques restent vraies dans le cas non orienté malgré la dépendance entre les envois.

Théorème 3.6.4. *Si $p+q \geq 1$, alors le temps de diffusion dans $\mathcal{M}_{n,p,q,g}$ est majorée par $O\left(\frac{1}{1-(1-p)^{n-1}} \log n\right)$ avec forte probabilité.*

Dans le cas non orienté, la probabilité qu'un nœud u appelle un nœud non informé v se calcule de la même manière que dans le cas orienté. Par contre, les probabilités ne sont plus indépendantes, par exemple, soient deux nœuds informés u et u', l'événement "u appelle u'" n'est pas indépendant de l'événement u' appelle u car ils sont tous les deux corrélés à l'événement "l'arête $\{u, v\}$" est présente. Nous pouvons, par contre, montrer qu'on ne perd, au plus, qu'un facteur 2 dans la diffusion.

Démonstration. Comme dans la preuve du cas orienté, nous commençons par le cas $p + q = 1$.

***Affirmation* 3.6.5.** *Pour tout $i \in [n]$, A_i, B_i, C_i sont des événements aléatoires formant une partition et dépendants des autres événements A_j, B_j, C_j pour $j \in [n]$. Soit a_i, b_i, c_i les indicatrices correspondantes. Si pour toute filtration F_i définie uniquement à partir des A_j, B_j, C_j pour $j < i$, nous avons*

$$\mathbb{P}(B_i \mid (B_i \cup C_i) \cap F_i) \geq r \tag{3.3}$$

et si pour i de 1 à n, $(B_i \vee C_i)$ survient au total au moins k fois, c'est-à-dire

$$\sum_i (b_i + c_i) \geq k. \tag{3.4}$$

Alors $\sum_i b_i$ domine stochastiquement $\mathrm{Bin}(k, r)$.

Preuve. Soit i_j le premier indice (aléatoire) pour lequel $\sum_{i \leq i_j} b_i + c_i = j$. Étudions $S_j = \sum_{i \leq i_j} b_i$ et montrons par récurrence que S_j domine stochastiquement $\mathrm{Bin}(j, r)$.

C'est évident pour $j = 0$, supposons-le alors vrai en $j-1$. Nous avons, par conséquence de l'hypothèse, que pour tout $s_1, s_2, \ldots, s_{j-1} \in \{0, 1\}$

$$\mathbb{P}(S_j - S_{j-1} = 1 \mid S_1 - S_1 = s_1, S_2 - S_1 = s_2, \ldots, S_{j-1} - S_{j-2} = s_{j-1}) \geq \rho \quad (3.5)$$

En effet, la condition est une combinaison des événements sur les A_i, B_i, C_i pour $i < i_j$ et comme nous avons aussi $B_{i_j} \cup C_{i_j}$, l'inégalité est vraie par hypothèse.

Et la condition 3.3 permet de montrer par récurrence que S_j domine stochastiquement $\mathrm{Bin}(j, r)$. Et nous concluons grâce au fait que, par hypothèse, S_k minore la somme de b_i. ◇

Notons $I = \{u_1, \ldots, u_{|I|}\}$ l'ensemble des nœuds informés à l'étape t. Nous appliquons l'affirmation précédente avec $n = |I|$ le nombre de nœuds informés, $k = \frac{n}{2}$, $r = \frac{\rho}{n-1} = \frac{1-(1-p)^{n-1}}{n-1}$ et

- A_i est l'événement "le nœud u_i appelle un nœud $j < i$ qui a appelé u_i au temps t",
- B_i est l'événement "le nœud u_i appelle le nœud v",
- C_i est le complémentaire de $A_i \cup B_i$.

Nous avons évidemment qu'au plus $\frac{|I|}{2}$ événements A_i se produisent. En effet, chaque événement A_i implique un événement C_j unique, car u_i et u_j se sont appelés mutuellement.

Il reste alors à montrer que l'hypothèse 3.5 est vérifiée. Pour le voir, partitionnons les événements A_j, B_j, C_j en événements "u_j informe w_j", noté $u_j \to w_j$. L'événement B_i est donc l'événement $u_i \to v$. Soit i fixé, nous définissons aussi l'événement $u_j \mapsto w_j$:

- si $w_j = u_i$, alors $u_j \mapsto u_i$ signifie $u_j \to u_i \cap \neg u_i \to u_j$,
- si $w_j \neq u_i$, alors $u_j \mapsto w_j$ signifie simplement $u_j \to w_j$.

Ainsi l'événement "$(B_i \cup C_i) \cap \forall j < i, u_j \to w_j$" est exactement l'événement "$\forall j < i, u_j \mapsto w_j$".

Le calcul étant le même que dans le cas orienté, nous savons que $\mathbb{P}(u_i \to v) \geq \rho$. Nous montrons donc que $\mathbb{P}(u_i \to v \mid \forall j < i, u_j \mapsto w_j) \geq \rho$ par récurrence, en montrant que $\mathbb{P}(u_i \to v \mid \forall j < k, u_j \mapsto w_j) \geq \mathbb{P}(u_i \to v \mid \forall j < k-1, u_j \mapsto w_j)$.

Soit $\widetilde{G}$ un événement qui impose la présence ou l'absence de toutes les arêtes différentes de $\{u_k, u_w\}$ et qui soit compatible avec $\forall j < k-1, u_j \to w_j$. Nous avons alors, si $w_k \neq u_i$,

$$\begin{aligned}
&\mathbb{P}(u_i \to v \mid \widetilde{G} \cap \forall j < k-1, u_j \mapsto w_j \cap u_k \mapsto w_k) \\
&= \mathbb{P}(u_i \to v \mid \widetilde{G} \cap \forall j < k-1, u_j \mapsto w_j \cap u_k \to w_k) \\
&= \mathbb{P}(u_i \to v \mid \widetilde{G} \cap \forall j < k-1, u_j \mapsto w_j \cap \{u_k, w_k\} \in G_{2t}) \\
&\leq \mathbb{P}(u_i \to v \mid \widetilde{G} \cap \forall j < k-1, u_k \mapsto w_k)
\end{aligned}$$

Si $w_k = \perp$, c'est-à-dire que u_k rate son envoi car il n'a pas de voisin, alors l'événement $\widetilde{G}$ doit, en particulier, contenir que toutes les arêtes incidentes à u_k sont absentes dans G_{2t}.

Si $w_k = u_i$,

$$\begin{aligned}
&\mathbb{P}(u_i \to v \mid \widetilde{G} \cap \forall j < k-1, u_j \mapsto w_j \cap u_k \mapsto w_k) \\
&= \mathbb{P}(u_i \to v \mid \widetilde{G} \cap \forall j < k-1, u_j \mapsto w_j \cap u_k \to u_k \cap \neg u_k \to u_k) \\
&\leq \mathbb{P}(u_i \to v \mid \widetilde{G} \cap \forall j < k-1, u_j \mapsto w_j)
\end{aligned}$$

C'est vrai pour tous les événements $\widetilde{G}$ compatibles avec l'événement $\forall j < k-1, u_j \to w_j$ c'est donc vrai pour la somme. Nous avons bien

$$\mathbb{P}(u_i \to v \mid \forall j < k, u_j \mapsto w_j) \geq \mathbb{P}(u_i \to v \mid \forall j < k-1, u_j \mapsto w_j)$$

et nous pouvons donc appliquer l'affirmation précédente. Nous avons montré que pour tout nœud non informé v, au moins la moitié des nœuds informés informe v avec probabilité ρ quels que soient les appels des nœuds informés d'indice inférieur. Pour conclure, il est important de remarquer que ces nœuds sont les mêmes quelque soient le nœud v non informé considéré. En effet A_i ne dépend pas de v, et donc $B_i \cup C_i = \neg A_i$ non plus. Ainsi, lorsque $p + q = 1$, le nombre de nouveaux nœuds informés en deux étapes domine stochastiquement le nombre de nœuds informés en une étape de la diffusion aléatoire, ayant une probabilité d'erreur $1 - \rho$ lors des communications, dans une clique.

Le cas $p + q > 1$, se fait comme dans la preuve du cas orienté en considérant deux étapes consécutives et en considérant les probabilités d'envoi selon les graphes aux temps pairs. Nous nous ramenons donc à la diffusion aléatoire avec erreur, dans une clique, en considérant au total quatre étapes consécutives.

□

3.7 Conclusion

3.7.1 Inondation

Par les résultats de ce chapitre et ceux de [CMM+10, CMPS11], l'inondation dans les graphes à évolution arête-markovienne homogène (où toutes les arêtes ont les mêmes p et q) est maintenant bien comprise en particulier pour les principaux cas, à savoir : lorsque le système est à l'état stationnaire (c'est-à-dire $G_0 \in \mathcal{G}_{n,\widehat{p}}$) ou lorsque le système démarre à partir de zéro (c'est-à-dire G_0 est le graphe vide). Par contre, déterminer les performances asymptotiques exactes de l'inondation pour une distribution initiale arbitraire, et pour des taux de naissance et de mort arbitraires semblent être un défi. Au-delà de ce problème et du cas orienté, nous identifions deux axes principaux pour une généralisation de nos résultats.

Le cas orienté. Bien que nous n'en n'ayons pas eu l'utilité, notre méthodologie se transpose directement dans le cas où les graphes sont orientés. Le graphe pondéré représentatif est alors défini de manière similaire. Chose surprenante, on peut remarquer que si, pour tout couple i, j, $p_{ij} = p_{ji}$ et $q_{ij} = q_{ji}$, alors on peut se contenter

d'un graphe représentatif non orienté. En effet, au lieu de considérer les variables aléatoires Z_{ij} et Z_{ji} donnant le poids des arcs respectivement ij et ji, on considère les deux variables aléatoires ayant la même distribution $Z_{\{i,j\}}$ et $Z'_{\{i,j\}}$ où, d'une part, le poids $Z_{\{i,j\}}$ est attribué à l'arc ij si i est informé avant j, à ji sinon, et, d'autre part, le poids $Z'_{\{i,j\}}$ est attribué à l'autre arc. Et on remarque que les poids Z' n'ont aucune utilité pour l'inondation puisque, par définition, quand l'arc correspondant devrait être utilisé les deux extrémités sont déjà informées.

Dépendance spatiale. Comme mentionné plus tôt, le lemme de réduction s'applique dans le contexte général des suites de graphes à évolution arête-markovienne, où les arêtes peuvent avoir des lois d'évolution markovienne différentes, mais où elles sont impérativement mutuellement indépendantes. Le fait que les arêtes soient mutuellement indépendantes est un élément clé pour l'application de notre méthodologie, il y a donc peu d'espoir d'avoir une technique aussi directe dans le cas où on souhaite ajouter de la dépendance spatiale.

Par contre l'hétérogénéité nous laisse quand même une certaine liberté pour l'application d'une telle méthodologie. Ceci soulève le problème de la construction d'un processus de Markov "naturel" décrivant l'apparition et la disparition des arêtes donnant une suite de graphes dont l'état stationnaire est $\mathcal{G}_{\mathbf{w}}$, pour toute suite $\mathbf{w}$ donnée. Bien évidemment, paramétrer les probabilités $p_{i,j}$ définies telles que dans l'équation 3.2, et paramétrer alors $q_{i,j} = 1 - p_{i,j}$ respecte ce critère, mais ignore toute dépendance temporelle puisque le graphe G_{t+1} est en fait indépendant de G_t. En "pratique", les réseaux dynamiques ont une dépendance temporelle. Ainsi la construction d'un modèle manipulable et réaliste, pour une suite markovienne de graphes pour, par exemple, la distribution en loi de puissance, est importante. Il n'est pas clair que la construction d'un tel modèle soit possible sans introduire aussi une dépendance spatiale (c'est-à-dire, une dépendance entre les arêtes). Le modèle arête-markovien se restreint à une dynamique sans dépendance spatiale. Savoir si cette restriction empêche la construction d'un modèle de graphe dynamique avec une distribution de degré donnée est une question intéressante.

Dépendance temporelle. Un autre axe de recherche pouvant se révéler fructueux serait de considérer le renforcement de la dépendance avec le temps, par exemple en considérant le cas où la probabilité pour chaque arête de mourir est une fonction du nombre d'étapes consécutives durant lesquelles cette arête a survécu. Ceci refléterait en quelque sorte la dépendance spatiale réaliste capturant ce qui se produit sur les réseaux radio ou dans les systèmes P2P : une paire de nœuds qui ont été connectés pendant une longue période de temps a plus de chance de rester connectée à l'étape suivante que deux nœuds qui viennent juste d'être connectés. Bien que la définition des graphes représentatifs du nouveau processus markovien soit alors plus compliquée, nous ne voyons pas de contre-indication majeure à la transposition du lemme de réduction à ce cadre.

3.7.2 Diffusion 1-Port

Les résultats que nous avons dans le cadre de la diffusion 1-Port aléatoire sur les graphes à évolution arête-markovienne ne sont pas aussi complets que ceux que nous avons pour l'inondation. Ils offrent toutefois une information importante dans le cas homogène. Nous savons que si $p + q \geq 1$, alors le temps de diffusion est $O\left(\frac{1}{1-(1-p)^{n-1}} \log n\right)$, avec forte probabilité. Nous apprenons ainsi que, si $p + q \geq 1$, le dynamisme a un impact relativement limité à condition que p ne soit pas trop petit. Par contre si p devient très petit, alors le degré moyen de chaque nœud à chaque étape est alors faible, dans ce cas, le facteur $\frac{1}{1-(1-p)^{n-1}}$ devient significatif et le temps de diffusion 1-Port et le temps d'inondation sont asymptotiquement les mêmes. Il est alors naturel de se demander quel est le temps de diffusion 1-Port dans le cas $p + q < 1$.

Ensuite, il peut être intéressant de considérer la diffusion dans les graphes à évolution arête-markovienne hétérogène et plus généralement sur tous les modèles de graphes dynamiques que nous avons mentionnés précédemment.

Chapitre 4

Diffusion spatiale

Ce chapitre est dédié à l'étude de la diffusion spatiale, il améliore les résultats de l'article *Sub-linear Universal Spatial Gossip Protocols* avec Pierre Fraigniaud, et présenté à *16th international colloquium on Structural Information and Communication Complexity (SIROCCO 2009)* [BF10].

Dans la diffusion spatiale, plutôt que de s'intéresser au temps mis par une information pour se propager à tous les nœuds selon un protocole donné, nous nous intéressons au temps mis par le protocole pour informer chaque nœud en fonction de son éloignement par rapport à la source. Nous montrons que, dans le cas des métriques de graphes, le protocole où chaque nœud appelle uniquement ses voisins dans le graphe est bien un protocole de diffusion spatiale. Nous donnons son temps de propagation. Nous établissons ensuite un lien entre la diffusion spatiale et les graphes augmentés. En effet, nous donnons est une méthode permettant de transposer certains résultats concernant le routage dans les graphes augmentés en des résultats sur le temps de propagation d'un protocole de diffusion spatiale. Enfin nous donnons un protocole de diffusion spatiale valable sur toutes les métriques de graphe. Nous calculons son temps de propagation dans une métrique de graphe quelconque. Enfin, nous décrivons l'adaptation à effectuer pour que le protocole soir applicable sur n'importe quelle métrique, et nous calculons ses performances dans les métriques à densité uniforme.

4.1 Introduction

Dans cette section nous introduisons formellement le problème de la diffusion spatiale. Nous listons les principaux résultats de la littérature. Enfin, nous résumons ensuite de nos principaux résultats.

4.1.1 Définition et état de l'art

Nous supposons que les nœuds du réseau sont placés dans un espace métrique. Dans le modèle de communication spatiale que nous considérons dans ce chapitre,

chaque nœud ne peut envoyer une information qu'à un seul autre nœud à chaque étape, comme dans le modèle 1-PORT. Nous considérons toutefois que la communication entre un nœud informé et le nœud sélectionné est réalisée via un protocole de communication point-à-point sous-jacente, qui permet à toute paire de nœuds distincts de communiquer dans le réseau. C'est-à-dire qu'ici, la métrique ne définit plus une notion de voisinage, puisque, au contraire, n'importe quels nœuds peuvent communiquer entre eux. En contrepartie nous imposons de pouvoir définir la vitesse de propagation en fonction de l'éloignement à la source.

Dans le protocole UNIFORM, à chaque étape, un nœud informé sélectionne le nœud qu'il appelle uniformément parmi l'ensemble des nœuds. Le protocole UNIFORM est équivalent au protocole de diffusion RANDOM dans la clique, et nous savons qu'une nouvelle information se diffuse, a.f.p., à tous les nœuds de V en $O(\log n)$ étapes, où $n = |V|$. Ce temps est asymptotiquement optimal car le nombre de nœuds au courant d'une information donnée peut tout au plus doubler à chaque étape, et donc il faut au moins $\Omega(\log n)$ étapes pour que tous les nœuds reçoivent une nouvelle information. Cependant, il a été remarqué, dans [KKD04], que UNIFORM n'est pas approprié dans les cas où les nœuds proches de la source d'une nouvelle information devraient préférentiellement recevoir cette information plus rapidement que les nœuds lointains. Une telle exigence se révèle importante dans le contexte de la localisation de ressources [PRR99] dans lequel les utilisateurs souhaitent trouver les plus proches copies des ressources partagées (par exemple, des films). Elle est également importante dans le cadre d'une propagation d'alarme [KKD04].

Dans [KKD04], les auteurs ont abordé la question de la conception de protocoles de diffusion satisfaisants une contrainte supplémentaire : plus un nœud est proche de la source, plus il doit recevoir l'information rapidement. Afin de mesurer la sensibilité à la distance d'un protocole de diffusion, ils ont examiné son ***temps de propagation***, c'est-à-dire le temps mis par un nœud pour être informé, en fonction de la distance à la source, dans un espace métrique (V, δ). Dans un tel espace métrique, V est l'ensemble (fini ou infini, mais dénombrable) de points, ou nœuds, et δ est la fonction de distance entre les nœuds. Kempe et al. ont conçu un protocole de diffusion spatiale, appelé ici DENSITY, vérifiant que si les nœuds dans V sont répartis avec une densité uniforme $\mathbb{R}^D$ avec la métrique $\mathcal{L}_k$, alors une nouvelle information se propage aux nœuds à distance d de n'importe quelle source s en $O(\log^{1+\epsilon} d)$ étapes, avec une probabilité au moins $1 - O(\frac{1}{\log d})$. Par densité uniforme, nous entendons qu'il existe deux constantes positives β_1 et β_2 telles que, pour tout $r \geq 1$, le nombre de nœuds dans toute boule de rayon R est au moins $\beta_1 r^D$, et au plus $\beta_2 r^D$.

Il est aussi montré, dans [KKD04], que pour $r \leq 1$, le temps de diffusion spatiale est minoré par une fonction de n. Et pour $r > 2$, par un polynôme en d. Le cas $r = 2$ est ouvert. Remarquons que, pour la diffusion spatiale il faut que $r \in]1, 2[$ pour avoir le résultat souhaité, mais dans le cas du routage glouton il faut $r = 1$. En fait, il faudra garder en tête que, bien qu'il soit naturel de vouloir rapprocher ces deux résultats, il faut plutôt considérer le diamètre du graphe augmenté, et non son diamètre glouton. Rappelons que de [MN04, NM05], nous savons que le diamètre est

polylogarithmique en n dès que $r < 2$, polynomial si $r > 2$ et n'est pas connu pour $r = 2$. Pour le cas $r \leq 1$, le problème principal vient du fait que dans la distribution φ_u d'envoi, le facteur de normalisation est fonction de n.

Les protocoles dont les performances dépendent de la distance entre les sources et les destinataires, sont appelés protocoles de diffusion *spatiale*. Afin de comparer les temps de propagation de différents protocoles de diffusion spatiale, dans différents espaces métriques (V, δ), nous devons prendre en compte le fait que le nombre de nœuds à une distance donnée à partir d'un nœud donné varie considérablement d'un espace métrique à un autre, et même au sein d'un même espace métrique. Dans une métrique (V, δ), il est alors plus commode de définir le temps de propagation en fonction des rangs des nœuds, où le ***rang*** du nœud u relativement à un autre nœud s est le nombre des nœuds dont les distances à partir de s sont plus petites que $\delta(s, u)$. En effet, un protocole de diffusion ne peut pas assurer qu'une information atteint un nœud proche de la source rapidement s'il y a un grand nombre d'autres nœuds qui sont encore plus proches de cette source.

Redéfinissons alors le temps de propagation en conséquence La boule de rayon d centrée en s est définie comme

$$B(s, d) = \{u \in V, \delta(s, u) \leq d\}.$$

Pour tout nœud s et toute boule B centrée en s et de taille $b \geq 1$, soit $T_s(B)$ la variable aléatoire égale au nombre d'étapes nécessaires pour qu'une nouvelle information provenant du nœud s atteigne tous les nœuds de la boule B. Nous considérons que le protocole de diffusion spatiale a un temps de propagation $f(B)$ pour une fonction f donnée si pour tout $s \in V$, $T_s(B) \leq O(f(B))$ avec forte probabilité. Ici, "avec forte probabilité" signifie avec probabilité au moins $1 - O(\frac{1}{b^\alpha})$ avec une constante $\alpha > 0$. Les paramètres de B pouvant par exemple être utilisés sont sa taille b, son rayon, et le degré maximal Δ_B des nœuds intérieurs de B. (Si la boule B considérée est évidente dans le contexte, nous utiliserons Δ sans indice B.) Idéalement, f ne serait une fonction que de b.

Par définition de la densité uniforme, les boules dans la sous-métrique $(V, \mathcal{L}_k)$ de $(\mathbb{R}^D, \mathcal{L}_k)$ induites par un ensemble V de nœuds placées avec une densité uniforme dans $\mathbb{R}^D$ ont une taille polynomiale en leur rayon. À partir de ce fait, nous pouvons montrer que le protocole de diffusion spatiale de [KKD04] a un temps de propagation de $O(\log^{2+\epsilon} b)$. Ce résultat inspire la question de l'existence de protocoles de diffusion efficaces (c'est-à-dire des protocoles avec un temps de propagation borné) dans métriques arbitraires, ou du moins dans métriques de graphes arbitraires.

Remarque. Formellement, dans [KKD04] il est montré que, si $1 < r < 2$ alors si un nœud u appelle v à distance $d = \delta(u, v)$ avec probabilité $c_u(d+1)^{rD}$ où c_u est un facteur de normalisation constant. Alors avec probabilité $1 - O(\log^{-\alpha} d)$, chaque nœud à distance d est informé en $O(\log^{\frac{1}{1-\log r}} d. \log\log d)$ étapes. À cause de la densité uniforme $\log d \in \Theta(\log b)$, et le terme $\log b$ supplémentaire permet d'obtenir la forte probabilité.

Protocole	Métrique	Temps de propagation	Temps de diffusion
UNIFORM	finie arbitraire		$O(\log n)$ [FG85, Pit87]
DENSITY	à densité uniforme	$O(\log^{2+\epsilon} b)$ [KKD04]	
RANDOM	de graphe arbitraire	$O(b \log b)$	$O\left(\Delta(D + \log n)\right)$ [FPRU90]
RANKSCALE	de graphe arbitraire	$O\left((b^{\frac{1}{3}} \log^{\frac{1}{2}} b + \Delta) \log b\right)$ et $o\left(\Delta 2^{(\log b)^{\frac{1}{2}+\epsilon}}\right)$	
	à densité uniforme	$O\left(\log^2 b \log\log b\right)$	

TABLE 4.1 – Temps de diffusion de différents protocoles. n désigne le nombre de nœuds, b la taille de la boule, Δ le degré maximum et D le diamètre.

Rappelons qu'une métrique de graphe (V, δ) est déterminée par un graphe non orienté $G = (V, E)$, où la distance $\delta(u, v)$ entre deux sommets u et v est la longueur d'un plus court chemin entre u et v dans G.

Dans les métriques de graphes, un candidat naturel pour un protocole de diffusion spatiale est celui qui n'utilise que les arêtes du graphe et où chaque nœud sélectionne le destinataire de sa prochaine communication aléatoirement uniformément parmi ses voisins dans le graphe. Ce protocole est alors équivalent au protocole RANDOM dans le graphe définissant la métrique, il termine en $O(\Delta(D + \log n))$ étapes, a.f.p., où Δ désigne le degré maximal des nœuds et D désigne le diamètre du graphe. De plus, il n'est pas difficile d'adapter le résultat de [FPRU90] afin de montrer que le temps de propagation du protocole RANDOM est $O(b \log b)$ (voir Section 4.2), ce qui prouve l'existence d'un protocole de diffusion spatiale pour toute métrique de graphes. Cette borne est exacte. En effet, rappelons que dans l'étoile à n nœuds (un arbre de n nœuds avec $n - 1$ feuilles et un seul nœud interne appelé centre), une information provenant du centre de l'étoile aura atteint tous les nœuds à distance 1 en temps $\Omega(n \log n)$, par l'équivalence avec le problème du collectionneur de coupon.

L'objectif principal de ce chapitre est de concevoir un protocole de diffusion spatiale dont le temps de propagation est sous-linéaire en la taille des boules.

4.1.2 Résumé des résultats présentés dans ce chapitre

Dans un premier temps nous montrons que le protocole classique de diffusion 1-PORT aléatoire RANDOM est bien un protocole de diffusion spatiale, en refaisant la preuve de [FPRU90].

Nous établissons ensuite un lien étroit entre la diffusion spatiale et les graphes augmentés. En particulier, grâce à ce lien, il est possible de transposer des résultats connus sur les distances dans un graphe augmenté en un temps de propagation.

Nous décrivons enfin un protocole de diffusion spatiale, appelé RANKSCALE qui est une simplification du protocole LOGSCALE décrit dans [BF10]. Nous prouvons que, dans les métriques de graphe de degré maximum Δ, son temps de diffusion est au plus $O\left((b^{\frac{1}{3}} \log^{\frac{1}{2}} b + \Delta) \log b\right)$ étapes, mais aussi au plus $o\left(\Delta 2^{(\log b)^{\frac{1}{2}+\epsilon}}\right)$ étapes. Les

performances de ce protocole par rapport aux protocoles mentionnés précédemment sont résumées dans la table 4.1. RANKSCALE a un temps de propagation beaucoup plus petit que RANDOM. Dans les métriques de graphe fini, il a, au plus, à une constante près, le même temps de diffusion que RANDOM, à cause du fait que, en moyenne, RANKSCALE agit comme RANDOM pendant la moitié des étapes. Pendant l'autre moitié, chaque nœud sélectionne le destinataire de la transmission avec une probabilité que les échelles avec le logarithme des rangs. En fait, en combinant RANKSCALE avec UNIFORM (chaque nœud agit comme dans un protocole avec la moitié de la probabilité, et comme dans l'autre protocole avec la moitié de probabilité), nous obtenons un protocole de diffusion spatiale avec, à une constante près, le temps de propagation identique à RANKSCALE mais avec le même temps de diffusion que UNIFORM. Bien que conçu pour les métriques de graphes, notre protocole RANKSCALE peut également être appliqué aux systèmes métriques arbitraires. Dans les espaces métriques à densité uniforme (c'est-à-dire, le même cadre que dans [KK02]), RANKSCALE atteint un temps de propagation polylogarithmique de $O(\log^2 b \log\log b)$ étapes.

4.2 Protocole RANDOM revisité

Dans cette section, nous revisitons brièvement le protocole RANDOM analysé dans [FPRU90], nous prouvons que son temps de propagation est $O(b \log b)$ et nous donnons par la même occasion un résultat préliminaire utile pour une démonstration de notre contribution.

Proposition 4.2.1. *Pour toute métrique de graphe (V, δ), et pour tout nœud source $s \in V$, le protocole* RANDOM *vérifie qu'un message introduit au nœud s atteint tous les nœuds d'une boule de taille b centrée en s en moins de $O(b \log b)$ étapes, avec forte probabilité. Plus généralement, le message atteint tout ensemble C ou toute union d'ensembles de taille totale b vérifiant (chacun) la propriété suivant : il existe une constante k telle que pour tout $t \in C$, il existe un chemin de s à t noté $s_0 = s, s_1, \ldots, s_\ell = t$, tel que*

$$\sum_{i=0}^{\ell-1} \deg(u_i) \leq kb \tag{4.1}$$

en au plus $O(b \log b)$ étapes, avec forte probabilité.

Démonstration. Nous utilisons ici la même structure de preuve que dans [FPRU90], dans un premier temps nous montrons qu'une boule vérifie la propriété 4.1 avec $k = 3$, puis démontrons le cas général, ce qui prouve qu'avec le protocole RANDOM, chaque nœud de B reçoit l'information de source s en au plus $O(b \log b)$ étapes, avec forte probabilité.

Soit B une boule centrée en s, contenant $b = |B|$ nœuds. Nous montrons que B vérifie la propriété 4.1.

Pour la première partie, nous utilisons la même observation que dans [FPRU90]. Elle établit que pour chaque nœud en dehors du plus court chemin dans le graphe ne peut être adjacent qu'à au plus 3 nœuds du chemin.

Soit $u \in B$ et $P = (u_0, u_1, \ldots, u_\ell)$ un plus court chemin de s vers u, avec $u_0 = s$ et $u_\ell = u$. Chaque nœud dans $B \setminus P$ ne peut être adjacent qu'à au plus 3 nœuds de P. Un nœud hors de B adjacent à P ne peut être adjacent qu'à u_ℓ puisque sinon il serait dans B. Ainsi

$$\sum_{i=0}^{\ell-1} \deg(u_i) \leq 3b.$$

Maintenant, soit C un ensemble de taille b, montrons que si pour une constante k, et pour tout $t \in C$, il existe un chemin de s à t noté $s_0 = s, s_1, \ldots, s_\ell = t$, tel que $\sum_{i=0}^{\ell-1} \deg(u_i) \leq kb$ alors le message atteint tout ensemble C en $O(b \log b)$ étapes, avec forte probabilité. Soit X_u la variable aléatoire égale au nombre d'étapes nécessaires pour que l'information atteigne u, alors nous avons $\mathbb{E}[X_u] \leq kb$.

Soit $\alpha > 1$. Pour $i = 1, \ldots, \alpha \log b$, soit Y_i des variables aléatoires indépendantes et identiquement distribuées comme X_u. Puisque le choix fait par chaque nœud dans le protocole Random ne tient pas compte du passé (protocole sans mémoire) et indépendant de la source du message et du choix fait par les autres nœuds, nous avons

$$\mathbb{P}(X_u \geq 2\ \alpha\ \log(b)\ \mathbb{E}[X_u]) \leq \Pi_{i=1}^{\log b} \mathbb{P}(Y_i \geq 2\mathbb{E}[X_u]).$$

Par l'inégalité de Markov, nous obtenons,

$$\mathbb{P}(X_u \geq 2\ \alpha\ \log(b)\ \mathbb{E}[X_u]) \leq \left(\frac{1}{2}\right)^{\alpha \log b} = \frac{1}{b^\alpha}.$$

Ainsi $\mathbb{P}[X_u \geq 2k\ \alpha\ b\ \log b] \leq \frac{1}{b^\alpha}$. Par l'inégalité de Boole, nous obtenons qu

$$\mathbb{P}(\exists u \in B, X_u \geq 2k\ \alpha\ b\ \log b) \leq \frac{1}{b^{\alpha-1}},$$

ce qui complète la preuve. □

4.3 Diffusion spatiale et routage glouton dans les graphes augmentés

Dans cette section, après une présentation de problème du routage, nous traçons un parallèle entre le routage et la diffusion. Et en particulier nous donnons un théorème général permettant, sous certaines contraintes, de transposer un résultat sur le routage dans les graphes augmentés en un résultat sur la diffusion spatiale.

4.3.1 Le routage

Dans le routage point-à-point, le but est d'envoyer (rapidement) une information d'une source à un destinataire en passant par les nœuds du réseau où chaque nœud transmet le message à un seul de ses voisins. Le routage est un cas très particulier de la (multi)diffusion où l'ensemble des cibles est de taille 1. Dans le cas des réseaux statiques, si les nœuds ont suffisamment d'information sur la structure du réseau,

il est alors possible de définir un protocole permettant au message d'atteindre le destinataire en un nombre d'étape optimal. Ce nombre optimal est la distance entre la source et le destinataire dans le réseau. En effet, il est possible d'avoir les mêmes performances en termes de temps que l'inondation avec un protocole où chaque nœud qui reçoit l'information ne la retransmet qu'une seule fois et à un seul de ses voisins, à condition que chaque nœud ait suffisamment d'information pour transférer l'information à un de ses voisins les plus proches de la cible. Dans un protocole optimal, à chaque envoi de l'information, la distance du dernier nœud informé à la cible est réduite de 1. Par exemple, c'est le cas si chaque nœud a la possibilité de calculer un plus court chemin comme le routage à état de lien *Open Shortest Path First* (OSPF) et *Intermediate System to Intermediate System* (IS-IS). C'est aussi possible si chaque nœud connaît la distance entre ses voisins et la cible. On parle alors de routage à vecteur de distances, utilisé dans les protocoles *Routing Information Protocol* (RIP, protocole d'information de routage) et *Interior Gateway Routing Protocol* (IGRP) et *Enhanced Interior Gateway Routing Protocol* (EIGRP).

Remarque. Si les nœuds n'ont aucune information sur la structure du réseau et qu'un nœud qui reçoit l'information la retransmet (même s'il l'avait déjà reçu) une seule fois à un seul de ses voisins choisi aléatoirement. Le calcul du nombre d'étapes nécessaires pour informer la cible se ramène à l'étude d'une marche aléatoire. Et nous savons [Fei95b] que le temps de couverture d'une marche aléatoire sur un graphe est plus $O(n^3)$ étapes et que cette borne est atteinte.

Le routage étant un processus très important il a fait l'objet de nombreuses études sous diverses contraintes et différentes variantes. Celle qui nous intéresse particulièrement est le cas du routage dans les graphes augmentés, où les nœuds ont une information qui est limitée au graphe de base (avant augmentation) et à leur voisinage.

4.3.2 Les graphes augmentés

La notion de graphe augmenté a été introduite par [WS98], puis développée par [Kle00], dans le but de modéliser le résultat empirique de l'expérience de Milgram [Mil67] portant sur les graphes "petit-mondes" [1] dont le résultat est connu sous le terme de "paradoxe de Milgram" ou de "six degrés de séparation" car elle démontre expérimentalement cette idée de Frigyes Karinthy popularisée par la pièce de théâtre *Six Degrees of Separation* de John Guare.

Définition. *Un graphe augmenté est un graphe G auquel sont ajoutés des liens longs, qui sont des arcs entre des pairs de sommets servant à accélérer les communications. S'il y a un lien long de u vers v, v est dit contact ou voisin lointain de u. Lorsqu'on ajoute un seul lien long par nœud $u \in G$ qui est l'extrémité sortante de l'arc, et que l'autre extrémité est choisie selon une loi de probabilité φ_u le graphe augmenté sera noté (G, φ).*

1. à ne pas confondre avec l'expérience de soumissions à l'autorité faite pas ce même psychologue.

Remarque. La métrique, c'est-à-dire les distance entre les nœuds, reste celle définie uniquement à l'aide des arêtes de G appelés liens courts. Lors de l'exécution du routage glouton, un nœud connaît la distance (dans G) de tous ses voisins à la source, y compris celle de ses voisins par les liens longs.

Le routage glouton dans un graphe augmenté a été défini dans [Kle00] afin de donner un modèle de protocole de routage décentralisé pour modéliser l'expérience de Milgram. C'est le protocole de routage sans mémoire où la décision de routage prise par un nœud u pour envoyer un message vers le destinataire t est constitué de fonctions simples, le nœud u :

(1) sélectionne un de ses voisins v (voisin dans G ou contact lointain) le plus proche de t pour la distance dans G,

(2) transfère le message à v.

Ce processus suppose que chaque nœud a une connaissance des distances dans G. Par contre un nœud n'a pas connaissance des liens longs ajoutés à G en dehors du ou des siens.

Par la suite, nous considérons en particulier les graphes augmentés où il y a un unique lien long par nœud. Lorsqu'on ajoute un lien long par nœud $u \in G$ qui est l'extrémité sortante de l'arc, et que l'autre extrémité v est choisie selon une loi de probabilité $\varphi(u, v)$ le graphe augmenté sera noté (G, φ).

Définition. *Le* **diamètre glouton***, noté* greedy*, est le temps moyen pour qu'un message arrive à destination, dans le graphe augmenté, lorsque tous les nœuds suivent l'algorithme de routage glouton.*

Remarquons que le diamètre glouton d'un graphe non augmenté exactement la distance moyenne dans ce graphe, et que

$$\mathrm{diam}(G, \varphi) \leq \mathrm{greedy}(G, \varphi) \leq \mathrm{diam}(G).$$

Dans [Kle00], l'auteur a étudié le cas des grilles $k \times k$. Lorsqu'il y a un lien long de u vers v avec probabilité $\varphi(u, v)$ proportionnelle à $d(u, v)^{-r}$ (on parle alors de loi r-harmonique), le temps moyen du routage glouton dans (G, φ) est :

- $O(\log^2 k)$ si $r = 2$
- $\Omega(k^{\frac{2-r}{3}})$ si $0 \leq r < 2$
- $\Omega(k^{\frac{r-2}{r-1}})$ si $r > 2$

Ces résultats peuvent êtres généralisés aux autres dimensions. Dans [BFKK01], les auteurs se sont intéressés au cas de l'anneau, et ont donné des bornes inférieures et supérieures au diamètre glouton pour toutes les distributions r-harmoniques. En particulier, si $r = 1$ alors le diamètre glouton est $\Theta(\log^2 n)$. Ce résultat est généralisé dans [MN04] où les auteurs montrent que, quelle que soit la dimension d, le diamètre glouton des grilles d-dimensionnelles augmentées selon la loi d-harmonique est effectivement $\Theta(\log^2 n)$, quand bien même le graphe augmenté est de diamètre $\Theta(\log n)$ (montré pour $d = 2$).

Ce modèle introduit par Jon Kleinberg a été utilisé en pratique dans des systèmes pair-à-pair [MNR02, ZGG04, MBR03].

en $O(\log^{1+\frac{1}{d}} n)$ avec de la mémoire supplémentaire

Une généralisation classique de la distribution d-harmonique pour la grille d-dimensionnelle est la distribution inverse au rang d'un nœud, c'est-à-dire au nombre de nœuds au moins aussi proches que lui, formellement $\varphi(u,v) = \frac{1}{|B(u,d(u,v)|}$, nous l'appelleront distribution par rang. De plus, cette distribution est en accord avec le résultat empirique faite sur les réseaux sociaux dans [LNNK+05] qui montre que deux tiers des liens d'amitiés sont distribués ainsi.

Depuis, de nombreuses études se sont intéressées aux augmentations de familles de graphes particulières, et ont entre autre montré l'existence d'une augmentation donnant un diamètre glouton polylogarithmique en la taille du graphe dans de nombreux cas.

Dans [Fra05], pour les graphes de largeur arborescente bornée par k, l'auteur a montré l'existence d'une augmentation donnant un diamètre glouton en $O(k \log^2 n)$ basé sur la décomposition arborescente. Ce problème étant NP-difficile, il a aussi donné un algorithme polynomial garantissant un diamètre glouton d'au plus $O(k\sqrt{\log k} \log^2 n)$. En tant que corollaire, il obtient l'existence d'une augmentation donnant un diamètre glouton de $O((k + \log n) \log n)$ pour les graphes de triangularité bornée par k.

Le cas de graphes à croissance bornée, c'est-à-dire lorsqu'il existe une constante $\alpha > 0$, telle que la taille de chaque boule de centre u dans le graphe de base peut être écrite comme $b_u(r) = r^{d_u(r)}$ où $d_u(r) : [2, \infty) \to \mathbb{R}$ est $\mathcal{C}^1$ et satisfait $\forall r \geq 2, d'_u(r) \leq \frac{\alpha}{r \ln r}$, a été étudié dans [DHLS06]. Il est montré qu'à l'aide d'une augmentation proche de l'augmentation par rang, c'est-à-dire à un facteur polylogarithmique près, le diamètre glouton devient polylogarithmique. Les auteurs montrent en fait un résultat plus fort, ils montrent que la distance gloutonne est une fonction polylogarithmique de la distance dans le graphe de base. Ce résultat y est aussi étendu aux produits cartésiens de deux graphes à croissance bornée, qui ne donne pas nécessairement un graphe à croissance bornée. Remarquons que le facteur polylogarithmique utilisé dans l'augmentation variante de celle par rang permet essentiellement d'avoir un facteur de normalisation borné par une constante au lieu de $\log n$ ce qui permet alors d'exprimer les résultats indépendamment de n.

Dans le cas des graphes à dimension doublante bornée, [Sli07] a montré qu'en rajoutant un nombre polylogarithmique d'arêtes, on peut rendre le diamètre glouton logarithmique.

Dans [AG06] il est montré que toute famille de graphes (pondérés) excluant un mineur fixe, il existe une augmentation telle que le diamètre glouton soit polylogarithmique. En particulier, il est montré que pour tout graphe k-chemin-séparable de ratio [2] ϱ, il existe une distribution des liens longs qui soit calculable en temps polynomial et telle que le routage glouton prenne un temps $O(k^2 \log^2 n \log^2 \varrho)$ en moyenne et si un mineur H est exclu d'un graphe G alors G est k-chemin-séparable

2. rapport entre le plus grand et le plus petit poids

avec une constante k dépendant uniquement de H.

Il est facile de démontrer que si φ est la loi uniforme alors le temps du routage glouton est $O(\sqrt{n})$, ce résultat est attribué à David Peleg dans [FGK+09] où on trouvera la preuve. Dans [FGK+09] les auteurs ont montré qu'il existe une augmentation, définie pour n'importe quel graphe, qui garantit que le routage glouton s'effectue en $\tilde{O}(n^{\frac{1}{3}})$ étapes.

Dans [FLL10], les auteurs ont donné une borne inférieure $\Omega(n^{\frac{1}{\sqrt{\log n}}})$ au temps moyen d'un routage glouton pour toute distribution des liens longs, cette borne est valable pour tous les algorithmes de routage réduisant à chaque envoi la distance au destinataire. Elle a été approchée dans [FG10] à l'aide d'une légère variante du routage glouton appelée routage glouton focalisé.

D'autres auteurs se sont intéressés à des algorithmes de routage non gloutons, dans [MN04], en rajoutant un peu d'information, dans le cas des grilles d-dimensionnelles avec l'augmentation d-harmonique ils parviennent à un temps de routage $O(\log^{1+\frac{1}{d}} n)$. Dans ce même article complété de [NM05] on apprend que bien que le diamètre glouton n'est polylogarithmique avec l'augmentation r-harmonique que lorsque $r = d$, le diamètre de la grille augmentée est $\Theta(\log n)$ si $r \leq d$, polylogarithmique si $d < r < 2d$ et exponentiel si $r > 2d$. Le cas $r = 2d$ est encore ouvert.

Dans [FGP06], la majoration $O(\log^{1+\frac{1}{d}} n)$ est aussi montrée pour un protocole sans information, mais qui permet au nœud de consulter ses voisins avant d'envoyer le message, dans ce cas cette borne est aussi montrée exacte.

4.3.3 Lien entre routage glouton et diffusion spatiale

Cette partie est dédiée à la démonstration d'un théorème général permettant de transposer certains résultats sur le routage dans les graphes augmentés en un résultat sur la diffusion spatiale de même métrique.

Théorème 4.3.1. *Soit (G, φ) un graphe augmenté. Si, pour tout s, t, la distance gloutonne de s à t est majorée par une fonction $f(B(s, \delta(s,t)))$ ou $f(B(t, \delta(s,t)))$ décroissante pour l'inclusion alors la diffusion spatiale de source s dans G du protocole où un nœud u informé appelle*

- *un voisin dans G avec probabilité $\frac{1}{2\deg(u)}$,*
- *un nœud v de G avec probabilité $\frac{\varphi_u(v)}{2}$,*

est telle que tous les nœuds v de $B(s,d)$ sont informés en au plus

$$O(\Delta_{B(s,d)} \log d(f(B(s,d)) + \log b))$$

étapes, où $\Delta_{B(s,d)}$ est le degré maximum des nœuds intérieurs à $B(s,d)$ c'est-à-dire les nœuds de $B(s, d-1)$ et $b = |B(s,d)|$.

Ce théorème s'applique difficilement lorsque $\varphi_u(v)$ est inversement proportionnel à $|B(u, \delta(u,v))|$ puisque que le facteur de normalisation peut devenir $\log n$, mais on

trouve des résultats convenables comme dans [DHLS06] où $\varphi_u(v)$ est inversement proportionnel à $|B(u, \delta(u, v))|(\log |B(u, \delta(u, v))|)^r$ pour $r > 1$.

Démonstration. Puisque le choix d'un destinataire par un "lien long" est indépendant, sans perte de généralité, nous pouvons considérer que pour chaque nœud $u \in V$ la liste des destinataires par liens longs est tirée aléatoirement dès le début du processus ; et qu'avec probabilité $\frac{1}{2}$ il appelle (uniformément) un voisin, et probabilité $\frac{1}{2}$ il appelle le suivant dans la liste.

Si nous considérons un processus secondaire, qui fait les mêmes choix aléatoires sauf que s'il doit appeler un nœud par un lien long, il rappelle toujours le premier dans la liste, nous pouvons aussi considérer qu'après le premier ces appels sont perdus. Un nœud v est nécessairement informé plus rapidement dans la diffusion spatiale (normale) que dans le processus secondaire.

Considérons alors un processus tierce, qui fait à nouveau les mêmes choix aléatoires sauf que lorsqu'un nœud u doit appeler par un lien long (c'est-à-dire avec une probabilité $\frac{1}{2}$), il n'effectue l'appel vers le premier élément qu'avec probabilité $\frac{1}{\deg(u)}$, les autres appels étant perdus. Un nœud v est nécessairement informé plus rapidement dans le processus secondaire que dans le processus tierce.

Or ce processus tierce est connu, en effet il correspond au protocole Random sauf qu'il y a des pertes (avec probabilité majorée par $\frac{1}{2}$), la même analyse que dans la preuve de [FPRU90] donnant la borne générale sur le temps de diffusion $O(\deg(G)(\mathrm{diam}(G) + \log n)))$ nous donne alors la majoration

$$O(\max\{\Delta_{B(v,d)}, \deg(s)\}(\mathrm{greedy}(s, v) + \log b))$$

sur le temps de diffusion à v, avec forte probabilité, en suivant le chemin glouton vers v.

Le terme $\max\{\Delta_{B(v,d)}, \deg(s)\}$ vient du fait que dans le routage glouton, l'information se rapproche toujours du destinataire. C'est donc un majorant du degré des nœuds rencontrés sur le chemin glouton de s à t.

Remarquons que $B(v, d)$ n'est pas nécessairement inclus dans $B(u, d)$, pour pallier cette difficulté, nous raisonnons par dichotomie. Sur un plus court chemin de s à v, notons v_i le nœud à distance $\frac{\delta(u,v)}{2^i}$ de v. En particulier $s = v_0$ et $v = v_{\log d}$. Et au lieu de considérer le chemin glouton de s à v, considérons le chemin constitué de la concaténation des $\log d$ sous-chemins gloutons de v_i à v_{i+1}, nous avons alors que

$$B(v_{i+1}, \delta(v_{i+1}, v_{i+2})) \subseteq B(v_i, \delta(v_i, v_{i+1}))$$

et

$$B(v_{i+2}, \delta(v_{i+1}, v_{i+2})) \subseteq B(v_{i+1}, \delta(v_i, v_{i+1})).$$

Nous pouvons alors conclure en appliquant le résultat précédent sur chacun des $\log d$ sous-chemins gloutons. □

Puisque les liens longs vers les voisins sont totalement inutiles pour le routage, nous obtenons facilement le corollaire suivant.

Corollaire 4.3.2. *Nous obtenons exactement le même résultat si un nœud u informé appelle :*

- *un voisin dans G avec probabilité $\frac{1}{2\deg(u)}$*
- *un nœud v non voisin dans G avec probabilité $\frac{\varphi'_u(v)}{2}$ où φ'_u est la distribution φ_u où les voisins ne peuvent pas recevoir de lien long. Formellement, $\varphi'_u(v)$ est proportionnel à $\varphi'_u(v)$ si v n'est pas un voisin de u et 0 sinon.*

La même preuve convient aussi pour adapter tout protocole de routage sur un graphe augmenté dans lequel, à chaque étape d'envoi, le message se rapproche de la cible.

4.4 Le protocole de diffusion Rankscale

Dans cette section nous définissons le protocole Rankscale. Comme c'est un protocole de diffusion spatiale 1-Port, la seule chose que nous avons besoin de spécifier est la façon dont un nœud u informé sélectionne un nœud v à chaque étape. Dans [BF10] nous nous étions inspirés du processus d'augmentation du graphe de [FGK+09], dans le sens où le processus utilise un ensemble de boules de taille croissante exponentiellement dans lequel les nœuds sont sélectionnés et qu'après avoir déterminé une boule, le nœud était choisi uniformément aléatoirement dans cette boule. Dans [FGK+09], le paramètre k correspondant à la boule de taille 2^k considéré était choisi aléatoirement uniformément dans $\{1, \ldots, \log n\}$, mais dans [BF10] il est choisi avec une probabilité décroissante proportionnelle à $\frac{1}{k}$ de plus le processus de sélection donne un point plus important aux nœuds directement voisins (dans [FGK+09] ceci était logiquement ignoré puisqu'il s'agit de l'augmentation d'un graphe pour du routage, et que chaque nœud contactera un de ses voisins directs si le voisin par le lien long n'est pas meilleur). Ici nous nous inspirons plutôt de l'augmentation basée sur le rang de [DHLS06] qui est un bon compromis entre les augmentations utilisées dans [KKD04] et [FG10].

À chaque étape, le nœud $v \neq u$ est choisi par u avec probabilité :

- $p_{u,v} = \frac{1}{\sigma\ |B(u,\delta(u,v)|\log^{\frac{3}{2}}(|B(u,\delta(u,v)|)}$ si v n'est pas voisin de u ($v \in N(u)$),
- $p'_{u,v} = \frac{1}{\sigma\ |B(u,\delta(u,v))|}$ si v est voisin de u ($v \notin N(u)$),

où σ est une constante globale majorant pour tout u le facteur de normalisation

$$\sigma_u = \sum_{v \in N(u)} \frac{1}{|B(u,\delta(u,v)) - 1|} + \sum_{v \notin N(u)} \frac{1}{|B(u,\delta(u,v)|\log^{\frac{3}{2}}(|B(u,\delta(u,v)|)}.$$

La constante σ est bien définie, en particulier nous prendrons la limite de la série $1 + \sum_{i \geq 2} \frac{1}{i \log^{\frac{3}{2}}(i)}$. En effet, comme dans [FG10], en triant les nœuds $v \neq u$ tels que $\delta(u, v_i) \leq \delta(u, v_{i+1})$, la première partie de la somme correspondant aux voisins de u vaut 1, et pour la seconde partie nous avons que $|B(u, \delta(u, v_i)| \geq |i + 1|$ et donc

$$\sigma_u \leq 1 + \sum_{i \geq 2} \frac{1}{i \log^{\frac{3}{2}}(i)}.$$

De plus nous avons

$$\sigma_u \geq \sum_{v \in N(u)} \frac{1}{|B(u,\delta(u,v)| \log^{\frac{3}{2}}(|B(u,\delta(u,v)|)} = \frac{|B(u,\delta(u,v)| - 1}{|B(u,\delta(u,v)|} \geq \frac{1}{2},$$

ce qui montre que prendre σ à la place de σ_u ne change la probabilité que par un facteur constant.

Notons que $\sum_{v \neq u} p_{u,v}$ peut être strictement plus petit que 1, dans ce cas nous ignorons ce qui reste ou nous considérons que u s'appelle inutilement avec probabilité $1 - \sum_{v \neq u} p_{u,v}$. Notons aussi que, si nécessaire, on peut facilement considérer des défaillances lors des transmissions en prenant une valeur plus grande que σ. En effet, pour considérer des pannes indépendantes survenant avec probabilité p lors des transmissions, il suffit de considérer $\sigma' = \frac{\sigma}{1-p}$ au lieu de σ.

Dans [FG10], la probabilité que u ait un lien long vers v est de l'ordre de $\frac{1}{\sigma\ |B(u,\delta(u,v))|}$ ce qui peut donner un facteur de normalisation de l'ordre de $\log n$ pour un graphe fini à n sommets. Ceci ne permet pas d'avoir une fonction de propagation bien définie en fonction des paramètres de la boule de centre s, et de plus n'est pas du tout défini pour les métriques infinies. Dans [KKD04] sur les graphes à densité uniforme, cela revient à ce que u appelle v avec une probabilité de l'ordre de $\frac{1}{\sigma\ |B(u,\delta(u,v))|^{1+\epsilon}}$ pour tout $\epsilon > 0$. Nous pouvons faire de même, mais on augmenterait alors le temps de propagation et σ devrait dépendre de ϵ.

Protocole Rankscale **:**
Soit $b = B(u,\delta(u,v))$,

$$\mathbb{P}(u \to v) = \begin{cases} p_{u,v} = \frac{1}{\sigma\ b \log^{\frac{3}{2}}(b)} & \text{si } u \text{ et } v \text{ sont voisins,} \\ p'_{u,v} = \frac{1}{\sigma\ b} & \text{sinon.} \end{cases}$$

Nous pouvons déjà faire deux constatations utiles par la suite :

- $p_{u,v} \leq p'_{u,v}$,
- si u et v sont voisins alors $p'_{u,v} \geq \frac{\deg(u)}{2\sigma}$.

Remarque. Nous pouvons aussi envisager de légères variantes. Nous pouvons remplacer l'exposant $\frac{3}{2}$ par n'importe quelle constante strictement supérieure à 1, nous avons choisi $\frac{3}{2}$ pour les simplifications dans la preuve. En prenant $\mathbb{P}(u \to v) = \frac{1}{\sigma\ b \log^{\frac{3}{2}}(b)}$ pour avoir un protocole défini plus simplement, dans ce cas, nous auvons essentiellement Δ qui est remplacé par $\Delta \log^{\frac{3}{2}} \Delta$.

4.4.1 Temps de propagation de Rankscale, première approche

Dans cette section, nous démontrons le résultat principal suivant :

Théorème 4.4.1. *Pour toute métrique de graphe (V,δ) et toute source $s \in V$, le protocole* Rankscale *assure qu'un message de source s atteint tous les nœuds d'une boule de taille b centrée en s en moins de $O\left((b^{\frac{1}{3}} \log^{\frac{1}{2}} b + \Delta) \log b\right)$ étapes, avec forte probabilité.*

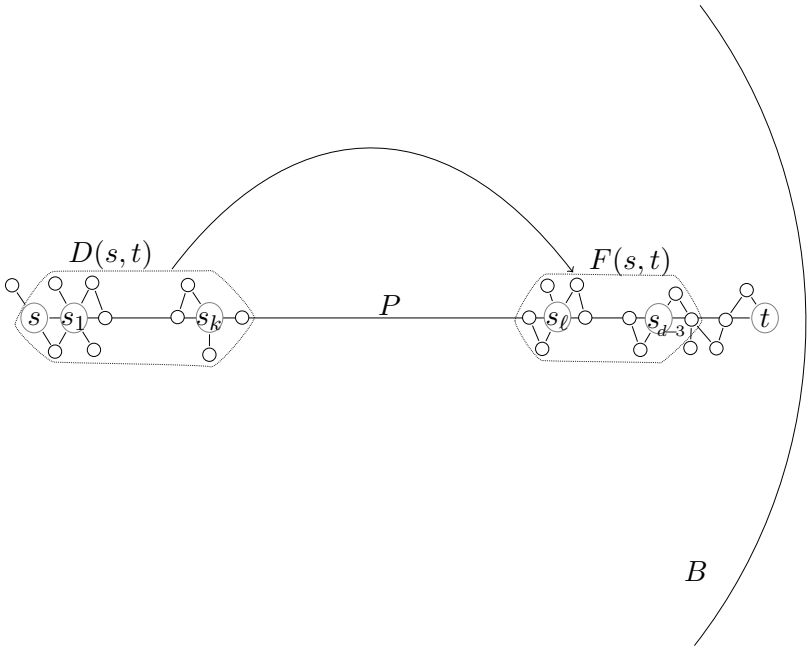


FIGURE 4.1 – Représentation graphique de la preuve du théorème 4.4.1

Démonstration. Soit (V, δ) une métrique de graphe, $s \in V$ et B une boule centrée en s contenant $b = |B|$ nœuds. Nous montrons que, avec forte probabilité, tous les nœuds de B reçoivent une information de s au plus $O\left((b^{\frac{1}{3}} \log^{\frac{1}{2}} b + \Delta) \log b\right)$ étapes après qu'elle soit apparue en s.

Soit $\nu = (\sigma b \log^{\frac{3}{2}}(b))^{\frac{1}{3}}$.

Soit t un nœud de B et $P = (s_0, s_1, \ldots, s_d)$ un plus court chemin de $s_0 = s$ à $s_d = t$ où $d = \delta(s,t)$. Nous définissons $D(s,t)$ et $F(s,t)$ comme suit :

- soit k le plus petit indice tel que $\bigcup_{i=1}^{k} s_i \cup N(s_i)$ soit de taille au moins ν, $D(s,t) = \bigcup_{i=1}^{k} s_i \cup N(s_i)$
- soit ℓ le plus grand indice tel que $\bigcup_{i=\ell}^{d-3} s_i \cup N(s_i)$ soit de taille au moins ν, $F(s,t) = \bigcup_{i=\ell}^{d-3} s_i \cup N(s_i)$

Une représentation graphique en est donnée à la figure 4.1.

Remarquons que $D(s,t)$ et $F(s,t)$ vérifient la propriété 4.1 pour n'importe quelle source qui est un nœud intérieur (c'est-à-dire, un nœud n'ayant pas de voisin hors de l'ensemble) et que l'intérieur de $D(s,t)$ ou de $F(s,t)$ est un sous-chemin (connexe) de P.

La démonstration se décompose en trois grandes parties, chacune nécessitant $O\left((b^{\frac{1}{3}} \log^{\frac{1}{2}} b + \Delta) \log b\right)$ étapes :

- d'abord, informer l'ensemble des nœuds de $D(s,t)$ à partir de s,
- ensuite, informer un nœud $v \in F(s,t)$ à partir d'un nœud $u \in D(s,t)$,
- enfin, informer t à partir de v.

***Affirmation* 4.4.2.** *Le nombre moyen d'étapes du protocole* RANKSCALE *pour qu'un nœud u quelconque de $D(s,t)$ reçoive l'information est borné par $\nu + 2\Delta$.*

Preuve. Soit un nœud u de $D(s,t)$, considérons le chemin simple de s à u dont tous les nœuds sauf éventuellement u sont sur le plus court chemin de s à t qui a servi à définir $D(s,t)$, c'est-à-dire $s_0, s_1, \ldots, s_i, u$ où $i \in \{1, \ldots, k\}$ tel que s_i soit un voisin de u

$$\begin{aligned}\sum_{h=0}^{i} \deg(s_h) &= \deg(s_0) + \sum_{h=1}^{i} \deg(s_h) \\ &\leq \Delta + |D(s,t)| \\ &\leq \nu + 2\Delta.\end{aligned}$$

De $s = s_0$ à u il existe donc un chemin dont la somme des degrés des nœuds est majorée par $\nu + 2\Delta$. Comme, par construction, un nœud x appelle son voisin avec probabilité au moins $\frac{1}{2\sigma \deg(x)}$, avec les mêmes arguments que pour la preuve de la proposition 4.2.1 u est informé en moyenne en moins de $2\sigma(\nu + 2\Delta)$ étapes. ◇

***Affirmation* 4.4.3.** *Chaque nœud $u \in D(s,t)$ appelle un nœud $v \in F(s,t)$ avec probabilité au moins $p_{s,t}$. Ainsi, si tous les nœuds de $D(s,t)$ sont informés, un nœud $v \in F(s,t)$ l'est en moyenne ν étapes plus tard.*

Preuve. Soit $u \in D(s,t)$ et $v \in F(s,t)$. Comme $p_{u,v} \leq p'_{u,v}$, pour montrer la première partie, nous n'avons qu'à montrer que $p_{u,v} \geq p_{s,t}$, et par définition de p il suffit de montrer que $B(u, \delta(u,v)) \subset B(s,d)$ où $d = \delta(s,t)$. Soit $i \in \{1, \ldots, k\}$ et $j \in \{\ell, \ldots, d-3\}$ tels que s_i soit un voisin de u et s_j soit un voisin de v. $\delta(u,v) \leq \delta(u,s_i) + \delta(s_i,s_j) + \delta(s_j,v) = j - i + 2$ donc $\delta(s,u) + \delta(u,v) \leq (i+1) + (j - i + 2) = j - 3 \leq d$ d'où nous déduisons $B(u, \delta(u,v)) \subset B(s,d)$ et finalement le résultat.

Dans un second temps il suffit de constater que $B(s,t)$ et $D(s,t)$ sont de taille au moins $\nu = (\sigma b \log^{\frac{3}{2}}(b))^{\frac{1}{3}}$ et $p_{u,v} \geq p_{s,t} \geq (\sigma b \log^{\frac{3}{2}}(b))^{-1} = \nu^{-3}$. ◇

***Affirmation* 4.4.4.** *Le nombre moyen d'étapes du protocole* RANKSCALE *pour que le nœud t reçoive l'information une fois qu'un nœud v quelconque de $F(s,t)$ est informé, est borné par $\nu + 4\Delta$*

Preuve. Nous suivons encore une fois la même idée de la démonstration, soit $j \in \{\ell, \ldots, d-3\}$ tel que s_j soit un voisin de v. Considérons le chemin

$$v, s_j, s_{j+1}, \ldots, s_{d-2}, s_{d-1}, s_d.$$

$$\begin{aligned}\deg(v) + \sum_{h=j}^{d-1} \deg(s_h) &= \deg(v) + \sum_{h=j}^{d-3} \deg(s_h) + \deg(s_{h-2}) + \deg(s_{h-1}) \\ &\leq |F(s,t)| + 3\Delta \\ &\leq \nu + 4\Delta.\end{aligned}$$

De v à t il existe donc un chemin dont la somme des degrés des nœuds est majorée par $\nu + 4\Delta$. Comme, par construction, un nœud x appelle son voisin avec probabilité au moins $\frac{1}{2\sigma \deg(x)}$, avec les mêmes arguments que pour la preuve de la proposition 4.2.1, une fois que v est informé, t est informé en moyenne en moins de $2\sigma(\nu + 4\Delta)$ étapes. $\diamond$

Nous complétons la preuve en multipliant les temps par $k \log(b)$ avec k bien choisi afin d'obtenir, à l'aide de l'inégalité de Markov et de l'indépendance temporelle, des résultats en probabilité suffisamment forte pour appliquer l'inégalité de Boole.

Soit $T_{s,u}$ la variable aléatoire comptant le nombre d'étapes du protocole Rankscale nécessaire pour qu'une information apparaissant en s atteigne $u \in B(s,t)$. Soit $\alpha > 2$, nous savons que $\mathbb{E}[T_{s,u}] \leq 2\sigma(\Delta + 2\nu)$, donc par l'inégalité de Markov nous obtenons que

$$\mathbb{P}(T_{s,u} > 2\,\alpha\, \log(b)\, \mathbb{E}[T_{s,u}]) < \frac{1}{2^{\alpha \log b}} \leq \frac{1}{b^{\alpha}}.$$

et donc par inégalité de Boole

$$\mathbb{P}(\forall u \in D(s,t),\ T_{s,u} \leq 2\,\alpha\, \log(b)\, 2\sigma(\Delta + 2\nu)) \geq 1 - \frac{1}{b^{\alpha-1}}$$

Après $2\alpha \log(b) 2\sigma(\Delta + 2\nu)$ étapes tous les nœuds de $D(s,t)$ ont eu l'information avec probabilité au moins $1 - \frac{1}{b^{\alpha-1}}$. De même une fois que tout $D(s,t)$ est informé, nous montrons (uniquement avec l'inégalité de Markov) qu'avec probabilité $1 - \frac{1}{b^{\alpha}}$ il existe un nœud $v \in F(s,t)$ ayant reçu l'information $2\alpha \log(b)\nu$ étapes plus tard et aussi qu'une fois qu'un v est informé t obtient reçoit l'information $2\alpha \log(b) 2\sigma(\Delta + 4\nu)$ étapes plus tard. Nous concluons en utilisant encore une fois l'inégalité de Boole pour obtenir une probabilité que tous nœuds $t \in B$ aient l'information dans le même temps, avec probabilité $1 - O(\frac{1}{b^{\alpha-2}})$. $\square$

4.4.2 Temps de propagation de Rankscale, seconde approche

Une autre approche donne un résultat intéressant, la fonction en b est bien inférieure, mais le degré maximal de la boule est alors un terme multiplicatif au lieu d'additif.

Théorème 4.4.5. *Pour toute métrique de graphe (V, δ) et toute source $s \in V$, le protocole* Rankscale *assure qu'un message de source s atteint tous les nœuds d'une boule de taille b centrée en s en $o\Big(\Delta 2^{(\log b)^{\frac{1}{2}+\epsilon}}\Big)$ étapes, avec forte probabilité.*

Démonstration. Le temps moyen pour un message en s pour atteindre un message en t est majoré par le temps mis par s pour informer un nœud intermédiaire v additionné du temps mis par un message en v pour parvenir à t. Notons $T(s,t)$ le temps moyen, nous cherchons à majorer ce temps par une fonction F qui ne dépend que de $B(s, \delta(s,t))$, et plus particulièrement de sa taille $|B(s, \delta(s,t))|$, son degré

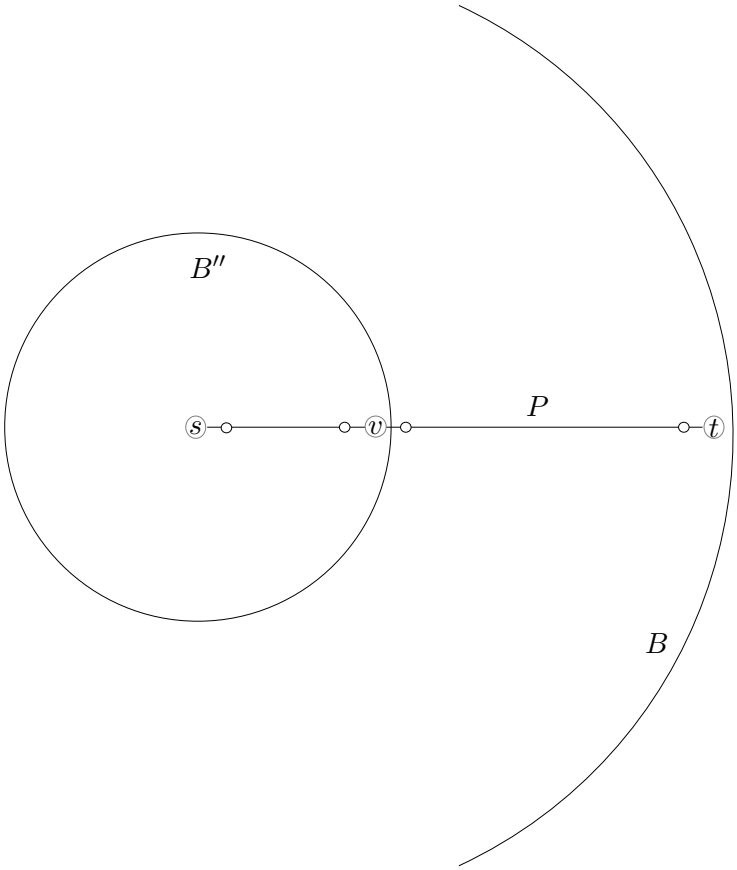


FIGURE 4.2 – Représentation graphique de la preuve du théorème 4.4.5

maximal $\Delta(B(s, \delta(s,t)))$ et la distance $\delta(s,t)$, avec F une fonction croissante de chacun de ces paramètres. Nous noterons les arguments de F dans l'ordre $F(d, b, \Delta)$.

Montrons alors, par récurrence sur d et b, que pour tout $\alpha > \frac{1}{2}$, il existe une constante γ, telle que si $f(b) = 2^{\gamma(\log b)^\alpha}$ alors $F(d, b, \Delta) = \Delta \log(d) f(b)$ majore tout $T(s,t)$ où s et t sont tels que $b \geq |B(s, \delta(s,t))|$, $\Delta \geq \Delta(B(s, \delta(s,t))$ et $d \geq \delta(s,t)$.

Soit B une boule de centre s. Notons $b = |B|$ et $\Delta = \Delta(B)$. Considérons $t \in B$, et notons $d = \delta(s,t)$. Par la suite, nous ne considérons que des boules B' incluses dans $B(s, \delta(s,t))$, ainsi d et Δ sont des majorants des valeurs $\delta(v,t)$ et $\Delta(B')$.

Nous avons que pour tout $b \leq \Delta$, l'inégalité $F(1, b, \Delta) \geq \Delta$ est vérifiée. Considérons un plus court chemin de s à t. Soit v le nœud sur ce plus court chemin à distance $\frac{d}{3}$ de s.

Comme $T(s,t) \leq T(s,v) + T(v,t)$, par récurrence,

$$T(s,t) \leq T(s,v) + F(\frac{2d}{3}, b', \Delta') \leq T(s,v) + F(\frac{2d}{3}, b, \Delta),$$

où b'' et Δ'' sont respectivement la taille et le degré maximum de $B'' = B(v, \frac{2d}{3})$.

Une représentation graphique en est donnée à la figure 4.2.

Nous n'appliquons en fait la récurrence que sur le second terme de la somme. Il suffit alors de montrer que $T(s,v) \leq \Delta f(b)$. Par récurrence, nous avons $T(s,t) \leq F(d,b',\Delta) = \Delta \log(d) f(b')$. Mais aussi, nous pouvons remarquer que pour tout $u \in B(s, \frac{\delta}{3})$, nous avons $B(u, \delta(u,v)) \subseteq B(s,\delta)$, et donc $p_{u,v} \geq p_{s,t} = \frac{1}{\sigma\ b \log^{\frac{3}{2}} b}$. Alors, soit $B' = B(s, \frac{d}{3})$ et b' sa taille, par l'affirmation qui suite nous avons $T(s,t) \leq c\frac{b \log^{\frac{5}{2}} b}{b'}$ pour une constante c.

***Affirmation* 4.4.6.** *Si chaque envoi se fait avec probabilité ρ, alors le temps moyen de diffusion aléatoire dans la clique est $2\frac{\ln n + \Theta(1)}{\rho}$.*

Preuve. En effet, s'il y a i nœuds informés, alors il faut en moyenne $\frac{n}{p(n-1)}$, c'est-à-dire $\frac{n}{ip(n-i)}$ étapes, pour informer un nouveau nœud. Par linéarité de l'espérance, le temps total moyen est donc

$$\begin{aligned}
\sum_{i=1}^{n-1} \frac{n}{ip(n-i)} &= \frac{1}{p} \sum_{i=1}^{n-1} \frac{n}{i(n-i)} \\
&= \frac{1}{p} \sum_{i=1}^{n-1} \frac{1}{i} + \frac{1}{(n-i)} \\
&= \frac{2}{p} \sum_{i=1}^{n-1} \frac{1}{i} \\
&= \frac{2}{p}(\ln(n) + \Theta(1))
\end{aligned}$$

◇

Si $b' \geq \frac{cb \log^{\frac{5}{2}} b}{f(b)}$ alors $\frac{cb'}{b \log^{\frac{5}{2}} b} \leq f(b) \leq \Delta f(b)$. Sinon, nous montrons que $T(s,t) \leq \Delta \log(d) f(b')$ d'une autre manière. Soit γ une constante (qui dépend de la constante α) telle que :

- $\gamma \geq \frac{1}{\alpha}$,
- $\frac{\gamma \log^{\alpha} x}{2} \geq \log c + \frac{5}{2} \log\log x$ dès que $x \geq 2$,
- et $\frac{\gamma \log^{2\alpha-1} x}{2} \geq \log\log x$ dès que $x \geq 2$.

Comme $\log d \leq \log b$ et $b' \leq \frac{cb \log^{\frac{5}{2}} b}{f(b)}$, nous montrons alors que

$$\log b \leq \frac{f(b)}{f\left(\frac{cb \log^{\frac{5}{2}} b}{f(b)}\right)}.$$

L'inégalité étant évidente pour $b = 1$ ou $b = 2$, nous montrons uniquement qu'elle est

vérifiée pour $b > 2$. Pour $b > 2$, nous avons

$$\begin{aligned}
&\gamma(\log b)^\alpha - \gamma(\log b + \log c + \frac{5}{2}\log\log b - \gamma(\log b)^\alpha)^\alpha \\
&= \gamma(\log b)^\alpha[1 - (\log b + \log c + \frac{5}{2}\log\log b - \gamma(\log b)^\alpha)^\alpha] \\
&= \gamma(\log b)^\alpha[1 - (1 + \frac{\log c + \frac{5}{2}\log\log b}{\log b} - \gamma(\log b)^{\alpha-1})^\alpha] \\
&\geq \alpha\gamma(\log b)^\alpha[\gamma(\log b)^{\alpha-1} - \frac{\log c + \frac{5}{2}\log\log b}{\log b}] \\
&= \alpha\gamma[\gamma(\log b)^{2\alpha-1} - (\log b)^{\alpha-1}(\log c + \frac{5}{2}\log\log b)] \\
&\geq \log\log b \text{ d'après les hypothèses sur } \gamma.
\end{aligned}$$

Comme

$$\log\log b \leq \gamma(\log b)^\alpha - \gamma(\log b + \log c + \log\log b - \gamma(\log b)^\alpha)^\alpha$$

et

$$f(b) = 2^{\gamma(\log b)^\alpha},$$

nous en déduisons l'inégalité souhaitée.

Nous concluons en multipliant par $\log b$ pour avoir la forte probabilité, ce terme étant masqué par l'expression de $f(b)$. □

Remarque. Nous pourrions obtenir le même résultat en appliquant le théorème 4.3.1 et en refaisant la preuve de [FG10] avec la même distribution que dans [DHLS06], qui ajoute uniquement un facteur polylogarithmique, et permet ainsi d'avoir un facteur de normalisation qui ne dépend plus de n. Plus formellement, la probabilité qu'il y ait le lien long de u vers v n'est alors plus $\frac{1}{\nu_u|B(u,\delta(u,v))|}$ avec un facteur de normalisation ν_u éventuellement logarithmique en n, mais avec probabilité $\Phi_u(v) = \frac{1}{|B(u,\delta(u,v))*(\log|B(u,\delta(u,v))|)^r}$ où $r > 1$, ce qui donne un facteur de normalisation ν_u borné par des constantes.

Quasiment toute la preuve reste inchangée. Notons $b = |B(t,\delta(s,t))|$. Essentiellement, le lemme 8 devient $\text{ecc}[||\sigma(v,u,r)||] = O(\kappa\log^r b)$. Ensuite dans la fin de la preuve du théorème 1, faisons attention en appliquant les lemmes 5 à 7. En effet, dans le lemme 5, $I \leq c\log^3 d_v \leq c\log^3 b$, dans le lemme 6, $I \leq \log_\kappa N(u,\hat{r}) \leq \log_\kappa b$. Pour le lemme 7, nous obtenons aussi $I \leq \log_\kappa b$ à condition de redéfinir le cas $\perp$ pour que $D_\perp = B(t,\delta(s,t)|$ ce qui ne change pas la preuve.

Nous obtenons ainsi que $\text{greedy}(s,t) \in o(2^{(\log|B(s,\delta(s,t))|)^{\frac{1}{2}+\epsilon}})$, pour tout $\epsilon > 0$. Puis nous concluons en appliquant le théorème 4.3.1.

4.4.3 Cas des métriques à densité uniforme

Le protocole RANKSCALE peut aussi s'appliquer aux métriques quelconques (et pas uniquement aux métriques de graphe) à condition d'apporter une petite modification au processus de sélection : soit en considérant que les voisins d'un nœud sont ceux à distance au plus 1, soit en définissant le processus de sélection uniquement à l'aide de la probabilité suivante :

$$\mathbb{P}(u \to v) = p_{u,v}$$

Ce protocole est une très légère amélioration de [KKD04], puisque que, dans la distribution de probabilité, nous pouvons avoir un exposant sur la taille de la boule qui est essentiellement 1 sans avoir un facteur de normalisation polylogarithmique en n. Nous obtenons un résultat similaire en suivant la même preuve.

Théorème 4.4.7. *Pour toute métrique (V, δ) à densité uniforme, et toute source $s \in V$, le protocole* RANKSCALE *assure qu'une information introduite au nœud s atteint tous les nœuds d'une boule B de taille b centrée en s en au plus $O(\log^2 b \log\log b)$ étapes, avec forte probabilité.*

Nous rappelons que (V, δ) une sous-métrique de $(\mathbb{R}^D, \mathcal{L}_\in)$ est dite ***à densité uniforme*** si toutes boules de rayon $d \geq 1$ contient entre $\beta_1 d^D$ et $\beta_2 d^D$ nœuds de V.

Démonstration. La preuve suit les mêmes idées que l'analyse du protocole DENSITY dans [KKD04]. Sans perte de généralité, dans cette preuve nous supposons qu'un nœud effectue un appel à chaque étape, même s'il n'est pas informé, et qu'il n'informe donc le nœud appelé que s'il est lui même informé. Soit $s \in V$, B une boule de centre s, de taille b et de rayon $d \geq 1$. Par l'hypothèse de densité uniforme,

$$\beta_1 d^D \leq b \leq \beta_2 d^B$$

donc la probabilité qu'un nœud appelle un nœud à distance au plus d est minoré par

$$x = \frac{1}{\sigma\ \beta_2 d^D \log^{\frac{3}{2}} (\beta_2 d^D)}.$$

Soit c_s un point de l'espace (pas nécessairement dans V) appartenant au segment $[s, t]$ le plus proche de s tel que $B(c_s, \delta(s, c_s))$ soit de taille $x^{-\frac{1}{2}}$, nous notons C_s cette boule. De même, soit c_t un point de l'espace (pas nécessairement dans V) appartenant au segment $[s, t]$ le plus proche de t tel que $B(c_t, \delta(s, c_s))$ soit de taille $x^{-\frac{1}{2}}$, C_t cette boule. Une représentation graphique en est donnée à la figure 4.3.

Notons $\mathcal{E}$ l'événement "il existe au moins un appel de C_s vers C_t survenant pendant au moins l'une des $\ln \gamma$ étapes consécutives".

***Affirmation* 4.4.8.** *Nous avons $\mathbb{P}[\mathcal{E}] \geq 1 - \frac{1}{\gamma(b)}$.*

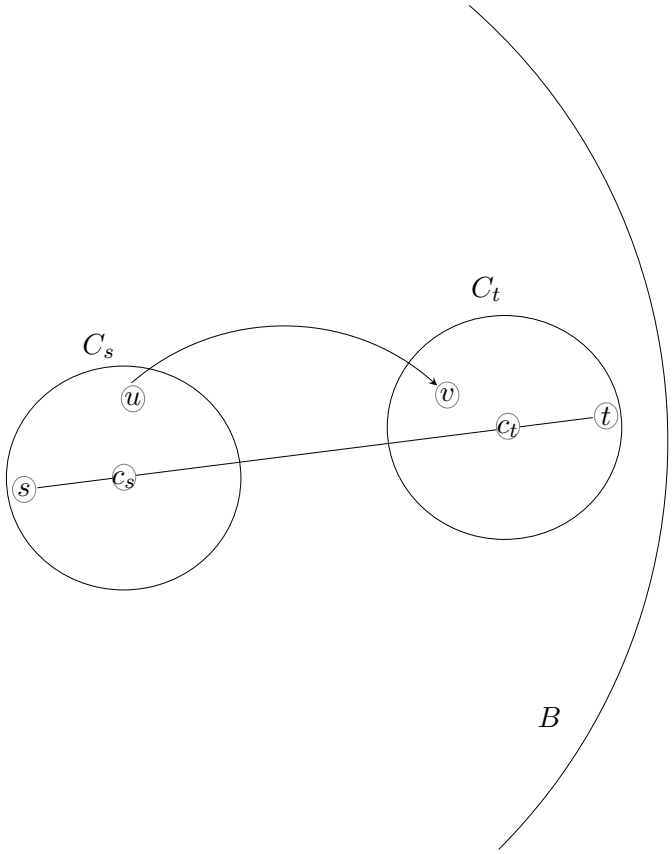


FIGURE 4.3 – Représentation graphique de la preuve du théorème 4.4.1

Preuve. Par construction, nous savons qu'un sommet de C_s est à distance au plus d d'un sommet de C_t, s'il est informé, il l'appelle donc avec probabilité au moins x. Ainsi, nous pouvons minorer la probabilité qu'il y ait un appel de C_s vers C_t. Nous avons

$$1 - p = \prod_{u \in U_s} (1 - \sum_{v \in U_t} p_{u,v})^{\ln(\gamma(b))}.$$

Puisque nous avons $p_{u,v} \geq x$, alors

$$1 - p \leq (1 - x|U_t|)^{\ln(\gamma)|U_s|} \leq e^{-x \ln(\gamma(b))\, |U_s|\, |U_t|}$$

et donc

$$1 - p \leq e^{-\ln(\gamma(b))} = 1 - \frac{1}{\gamma(b)}.$$

◇

À présent, nous montrons l'affirmation suivante.

***Affirmation* 4.4.9.** *Pour tout $t \in B$, t reçoit une information initiée en s en au plus $\gamma(b)g(b)$ étapes, avec probabilité au moins $1 - \frac{g(b)}{\gamma(b)}$ où $g(b)$ est une solution de la récurrence*

$$g(k) = 1 + 2g(\beta' \, b^{\frac{1}{2}} \log^{\frac{3}{4}}(\beta b)).$$

où β et β' sont des constantes valant respectivement $\frac{\beta_2}{\beta_1}$ et $\left(\frac{\beta_2}{\beta_1}\right)^{\frac{3}{2}} 2^D \sigma^{\frac{1}{2}}$.

Preuve. Nous établissons l'affirmation par induction. Considérons trois intervalles de temps consécutifs :

$$\begin{aligned} I_s &= [1, \gamma(b) \; g(\beta \, b^{\frac{1}{2}} \log^{\frac{3}{4}}(b))] \\ I &= [|I_s| + 1 \; , \; |I_s| + \gamma(b)] \\ I_t &= [|I_s| + |I| + 1 \; , \; |I_s| + |I| + \gamma(b) \; g(\beta \, b^{\frac{1}{2}} \log^{\frac{3}{4}}(b))]. \end{aligned}$$

Par l'inégalité, sur la probabilité de l'événement $\mathcal{E}$, que nous avons établi dans l'affirmation 4.4.8, nous savons que durant l'intervalle de temps I, il existe un nœud $u \in C_s$ qui appelle un nœud de $v \in C_t$ avec probabilité au moins $1 - \frac{1}{\gamma(b)}$. Par l'hypothèse de densité uniforme, comme s est dans une boule de centre c_s et de taille $x^{-\frac{1}{2}}$, $\delta(s, c_s)^D \leq \frac{x^{-\frac{1}{2}}}{\beta_1}$, d'où $\delta(s, u)^D \leq 2^D \frac{x^{-\frac{1}{2}}}{\beta_1}$, et donc c_s est dans une boule de centre s et de taille

$$2^D \frac{\beta_2}{\beta_1} x^{-\frac{1}{2}} = 2^D \frac{\beta_2}{\beta_1} \sigma^{\frac{1}{2}} (\beta_2 d^D)^{\frac{1}{2}} \log^{\frac{3}{2}} \left(\beta_2 d^D\right).$$

En appliquant une nouvelle fois l'hypothèse de densité uniforme , en particulier le fait que $d^D \leq \frac{b}{\beta_1}$, nous obtenons

$$2^D \frac{\beta_2}{\beta_1} \sigma^{\frac{1}{2}} (\beta_2 d^D)^{\frac{1}{2}} \log^{\frac{3}{4}} \left(\beta_2 d^D\right) \leq \beta' b^{\frac{1}{2}} \log^{\frac{3}{4}} (\beta b)$$

Ainsi, par l'hypothèse d'induction, le nœud u a reçu l'information de la source s durant l'intervalle de temps I_s avec probabilité $1 - \frac{g(\beta' \, b^{\frac{1}{2}} \log^{\frac{3}{4}}(\beta b))}{\gamma(b)}$. De même, par l'hypothèse d'induction, le nœud t reçoit le message de v, vu comme une source secondaire, durant l'intervalle de temps I_t avec probabilité $1 - \frac{g(\beta' \, b^{\frac{1}{2}} \log^{\frac{3}{4}}(\beta b))}{\gamma(b)}$. Par conséquent, durant l'intervalle de temps de $|I_s| + |I| + |I_t| = \gamma(b)g(b)$ étapes, le nœud t reçoit l'information de la source s avec probabilité au moins

$$\left(1 - \frac{1}{\gamma(b)}\right) \left(1 - \frac{g(\beta' \, b^{\frac{1}{2}} \log^{\frac{3}{4}}(\beta b))}{\gamma(b)}\right) \left(1 - \frac{g(\beta' \, b^{\frac{1}{2}} \log^{\frac{3}{4}}(\beta b))}{\gamma(b)}\right)$$

c'est-à-dire avec probabilité au moins $1 - \frac{g(b)}{\gamma(b)}$. $\diamond$

Comme $g(k) \leq O(k \log k)$, en appliquant l'affirmation 4.4.9, nous obtenons que pour tout $t \in B$, t reçoit l'information initiée en s en au plus $O(\log^2 b \log\log b)$ étapes, avec probabilité au moins $1 - O(\frac{\log b \log\log b}{\gamma(b)})$. Le théorème suit en prenant $\gamma(b) = b^\alpha$ avec $\alpha > 1$, puis en appliquant l'inégalité de Boole pour chaque nœud de B. □

Remarque. En prenant $\gamma(b) = \log^{1+\alpha}(b)$, nous obtenons un résultat comparable à celui de [KKD04], un nœud de la boule B de taille b est informé en $O(\log b(\log\log b)^2)$ étapes, avec probabilité $1 - (\log b)^{-\alpha}$.

4.5 Conclusion

Dans ce chapitre, nous avons prouvé qu'il existe un protocole de diffusion spatiale universel pour toutes les métriques de graphes, dont le temps propagation est $O\left((b^{\frac{1}{3}} \log^{\frac{1}{2}} b + \Delta) \log b\right)$ et $o\left(\Delta 2^{(\log b)^{\frac{1}{2}+\epsilon}}\right)$, pour tout $\epsilon > 0$. Une question naturelle est de savoir si cette borne peut être améliorée. Nous pensons qu'en combinant les deux approches il est possible d'améliorer les résultats sans toutefois pouvoir s'affranchir complètement du terme Δ. En particulier, il serait très intéressant de prouver ou réfuter l'existence d'un protocole de diffusion universel dont le temps de propagation est en $O(\Delta \log^\alpha b)$ ou même en $O(\Delta + \log^\alpha b)$ pour un $\alpha \geq 1$ quelconque. Une telle majoration polylogarithmique du temps de diffusion peut être obtenue dans des métriques spécifiques, dont en particulier celles à densité uniforme [KKD04]. Une autre extension naturelle de notre travail serait donc d'étendre nos résultats aux systèmes métriques arbitraires sans aucune hypothèse sur la densité. En fait, l'existence même d'un protocole dont le temps de propagation est fini n'est pas clair dans ce contexte général : Uniform a un temps de propagation non borné, et Density et Rankscale ont tout deux des temps de propagation polylogarithmiques dans les métriques à densité uniforme, mais leurs performances dans des métriques arbitraires ne sont pas connues.

Dans le cas des métriques de graphes, une variante pourrait être de supposer que les nœuds ont deux mécanismes de transmission d'information : un premier qui lui permet d'envoyer l'information à tous les nœuds à distance 1 en une étape (comme dans l'inondation), et un second lui permettant d'informer un seul autre nœud dans le graphe. Avec cette hypothèse, suggérée par George Giakkoupis, nos résultats se transposent en supprimant le terme Δ.

Chapitre 5

Conclusion et perspectives

5.1 Conclusion

Dans cette thèse, nous avons mis en évidence différents résultats sur le temps de diffusion de protocoles décentralisés, avec pour but d'améliorer la compréhension de ces protocoles et d'en proposer de nouveaux, plus rapides.

Dans le chapitre 2, nous nous sommes intéressés aux pires temps de diffusion 1-Port selon trois hypothèses sur l'information possédée par chaque nœud à propos de ses voisins, tels qu'introduits dans [HL98]. Lorsque chaque nœud ne se souvient que d'à qui il a déjà envoyé l'information, nous avons montré que le calcul du pire temps de diffusion est un problème polynomial. Lorsque chaque nœud se souvient aussi des voisins qui lui ont transmis l'information, le problème est alors NP-difficile et même $(\frac{1}{2}+\epsilon)$-inapproximable, pour tout $\epsilon > 0$, même dans le cas des graphes bipartis. Dans l'hypothèse considérée la plus forte, où chaque nœud sait quels sont ses voisins qui possèdent déjà l'information, le problème est également NP-difficile et même $((\frac{1}{n})^{\frac{1}{2}+\epsilon})$-inapproximable, pour tout $\epsilon > 0$. Nous avons aussi montré d'autres différences importantes entre ces modèles, telles que des écarts exponentiels entre les temps de diffusion, pour un même réseau, selon l'hypothèse considérée, ou bien la présence ou non de diodes. Nous avons également décrit différents outils pour une future exploration de ce problème. L'analyse de ce problème est en effet intéressante car il fournit une borne sur le temps de diffusion pour des protocoles de diffusion ListBased et QuasiRandom. Cette borne peut se révéler importante pour calibrer une application qui serait très sensible au délai d'arrivée des paquets comme la diffusion en flux continu. De ce point de vue, l'algorithme polynomial dans la première variante est un résultat positif. Par contre, le fait que ce calcul soit difficilement approximable si chaque nœud sait quels sont ses voisins informés est un résultat négatif. Toutefois, ce même résultat peut également être vu comme un résultat positif dans un autre contexte. En effet, il nous garantit qu'un adversaire ne peux pas facilement imposer des listes impliquant de trop mauvaises performances. Quel que soit le point de vue considéré, il serait intéressant d'améliorer nos résultats dans le cas où chaque nœud se rappelle avec qui il a déjà interagi, afin de savoir s'il

est possible ou non de calculer une "bonne" approximation du temps de diffusion et un schéma associé.

Dans le chapitre 3, nous avons considéré le cas d'un réseau dynamique modélisé par un graphe à évolution arête-markovienne. Nous avons décrit une technique générale permettant, dans le cas stationnaire, de réduire le temps d'inondation au calcul du diamètre d'un graphe pondéré aléatoirement mais statique. Nous avons également donné une technique permettant de majorer et minorer le temps d'inondation dans le cas général, en se ramenant au cas stationnaire. Nous avons appliqué ces techniques au cas de l'évolution arête-markovienne homogène (afin de compléter les résultats de [CMM+10, CMPS11]), ainsi qu'à certains cas hétérogènes. Finalement, nous donnons une borne sur le temps de diffusion aléatoire 1-Port pour certaines valeurs du taux de naissance et de mort. Grâce à nos résultats, nous avons amélioré la compréhension des réseaux dynamiques à évolution arête-markovienne. En effet, le temps d'inondation dans un réseau dynamique est une borne inférieure sur le temps de diffusion de tout protocole où les nœuds ne communiquent qu'avec leurs voisins directs. De plus, ce temps d'inondation donne une mesure des graphes dynamiques aussi importante que l'est le diamètre pour les graphes statiques. Même le résultat plus faible concernant la diffusion 1-Port permet de montrer que l'algorithme aléatoire Random est peu sensible à un fort dynamisme du réseau.

Dans le chapitre 4, nous avons étudié la diffusion spatiale. Dans la diffusion spatiale, introduite dans [KKD04], les communications peuvent se faire entre tous les nœuds. Par contre, le but n'est pas uniquement d'informer tous les nœuds rapidement (nous savons déjà que c'est faisable en temps logarithmique en la taille du réseau si le réseau est fini), mais d'informer chaque nœud en un nombre d'étapes dépendant de sa proximité à la source, indépendamment de la taille du réseau. Nous avons décrit un protocole, appelé Rankscale, dont le nombre d'étapes nécessaires pour informer un nœud est majoré en fonction des paramètres de la plus petite boule contenant ce nœud et centrée en la source. Le protocole Rankscale informe tous les nœuds dans une boule B de taille b et de degré maximal Δ_B en un nombre d'étapes majoré par le minimum entre $O\left((b^{\frac{1}{3}}\log^{\frac{1}{2}} b + \Delta)\log b\right)$ et $o(\Delta 2^{(\log b)^{\frac{1}{2}+\epsilon}})$, avec forte probabilité. Nous avons également donné une méthode permettant de transposer certains résultats concernant le routage dans les graphes augmentés en des résultats sur la diffusion spatiale. Les protocoles de diffusion spatiale ont au moins deux avantages. Ils peuvent, d'une part, servir à diffuser rapidement une information en privilégiant les nœuds proches de la source, ce qui est crucial aussi bien pour diffuser une alerte que pour rechercher des ressources proches. Le second avantage offert par les protocoles de diffusion spatiale est, d'autre part, celui de la modélisation. Ils pourraient en effet servir à modéliser la propagation d'un virus ou d'un "*buzz*".

Pour conclure, la diffusion est un problème clé en informatique dont les contraintes sont multiples. Les protocoles aléatoires de diffusion décentralisée répondent à certaines d'entre elles. Ils ont la capacité de s'adapter facilement à une évolution du réseau, ils nécessitent moins de ressources de calcul que les protocoles déterministes, et, dans de nombreux cas, leurs performances sont proches de l'optimal. Nous re-

tiendrons par ailleurs qu'ils peuvent également servir de modèle de propagation virale.

5.2 Perspectives

Outre les problèmes ouverts directement liés à nos travaux que nous avons mentionnés dans la dernière section de chacun des chapitres 2, 3 et 4, de nombreux problèmes concernant la diffusion décentralisée d'information demeurent. Nous identifions ci-dessous trois axes de perspectives pour nos recherches : d'autres mesures, d'autres tâches proches de la diffusion, et d'autres contraintes.

5.2.1 Autres mesures

Dans cette thèse nous n'avons considéré les performances des protocoles de diffusion qu'en termes de temps, c'est-à-dire du nombre d'étapes nécessaires pour informer tous les nœuds d'un réseau. L'analyse des performances peut évidemment être posée sous un autre angle. Par exemple, celui de la quantité d'information transmise, en termes du nombre total de messages ou de bits transmis. Dans [KSSV00], le protocole Random dans le modèle 1-Port et Push prend $\Theta(\log n)$ étapes pour diffuser l'information dans la clique. Toutefois, au total, l'information est envoyée $\Theta(n \log n)$ fois. Dans la mesure où les voisins sont appelés indépendamment et aléatoirement, il n'est pas surprenant que le nombre de communications nécessaires à la diffusions soit $\Theta(n \log n)$ puisque c'est le nombre de tirages nécessaires pour obtenir tous les coupons dans le problème du collectionneur de coupons. Dans la version Pull, le résultat est le même [1]. Dans [KSSV00], à chaque étape, chaque nœud établit une connexion avec un de ses voisins choisi aléatoirement, puis chacun des deux nœuds impliqués dans une communication détermine s'il envoie ou non l'information à son correspondant. Les auteurs proposent deux algorithmes ne nécessitant que $\Theta(n \log \log n)$ envois de l'information, et montrent qu'il faut toujours $\Omega(n \log \log n)$ transmissions avant que tous les nœuds soient informés par un algorithme ne distinguant pas les nœuds (*adress oblivious*), quelque soit son temps de diffusion [2].

D'autres études ont été effectuées sur ce problème. Dans [Els06], l'auteur propose un autre algorithme de type PushPull qui, sur les graphes aléatoires d'Erdős-Rényi $\mathcal{G}_{n,p}$ où $p \geq \frac{\ln^\delta n}{n}$ pour une constante $\delta > 2$, termine en temps $O(\log n)$

1. Évidement, si l'on suppose qu'un nœud ne fait plus d'appels une fois qu'il a obtenu l'information souhaitée, il y aurait alors uniquement $n-1$ transmission de l'information. Cette hypothèse est néanmoins peu réaliste. En effet, un nœud ne sachant pas à quel moment la nouvelle information apparaît sur le réseau, dans un modèle Pull, un nœud doit perpétuellement contacter ses voisins. Dans le cas où un nœud appelé renvoie systématiquement l'information, le nombre d'échanges serait encore $\Theta(n \log n)$.

2. Si l'algorithme peut distinguer les nœuds, il est possible d'avoir uniquement $n-1$ transmissions mais avec un temps de diffusion $\Theta(n \log n)$. Il suffit que seul le nœud à l'origine de l'information s'occupe des transmissions et ne la transmette que si un de ses voisins la lui demande pour la première fois ; mais en imposant que tous les nœuds doivent avoir reçu l'information en $O(\log n)$ étapes alors, un protocole nécessite au moins $\Omega(n \log n)$ transmissions.

avec $O(n \max\{\log\log n, \frac{\log n}{\log(np)}\})$ transmissions. Il montre de plus que cette borne est optimale dans le sens où tout algorithme dans lequel chaque nœud établit la communication aléatoirement avec l'un de ses voisins et terminant en temps $\Theta(\log n)$ effectue au moins $\Omega(n \max\{\log\log n, \frac{\log n}{\log(np)}\})$ communications.

D'après [ES08], en rajoutant de la mémoire aux nœuds, c'est-à-dire si on leur permet de se rappeler des 3 derniers nœuds appelés ou bien si on leur permet d'appeler directement 4 voisins distincts, alors il existe un algorithme de type PushPull qui permet d'effectuer $O(n \log\log n)$ transmissions. De plus, [ES08] montre qu'un algorithme où chaque nœud établit aléatoirement les connections parmi tous ses voisins qu'il n'a pas déjà contactés nécessite au plus $O(n \log\log n)$ transmissions.

De manière similaire, dans [BEF08], les auteurs ont montré qu'un algorithme qui ne distingue pas les nœuds nécessite $\Omega(\frac{n \log n}{\log d})$ transmissions pour diffuser le message en $O(\log n)$ étapes dans un graphe aléatoire d-régulier. Ces mêmes auteurs décrivent également un algorithme proche de celui PushPull où chaque nœud ouvre 4 communications vers des voisins distincts. Cet algorithme s'exécute en temps $O(\log n)$ et nécessite $\Omega(n \log\log n)$ transmissions pour $\delta \leq d \leq \delta \log n$ où δ est une constante suffisamment grande.

Des études similaires ont été menées sur les protocoles QuasiRandom en considérant la quantité de messages transmis, comme l'a fait [KSSV00] dans le cas aléatoire. Dans [BES10], il est montré qu'avec un algorithme PushPull il est possible d'informer tous les nœuds en $O(\log n)$ étapes et en n'utilisant que $O(n \log\log n)$ transmissions dans les graphes aléatoires $\mathcal{G}_{n,p}$ (pour $\frac{\log^2 n}{n} \leq p \leq 2^{o(\sqrt{\log n})}$). Ce résultat est à comparer avec celui de [Els06] qui stipule que, dans le modèle de communication Push, un protocole de diffusion aléatoire Random nécessite au moins $\Omega(\frac{n \log n}{\log(np)})$ étapes. Sur l'hypercube, les auteurs de [BES10] obtiennent $O(n(\log\log n)^2)$ transmissions. Ces résultats rappellent ceux du cas aléatoire auquel est ajoutée une mémoire des appels précédents, ce qui n'est pas si surprenant puisque que dans le cas QuasiRandom, un nœud n'appelle pas deux fois un même voisin.

Dans la mesure où la capacité des réseaux réels est limitée, il est naturel de s'intéresser à la quantité d'information transmise sur le réseau par un protocole de diffusion. Nous nous sommes intéressés au pire temps de la diffusion aléatoire, il peut être intéressant de connaître la quantité totale d'information envoyée. Afin d'économiser les ressources, dans le cadre de l'inondation dans les réseaux dynamiques, nous avons introduit l'*activité* d'un nœud afin de rendre l'inondation parcimonieuse. Une autre possibilité, suggérée par les études mentionnées précédemment, serait d'autoriser les nœuds à ne pas transmettre l'information.

5.2.2 Autres tâches

Dans cette thèse, nous ne nous sommes intéressés qu'à la diffusion d'une seule information. Nous savons qu'en termes de temps, sous l'hypothèse que les nœuds puissent envoyer toutes les informations en leur possession en une seule communication, les résultats sur la diffusion se transposent (à une constante près) à l'échange

total, où chaque nœud est la source d'un message. Il est alors naturel de se demander ce qui se passerait dans le cas plus réaliste où les liens ont une capacité bornée. Il pourrait aussi être intéressant d'étudier le cas très général où plusieurs nœuds ont plusieurs informations à transmettre au cours du temps, comme cela peut être le cas dans des réseaux distribués tels que les réseaux de capteurs ou dans réseaux pair-à-pair. Dans [FG10], les auteurs ont considéré un protocole PushPull ne distinguant pas les nœuds, et où les nœuds communiquent avant de transmettre un message. Ils ont considéré la quantité précise d'information transmise mesurée en bit. Leur protocole diffuse chaque message dans une clique de taille n en $O(\log n)$ étapes en utilisant au total $O(n(b + \log\log n \log b))$ bits par message, où b est la taille en bits du message. L'étude d'un tel protocole de diffusion 1-Port, où les nœuds négocient l'envoi de l'information, sur d'autres graphes ou, dans le contexte de la diffusion spatiale, sur d'autres métriques, peut se révéler fondamentale pour une application à des réseaux dont la capacité est faible ou sur lesquels il y a beaucoup de messages à transmettre.

Lorsqu'il y a plusieurs messages à transmettre, une autre manière d'économiser de la bande passante, qui fourni un axe d'étude potentiel, est de combiner les messages en faisant du codage de réseaux (*network coding*) tel qu'étudié dans [DMC06, ABCHL11, Hae11]. Nous pouvons nous demander comment le codage de réseaux peut s'appliquer à des réseaux dynamiques ou à la diffusion spatiale.

5.2.3 Autres contraintes

Le protocole d'inondation semble parfaitement adapté aux réseaux radios, dans la mesure où les ondes radios se propagent dans toutes les directions. Dans ce cas, il peut toutefois survenir des interférences. Ces interférences peuvent êtes négligées si les nœuds communiquent sur des fréquences différentes, ou tout du moins si tous les voisins de chaque nœud communiquent sur des fréquences différentes. En revanche, si tous les nœuds communiquent sur les mêmes fréquences, alors il est nécessaire de rajouter un mécanisme régulant les envois de chaque nœud informé afin de limiter les interférences [BYGI92]. Les auteurs de [CMPS09] se sont intéressés à l'inondation avec interférence dans un réseau dynamique modélisé par une suite de graphes aléatoires d'Erdős-Rényi indépendants. Nous pouvons nous intéresser au cas plus général des réseaux à évolution arête-markovienne. Nous pouvons également considérer les interférences dans le cas des protocoles de diffusion 1-Port.

Dans cette thèse, nous n'avons considéré que des protocoles dans lesquels les communications se limitent à l'information à transmettre, les nœuds ne se transmettant pas d'information concernant leur voisinage. Ces protocoles ont l'avantage de nécessiter moins de ressources que des protocoles où les nœuds communiquent leur connaissance du réseau afin de pouvoir ensuite calculer un schéma de diffusion optimal. En particulier, après une défaillance, un nœud est plus rapidement actif. Une autre piste de recherche, dans le cadre général de la diffusion d'information, serait l'étude de protocoles ayant la possibilité de s'échanger des informations sur la structure du réseau, comme c'est le cas dans les protocoles de routage à état de lien

Open Shortest Path First (OSPF) et *Intermediate System to Intermediate System* (IS-IS). Nous pouvons alors nous demander quel est l'impact sur le temps de diffusion de la quantité d'information dont chaque nœud dispose à propos du réseau.

Bibliographie

[AAG87] Y. Afek, B. Awerbuch et E. Gafni : Applying static network protocols to dynamic networks. Dans les actes de : *28th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 358–370, 1987.

[ABCHL11] C. Avin, M. Borokhovich, K. Censor-Hillel et Z. Lotker : Order optimal information spreading using algebraic gossip. Dans les actes de : *30th annual ACM SIGACT-SIGOPS symposium on Principles of Distributed Computing (PODC)*, pages 363–372, 2011.

[ACL01] W. Aiello, F. Chung et L. Lu : A random graph model for power law graphs. *Experimental Mathematics*, 10:53–66, 2001.

[ADHP09] S. Angelopoulos, B. Doerr, A. Huber et K. Panagiotou : Tight bounds for quasirandom rumor spreading. *The Electronic Journal of Combinatorics*, 16(1), 2009.

[AG06] I. Abraham et C. Gavoille : Object location using path separators. Dans les actes de : *25th annual ACM symposium on Principles of distributed computing (PODC)*, pages 188–197, 2006.

[AHK94] R. Ahlswede, H. S. Haroutunian et L. H. Khachatrian : *Messy Broadcasting in Networks*. 1994.

[AKL08] C. Avin, M. Koucký et Z. Lotker : How to explore a fast-changing world (cover time of a simple random walk on evolving graphs). Dans les actes de : *35th International Colloquium on Automata, Languages and Programming (ICALP)*, volume 5125, 2008.

[APRU12] J. Augustine, G. Pandurangan, P. Robinson et E. Upfal : Towards robust and efficient computation in dynamic peer-to-peer networks. Dans les actes de : *23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 551–569, 2012.

[AS00] N. Alon et J. H. Spencer : *The probabilistic method*. 2000.

[BCF11] H. Baumann, P. Crescenzi et P. Fraigniaud : Parsimonious flooding in dynamic graphs. *Distributed Computing*, 24:31–44, 2011.

[BEF08] P. Berenbrink, R. Elsässer et T. Friedetzky : Efficient randomised broadcasting in random regular networks with applications in peer-to-peer systems. Dans les actes de : *27th ACM symposium on Principles of Distributed Computing (PODC)*, pages 155–164, 2008.

[BES10] P. Berenbrink, R. Elsässer et T. Sauerwald : Communication complexity of quasirandom rumor spreading. Dans les actes de : *18th Annual European Symposium (ESA)*, pages 134–145, 2010.

[BF03] S. Bhadra et A. Ferreira : Complexity of connected components in evolving graphs and the computation of multicast trees in dynamic networks. Dans les actes de : *Second International Conference on Ad-Hoc, Mobile, and Wireless Networks (ADHOC-NOW)*, volume 2865 de *Lecture Notes in Computer Science*, pages 259–270, 2003.

[BF10] H. Baumann et P. Fraigniaud : Sub-linear universal spatial gossip protocols. Dans les actes de : *16th International Colloquium Structural Information and Communication Complexity (SIROCCO)*, volume 5869 de *Lecture Notes in Computer Science*, pages 44–56, 2010.

[BFHdV12] H. Baumann, P. Fraigniaud, H. A. Harutyunyan et R. de Verclos : The worst case behavior of randomized gossip. Dans les actes de : *9th annual conference on Theory and Applications of Models of Computation (TAMC)*, pages 330–345, 2012.

[BFKK01] L. Barrière, P. Fraigniaud, E. Kranakis et D. Krizanc : Efficient routing in networks with long range contacts. Dans les actes de : *15th International Conference on Distributed Computing (DISC)*, volume 2180 de *Lecture Notes in Computer Science*, pages 270–284, 2001.

[BGPS06] S. Boyd, A. Ghosh, B. Prabhakar et D. Shah : Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.

[BNGNS00] A. Bar-Noy, S. Guha, J. S. Naor et B. Schieber : Message multicasting in heterogeneous networks. *SIAM Journal on Computing*, 30(2):347–358, 2000.

[Bol01] B. Bollobás : *Random Graphs.* 2001.

[BYGI92] R. Bar-Yehuda, O. Goldreich et A. Itai : On the time-complexity of broadcast in multi-hop radio networks : An exponential gap between determinism and randomization. *Journal of Computer and System Sciences*, 45(1):104–126, août 1992.

[CHS10] K. Censor-Hillel et H. Shachnai : Partial information spreading with application to distributed maximum coverage. Dans les actes de : *29th Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 161–170, 2010.

[CL01] F. Chung et L. Lu : The diameter of sparse random graphs. *Advances in Applied Mathematics*, 26(4):257–279, 2001.

[CL02] F. Chung et L. Lu : Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6:125–145, 2002.

[CL03] F. R. K. Chung et L. Lu : The average distance in a random graph with given expected degrees. *Internet Mathematics*, 1(1):91–113, 2003.

[CL06a] F. Chung et L. Lu : *Complex Graphs and Networks*. 2006.

[CL06b] F. Chung et L. Lu : The volume of the giant component of a random graph with given expected degrees. *SIAM Journal on Discrete Mathematics*, 20(2):395–411, 2006.

[CL07] N. B. Chang et M. Liu : Controlled flooding search in a large network. *IEEE/ACM Transactions on Networking*, 15(2):436–449, 2007.

[CLF+07] D. Chakrabarti, J. Leskovec, C. Faloutsos, S. Madden, C. Guestrin et M. Faloutsos : Information survival threshold in sensor and P2P networks. Dans les actes de : *26th IEEE International Conference on Computer Communications (INFOCOM)*, pages 1316–1324, 2007.

[CLP10a] F. Chierichetti, S. Lattanzi et A. Panconesi : Almost tight bounds for rumour spreading with conductance. Dans les actes de : *42nd ACM Symposium on Theory of Computing (STOC)*, pages 399–408, 2010.

[CLP10b] F. Chierichetti, S. Lattanzi et A. Panconesi : Rumour spreading and graph conductance. Dans les actes de : *21rst Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1657–1663, 2010.

[CMM+10] A. E. F. Clementi, C. Macci, A. Monti, F. Pasquale et R. Silvestri : Flooding time of edge-Markovian evolving graphs. *SIAM Journal on Discrete Mathematics*, 24(4):1694–1712, 2010.

[CMPC04] P. Costa, M. Migliavacca, G. P. Picco et G. Cugola : Epidemic algorithms for reliable content-based publish-subscribe : An evaluation. Dans les actes de : *24th International Conference on Distributed Computing Systems (ICDCS)*, pages 552–561, 2004.

[CMPS09] A. E. F. Clementi, A. Monti, F. Pasquale et R. Silvestri : Broadcasting in dynamic radio networks. *Journal of Computer and System Sciences*, 75(4):213–230, 2009.

[CMPS11] A. E. F. Clementi, A. Monti, F. Pasquale et R. Silvestri : Information spreading in stationary Markovian evolving graphs. *IEEE Transactions on Parallel and Distributed Systems*, 22(9), 2011.

[CPS09] A. E. F. Clementi, F. Pasquale et R. Silvestri : MANETS : High mobility can make up for low transmission power. Dans les actes de : *36th International Colloquium on Automata, Languages and Programming (ICALP)*, volume 5556 de *Lecture Notes in Computer Science*, pages 387–398, 2009.

[CR06] A. Czumaj et W. Rytter : Broadcasting algorithms in radio networks with unknown topology. *Journal of Algorithms*, 60(2):115–143, 2006.

[CS11] A. E. F. Clementi et R. Silvestri : Parsimonious flooding in geometric random-walks. Dans les actes de : *25th international conference on Distributed computing (DISC)*, pages 298–310, 2011.

[DFKS11] B. Doerr, T. Friedrich, M. Künnemann et T. Sauerwald : Quasi-random rumor spreading : An experimental analysis. *ACM Journal of Experimental Algorithmics*, 16, 2011.

[DFS08] B. Doerr, T. Friedrich et T. Sauerwald : Quasirandom rumor spreading. Dans les actes de : *19th annual ACM-SIAM symposium on Discrete algorithms (SODA)*, pages 773–781, 2008.

[DFS09] B. Doerr, T. Friedrich et T. Sauerwald : Quasirandom rumor spreading : Expanders, push vs. pull, and robustness. Dans les actes de : *36th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 366–377, 2009.

[DGH+88] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart et D. Terry : Epidemic algorithms for replicated database maintenance. *ACM Operating Systems Review, SIGOPS*, 22(1):8–32, 1988.

[DGM08] M. Draief, A. Ganesh et L. Massoulié : Thresholds for virus spread on networks. *The Annals of Applied Probability*, 18(2):359–378, 2008.

[DHLS06] P. Duchon, N. Hanusse, E. Lebhar et N. Schabanel : Could any graph be turned into a small-world ? *Theoretical Computer Science*, 355(1):96–103, 2006.

[DMC06] S. Deb, M. Médard et C. Choute : Algebraic gossip : a network coding approach to optimal multiple rumor mongering. *IEEE Transactions on Information Theory*, 52(6):2486–2507, 2006.

[EK03] M. Elkin et G. Kortsarz : Sublogarithmic approximation for telephone multicast : path out of jungle. Dans les actes de : *14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 76–85, 2003.

[EK05] M. Elkin et G. Kortsarz : A combinatorial logarithmic approximation algorithm for the directed telephone broadcast problem. *SIAM Journal on Computing*, 35(3):672–689, 2005.

[Els06] R. Elsässer : On the communication complexity of randomized broadcasting in random-like graphs. Dans les actes de : *18th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, pages 148–157, 2006.

[ER59] P. Erdős et A. Rényi : On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–291, 1959.

[ES08] R. Elsässer et T. Sauerwald : The power of memory in randomized broadcasting. Dans les actes de : *19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 218–227, ACM/SIAM, 2008.

[ES09] R. Elsässer et T. Sauerwald : On the runtime and robustness of randomized broadcasting. *Theoretical Computer Science*, 410(36):3414–3427, 2009.

[Fei95a] U. Feige : A tight lower bound on the cover time for random walks on graphs. *Random Structures and Algorithms*, 6(4):433–438, 1995.

[Fei95b] U. Feige : A tight upper bound on the cover time for random walks on graphs. *Random Structures and Algorithms*, 6(1):51–54, 1995.

[Fer02] A. Ferreira : On models and algorithms for dynamic communication networks : The case for evolving graphs. Dans les actes de : *4e Rencontres francophones sur les aspects algorithmiques des télécommunications (ALGOTEL)*, pages 155–161, 2002.

[FG85] A. M. Frieze et G. R. Grimmett : The shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics and Combinatorial Operations Research and Computer Science*, 10:57–77, 1985.

[FG10] P. Fraigniaud et G. Giakkoupis : On the bit communication complexity of randomized rumor spreading. Dans les actes de : *22nd ACM symposium on Parallelism in algorithms and architectures (SPAA)*, pages 134–143, 2010.

[FGK+09] P. Fraigniaud, C. Gavoille, A. Kosowski, E. Lebhar et Z. Lotker : Universal augmentation schemes for network navigability. *Theoretical Computer Science*, 410(21–23):1970–1981, 2009.

[FGKM10] D. Frey, R. Guerraoui, A.-M. Kermarrec et M. Monod : Boosting gossip for live streaming. Dans les actes de : *10th IEEE International Conference on Peer-to-Peer Computing (P2P)*, pages 1–10, 2010.

[FGP06] P. Fraigniaud, C. Gavoille et C. Paul : Eclecticism shrinks even small worlds. *Distributed Computing*, 18:279–291, 2006.

[FH09] N. Fountoulakis et A. Huber : Quasirandom rumor spreading on the complete graph is as fast as randomized rumor spreading. *SIAM Journal on Discrete Mathematics*, 23(4):1964–1991, 2009.

[FL94] P. Fraigniaud et E. Lazard : Methods and problems of communication in usual networks. *Discrete Applied Mathematics*, 53(1-3):79–133, 1994.

[FLL10] P. Fraigniaud, E. Lebhar et Z. Lotker : A lower bound for network navigability. *SIAM Journal on Discrete Mathematics*, 24(1):72–81, 2010.

[FPRU90] U. Feige, D. Peleg, P. Raghavan et E. Upfal : Randomized broadcast in networks. *Random Structures & Algorithms*, 1(4):447–460, 1990.

[FPS12] N. Fountoulakis, K. Panagiotou et T. Sauerwald : Ultra-fast rumor spreading in social networks. Dans les actes de : *23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1642–1660, 2012.

[Fra05] P. Fraigniaud : Greedy routing in tree-decomposed graphs. Dans les actes de : *13th Annual European Symposium on Algorithms (ESA)*, volume 3669 de *Lecture Notes in Computer Science*, pages 791–802, 2005.

[Gia11] G. Giakkoupis : Tight bounds for rumor spreading in graphs of a given conductance. Dans les actes de : *28th International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 9 de *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 57–68, 2011.

[Gil59] E. N. Gilbert : Random graphs. *Annals of Mathematical Statistics*, 30:1141–1144, 1959.

[GJ79] M. R. Garey et D. S. Johnson : *Computers and Intractability.* 1979.

[GKG06] I. Gupta, A.-M. Kermarrec et A. J. Ganesh : Efficient and adaptive epidemic-style protocols for reliable and scalable multicast. *IEEE Transactions on Parallel and Distributed Systems*, 17(7):593–605, 2006.

[Gnu] Gnutella Protocol Specification 0.6 : http ://rfc-gnutella.sourceforge.net.

[GT02] M. Grossglauser et D. N. C. Tse : Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Transactions on Networking*, 10(4):477–486, 2002.

[Hae11] B. Haeupler : Analyzing network coding gossip made easy. Dans les actes de : *43rd ACM Symposium on Theory of Computing (STOC)*, pages 293–302, 2011.

[HG97] F. Harary et G. Gupta : Dynamic graph models. *Mathematical and Computer Modelling*, 25(7):79–87, 1997.

[HH02] T. E. Hart et H. A. Harutyunyan : Improved messy broadcasting in hypercubes and complete bipartite graphs. *Congressus Numerantium*, 156:181–195, 2002.

[HHL88] S. M. Hedetniemi, S. T. Hedetniemi et A. L. Liestman : A survey of gossiping and broadcasting in communication networks. *Networks*, 18(4):319–349, 1988.

[HHL11] H. A. Harutyunyan, P. Hell et A. L. Liestman : Messy broadcasting - decentralized broadcast schemes with limited knowledge. *Discrete Applied Mathematics*, 159(5):322–327, 2011.

[HKMP96] J. Hromkovič, R. Klasing, B. Monien et R. Peine : *Dissemination of Information in Interconnection Networks (Broadcasting & Gossiping).* 1996.

[HL98] H. A. Harutyunyan et A. L. Liestman : Messy broadcasting. *Parallel Processing Letters*, 8(2):149–159, 1998.

[JL04] A. Jarry et Z. Lotker : Connectivity in evolving graph with geometric properties. Dans les actes de : *2004 joint workshop on Foundations of mobile computing (DIALM-POMC)*, pages 24–30, 2004.

[JLR00] S. Janson, T. Luczak et A. Rucinski : *Random Graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. 2000.

[JM95] K. Jansen et H. Müller : The minimum broadcast time problem for several processor networks. *Theoretical Computer Science*, 147(1–2):69–85, 1995.

[JRS95] A. Jakoby, R. Reischuk et C. Schindelhauer : The complexity of broadcasting in planar and decomposable graphs. Dans les actes de : *14th International Workshop on Graph-Theoretic Concepts in Computer Science (WG)*, 1995.

[KCV92] D. W. Krumme, G. Cybenko et K. N. Venkataraman : Gossiping in minimal time. *SIAM Journal on Computing*, 21(1):111–139, 1992.

[KDG03] D. Kempe, A. Dobra et J. Gehrke : Gossip-based computation of aggregate information. Dans les actes de : *44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 482–491, 2003.

[KK02] D. Kempe et J. Kleinberg : Protocols and impossibility results for gossip-based communication mechanisms. Dans les actes de : *43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 471–480, 2002.

[KKD04] D. Kempe, J. Kleinberg et A. Demers : Spatial gossip and resource location protocols. *J. ACM*, 51(6):943–967, 2004.

[Kle00] J. Kleinberg : The small-world phenomenon : an algorithm perspective. Dans les actes de : *Thirty Second Annual ACM Symposium on Theory of Computing (STOC)*, pages 163–170, 2000.

[KLO10] F. Kuhn, N. Lynch et R. Oshman : Distributed computation in dynamic networks. Dans les actes de : *42nd ACM Symposium on Theory of Computing (STOC)*, pages 513–522, New York, NY, USA, 2010.

[KM98] E. Kushilevitz et Y. Mansour : An $\omega(d \log(n/d))$ lower bound for broadcast in radio networks. *SIAM Journal on Computing*, 27(3):702–712, 1998.

[KP95] G. Kortsarz et D. Peleg : Approximation algorithms for minimum-time broadcast. *SIAM Journal on Discrete Mathematics*, 8(3):401–427, 1995.

[KSSV00] R. Karp, C. Schindelhauer, S. Shenker et B. Vocking : Randomized rumor spreading. Dans les actes de : *41st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 565–574, 2000.

[KSW05a] F. Kuhn, S. Schmid et R. Wattenhofer : *A Self-repairing Peer-to-Peer System Resilient to Dynamic Adversarial Churn*, volume 3640 de *Lecture Notes in Computer Science*. 2005.

[KSW05b] F. Kuhn, S. Schmid et R. Wattenhofer : *A Self-repairing Peer-to-Peer System Resilient to Dynamic Adversarial Churn*, volume 3640 de *Lecture Notes in Computer Science*. 2005.

[LCC+02] Q. Lv, P. Cao, E. Cohen, K. Li et S. Shenker : Search and replication in unstructured peer-to-peer networks. Dans les actes de : *16th international conference on Supercomputing (ICS)*, pages 84–95, 2002.

[LEH03] J. Luo, P. Eugster et J.-P. Hubaux : Route driven gossip : probabilistic reliable multicast in ad hoc networks. Dans les actes de : *22nd Annual Joint Conference of the IEEE Computer and Communications (INFOCOM)*, volume 3, pages 2229–2239, 2003.

[LNNK+05] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins et R. L. Graham : Geographic routing in social networks. *National Academy of Sciences of the United States of America*, 102(33):11623–11628, 2005.

[Lu01] L. Lu : The diameter of random massive graphs. Dans les actes de : *12th ACM-SIAM Annual Symposium on Discrete Algorithms (SODA)*, pages 912–921, 2001.

[MBR03] G. S. Manku, M. Bawa et P. Raghavan : Symphony : Distributed hashing in a small world. Dans les actes de : *USENIX Symposium on Internet Technologies and Systems (USITS)*, 2003.

[Mid93] M. Middendorf : Minimum broadcast time is np-complete for 3-regular planar graphs and deadline 2. *Information Processing Letters*, 46(6):281–287, 1993.

[Mil67] S. Milgram : The small world problem. *Psychology Today*, 1(1):60–67, 1967.

[MN04] C. U. Martel et V. Nguyen : Analyzing Kleinberg's (and other) small-world models. Dans les actes de : *23rd Annual ACM Symposium on the Principles of Distributed Computing (PODC)*, pages 179–188, 2004.

[MNR02] D. Malkhi, M. Naor et D. Ratajczak : Viceroy : a scalable and dynamic emulation of the butterfly. Dans les actes de : *21st annual symposium on Principles of distributed computing (PODC)*, pages 183–192, 2002.

[MRS+98] M. V. Marathe, R. Ravi, R. Sundaram, S. Ravi, D. J. Rosenkrantz et H. B. H. III : Bicriteria network design problems. *Journal of Algorithms*, 28(1):142–171, 1998.

[NM05] V. Nguyen et C. Martel : Analyzing and characterizing small-world graphs. Dans les actes de : *16th annual ACM-SIAM symposium on Discrete algorithms (SODA)*, pages 311–320, 2005.

[Ola91] S. Olariu : An optimal greedy heuristic to color interval graphs. *Information Processing Letters*, 37(1):21–25, 1991.

[Pit87] B. Pittel : On spreading a rumor. *SIAM Journal on Applied Mathematics*, 47(1):213–223, 1987.

[PRR99] C. G. Plaxton, R. Rajaraman et A. W. Richa : Accessing nearby copies of replicated objects in a distributed environment. *Theory of Computing Systems*, 32(3):241–280, 1999.

[Rav94] R. Ravi : Rapid rumor ramification : approximating the minimum broadcast time. Dans les actes de : *35th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 202–213, 1994.

[Sau07] T. Sauerwald : On mixing and edge expansion properties in randomized broadcasting. Dans les actes de : *18th International Symposium on Algorithms and Computation (ISAAC)*, volume 4835 de *Lecture Notes in Computer Science*, pages 196–207, 2007.

[SCH81] P. J. Slater, E. J. Cockayne et S. T. Hedetniemi : Information dissemination in trees. *SIAM Journal on Computing*, 10(4):692–701, 1981.

[SD12] A. D. Sarwate et A. G. Dimakis : The impact of mobility on gossip algorithms. *IEEE Transactions on Information Theory*, 58(3):1731–1742, 2012.

[Sli07] A. Slivkins : Distance estimation and object location via rings of neighbors. *Distributed Computing*, 19:313–333, 2007.

[SMZ03] K. Sripanidkulchai, B. Maggs et H. Zhang : Efficient content location using interest-based locality in peer-to-peer systems. Dans les actes de : *22nd Annual Joint Conference of the IEEE Computer and Communications (INFOCOM)*, volume 3, pages 2166–2176, 2003.

[SS11] T. Sauerwald et A. Stauffer : Rumor spreading and vertex expansion on regular graphs. Dans les actes de : *22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 462–475, 2011.

[WS98] D. J. Watts et S. H. Strogatz : Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[ZGG04] H. Zhang, A. Goel et R. Govindan : Using the small-world model to improve freenet performance. *Computer Network*, 46(4):555–574, 2004.