HAL
open science

Loi de puissance et caractérisation des réseaux dynamiques

Eric Fleury, Jean-Loup Guillaume, Céline Robardet, Antoine Scherrer

To cite this version:

Eric Fleury, Jean-Loup Guillaume, Céline Robardet, Antoine Scherrer. Loi de puissance et caractérisation des réseaux dynamiques. 9ème Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2007, Ile d'Oléron, France. pp.17-20. inria-00176964

HAL Id: inria-00176964

https://inria.hal.science/inria-00176964

Submitted on 5 Oct 2007

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.

Loi de puissance et caractérisation des réseaux dynamiques

E. Fleury[1], J.-L. Guillaume[1], C. Robardet[2] et A. Scherrer[3]

[1] *INRIA ARES - CITI, INSA de Lyon, 69621 Villeurbanne, France*
[2] *LIRIS / CNRS UMR 5205, INSA de Lyon, 69621 Villeurbanne, France*
[3] *INRIA COMPSYS - CITI, INSA de Lyon, 69621 Villeurbanne, France*

Caractériser la dynamique et l'évolution au cours du temps d'un réseau sans fil multi sauts (réseaux *ad hoc*, réseaux de senseurs) apparaît comme un problème important car cela doit permettre de comprendre, de modéliser et de développer des algorithmes et des protocoles adaptés. À partir des traces expérimentales collectées lors de la conférence Infocom 2005 [1], nous proposons une analyse plus avancée de la structure évolutive de ce genre de réseaux. Plus précisément, nous montrons que la seule caractérisation au travers d'une loi de puissance des contacts et inter-contacts entre individus/nœuds n'est pas suffisante pour capturer et rendre l'évolution du réseau. À partir d'une analyse combinant trois approches [2] qui se révèlent très complémentaires (théorie des graphes, processus aléatoires et fouille de données), nous proposons un modèle simple mettant en évidence la complexité de la structure évolutive.

Mots clefs : réseaux dynamiques, mobilité, modélisation

1 Introduction

Il existe un très grand nombre de travaux à l'heure actuelle sur les réseaux radio multi sauts qui couvrent les réseaux ad hoc, les réseaux de capteurs, les DTN (*Delay Tolerant Network*), les systèmes d'informations spontanés (SIS)... Une caractéristique commune à tous ces types de réseaux est que la connectivité entre leurs entités s'avère intermittente, sporadique et variable en performance et en fiabilité. Néanmoins, à l'inverse des réseaux d'infrastructure de communication (filaire pour le cœur de l'Internet et/ou sans fil pour le GSM/UMTS), la perte d'un lien de communication ne doit pas être considérée comme une défaillance du système mais bien comme un comportement normal. L'un des défis est donc de concevoir et de mettre en œuvre des protocoles et des applications (Livraison asynchrone de SMS, dissémination d'alertes) qui supportent la mobilité et donc des connexions intermittentes. Afin de modéliser ce type de réseau pour en dériver des bornes d'optimalité, des protocoles robustes ou des heuristiques optimisées, il est crucial de pouvoir caractériser finement la dynamique et l'évolution des connexions au cours du temps, en particulier il faut arriver à caractériser la durée des contacts (temps pendant lequel deux nœuds sont à porté radio) et inter-contacts (temps écoulé entre deux périodes de contact successives) entre les entités du réseau.

Des mesures récentes [1] donnant la durée des contacts ainsi que la durée entre deux contacts successifs ont été collectées lors de la conférence Infocom 2005. Des petits objets communicants (IMOTE), émettant périodiquement un message de présence et enregistrant les messages de présence des autres nœuds à portée radio, ont été distribués aux participants. Ceux-ci les ont gardés sur eux tout au long de la conférence[†]. À partir de ces données, l'analyse conduite dans [1] suggère de « *caractériser* » les durées des contacts et d'inter-contacts entre individus par une fonction de distribution cumulative qui suit une loi de puissance. Dans un article polémique, F. Keller [3] revisite un peu la popularité et l'omniprésence des lois de puissance et revient sur le fait qu'il n'est pas si surprenant d'en trouver « *un peu partout* ». La question que nous posons dans cet article est de savoir si cette simple caractérisation des durées est suffisante pour décrire l'évolution

[†] 41 conférenciers à Infocom, Hyatt Regency hotel à Miami, USA sur une durée de 254 152 secondes, soit environ 3 jours.

des réseaux dynamiques. À partir d'un modèle simple générant des réseaux dynamiques aléatoires ayant exactement les mêmes propriétés en terme de distribution des contacts nous montrons que cela n'est pas suffisant pour capturer la complexité intrinsèque liée à l'évolution des réseaux dynamiques. Cette différence est mise en lumière au travers notamment de l'étude des composantes connexes, mais nous développons aussi une analyse à partir d'arguments de processus stochastiques et de fouille de données pour mettre en lumière les éléments structurels qui ne sont pas pris en compte par la simple description des durées.

2 Méthodologie

Nous allons décrire brièvement les trois approches que nous avons employées pour étudier les réseaux dynamiques. La première approche étudie l'évolution de paramètres de graphes très classiques comme le degré, le diamètre... En plus de ces paramètres, nous avons aussi étudié l'évolution des composantes connexes (sous-graphes connexes maximaux : CC), en supposant que deux ensembles de sommets identiques connectés par des ensembles différents d'arêtes sont des composantes connexes distinctes. La densité et la structure des CC influencent directement la façon dont un réseau peut transmettre de l'information. C'est donc une mesure qui se révèle utile pour étudier et comparer des réseaux radio.

La seconde approche est basée sur la théorie des processus stochastiques. On considère alors certains paramètres du réseau dynamique (évolution du degré d'un nœud, durées des contacts entre une paire de sommets, nombre de CC) comme des séries temporelles que l'on cherche à caractériser à l'aide de processus stochastique, c'est-à-dire comme le résultat d'expériences aléatoires. Cette hypothèse n'est pas fondée pour les données réelles, mais de par la complexité globale du système, il apparaît un phénomène de *chaos déterministe* qui rend très difficile la mise en relation des données brutes collectées avec la physique du système. Nous allons ici employer des analyses stochastiques simples pour caractériser la variabilité des différentes séries temporelles au moyen de la fonction de distribution cumulative complémentaire, définie pour une variable aléatoire X par $P[X > x]$ (CCDF, plus discriminante que la PDF [4]).

La dernière approche se base sur des techniques de fouille de données qui permettent d'identifier tous les groupes d'individus qui se forment fréquemment, et de décrire les trajectoires des individus parmi ces groupes. L'identification des groupes se fait à l'aide d'un algorithme d'extraction de concepts formels qui, dans ce contexte, permet d'identifier tous les ensembles d'arêtes qui existent simultanément assez fréquemment. Les contraintes de taille minimum pour les ensembles d'arêtes et de fréquence minimum permettent de réduire l'espace de recherche et ainsi rendre l'énumération complète possible [5, 6]. Ce calcul est effectué au moyen de D-MINER [7].

3 Comparaison entre les données IMOTE et un modèle aléatoire

Afin de montrer que la seule caractérisation des durées des contacts et inter-contacts n'est pas suffisante pour capturer toute la complexité de l'évolution d'un réseau dynamique, nous avons défini un modèle de génération aléatoire très simple inspiré du modèle configurationnel [8]. Soit $d^+_{(u,v)}$ la distribution des durées des contacts pour le lien (u,v) et $d^+ = \cup_{\forall(u,v)} d^+_{(u,v)}$ (on définit de la même manière $d^-_{(u,v)}$ and d^- pour les distributions des durées inter-contacts). À partir de ces définitions, une séquence *on/off* pour un lien (u,v) générée par un modèle aléatoire pourra suivre soit une combinaison de $d^+_{(u,v)}$ et $d^-_{(u,v)}$ que nous appelons *modèle* LINK, soit une combinaison de d^+ and d^- que nous nommons *modèle* GLOBAL. Les distributions sont bien entendu extraites des données réelles. La figure 1(a) prouve que les 2 modèles aléatoires que nous avons définis collent parfaitement aux données réelles pour les distributions des contacts et inter-contacts (en terme de CCDF).

Néanmoins, si on regarde par exemple l'évolution du degré moyen (*i.e.* le degré moyen des nœuds vu comme une série temporelle), on s'aperçoit que les données n'ont plus du tout le même comportement, c'est-à-dire que les modèles ne reproduisent pas ces évolutions. La figure 1(b) illustre ce fait : la variabilité (représentée par la CCDF) est très différente, et ni la moyenne, ni la forme de la distribution ne sont reproduites par les modèles. Ceci met en lumière que les durées des contacts et inter-contacts n'ont qu'un faible impact

sur l'évolution et la variabilité des degrés et donc de la connectivité. La plus forte variabilité des données IMOTE vient notamment de l'existence de cycles humains (jour/nuit, repas, pauses) qui sont absents des données générées.

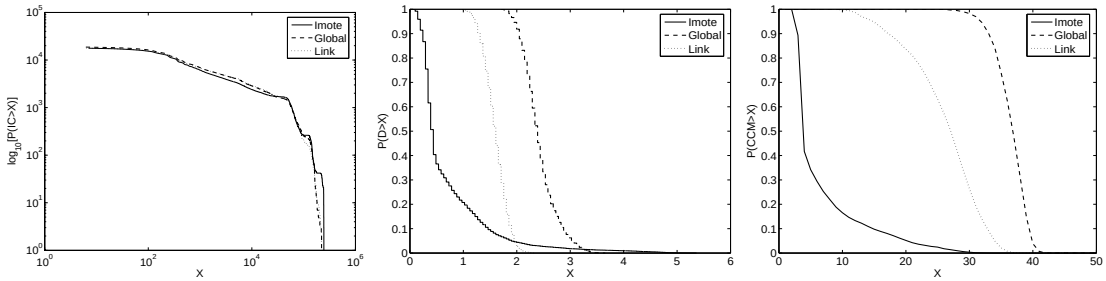

FIG. 1: (a-gauche) CCDF des durées inter-contacts pour les données IMOTE, et les modèles GLOBAL et LINK. (b-milieu) CCDF du degré moyen. (c-droite) CCDF de la taille maximum des CC.

En regardant maintenant les structures des CC, on s'aperçoit que là aussi les modèles générés n'arrivent pas à capturer certaines évolutions structurelles du réseau dynamique. Considérons tout d'abord la taille de la plus grande CC et son évolution au cours du temps. Comme l'illustre la figure 1(c), la CCDF est très différente entre le jeu de données et les deux modèles qui surestiment la taille maximum des CC. Cet aspect est encore plus marqué pour le modèle GLOBAL.

Plus précisément, concernant la distribution des tailles des CC, on retrouve aussi une grande disparité entre les données et les modèles ne prenant en compte que la distribution des durées. Sur la figure 2(a) qui concerne les données IMOTE, on observe une forte corrélation entre le nombre de sommets et le nombre de liens avec néanmoins une variation non constante dont le coefficient est proche de 4.5, *i.e.* pour un nombre donné de sommets on peut s'attendre à avoir une variation de la densité des CC de l'ordre de 4.5. On ne retrouve pas cette forte variation pour les modèles (figures 2(b) et (c)) qui offrent un nombre de liens globalement linéaire en fonction du nombre de sommets pour les CC qui s'apparentent donc à des arbres. Il faut noter que le trou pour le modèle GLOBAL est dû à la courte durée de l'expérience et se trouve comblé en générant des graphes dynamiques plus longs, pour lesquels on retrouve ainsi des composantes de taille moyenne. Mais la forme de la courbe reste inchangée et la probabilité d'apparition de CC denses reste très faible pour les deux modèles.

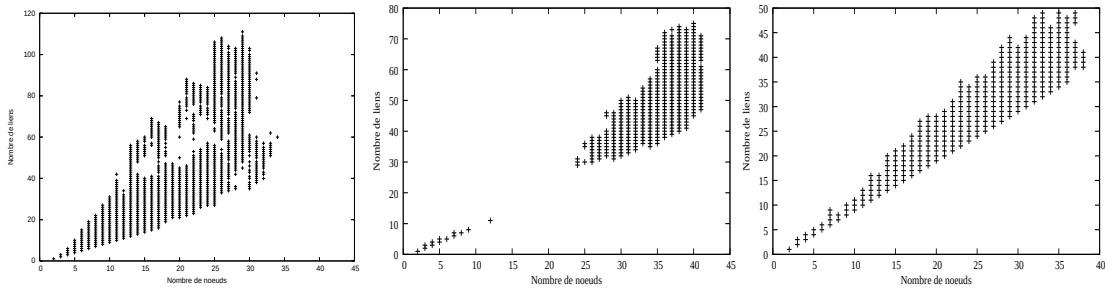

FIG. 2: Nombre de liens en fonction du nombre de sommets pour toutes les CC. (a-gauche) Données IMOTE (b-milieu) modèle GLOBAL, (c-droite) modèle LINK

Notre dernier angle d'analyse est basé sur l'utilisation de techniques de fouille de données permettant d'identifier des groupes d'individus bien établis et de décrire la dynamique des individus entre ceux-ci. Dans un premier temps, on calcule tous les groupes d'au moins 5 arêtes connexes présentes simultanément pendant au moins 10 pas de temps (1 pas de temps = 120s). On sélectionne ensuite ceux qui sont suffisamment dense (supérieur à 0.8), car, comme le montre la figure 3(b), la majorité de ces sous-graphes ne couvrent que peu d'individus et ont une densité faible. Lors d'une seconde phase, on fusionne les groupes

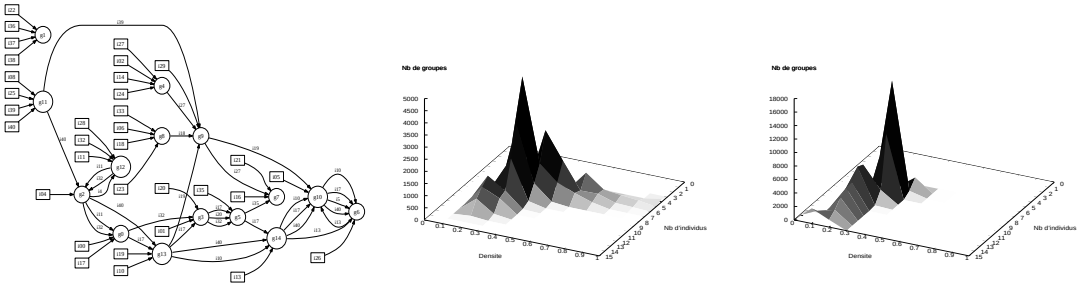

FIG. 3: (a-gauche) Trajectoires ordonnées dans le temps des individus entre les groupes sociaux pour les données IMOTE. Nombre de CC fréquentes en fonction de leur densité et du nombre d'individus : (b-milieu) pour les données IMOTE et (c-droite) pour le modèle LINK.

qui concernent les mêmes individus et des pas de temps similaires. On obtient ainsi des groupes « *sociaux* ». La dernière étape consiste à reconstruire les trajectoires des individus entre ces groupes. La figure 3(a) représente les trajectoires des individus (les arêtes du graphe étiquetées par les individus) parmi les groupes (les sommets). Les rectangles indiquent l'affectation initiale des individus. Cette représentation montre bien que la méthode permet d'identifier des groupes distincts (à l'exception des groupes 10 et 6) tout en étant fortement inter-connectés. Si l'on applique cette approche sur les données générées par les modèles aléatoires, on ne retrouve pas la même structure car les ensembles fréquents d'arêtes sont très peu denses (voir figure 3(c)). Ceci montre encore une fois que les distributions des durées des (inter)-contacts ne permettent de capturer ni la complexité de sous-structures existantes dans les données réelles ni leur dynamique.

4 Conclusion

Nous avons introduit une mesure basée sur les composantes connexes permettant de capturer des propriétés structurelles fortes des réseaux dynamiques. Nous avons aussi proposé deux variations d'un modèle très simple pour les réseaux dynamiques montrant que la seule prise en compte des durées des contacts et inter-contacts est loin d'être suffisante pour capturer tous les paramètres d'évolution des réseaux dynamiques. L'approche proposée qui combine différents points de vue complémentaires (théorie des graphes, analyse stochastique et fouille de données) nous semble aussi très importante : nous pensons en effet qu'une telle approche multi disciplinaire est fructueuse et peut même être étendue à d'autres domaines d'étude qui peuvent tirer profit des réseaux dynamiques. Pour finir, nous souhaitons élargir cette étude à des réseaux plus grands et sur une période plus longue via une expérimentation *in situ* au sein de l'INSA de Lyon[‡]

Références

[1] A. Chaintreau, J. Crowcroft, C. Diot, R. Gass, P. Hui, and J. Scott. Impact of human mobility on the design of opportunistic forwarding algorithms. In *INFOCOM 2006*, 2006.

[2] Éric Fleury, Jean-Loup Guillaume, Céline Robardet, and Antoine Scherrer. Analysis of dynamic sensor networks : power law then what ? In *COMSWARE*, Bangalore, India, January 2007.

[3] Keller Evelyn Fox. Revisiting "scale-free" networks. *BioEssays*, 27(10) :1060–1068, 2005.

[4] Lun Li, David Alderson, Reiko Tanaka, John C. Doyle, and Walter Willinger. Towards a theory of scale-free graphs : Definition, properties, and implications. *Internet Mathematics*, 2(4) :431–523, 2005.

[5] Rakesh Agrawal, Tomasz Imielinski, and Arun Swami. Mining association rules between sets of items in large databases. In *Proceedings ACM SIGMOD'93*, pages 207–216. ACM Press, 1993.

[6] H. Mannila and H. Toivonen. Levelwise search and borders of theories in knowledge discovery. *Data Mining and Knowledge Discovery journal*, 1(3) :241–258, 1997.

[7] Jéremy Besson, Céline Robardet, Jean-François Boulicaut, and Sophie Rome. Constraint-based concept mining and its application to microarray data analysis. *Intelligent Data Analysis*, 9(1) :59–82, 2005.

[8] B. Bollobas. *Random graphs*. Academic Press, 1985.

[‡] 250 étudiants sur une période de 3 à 4 semaines.