



**HAL**  
open science

## Analyse de la dynamique de réseaux de contacts

Thibault Rieutord

► **To cite this version:**

Thibault Rieutord. Analyse de la dynamique de réseaux de contacts. Cryptographie et sécurité [cs.CR]. 2012. dumas-00734459

**HAL Id: dumas-00734459**

**<https://dumas.ccsd.cnrs.fr/dumas-00734459v1>**

Submitted on 21 Sep 2012

**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.

# Rapport de stage : Analyse de la dynamique de réseaux de contacts

Thibault Rieutord

ENS Cachan - Antenne de Bretagne \ Université de Rennes 1 \ LIP6 - CNRS, UPMC

Encadrants : Clémence Magnien, Fabien Tarissan (LIP6 - CNRS, UPMC)

juin 2012

## Résumé

En munissant des individus de capteurs, il est possible de mesurer à quels moment deux personnes sont physiquement proches. L'ensemble de ces contacts, permettant l'échange d'information sans passer par une infrastructure existante, forme ce qu'on appelle un réseau de contacts. L'étude de tels réseaux se diversifie dans la littérature, mais la structure dynamique inhérente de ces réseaux complexifie grandement leur étude. Les méthodes classiques d'étude des graphes doivent donc être repensées.

Dans ce rapport de stage, on détaille des propositions pour adapter la notion de centralité, classique pour les graphes statiques, afin de prendre en compte l'aspect dynamique de tels réseaux. Le but est de quantifier l'importance des différents noeuds du réseau lors de diffusion d'information. Le rôle des noeuds évoluant au cours du temps, la métrique que nous proposons permet de quantifier l'importance d'un noeud donné en fonction du temps de départ de la diffusion d'information et ce pour tous les temps de départ possibles. Les résultats montrent que les métriques proposées permettent d'observer des nuances non négligeables entre les différents noeuds. L'évolution de l'importance d'un noeud reste toutefois difficile à bien caractériser, les résultats présentant une grande variabilité au cours du temps.

*Mot clés* : Graphe dynamique, Centralité dynamique, Réseau de contacts, Diffusion par inondation, Plus courts chemins.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>État de l'art</b>	<b>3</b>
2.1	Expériences en environnement réel . . . . .	3
2.2	Représentation et propriétés des réseaux de contacts . . . . .	5
2.2.1	Résultats sur les graphes dynamiques . . . . .	5
2.2.2	Représentation sous forme de séquence de graphes statiques . . . . .	6
2.2.3	Analyse simultanée poussée de deux expérimentations . . . . .	7
2.3	Modélisations de réseaux de contacts . . . . .	8
2.4	Évaluation de protocoles de routage pour réseaux de contacts . . . . .	10
<b>3</b>	<b>Analyse des propriétés générales des jeux de données utilisés</b>	<b>12</b>
3.1	Rollernet . . . . .	12
3.2	Infocom . . . . .	13
3.3	Pnas . . . . .	13
<b>4</b>	<b>Étude de l'importance des nœuds dans la diffusion</b>	<b>14</b>
4.1	Étude détaillée pour un temps de départ fixé . . . . .	14
4.1.1	Résultats globaux des phénomènes de diffusion . . . . .	14
4.1.2	Résultats obtenus par isolement des différentes sources . . . . .	17
4.1.3	Résultats obtenus par isolement des différents destinataires . . . . .	20
4.2	Étude généralisée à tous temps de départ . . . . .	20
4.2.1	Algorithme développé . . . . .	23
4.2.2	Évolution de l'importance des nœuds au cours du temps . . . . .	23
4.3	Calcul du nombre de messages en transit au cours du temps . . . . .	26
<b>5</b>	<b>Analyse des résultats et discussion</b>	<b>30</b>
5.1	Résultats pour différents jeux de données . . . . .	30
5.1.1	Infocom . . . . .	30
5.1.2	Pnas . . . . .	32
5.2	Comparaison entre différentes mesures de centralités dynamiques . . . . .	34
<b>6</b>	<b>Conclusion et perspectives</b>	<b>38</b>

# Chapitre 1

## Introduction

Ces dernières années, avec la prolifération de systèmes embarqués communicants, de nouveaux types de réseaux distribués ont émergé. Parmi eux on peut retrouver les réseaux de communications de véhicules, les réseaux de capteurs ou encore les réseaux de contacts. Chacun de ces réseaux présente une forme nouvelle de topologie où les agents communiquant se déplacent en changeant d'environnement et d'interlocuteurs, ces types de réseaux sont dénommés *Delay Tolerant Networks (DTN)*. Dans ces réseaux, l'existence d'un canal de communication entre deux nœuds donnés est rarement disponible. La nature dynamique de ces réseaux apporte de nouvelles perspectives encore peu explorées comme l'utilisation de la mobilité des agents pour transporter de l'information sans dépendre d'une infrastructure de communication, mais cela complexifie grandement les problématiques associées. Une des principales difficultés de l'étude de ces réseaux provient de la nécessité de comprendre au mieux l'évolution du réseau afin de proposer des algorithmes de routages efficaces. La modélisation et la caractérisation de paramètres pertinents est pour cela primordial.

Un réseau de contacts est l'un des principaux exemples dans le domaine des *DTNs*. Grâce à des capteurs transportés par des personnes, il est possible de mesurer à quels moments deux personnes sont physiquement proches. Ces capteurs explorent leur entourage en envoyant périodiquement des paquets et en écoutant les paquets envoyés par les autres capteurs, ils peuvent ainsi repérer d'autres capteurs à proximité. Ce type de méthode permet de mesurer les contacts à l'intérieur d'un groupe de personnes. Grâce à de tels réseaux, des protocoles de communications dits *ad hoc* peuvent être créés afin de transmettre de l'information entre les agents sans passer par une infrastructure existante. Mais comme le montrent les expérimentations, les réseaux de contacts présentent de nombreuses spécificités vis à vis d'autres *DTNs* ou même vis à vis de différents scénarios de dynamiques de contacts, et les études montrent l'importance de les prendre en compte lors de la création d'algorithmes de routage.

Lors de ce stage, nous nous sommes focalisés sur la caractérisation du comportement des différents nœuds de réseaux de contacts. Le but était d'essayer de trouver les nœuds importants pour la communication dans le réseau. La structure dynamique de ces graphes fait l'originalité de ce travail, alors que peu d'études sont encore disponibles. Dans le deuxième chapitre sera détaillé un état de l'art. Ensuite dans le troisième chapitre, les jeux de données qui seront utilisés seront brièvement présentés. La quatrième chapitre présente la description des méthodologies de mesure proposées. La cinquième partie consiste en une analyse des résultats obtenus pour les différents jeux de données présentés dans ce rapport. Le tout sera conclu par une brève discussion sur les méthodologies proposées et les perspectives envisagées.

## Chapitre 2

# État de l'art

Les graphes sont utilisés pour représenter bien des problèmes, et il s'agit d'un domaine bien étudié en informatique. Pour représenter ces réseaux de contacts un graphe dynamique est souvent utilisé, dans lequel les personnes sont représentées par des nœuds et leurs interactions par des arêtes bidirectionnelles entre les nœuds, associées au moment où a lieu le contact. Mais l'évolution de tels graphes est longtemps restée peu étudiée. La question se pose de savoir quelles propriétés des graphes statiques restent valides pour ces graphes dynamiques. À l'heure actuelle, nous manquons encore d'outils et de notions permettant de décrire la dynamique de tels graphes. Ceci est cependant crucial et pas uniquement pour des applications de communication *ad hoc* mais aussi dans un contexte bien plus large comme l'analyse de la topologie de l'internet, les graphes du web ou encore certains réseaux sociaux.

Dans un premier temps nous explorerons les différentes expérimentations en environnement réel qui ont été menées ces dernières années. Puis nous étudierons les différentes représentations utilisés dans la littérature pour les réseaux de contacts ainsi que les analyses qui ont été faites sur les données expérimentales. Dans la partie 2.3 les problématiques de modélisation de ces graphes seront discutées. Enfin une revue des propositions de protocoles de routage adaptés aux scénarios étudiés ainsi que des techniques d'évaluations de protocoles de routages sera présentée.

### 2.1 Expériences en environnement réel

Pour être capable de représenter des cas réels d'application, l'étude et l'obtention de données réelles par de larges expérimentations est nécessaire. Dans cette optique de nombreux travaux ont été réalisés, où un ensemble de personnes étaient suivies pendant des durées très variables. Au MIT un groupe d'élèves et de professeurs d'informatique a été équipé de capteurs communicants traçant l'historique de leurs contacts pendant 9 mois. Pendant plusieurs années une expérience similaire a été effectuée à l'université Cambridge. D'autres expériences ont été menées dans des environnements plus restreints, montrant des comportements parfois bien spécifiques.

Pour la réalisation de ces expériences des petits systèmes embarqués capables de communiquer entre eux et de stocker l'historique des contacts aperçus au cours du temps sont nécessaires. Dans les deux études détaillées dans [7] et [17] des *Intel Imotes* sont utilisés. Il s'agit d'un système embarqué sur puce possédant la technologie Bluetooth, une mémoire et un processeur

capable de traiter les contacts recensés au cours du temps. En plus de ces appareils distribués à quelques personnes du groupe, les téléphones portables possédant la technologie Bluetooth sont aussi recensés par les *Imotes* afin d'élargir l'ensemble des observations. Les *Imotes* listent les appareils périodiquement et repèrent les nouveaux arrivants et les disparitions et archivent des dates de début et fin du contact. Les traces fournies par l'ensemble des *Imotes* sont regroupées à la fin de l'expérience afin de modéliser un graphe dynamique de l'expérience pour l'analyser et faire des simulations sur ce scénario.

### **Analyse de l'expérience *Infocom***

Dans l'article [7], une expérience est présentée et commentée. Quelques dizaines de personnes assistant à la conférence *Infocom 2005* ont été suivies pendant 3 jours. Les auteurs étudient plusieurs aspects de l'expérience.

En premier lieu, une étude approfondie est faite sur les durées des contacts et les durées entre des contacts successifs. L'observation essentielle qui en résulte est que ces distributions suivent des lois de puissance que ce soit pour les contacts entre *Imotes* ou avec les téléphones portables. De plus il est montré que la distribution du nombre de contacts parmi les membres de l'expérience n'est pas uniforme, certains participants sont repérés de manière bien plus fréquente que d'autres. Or, bien que les simulateurs existants prennent en compte la distribution en loi de puissance des durées de contact et inter-contact, ils font l'hypothèse en général d'un comportement uniforme parmi les agents.

Un autre point intéressant est discuté dans l'article. Lors de l'expérience, des communautés plus ou moins restrictives ont été observées, dans lesquelles la fréquence des contacts des membres est plus élevée que la moyenne. Ceci est un point clef à prendre en compte pour créer des algorithmes de routage efficaces, car en effet transmettre l'information à un membre de la même communauté que le destinataire de manière privilégiée permet d'obtenir une plus grande efficacité que lors d'un choix aléatoire de l'intermédiaire.

Enfin, une dernière étude est faite sur l'évolution quotidienne des contacts. Elle montre la présence d'une forte corrélation entre la dynamique des différents jours vis à vis des différentes périodes de la journée. Ce dernier point n'est en général pas modélisé dans l'étude des réseaux de contacts, pourtant il est important de prendre en compte ces aspects pour la mise au point d'algorithmes de routage adaptés à des usages réalistes.

L'expérience fut réitérée l'année suivante à la conférence *Infocom 2006* [15]. Pour cette seconde édition, le nombre d'*Imotes* distribués à presque doublé et en plus une vingtaine d'*Imotes* ont été installés à des positions fixes dans les principaux lieux de passage de l'hôtel accueillant la conférence.

### ***RollerNet***

L'article [17] présente l'expérience *RollerNet* où quelques dizaines de participants d'une randonnée roller se passant à Paris un dimanche après-midi sont équipés d'*Imotes*.

Un phénomène très spécifique au scénario est observé, le phénomène dit d'accordéon. Au cours de la randonnée des phases d'accélération et de ralentissement de la tête de la randonnée provoquent des étirements et rétrécissements du groupe des randonneurs. Ce phénomène s'observe bien par la variation du nombre de contacts recensés au cours de ces différentes

phases ainsi que par l'augmentation du nombre de composantes connexes. Une étude plus détaillée du phénomène montre qu'il est bien plus marqué dans les randonneurs proches de la tête de la randonnée alors qu'il est bien moins prononcé pour ceux à l'arrière. Ceci s'explique facilement par l'effet d'accordéon du fait des pauses et des accélérations successives effectuées par le front de tête de la randonnée pour respecter les contraintes de circulation.

Une brève étude des distributions des temps de contacts et d'inter-contacts viennent confirmer que ces distributions respectent une loi de puissance comme précédemment observé dans l'expérience faite à la conférence *Infocom 2005* avec cependant des coefficients différents.

## 2.2 Représentation et propriétés des réseaux de contacts

Avant tout travail d'analyse et de modélisation, la question de la structure de données à utiliser afin de proposer une représentation adéquate de réseaux de contacts est nécessaire. Plusieurs approches différentes sont étudiées dans la littérature. Un graphe dynamique, où les arêtes sont étiquetées par les temps discrets ou intervalles des moments d'activité du contact associé, permet de représenter sans pertes d'information ces réseaux. D'autres représentations plus proches des données expérimentales sont aussi étudiées, consistant généralement en une liste des contacts observés, soit de manière centralisée, soit distribuée relativement à chaque agent. Cette dernière représentation étant généralement peu adaptée à l'analyse des dynamiques et propriétés du réseau, mais pouvant être notamment utilisée pour effectuer des simulations.

### 2.2.1 Résultats sur les graphes dynamiques

Quelques rares travaux proposent des résultats théoriques sur des graphes dynamiques. Dans [8], la validité dans le cas dynamique de propriétés des graphes statiques est étudiée. Plus précisément les auteurs démontrent que le théorème de Menger, qui stipule que le nombre maximum de chemins qui ne se recoupent pas entre deux nœuds  $s$  et  $t$  est strictement égal au nombre minimal de nœuds à retirer du graphe pour déconnecter  $s$  et  $t$ , n'est plus valable. Les auteurs montrent que le nombre de chemins respectant l'ordre temporel entre deux nœuds, utilisant une séquence d'arêtes existant à des temps croissants, peut être strictement inférieur au nombre de nœuds à retirer pour les déconnecter. Un exemple est présenté dans la figure 2.1. Les auteurs vont plus loin dans l'étude de la propriété et ils arrivent à prouver qu'un graphe ne respecte pas le théorème de Menger si et seulement si il possède une sous division du graphe  $\mathcal{Z}$  de la figure 2.1, c'est à dire s'il ne possède pas un sous graphe qui peut s'obtenir en rajoutant des nœuds sur les arêtes du graphe  $\mathcal{Z}$ .

Dans le même article est proposé un algorithme résolvant un problème d'inférence sur un graphe dynamique à information partielle. Plus précisément le problème résolu est de savoir si dans un graphe dynamique où le moment précis des contacts est méconnu (allant d'une date quelconque à une date restreinte à un intervalle), on peut déterminer si pour deux partitions données un scénario est possible dans lequel une source fixe peut joindre par un chemin respectant l'ordre causal l'ensemble des nœuds d'une partition mais aucun de l'autre. Ce problème de connectivité n'est pas posé dans les expériences décrites dans la section suivante. Ce problème est pertinent lors de cas de données avec des observations à information partielles afin de déterminer si un scénario précis de dissémination est possible. Des données de ce type sont fréquemment trouvées aussi bien pour des raisons techniques d'acquisition que dues à

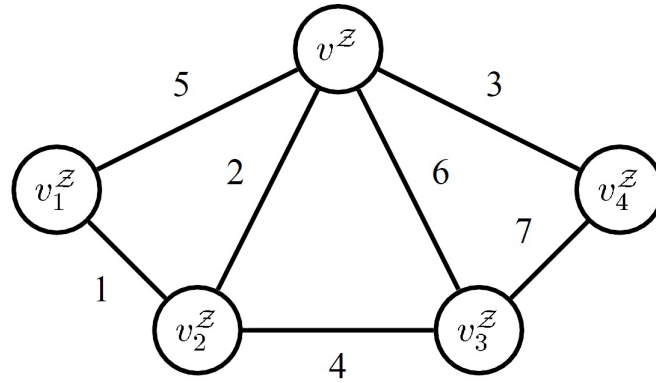


FIGURE 2.1 – Un graphe  $\mathcal{Z}$  avec un étiquetage invalidant le théorème de Menger

la représentation sous forme de séquence de graphes statiques, où une partie de l'information est perdue.

### 2.2.2 Représentation sous forme de séquence de graphes statiques

Dans de nombreux travaux, les réseaux de contacts sont représentés comme une séquence de graphes statiques. La dynamique étant représentée par cette série de graphes statiques. Chaque graphe représente une période de temps où sont agglomérés tous les contacts qui ont eu lieu. Cette représentation présente l'avantage de transformer l'analyse de graphes dynamique en une séquence d'analyse de graphes statiques, ce qui est bien documenté dans la littérature. Un des points épineux de cette représentation est le choix de la durée entre deux graphes successifs. En effet, un temps trop court et le graphe se retrouve presque sans observations de contacts, sans apporter d'informations pertinentes, et un temps trop long implique que la dynamique du réseau de contact est totalement perdue.

Une étude autour de la sélection appropriée de ce temps est proposée dans l'article [4] afin de limiter la perte d'informations pertinentes sur la dynamique de graphe. Les auteurs se servent des données produites par une expérience réalisée au sein du MIT pour effectuer leur analyse. Ils analysent les rythmes d'évolution des communautés au sein du réseau de contacts pour déterminer un temps caractéristique d'évolution du graphe. Afin de mesurer l'évolution du graphe trois mesures sont utilisées :

- le degré moyen des nœuds du graphe représentant la densité ;
- la valeur moyenne du coefficient de *clustering* mesurant l'importance des communautés au sein du graphe ;
- une mesure de corrélation des voisinages des nœuds mesurant les ressemblances des contacts actifs des différents nœuds.

Grâce à l'étude extensive des auto-corrélations de ces valeurs pour différents intervalles de temps entre deux états considérés du graphe, les auteurs arrivent à repérer des motifs de ressemblance de l'état du graphe tous les 8, 16 et 24 heures. Le motif journalier étant largement le plus évident. Ils proposent donc en conséquence une valeur d'environ 4 heures comme temps caractéristique de l'évolution du graphe à utiliser pour séquencer le graphe. Cette valeur permet de bien capturer les différentes phases d'évolutions et motifs journaliers.



### 2.2.3 Analyse simultanée poussée de deux expérimentations

A. Scherrer et Coll. présentent dans l'article [14] une analyse poussée prenant en compte de nombreux paramètres et propriétés des réseaux de contacts étudiés. Leur analyse est effectuée conjointement sur deux ensembles de données d'expériences, ceux de la conférence *Infocom 2005* détaillée précédemment et sur du *MIT Realty Mining* [5] mentionné précédemment et résultant des traces des contacts d'une centaine de membres du département d'informatique de l'université pendant 9 mois.

Lors de leur étude un très large spectre de propriétés est étudié ; les distributions des durées des contacts et inter-contacts ainsi que des coefficients de *clustering* mais aussi de manière originale des taux de suppression et d'apparition d'arêtes dans le réseaux ainsi que le nombre de triangles dans le graphe (trois personnes en contact simultané). Leur étude utilise une représentation sous forme de séquence de graphes statiques, mais pour être sûr de ne pas perdre d'information le temps d'échantillonnage est fixé à 1s afin de capturer de manière certaine toute évolution. L'étude sur les distributions des durées des contacts et inter-contacts vient confirmer les résultats existants stipulant que ces distributions suivent bien une loi de puissance.

#### Analyse statistique des propriétés structurelles

De manière similaire à l'étude effectuée dans [4], les auto-corrélations temporelles des différentes propriétés des graphes statiques sont calculées pour estimer leur temps moyen d'évolution. Des résultats similaires sont trouvés pour l'ensemble des mesures sauf en ce qui concerne la création et suppression d'arêtes qui évoluent de manière bien plus rapide. Ensuite des calculs des moyennes des corrélations croisées des différentes propriétés sont effectués et il en ressort une très forte corrélation des différentes propriétés avec des valeurs allant de 0.5 à 0.9, hors ajout et suppression d'arêtes qui sont totalement dé-corrélés. Cette étude montre que les ajouts et suppressions d'arêtes peuvent être considérés comme indépendants, mais qu'une telle hypothèse pour les autres propriétés serait fortement dommageable notamment pour effectuer une simulation. Les résultats obtenus relativement aux ajouts et suppressions d'arêtes montrent des temps caractéristiques d'évolution et des taux de corrélation tellement faibles qu'ils peuvent être considérés comme des processus aléatoires sans mémoire, et donc Markovien, sans perte significative d'information sur le réseau.

Enfin, une analyse sur l'évolution de différents paramètres de la structure du graphe représentant le réseau de contacts est effectuée. Dans un premier temps les auteurs caractérisent la stabilité temporelle des composantes connexes. Globalement, ils montrent que les trois quarts des ajouts et suppressions d'arêtes correspondent à des modifications de composantes connexes existantes. Le quart restant correspondant à une disparition d'une composante connexe ou à l'apparition d'une nouvelle composante connexe ; une autre possibilité étant aussi la séparation en deux d'une composante connexe ou la fusion de deux composantes connexes.

Dans un second temps, la densité de triangles dans le graphe est étudiée. Les auteurs montrent que pour les réseaux de contacts étudiés l'ajout d'une arête a quatre fois plus de chances d'ajouter un triangle dans le graphe que pour un graphe aléatoire. Ce résultat caractérise bien le fait que deux personnes en contact, et donc à proximité d'une troisième, ont de fortes chances d'être en contact entre elles.

## Émergence de communautés

Les auteurs s'intéressent aussi aux communautés présentes dans le graphe, une communauté étant un ensemble de personnes ayant plus d'interactions entre elles que la moyenne. Ces communautés ne sont pas capturées par l'étude des composantes connexes, bien que les membres d'une même communauté se retrouveront en général régulièrement au sein d'une même composante connexe. Trouver les communautés au sein du réseau de contacts n'est pas facile, trouver les communautés d'un graphe étant *NP-difficile*. Bien que la plupart des approches dans la littérature utilisent des heuristiques, une résolution exacte est privilégiée dans cet article. La méthode utilisée consiste à trier les couples de nœuds communiquant plus qu'une limite fixée et ensuite de les regrouper par composantes connexes, les composantes connexes n'ayant pas suffisamment d'interconnexions étant retirées ainsi que les composantes connexes qui ne sont pas assez denses. Ces communautés sont repérées pour de petits intervalles de temps, les résultats des différents intervalles étant ensuite réunis en regroupant les communautés qui diffèrent pour peu d'individus et plages temporelles. Les communautés trop peu représentées dans le temps ne sont pas retenues. Les résultats obtenus sur les expériences de réseaux de contacts étudiées repèrent une dizaine de communautés, certains individus pouvant faire partie de plusieurs communautés à la fois.

## 2.3 Modélisations de réseaux de contacts

### CTG : Générateur de traces de contacts

Une autre modélisation est proposée dans [3]. Le but est de simuler un réseau de contact des propriétés choisies et de générer des traces ressemblant à celles d'une expérience qui est donnée en entrée au simulateur. La modélisation du réseau de contacts est relativement simple, un graphe de taille et de distribution de degrés de nœuds données est généré aléatoirement. Ce graphe représente le réseau de contacts sans informations sur les moments où les contacts sont actifs. Ensuite la simulation du réseau de contact est effectuée à partir d'un état initial sans contact actif et pour chaque contact la durée avant commutation de actif vers inactif et inversement est choisie aléatoirement en respectant les distributions des temps de contacts et inter-contacts fournies. Une modélisation en loi de puissance est utilisée pour ces distributions dans le simulateur.

En utilisant des paramètres extraits d'une expérience réelle, les auteurs montrent que les paramètres extraits du résultat de la simulation correspondent bien à ceux de l'expérience réelle utilisée. Cette comparaison est uniquement effectuée sur les distributions fournies en entrée au simulateur et sans utiliser aucune autre propriété des deux réseaux de contacts. Les variables aléatoires étant choisies pour respecter ces distributions, le résultat n'est pas très pertinent en soi car il est imposé par la simulation. Aucune autre validation du modèle présenté n'est proposée.

Le générateur de traces ainsi présenté permet d'effectuer de nombreuses simulations répliquant des scénarios d'expériences réelles. De plus le simulateur permet de faire varier indépendamment les ordres de grandeur des durées de contact et inter-contact en faisant varier le coefficient de la loi de puissance utilisée. Enfin le nombre de participants à la simulation peut aussi être modifié à souhait. Pour que la distribution des degrés des nœuds du graphe reste cohérente, il est proposé d'appliquer une homothétie sur la distribution utilisée. Une étude sur

le passage à l'échelle des applications utilisant des réseaux de contacts peut donc être menée avec ce simulateur et cela indépendamment pour les trois propriétés traitées.

Dans le papier [14], dont l'analyse des réseaux de contacts a déjà été discutée en section 2.2.3, les résultats de l'analyse sont utilisés dans le but de proposer et d'évaluer des modélisations de réseaux de contacts similaires à ceux des expériences étudiées.

## Description des modèles proposés

Tous les modèles qui sont proposés possèdent le même squelette. Leur approche se base en tout premier sur un résultat très utile de leur analyse ; l'ajout et la suppression d'arêtes peuvent être considérés comme un processus markovien. Ceci implique pour la simulation de la dynamique du graphe que les changements d'état des arêtes sont indépendants entre eux pour chaque pas de temps. La modélisation consiste à calculer une probabilité  $P_{tr}(e, G_t)$  à chaque pas de temps et pour chaque arête, l'état de l'arête étant permuté si cette valeur dépasse un seuil. Dans un premier temps, cette probabilité est calculée de manière à respecter les distributions des durées de contacts et d'inter-contacts.

Dans un second temps, des pondérations sont rajoutées au calcul de la probabilité de transition pour forcer le graphe simulé à respecter les distributions des propriétés structurelles telles que le nombre d'arêtes, le nombre de nœuds connectés, le nombre de composantes connexes ou le degré moyen des nœuds connectés. Ceci étant fait en multipliant par un poids inférieur ou égal à 1 le précédent calcul de la probabilité de changement d'état. Ce poids est déterminé par le ratio entre les mesures du respect des propriétés structurelles à imposer avec ou sans changement de l'arête. Enfin un second poids est ajouté pour favoriser la création d'arête en ajoutant un triangle au graphe, en fonction des résultats mesurés dans l'analyse sur les expérimentations.

## Évaluation des modèles proposés

Au total six modélisations différentes sont évaluées. Les propriétés structurelles étant fortement corrélées, il n'est pas très pertinent de chercher à toutes les imposer en même temps. Sont étudiées les modélisations sans propriétés structurelles, avec la propriété du nombre de nœuds connectés et avec la propriété du nombre de composantes connexes. Les modèles imposant le nombre d'arêtes ou le degré moyen des nœuds connectés ne sont pas évalués, ces propriétés étant très fortement corrélées avec l'une des deux propriétés structurelles déjà évaluées. Enfin pour les trois cas décrits, les mêmes modélisations, avec en plus une pondération afin de respecter le taux de création d'arêtes créant de nouveaux triangles dans le graphe, sont aussi évaluées.

La première observation faite sur les résultats des simulations est la disparition des phénomènes de variance de la densité du graphe causés par des événements propres à l'expérimentation tels que les pauses repas ou les déplacements du groupe suivi. Ensuite lorsque l'on s'intéresse au respect des distributions des différentes mesures structurelles du graphe, le modèle prenant en compte uniquement le respect des durées de contacts et inter-contacts montre des limitations. Le nombre total de nœuds connectés est surévalué, alors que le nombre de composantes connexes et le nombre de triangles sont sous évalués. Ensuite, et cela est valable pour les trois modèles, les distributions des différentes mesures structurelles sont bien moins variables. Cette observation s'explique aisément par la non stationnarité des données de l'ex-

périence pour des raisons identiques à celles expliquant la première observation. Un dernier point en faveur des modèles pondérés est le non-respect total de la distribution de la densité parmi les composantes connexes fréquentes dans le premier modèle alors que les autres modèles fournissent eux un très bon respect de la distribution observée expérimentalement. Dans l'ensemble aucune observation ne permet de privilégier l'un des deux modèles pondérés.

Des observations sur la structure des graphes modélisés montrent certaines limitations des trois modèles proposés. Le nombre de triangles dans le graphe est fortement sous évalué et la densité des arêtes au sein des composantes connexes est relativement constante alors qu'elle devrait croître avec la taille de la composante connexe, les densités observées représentant uniquement les cas minimaux observés dans les expériences. Lorsque les modèles avec prise en compte du nombre de triangles sont testés, la sous-évaluation disparaît presque totalement. D'un autre côté, ces modèles surévaluent légèrement le nombre total de composantes connexes.

## 2.4 Évaluation de protocoles de routage pour réseaux de contacts

### Evaluation à l'aide de CTG

Dans le papier [3] décrivant CTG (Connectivity Trace Generator), une analyse de différents protocoles de routage est effectuée sur les simulations générées par l'outil à partir des traces de l'expérience menée à l'université de Dartmouth [11]. Quatre protocoles sont testés : les protocoles de références d'inondation et de diffusion épidémique ainsi que le protocole de *Context-Aware Adaptive Routing (CAR)* [12]. Ce dernier sélectionne le porteur ayant la plus grande probabilité de délivrer le message au destinataire parmi les contacts à portée. Enfin une version simplifiée où le porteur est choisi aléatoirement est également étudiée pour fournir une référence.

Grâce aux possibilités fournies par le simulateur CTG, l'efficacité en termes de délais, taux de livraison et nombre de messages transmis est calculée en faisant varier différentes caractéristiques du réseau. Les résultats sont fournis pour des variations du coefficient de la loi de puissance utilisée pour les distributions des durées moyennes des contacts et inter-contacts, ainsi que pour la distribution de la densité des contacts simultanés.

Les résultats obtenus pour les variations de durée moyennes des contacts et inter-contacts sont symétriques. Le protocole de dissémination épidémique donne les meilleurs résultats quelles que soient les distributions des durées des contacts et inter-contacts, avec un taux de délivrance autour de 60% dans tous les cas testés. Cependant le nombre de messages envoyés est de l'ordre de cinq fois supérieur aux protocoles utilisant un porteur. Les délais de livraisons augmentent logiquement pour tous les protocoles (sauf inondation) lorsque les durées moyennes de contacts et inter-contacts augmentent et que pour ces protocoles l'on observe dans tous les cas testés l'ordre croissant aléatoire-épidémique-CAR pour le taux de délivrance. En termes de taux de délivrance les protocoles utilisant un porteur deviennent de moins en moins efficaces, de 50% à 40% pour CAR et de 40% à moins de 10% pour la version aléatoire. Cette décroissance résulte de la difficulté à prédire le bon porteur pour des évolutions lentes du réseau. Le protocole d'inondation de son côté utilise un nombre extrêmement grand de messages et un taux de délivrance extrêmement faible dans tous les cas à cause de la petite taille des composantes connexes, montrant son inadaptation à de tels réseaux de type DTN ne possédant pas de composante majoritaire connexe. Enfin les résultats sur les différentes distri-

butions de densité montrent uniquement une amélioration des performances quand la densité du réseau augmente, sans différencier davantage les comportements des différents protocoles.

Cette évaluation de protocoles sur de nombreux paramètres à l'aide de *CTG* montre l'efficacité et la robustesse du protocole épidémique dans de tels réseaux bien qu'elle montre qu'un nombre important de messages redondants est transmis. Le protocole *CAR* se démarque bien aussi car il a l'avantage d'être économe en messages envoyés et d'impliquer de faibles baisses du taux de délivrance, cependant une augmentation significative du délai est observée. Les résultats de *CAR* sont bien meilleurs que ceux d'un protocole similaire avec un choix aléatoire du porteur utilisé. Le choix d'utiliser plutôt le protocole épidémique que le protocole *CAR* doit donc se baser sur l'importance des différentes caractéristiques, à savoir les performances en termes de délai et taux de délivrance ou l'utilisation économique des ressources de communication respectivement.

## Analyse de protocoles pour *RollerNet*

Le comportement de quatre protocoles est étudié dans l'article [17] sur le scénario de l'expérience *RollerNet*. Le cas étudié est celui de la transmission d'une information depuis l'arrière de la randonnée vers un destinataire en tête.

Le premier protocole est le protocole épidémique qui consiste en une transmission instantanée de l'information à tous ses contacts. Ce protocole est mal adapté au scénario : lors des phases d'étirement du groupe l'information ne se répand pas à travers l'ensemble des randonneurs et n'atteint pas le devant de la randonnée. De plus, notamment lors des phases denses de la randonnée, de nombreux messages redondants sont envoyés. Ce protocole sert avant tout de base de comparaison avec des protocoles plus complexes.

Ensuite le protocole de *Spay-and-Wait* est testé. Ce protocole consiste en la création d'un nombre limité de copies de l'information. Tant que plusieurs copies sont détenues, l'information est transmise périodiquement à ses contacts en partageant les copies. Si une seule copie est détenue, l'information est transmise uniquement au destinataire du message. Différents nombres de copies sont testés avec le scénario. Suivant les différentes phases du phénomène d'accordéon le délai de transmission du message est très variable. Plus le nombre de copies est élevé et plus le délai de transmission est faible, avec en contrepartie un nombre de redondances bien plus élevé dans les messages transmis.

Enfin, ce dernier protocole est modifié afin de s'adapter dynamiquement au scénario de *RollerNet*. Pour cela l'algorithme de *Spay-and-Wait* est modifié pour choisir en fonction de la densité de son voisinage le nombre de copies à diffuser pour avoir un délai constant. Le nombre de copies à envoyer étant choisi en fonction. Cette correspondance entre nombre de copies à transmettre et densité du voisinage est calibrée grâce aux traces de l'expérience. Ce protocole permet une bonne adaptation au phénomène d'accordéon observé, les tests effectués montrent que l'information est transmise en temps presque constant et que la redondance est bien limitée dans les phases de rétrécissement de la randonnée.

L'étude des protocoles de transmission d'information effectuée dans l'article [17] montre l'importance de la prise en compte du scénario caractérisant le réseau de contacts pour le choix d'un protocole adapté. Il propose une adaptation de protocoles existant pour les besoins spécifiques du scénario de *RollerNet* montrant des performances bien meilleures que les autres protocoles qui ne prennent pas en compte le phénomène d'accordéon.

## Chapitre 3

# Analyse des propriétés générales des jeux de données utilisés

Trois jeux de données seront présentés dans ce rapport.

Il s'agit de données déjà largement étudiées, notamment par une étudiante en thèse du LIP6 à l'UPMC, Lamia Benamara, et je me réfère à sa thèse soutenue en novembre 2011 [1] car de nombreuses analyses des propriétés générales des trois jeux de données ont déjà présentées dans son mémoire. Je redonne ici les principales caractéristiques de ces trois jeux de données de réseaux de contacts.

Les fichiers de données contiennent un nombre variables de contacts, listés sous la forme de quadruplets  $(I1, I2, Td, Tf)$  où  $I1$  et  $I2$  sont les identifiants des personnes pour lesquelles un contact a été enregistré entre un temps de début  $Td$  et un temps de fin  $Tf$ .

J'ai aussi recherché d'autres jeux de données de réseaux de contacts, et conduit sur ces réseaux des analyses du même type que celles qui sont présentées dans les chapitres suivants. Mais j'ai préféré mettre ici l'accent sur la méthodologie sans inclure une longue annexe avec une série de résultats sous forme de graphiques supplémentaires. Les résultats ne seront donc présentés que pour ces trois jeux de données étant assez représentatifs des réseaux de contacts disponibles.

### 3.1 Rollernet

Comme indiqué en introduction, il s'agit de données recueillies lors d'une randonnée en roller organisée à Paris en 2009 [17], où 62 participants ont été équipés de capteurs iMote. Chaque capteur effectue des scans toutes les 15 secondes. La durée totale de la randonnée est d'environ 3 heures, avec deux sessions de 80 minutes et une pause de 20 minutes. Le jeu de données contient environ 60 000 contacts. Pour avoir un aperçu de l'évolution du jeu de données, la figure 3.1a présente la moyenne du temps minimal requis pour transmettre un message dans le réseau.

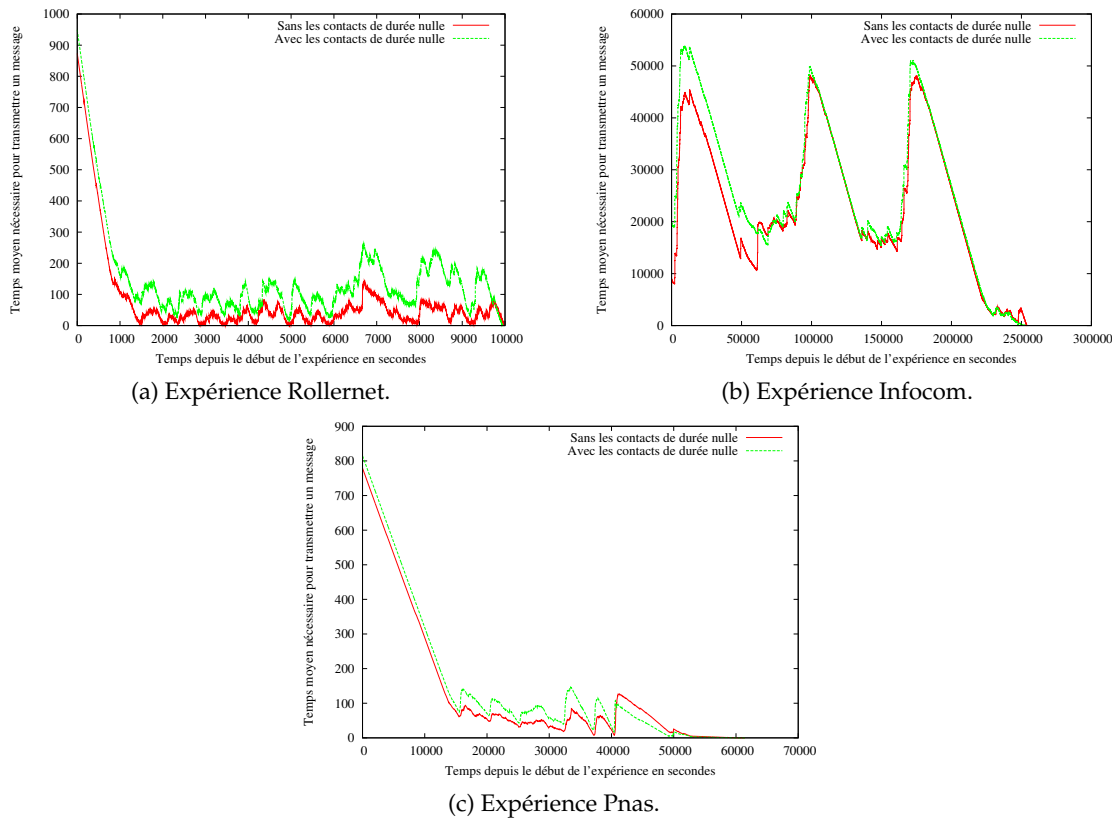


FIGURE 3.1 – Temps moyen requis pour transmettre un message dans le réseau, en utilisant les plus courts chemins.

## 3.2 Infocom

Nous avons aussi décrit dans l'introduction ces données collectées lors de la conférence Infocom 2005 [7], en distribuant des capteurs iMote à 41 étudiants participant à un workshop. Chaque capteur effectue des scans toutes les 120 secondes. La durée totale du workshop est de 4 jours et 3 nuits. Le jeu de données contient environ 30 000 contacts. Le moyenne du temps minimal nécessaire à la transmission d'un message dans le réseau est disponible en figure 3.1b

## 3.3 Pnas

Nous utilisons ici des données provenant d'une expérience conduite dans une école [13]. 800 personnes sont équipées de capteurs, dont 655 étudiants, 73 enseignants, 55 employés et 5 autres personnes. Chaque capteur effectue des scans toutes les 20 secondes. La durée de l'étude est d'1 jour. Le jeu de données contient environ 760 000 contacts. La figure 3.1c représente la moyenne du temps minimal pour transmettre un message dans le réseau.

## Chapitre 4

# Étude de l'importance des nœuds dans la diffusion

### 4.1 Étude détaillée pour un temps de départ fixé

Cette étude a pour but de caractériser dans le plus grand détail l'importance que peut avoir un nœud dans une transmission de message sur le réseau.<sup>1</sup>

Le temps de transmission n'est pas pris en compte dans l'étude et un contact ayant une durée nulle peut servir à la transmission d'un message. De plus le protocole utilisé est le protocole dit d'inondation, ce protocole consiste à transmettre l'information dès que possible à tous les nœuds à portée qui ne connaissent pas encore l'information. Ce protocole est très coûteux en nombre de messages transmis mais est optimal en temps nécessaire pour transmettre un message. Ce protocole permet donc de repérer les chemins les plus courts pour transmettre l'information d'un nœud donné à un autre.

La méthodologie utilisée pour caractériser l'importance d'un nœud pour la diffusion d'information consiste à comparer le phénomène de diffusion dans le réseau complet et dans le réseau sans le nœud étudié. Pour cela le temps minimal nécessaire pour transmettre un message depuis chaque nœud du réseau vers l'ensemble des autres est calculé pour tous les retraits possibles d'un nœud du réseau.

#### 4.1.1 Résultats globaux des phénomènes de diffusion

La figure 4.1 présente la distribution cumulative normalisée du temps de diffusion entre toutes les paires de nœuds pour le réseau complet, ainsi que les réseaux obtenus quand on enlève les nœuds 3, 29, 32 et 53 respectivement. Ces nœuds ont été sélectionnés car ils sont représentatifs des types de comportements observés et des catégories de participants. Les nœuds 29 et 32 font partie des organisateurs et se situent à l'avant de la randonnée, respectivement sur le côté gauche et le côté droit. Le nœud 53 est un ami de l'auteur, rollerman occasionnel, alors que le nœud 3 fait partie d'une association d'habitues ayant plus de chance de faire des déplacements rapides.

---

1. Lors de l'explication de la méthode utilisée, le jeu de donnée rollernet sera utilisé en exemple tout au long de cette partie.



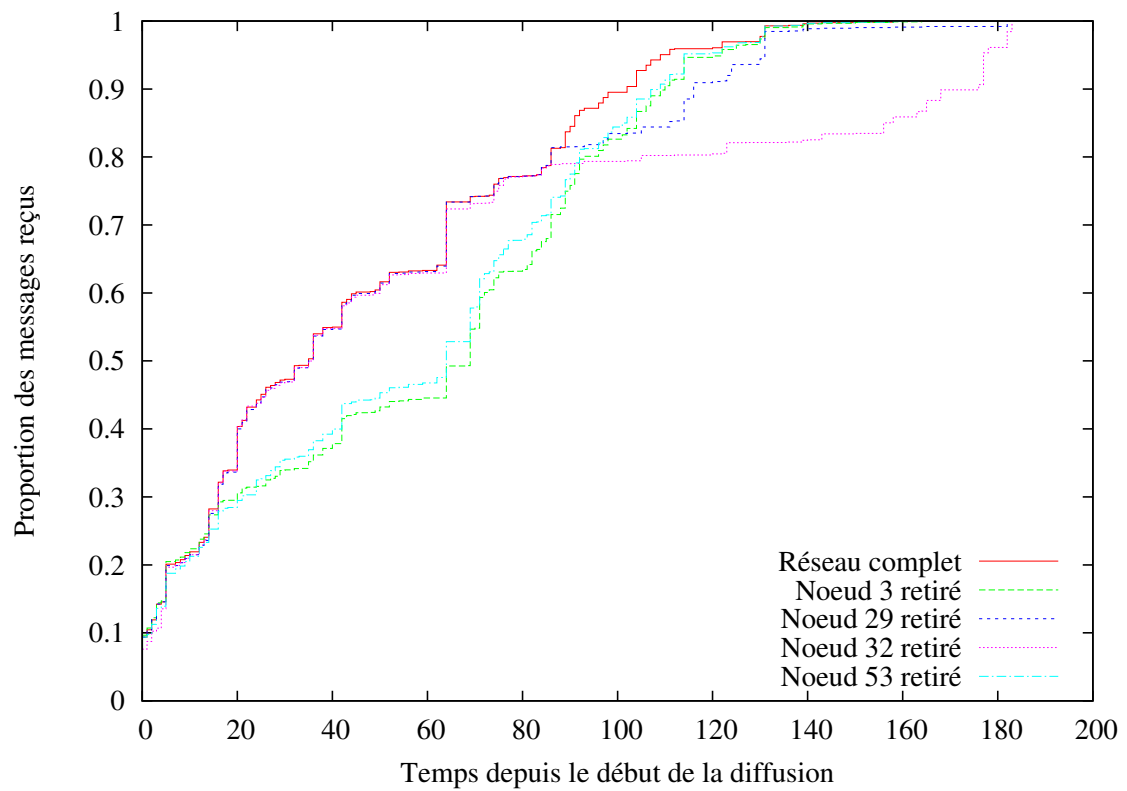


FIGURE 4.1 – Distribution cumulative normalisée des temps requis pour transmettre l'information de tout le monde vers tout le monde, avec tous les nœuds et sans les nœuds 3, 29, 32 et 53.

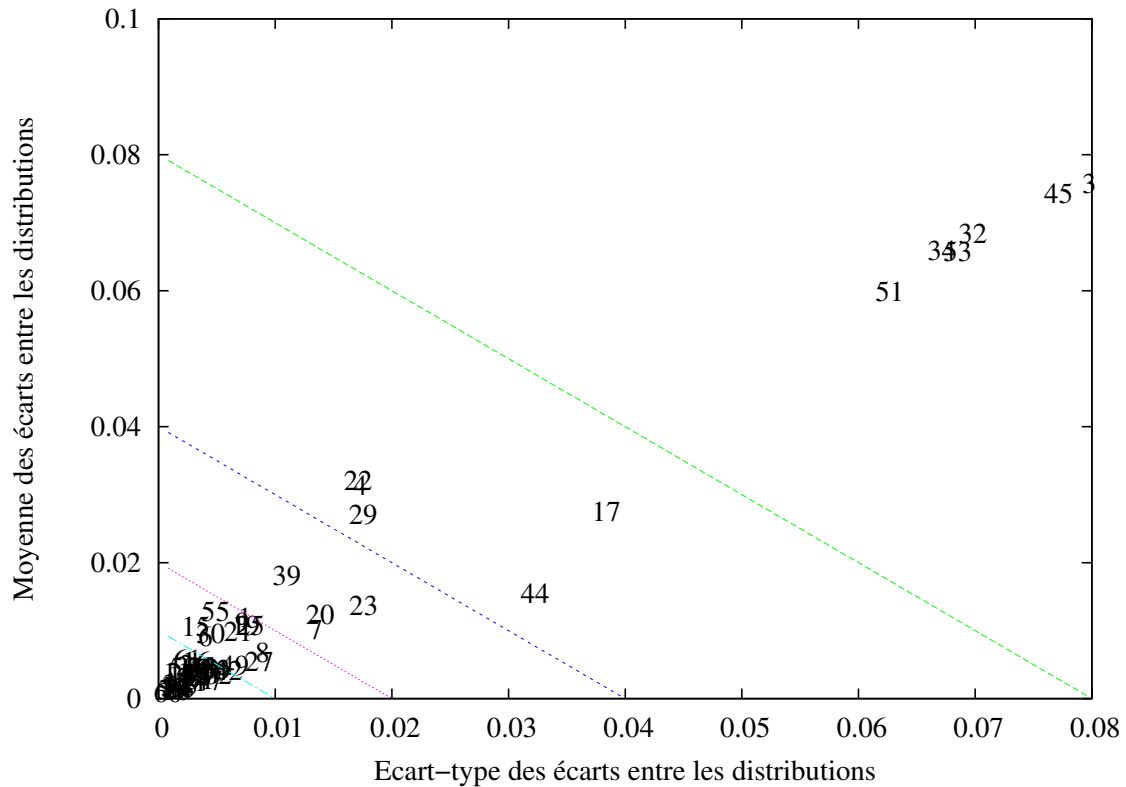


FIGURE 4.2 – Représentation conjointe de la moyenne et de l'écart-type des différences entre la distribution cumulative du réseau complet et celle du réseau amputé d'un de ses nœuds. Le label correspond au numéro du nœud retiré du réseau.

Pour le temps de départ choisi et les nœuds retirés sélectionnés dans la figure 4.1, on peut observer des écarts importants entre les différentes distributions observées. Afin de repérer et de catégoriser plus en détail et par un traitement automatisé les nœuds ayant un impact significatif relativement au phénomène de diffusion, on calcule l'écart moyen et l'écart-type des écarts entre les distributions cumulatives normalisées des temps de diffusions, dans le réseau complet et dans celui où on retire le nœud considéré. On peut observer dans la figure 4.2 la représentation de ces valeurs pour chaque nœud retiré du réseau. L'abscisse représente l'écart-type des différences et l'ordonnée la moyenne de ces écarts. Le label correspond au nœud retiré du réseau. On observe une certaine tendance linéaire entre les moyennes et écarts-types, les points étant tous dans un cône relativement étroit. De plus on observe un groupement naturel des points à différentes échelles. Nous proposons une catégorisation en 5 niveaux logarithmiques de l'impact, les droites représentées correspondent aux limites entre les différentes catégories. Les seuils sont fixés suivant la droite  $moyenne + \text{ecart} - \text{type} = 0.01, 0.02, 0.04$  et  $0.08$ . Ces seuils ont été fixés de manière arbitraire pour correspondre aux groupements naturellement observés, néanmoins les échelles de grandeurs utilisées restent cohérentes pour n'importe quel temps sélectionné dans le jeu de données rollernet.

Afin d'approfondir cette étude de l'impact du retrait d'un nœud, les comparaisons entre les distributions cumulatives de l'ensemble des temps de diffusion du réseau est insuffisant. En effet il se peut que les nœuds impactent seulement un nombre réduit de transmissions de messages mais apportent un délai supplémentaire important pour ces quelques cas, ou alors

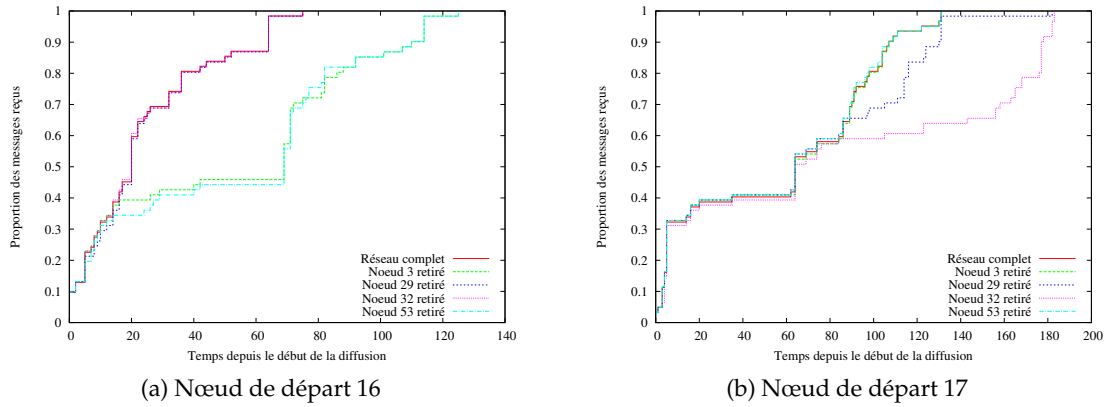


FIGURE 4.3 – Distributions cumulatives normalisées des temps requis pour transmettre l’information depuis des nœuds uniques vers tout le monde, avec et sans les nœuds 3, 29, 32 et 53.

ils peuvent apporter un délai supplémentaire faible identique dans toutes les communications. Ces deux impacts sont très différents et ne peuvent pas être différenciés par les observations globales faites dans les figures 4.1 et 4.2.

#### 4.1.2 Résultats obtenus par isolement des différentes sources

Afin d’obtenir une compréhension plus fine de l’impact d’un nœud donné lors de la diffusion, une étude poussée a été effectuée en isolant les sources d’émission de l’information. Ceci augmente fortement le nombre de distributions obtenues étant donné que pour chaque nœud retiré, l’ensemble des sources possible a été étudiée indépendamment.

La figure 4.3 montre les distributions cumulatives obtenues pour deux nœuds de départs différents. Les distributions correspondent aux mêmes nœuds retirés que pour la figure 4.1. On peut remarquer que pour les deux nœuds de départs sélectionnés les distributions sont très différentes. De plus les nœuds impactant de manière sensible ne se recoupent absolument pas pour les deux nœuds de départ sélectionnés.

Afin d’automatiser le procédé de comparaison des distributions et de catégoriser les différentes distributions obtenues, la même méthodologie que pour les distributions globales a été utilisée. La figure 4.4 représente les résultats obtenus dans ce cas, de manière similaire à la figure 4.2. De même une échelle logarithmique est utilisée pour catégoriser les distributions obtenues, les limites étant fixées dans ce cas à  $moyenne + \text{ecart-type} = 0.04, 0.08, 0.16$  et  $0.32$ .

La visualisation proposée pour observer les résultats obtenus de manière claire et pertinente est d’oublier les valeurs précises obtenues pour chaque couple possible (nœud de départ, nœud retiré du réseau) pour en retenir l’essentiel, qui en est capturé par le classement en cinq catégories telles que présentées précédemment. La figure 4.5 présente sous forme matricielle les résultats obtenus. En abscisse correspond le nœud qui a été retiré du réseau et en ordonnée le nœud de départ de la diffusion. Pour chaque coordonnée correspond un symbole suivant la catégorie correspondante, la catégorie correspondant aux écarts les plus faibles étant représentée par un blanc. De plus pour l’ordonnée 70 est affichée la catégorie correspondante déterminée lors de l’étude globale.

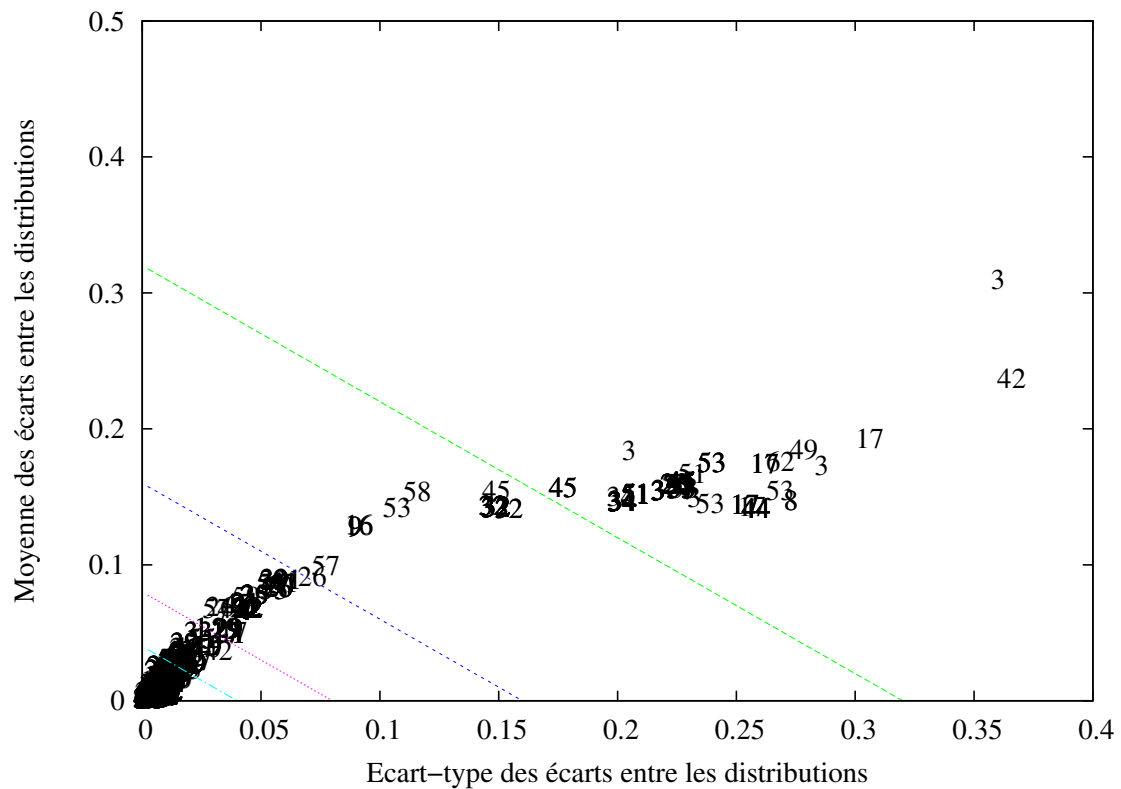


FIGURE 4.4 – Représentation conjointe de la moyenne et l'écart-type des différences entre la distribution cumulative du réseau complet et celle du réseau amputé d'un de ses nœuds pour une diffusion depuis une source unique. Le label correspond au numéro du nœud retiré du réseau, le nœud source de la diffusion n'est pas spécifié dans cette représentation.

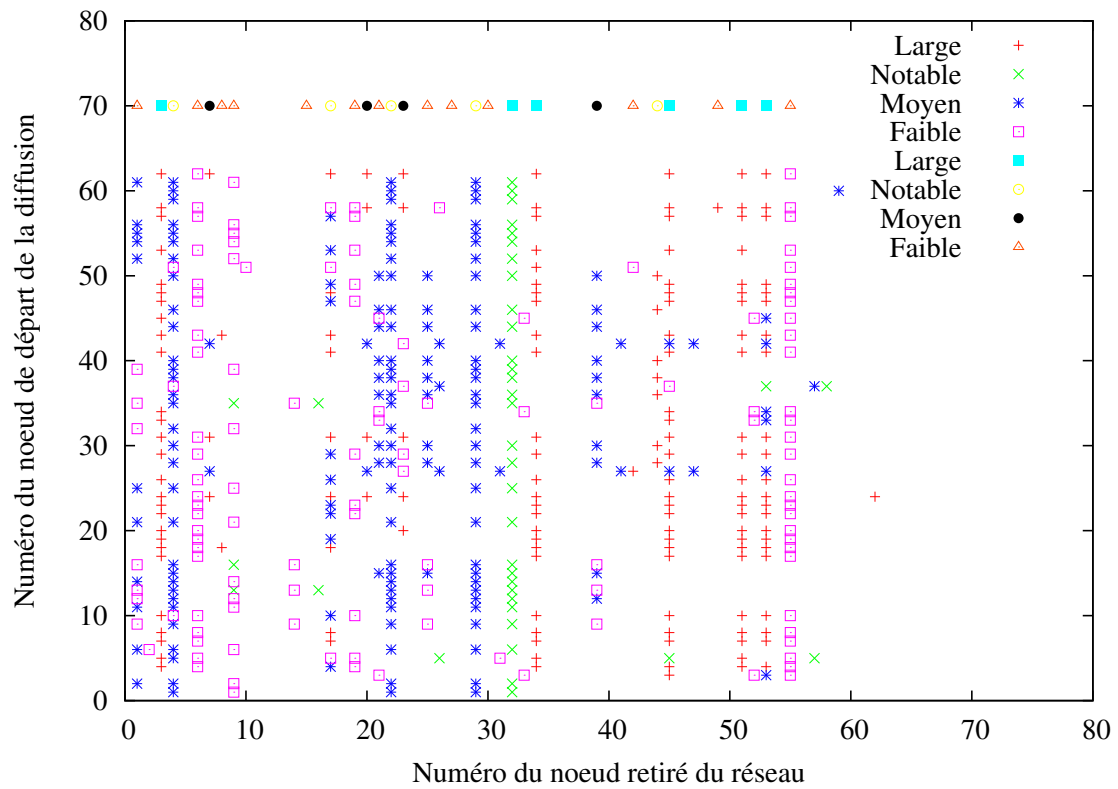


FIGURE 4.5 – Matrice représentant la catégorie associée pour tout couple donné de nœud retiré du réseau et nœud source de la diffusion. La catégorie correspondant à des variations très faibles vis à vis du réseau complet est représentée par un blanc. La ligne d'ordonnée 70 correspond à la catégorie obtenue pour une diffusion de tout le monde vers tout le monde. A comparer avec la figure 4.7 où les mêmes données sont représentées pour le numéro du nœud de destination de la diffusion.

Plusieurs observations intéressantes peuvent être faites sur les résultats présentés. Tout d'abord, tout nœud ayant un impact sur les phénomènes de diffusion affecte seulement un ensemble restreint de points de départ de la diffusion. De plus, on peut observer que les ensembles de nœuds de départ de la diffusion impactés se limitent à exactement deux groupes. Un certain nombre des nœuds dont le retrait impacte globalement la diffusion impacte le premier groupe, et les autres le second groupe mais aucun n'impacte même partiellement une partie des deux.

De plus, ces deux groupes sont strictement disjoints. On choisira comme représentant des deux groupes les nœuds 32 et 34 dont le retrait impacte fortement l'un des deux groupes respectivement. On peut ainsi associer au nœud 32 les nœuds 4, 22 et 29 ayant tous trois un impact presque identique mais plus faible que pour le nœud 32, et les nœuds 1, 9, 21, 25 et 39 dans une encore moindre mesure. Le nœud 44 se détache par un impact très fort sur un nombre limité des nœuds de départ affectés par le nœud 32. Le deuxième groupe du nœud 34 se comporte de manière presque identique sur les nœuds 3, 45, 51 et 53 et dans une moindre mesure avec le nœud 17 et encore plus faiblement pour les nœuds 7, 20 et 23.

### 4.1.3 Résultats obtenus par isolement des différents destinataires

Les résultats de la diffusion sont aussi regroupés par destinataire. Cela permet, au même titre qu'en isolant les sources, par un raffinement des comparaisons entre les distributions, de savoir quel nœud impacte fortement sur la réception des messages des autres. Ainsi, en recoupant avec la séparation par source, on peut identifier pour chaque nœud retiré du réseau les émetteurs et les récepteurs étant fortement impactés.

Au regard de la figure 4.6 on observe une grande similarité avec les résultats obtenus pour les différentes sources. Les nœuds impactant restent les mêmes et la distribution des moyennes et écart-types reste semblable. On peut noter une légère différence au niveau des différents points correspondant à un nœud retiré : au lieu d'être bien regroupés, ils forment une courbe assez régulièrement espacée.

La figure 4.6 présente un motif tout à fait similaire à la figure 4.4 obtenue en isolant les sources. La principale observation est la complémentarité entre les deux figures. On peut remarquer qu'un nœud est impacté par le retrait d'un autre uniquement pour les messages qu'il envoie ou ceux qu'il reçoit. On ne trouve presque pas de cas où les deux directions sont perturbées. De plus pour un nœud dont le retrait dégrade la transmission, les autres nœuds sont presque tous pénalisés en réception ou en émission.

Dans l'ensemble on retrouve bien quatre types de comportement. En vue de l'expérience, les participants se classent selon : être à l'arrière, à l'avant, impactant les transmissions vers l'avant ou impactant les transmissions vers l'arrière. Il serait intéressant de voir comment les comportements évoluent au cours du temps.

## 4.2 Étude généralisée à tous temps de départ

Les résultats obtenus pour différents temps de départ de la diffusion montrent une très grande variabilité du rôle joué par les nœuds. De plus, on observe des résultats très différents pour des temps de départ de la diffusion voisins. Il en résulte que l'on ne peut pas évaluer l'impact du retrait d'un nœud sur la diffusion sur l'étude de quelques temps de départ aussi

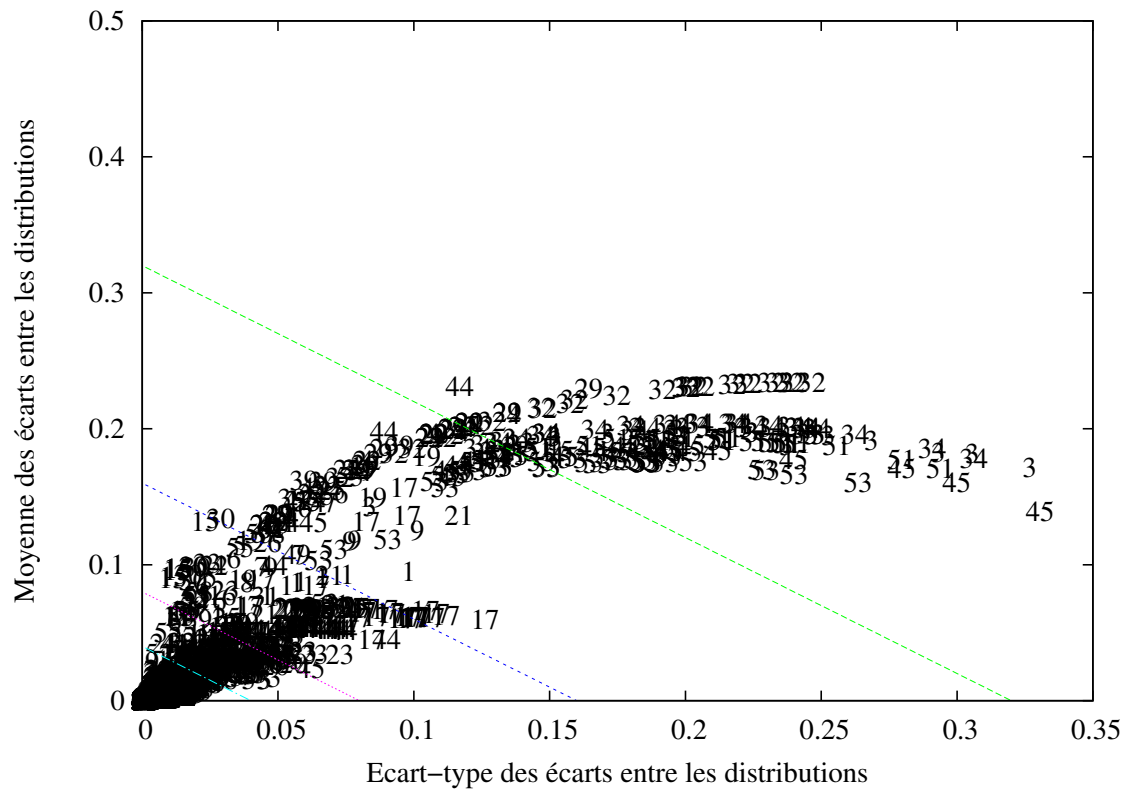


FIGURE 4.6 – Représentation conjointe de la moyenne et l'écart-type des différences entre la distribution cumulative du réseau complet et celle du réseau amputé d'un de ses nœuds pour un envoi de message de tout le monde vers un unique récepteur. Le label correspond au numéro du nœud retiré du réseau, le nœud de destination des messages n'est pas spécifié dans cette représentation.

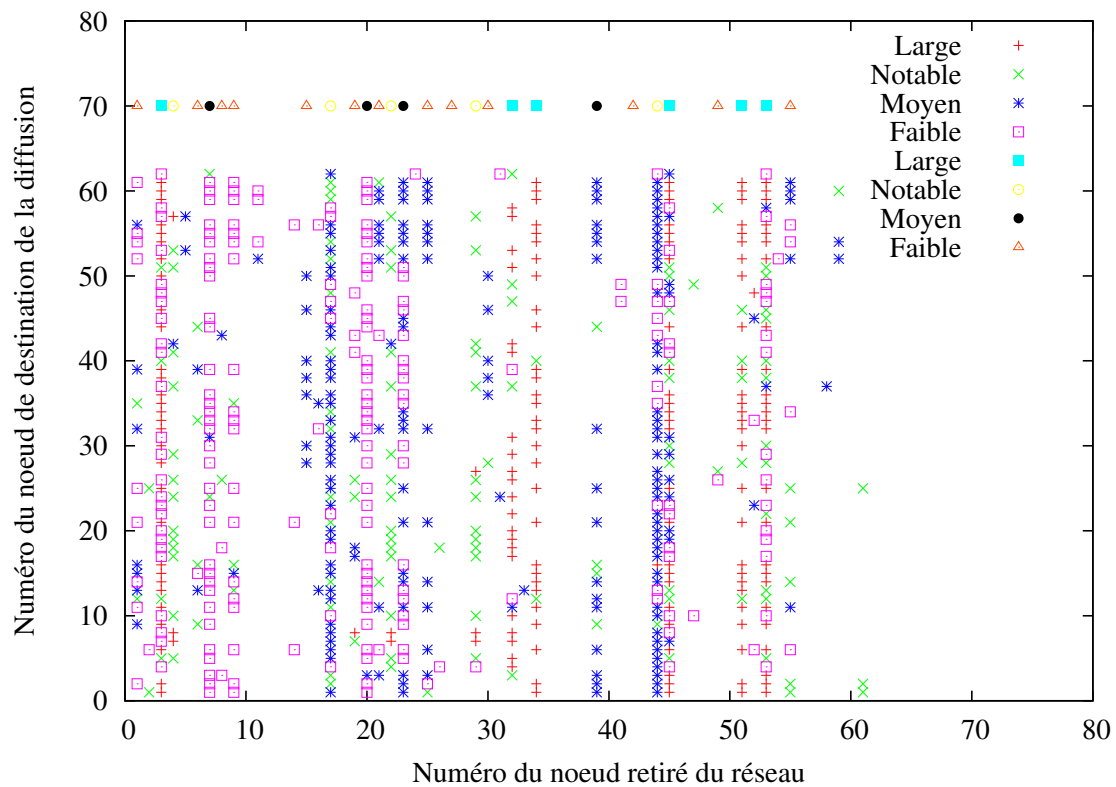


FIGURE 4.7 – Matrice représentant la catégorie associée pour tout couple donné de nœud retiré du réseau et nœud destinataire des messages. La catégorie correspondant à des variations très faibles vis à vis du réseau complet est représentée par un blanc. La ligne d’ordonnée 70 correspond à la catégorie obtenue pour une diffusion de tout le monde vers tout le monde.



bien choisis soient-ils. Il est donc nécessaire d'étudier ce phénomène tout au long de l'expérimentation et donc de mesurer l'impact du retrait d'un nœud au cours du temps.

### 4.2.1 Algorithme développé

La méthode utilisée précédemment consistait à choisir un temps de départ pour ensuite faire une recherche des plus courts chemins en simulant un processus d'inondation. Le temps requis par cette recherche ne permet pas d'en faire une exécution exhaustive pour tous les temps de départ possibles. On peut toutefois intuitivement que les calculer indépendamment pour chaque temps de départ n'est pas économe, on peut tirer avantage du calcul d'un temps donné pour le temps voisin. A partir de cette idée, on peut remarquer que la connaissance des temps requis pour le temps de départ  $t + 1$  et la connaissance de l'état du réseau au temps  $t$  suffit pour calculer les temps requis pour le temps de départ  $t$ . En effet si il existe un chemin de la source à la destination, le temps requis est 0, sinon il s'agit du temps minimal requis parmi les nœuds de sa composante connexe au temps  $t + 1$  plus une seconde. Une version en pseudo-code est disponible ci-après (Algorithme 1).<sup>2</sup>

Pour comprendre plus aisément comment se déroule l'algorithme 1, nous allons suivre l'évolution du programme sur un exemple simple. Le réseau considéré comprend 4 nœuds, numérotés de 1 à 4, et la durée totale de l'évolution du réseau limitée à 12 secondes ou unités de temps. Nous donnerons l'état pour ces 12 temps de la matrice *CommunicationTime* où on peut lire le temps nécessaire pour aller du nœud  $i$  au nœud  $j$  en ligne  $i$  et colonne  $j$  à l'instant considéré. Les 12 matrices sont disponibles dans la figure 4.9, en légende sous chaque matrice est donné le temps correspondant et les composantes connexes pour ce temps donné (sauf singletons). De plus une visualisation du réseau est disponible en figure 4.8, où pour chaque arête est indiqué le temps d'activité. On remarque aisément que les lignes correspondant à des nœuds d'une composante connexe sont identiques entre elles et que l'on a bien toujours 0 sur la diagonale, temps pour recevoir son propre message, et que la ligne correspondant à un nœud isolé au temps  $t$  est celle du temps  $t + 1$  avec les valeurs incrémentées de 1 sauf pour l'élément sur la diagonale.

### 4.2.2 Évolution de l'importance des nœuds au cours du temps

La figure 4.10 montre des exemples de l'importance d'un nœud donné au cours du temps. Cette importance est quantifiée sous forme de valeur d'impact calculée comme expliqué précédemment. Il s'agit de la moyenne des écarts entre les distributions cumulatives du réseau complet et du réseau amputé d'un de ses nœuds (Voire figure 4.2). Il s'agit des distributions des temps requis pour transmettre un message de tout le monde vers tout le monde comme présenté en figure 4.1. Cette valeur est calculée indépendamment pour chaque temps de départ possible.

Dans la figure 4.10, on peut observer pour les deux exemples donnés présenter un comportement apparemment erratique. Pour les deux nœuds, l'impact mesuré est négligeable pendant de nombreuses périodes, interrompus par de courtes périodes d'impact important. En comparaison, le nœud 53 présente une activité plus importante que le nœud 3. La différence se porte

---

2. Dans le cadre d'un projet ANR *DynGraph* de l'équipe d'accueil, le code source en python de l'algorithme (ainsi que pour certains traitements des données de sortie) est disponible sur internet à l'adresse suivante : <http://www-rp.lip6.fr/~magnien/DynGraph/Software/Centralite/>.

---

**Algorithm 1:** Calcul des temps de communication pour tout temps et tout couple d'émetteur et destinataire.

---

**input** : *Nodes* : A list of the all nodes of the graph.

*Connections* : An array of the list of edges created at the given time.

*Deconnections* : An array of the list of edges suppressed at the given time.

**output:** A sequence of *CommunicationTime* states for all times between first and last contact. The value at row  $i$  and column  $j$  corresponds to the current time needed to send a message from node  $i$  to node  $j$ .

*CommunicationTime* = Array[length(*Nodes*)] [length(*Nodes*)] ; // initialized to  $+\infty$

**foreach**  $n$  in *Nodes* **do** *CommunicationTime*[ $n$ ][ $n$ ] = 0;

**for**  $t \leftarrow$  *last\_contact\_time* **to** *first\_contact\_time* **do**

**foreach**  $c$  in *Deconnections*[ $t$ ] **do** add  $c$  **to** *ActiveLinks*;

**foreach**  $i, j$  in *Nodes*  $\times$  *Nodes*,  $i \neq j$  **do** *CommunicationTime*[ $i$ ][ $j$ ] + = 1;

*CCs* = ;

**foreach** ( $n1, n2$ ) in *ActiveLinks* **do**

$new\_cc = n1, n2$ ;

**foreach**  $cc$  in *CCs* **do**

**if**  $n1$  OR  $n2$  in  $cc$  **then**

$new\_cc = \text{Union}(new\_cc, cc)$ ;

                remove  $cc$  from *CCs*;

        add  $new\_cc$  **to** *CCs*;

**foreach**  $cc$  in *CCs* **do**

$minTime = \text{Array}[\text{length}(\text{Nodes})]$ ; // initialized to  $+\infty$

**foreach**  $i, dest$  in  $cc \times \text{Nodes}$  **do**

$minTime[dest] = \min(minTime[dest], \text{CommunicationTime}[i][dest])$ ;

**foreach**  $i, dest$  in  $cc \times \text{Nodes}$  **do**

$\text{CommunicationTime}[i][dest] = minTime[dest]$ ;

    export *CommunicationTime*;

**foreach**  $c$  in *Connections*[ $t$ ] **do** remove  $c$  from *ActiveLinks*;

---

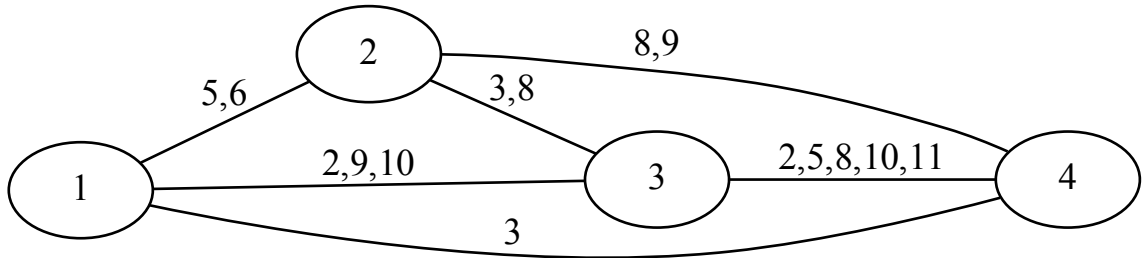


FIGURE 4.8 – Réseau utilisé pour l'exemple détaillé du déroulement de l'algorithme 1. Pour chaque arête sont indiqués les temps d'activité de celle-ci.

$\begin{pmatrix} 0 & * & * & * \\ * & 0 & * & * \\ * & * & 0 & * \\ * & * & * & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & * & * & * \\ * & 0 & * & * \\ * & * & 0 & 0 \\ * & * & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & * & 0 & 0 \\ * & 0 & * & * \\ 0 & * & 0 & 0 \\ 0 & * & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & * & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & * & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$
(a) 12. $\{\}$	(b) 11. $\{(3,4)\}$	(c) 10. $\{(1,3,4)\}$	(d) 9. $\{(1,3), (2,4)\}$
$\begin{pmatrix} 0 & * & 1 & 2 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & * & 2 & 3 \\ 2 & 0 & 1 & 1 \\ 2 & 1 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 3 & 2 & 0 & 2 \\ 3 & 2 & 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 3 & 3 \\ 0 & 0 & 3 & 3 \\ 4 & 3 & 0 & 0 \\ 4 & 3 & 0 & 0 \end{pmatrix}$
(e) 8. $\{(2,3,4)\}$	(f) 7. $\{\}$	(g) 6. $\{(1,2)\}$	(h) 5. $\{(1,2), (3,4)\}$
$\begin{pmatrix} 0 & 1 & 4 & 4 \\ 1 & 0 & 4 & 4 \\ 5 & 4 & 0 & 1 \\ 5 & 4 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 2 & 2 & 0 \\ 2 & 0 & 0 & 2 \\ 2 & 0 & 0 & 2 \\ 0 & 2 & 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 3 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 2 & 1 & 1 \\ 4 & 0 & 2 & 4 \\ 1 & 2 & 0 & 1 \\ 1 & 2 & 1 & 0 \end{pmatrix}$
(i) 4. $\{\}$	(j) 3. $\{(1,4), (2,3)\}$	(k) 2. $\{(1,3,4)\}$	(l) 1. $\{\}$

FIGURE 4.9 – Valeurs en fonction du temps, dans l'ordre décroissant calculé par le programme, de la matrice *CommunicationTime*. À la ligne  $i$  et colonne  $j$  on peut lire le temps nécessaire pour transmettre un message du nœud  $i$  vers le nœud  $j$ . En légende sous chaque matrice est donné le temps correspondant et les composantes connexes pour ce temps donné (sauf singletons).

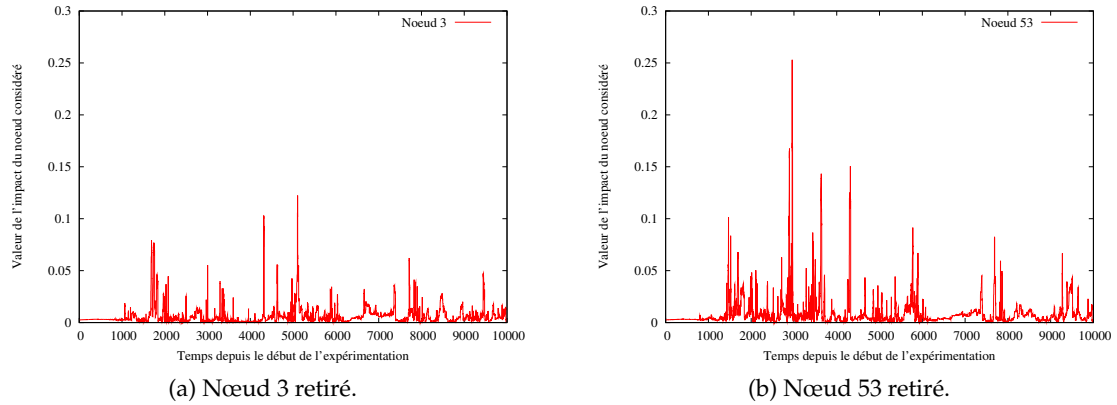


FIGURE 4.10 – Évolution de la valeur mesurée de l’impact du nœud retiré sur la diffusion selon la méthode proposée.

avant tout sur l’amplitude plus que sur la durée ou le nombre des pics d’importance. On peut remarquer que de manière contre intuitive, le nœud 53, qui est un amateur, présente un impact plus important que le nœud 3, qui est un habitué. Bien qu’une différence soit visible entre ces nœuds, le comportement erratique rend difficilement possible une quantification généralisée à partir de la simple lecture de graphiques obtenus pour des cas particuliers.

Tous les nœuds présentent le même type de graphique. Pour comparer l’impact global de l’ensemble des nœuds du réseau, il est possible de mesurer la valeur moyenne au cours du temps. Les moyennes mesurées sont réparties à peu près uniformément, malgré un amas de valeurs assez faibles, dans un intervalle allant du simple au quadruple. On arrive donc à mesurer une différence d’importance sur l’ensemble de l’expérimentation. La figure 4.11 présente les valeurs moyennes pour l’ensemble des nœuds composant le réseau en fonction de l’écart-type de l’importance au cours du temps. Il est difficile de le voir, mais le nœud 3 donné en exemple présente des valeurs assez faibles dans le groupement majoritaire. Le nœud 53 est lui à peu près le dixième en terme d’importance, au trois quart du maximum obtenu par le nœud 49. Autre remarque possible, l’écart-type est fortement corrélé linéairement avec la moyenne.

Cette observation permet de confirmer l’impression de régularité entre les graphiques des différents nœuds, l’impact mesuré variant par l’intensité plus que par la forme. Cette représentation ne permet pas de bien évaluer l’évolution de la mesure, ni de déterminer si l’impact des différents nœuds varie de manière significative entre les différentes phases de l’expérimentation.

### 4.3 Calcul du nombre de messages en transit au cours du temps

Une des mesures de centralité classique pour un graphe statique est la *Shortest-Path Betweenness centrality*. Elle essaye de caractériser les nœuds servant souvent d’intermédiaire aux autres en tenant compte des plus courts chemins. Plus formellement elle s’écrit :

$$c_B(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

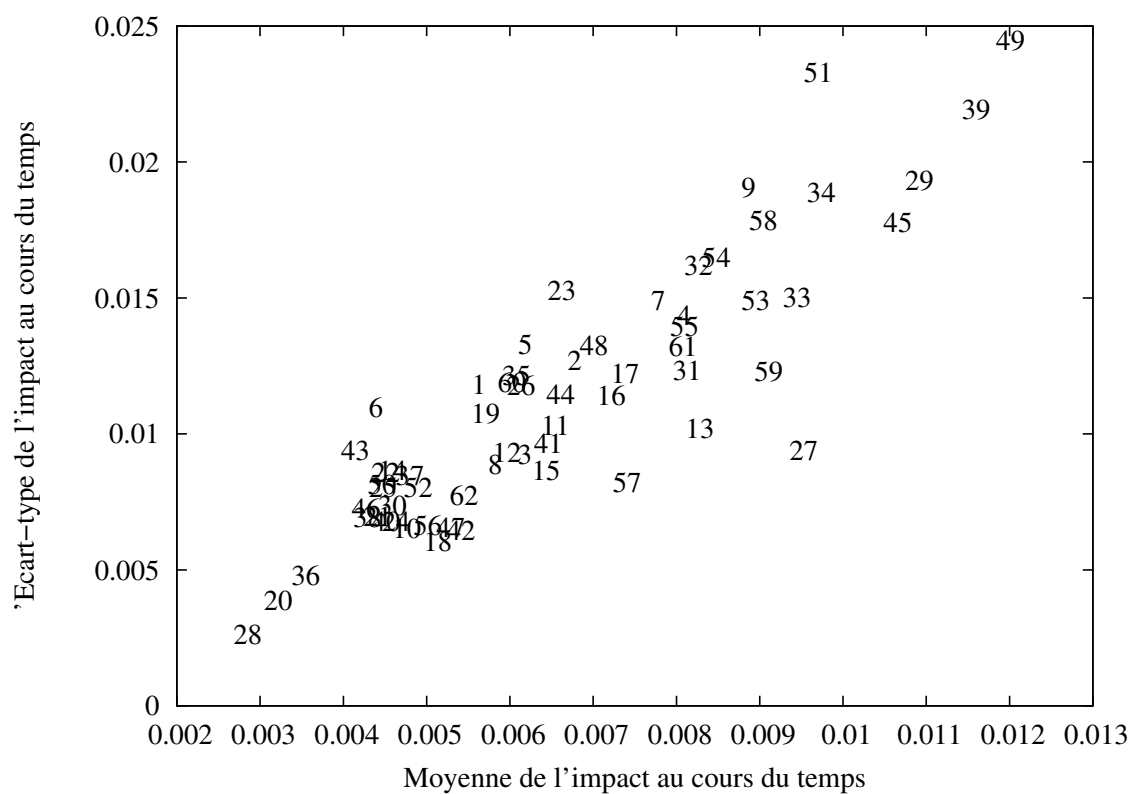


FIGURE 4.11 – Représentation des moyennes et écarts-types de la valeur de l'impact du retrait des nœuds au cours de l'expérimentation rollernet. Le label indique le nœud retiré correspondant à la mesure.

Avec  $v$  le nœud considéré,  $V$  l'ensemble des nœuds du graphe,  $\sigma_{st}$  correspond au nombre de plus courts chemins (en fonction du nombre d'intermédiaires) entre  $s$  et  $t$ , et  $\sigma_{st}$  correspond au nombre de ces chemins contenant  $v$ . Cela mesure le nombre de fois où le nœud considéré est sur le plus court chemin des autres, normalisé par le nombre de chemins possibles si il y en a plusieurs pour un couple donné. Cette notion de centralité se rapproche de notre mesure de l'impact du retrait d'un nœud lors de communications utilisant les plus courts chemins.

Cette mesure ne peut pas être directement appliquée à des graphes dynamiques. Si on l'applique à un instantané du réseau, la plupart des nœuds sont déconnectés et il n'existe pas de plus court chemin. Si on prend le graphe agrégé à ce moment on se retrouve avec presque l'ensemble des couples de nœuds directement connectés et il y a trop peu d'intermédiaires. Une manière d'adapter cette mesure au cas dynamique est de considérer les plus courts chemins pour transmettre un message. Regarder le plus court chemin de toute l'expérimentation n'est pas non plus souhaitable comme pour le cas des plus courts chemins en terme de nombre d'intermédiaires. Mais on peut calculer cette mesure pour un temps donné comme point de départ des chemins possibles. On peut l'écrire ainsi :

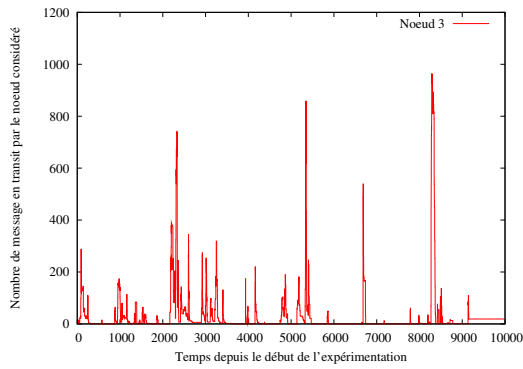
$$c_D(v, t) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \frac{\sigma'_{st}(t)(v)}{\sigma'_{st}(t)}$$

Avec  $\sigma'_{st}(t)$  le nombre de plus courts chemins en terme de délai de transmission à partir du temps  $t$  et  $\sigma'_{st}(t)(v)$  le nombre de ces plus courts chemins passant par  $v$ . Les boucles possibles ne sont pas prises en compte bien que pouvant ne pas altérer le délai de transmission.

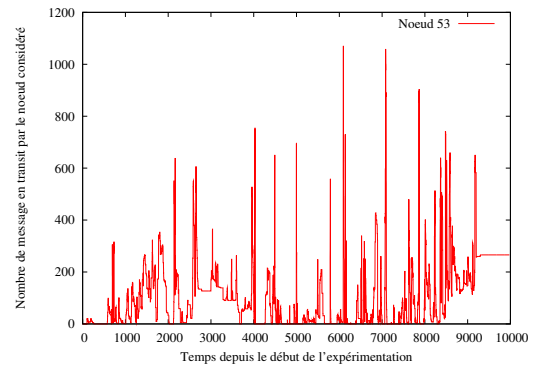
À partir de l'exécution du programme calculant le temps minimum pour transmettre un message à partir d'un temps donné, il est facile de déterminer en plus les chemins que le message a pu prendre. Lorsque à temps donné un nouveau contact fournit un temps plus court, on propage les anciens intermédiaires et on l'ajoute aussi. Cela correspond à compter le nombre de messages en transit qui seront propagés utilement pour fournir le message à un nouveau destinataire ne l'ayant pas encore.

La figure 4.12 présente la valeur de cette centralité au cours du temps pour les nœuds 3 et 53. Le graphique obtenu est similaire à celui de l'impact sur la diffusion. On observe d'importants pics de nombre de messages en transit sur de courtes périodes, séparés par des périodes avec peu de messages. On observe aussi que pour le nœud 53 le nombre de messages en transit paraît plus important que pour le nœud 3 sur l'ensemble de l'expérimentation. L'observation de graphiques obtenus pour d'autres nœuds montre bien la même typologie, on observe néanmoins une bien plus grande variabilité.

L'observation de la valeur moyenne de la valeur sur l'ensemble de l'expérimentation montre bien cette plus grande variabilité. La figure 4.13 représente les valeurs moyennes en fonction de l'écart-type mesuré sur l'ensemble de l'expérimentation. On peut voir que bon nombre de nœuds possèdent des valeurs relativement bien plus faibles et proches de 0, comparé avec l'impact du retrait sur la diffusion. Quatre nœuds se détachent un peu du lot avec un nombre élevé de messages en transit, mais l'ensemble est réparti assez uniformément et présente aussi une forte corrélation entre l'écart-type et la moyenne.



(a) Nœud 3 retiré.



(b) Nœud 53 retiré.

FIGURE 4.12 – Évolution du nombre de messages en transit par nœud lors d’une diffusion utilisant les plus courts chemins.

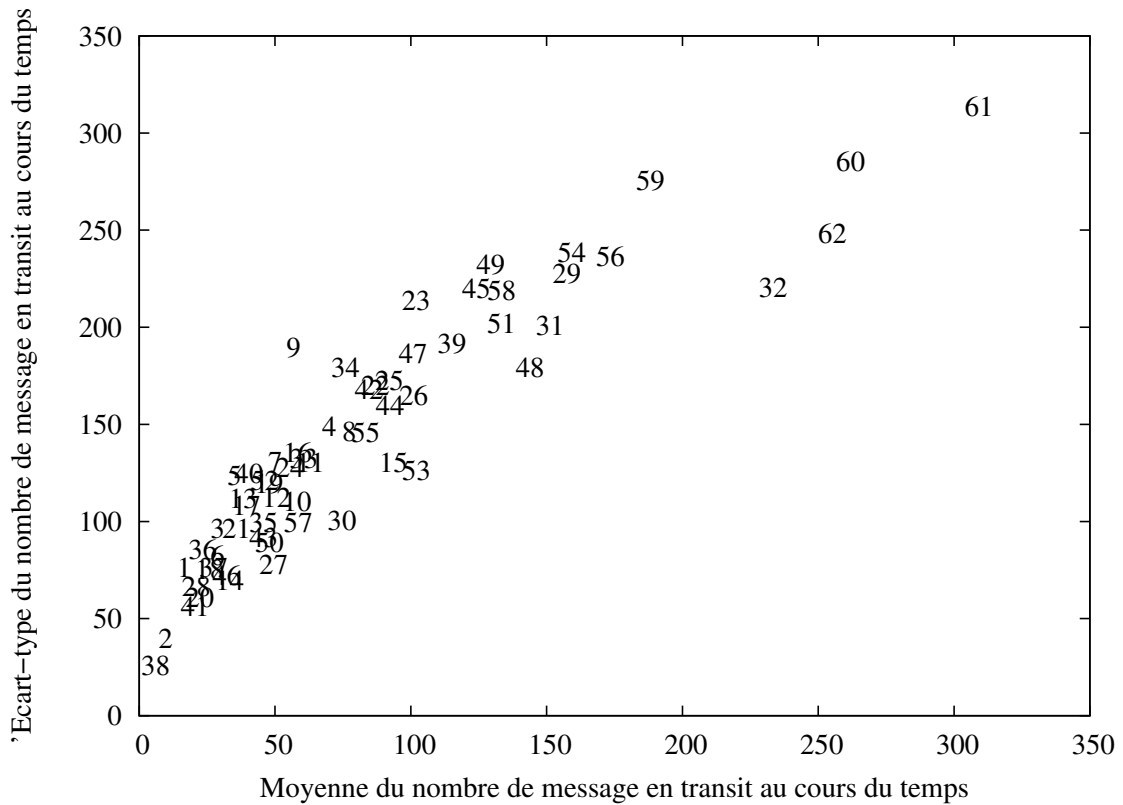


FIGURE 4.13 – Représentation des moyennes et écarts-types du nombre de messages en transit par nœud lors de la diffusion au cours de l’expérimentation rollernet. Le label indique le nœud correspondant à la mesure.

## Chapitre 5

# Analyse des résultats et discussion

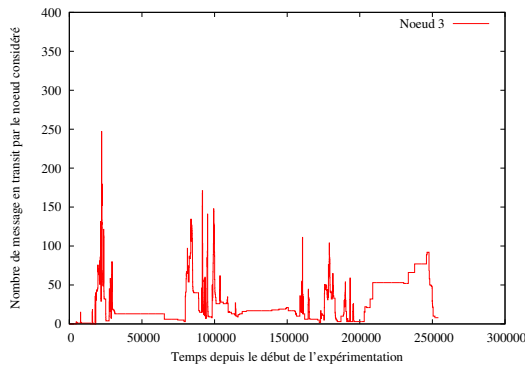
### 5.1 Résultats pour différents jeux de données

#### 5.1.1 Infocom

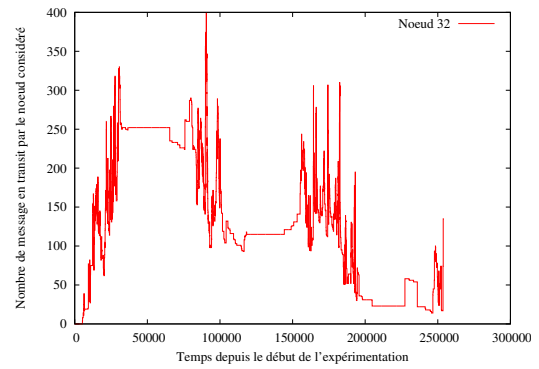
Pour le jeu de données Infocom, les résultats obtenus pour un temps de départ fixé étaient décevants. Peu d'information pouvait en être retirée, le retrait des nœuds n'impactant pas sur la diffusion de manière notable (de l'ordre de l'impact de la normalisation). En effet, l'expérience se déroule sur plusieurs jours, et les longs temps d'attente pour transmettre un message causent quelques valeurs extrêmes avec des retards de transmission dépassant plus d'un jour dans certains cas. De plus, sans doute à cause de la structure de la conférence réunissant des groupes de personnes pendant de longues périodes le jour, se créent des phénomènes de communauté dans laquelle chacun est remplaçable et des moments d'interaction forts comme aux repas permettant un échange important dans le réseau. Malheureusement, à cause d'un manque de temps face à de longues exécutions de programmes, les résultats de la valeur de l'impact au cours du temps n'est pas disponible. On ne peut pas dire grand chose et encore moins avec de bons arguments sur l'évolution de cet impact et si une information utile sur le comportement des nœuds peut en être retirée. Les résultats sont en cours de calcul mais ne seront pas disponibles prochainement, la durée importante de l'expérimentation décuplant le temps de travail.

Par ailleurs les résultats pour la centralité dynamique proposée sont eux disponibles. La figure 5.1 montre l'exemple de l'évolution du nombre de messages en transit au cours de l'expérimentation pour les nœuds 3 et 32. Le nœud 32 présente la plus grande moyenne du groupe, alors que le nœud 3 a une valeur relativement faible. On observe bien les périodes de jour et de nuit, le jour montrant une grande variabilité et la nuit peu d'évolution. Pour le nœud 32 on observe que le nombre de messages en transit reste élevé sur de grandes périodes, contrairement aux résultats obtenus pour Rollernet. L'observation de la moyenne du nombre de messages au cours de l'expérimentation en fonction de l'écart-type est similaire aux résultats obtenus pour Rollernet. On observe une grande variabilité avec des résultats proches de zéro, mais avec une répartition relativement homogène. De même il existe une bonne corrélation linéaire entre la moyenne et l'écart-type.





(a) Nœud 3 retiré.



(b) Nœud 32 retiré.

FIGURE 5.1 – Évolution du nombre de messages en transit par nœud lors d’une diffusion utilisant les plus courts chemins.

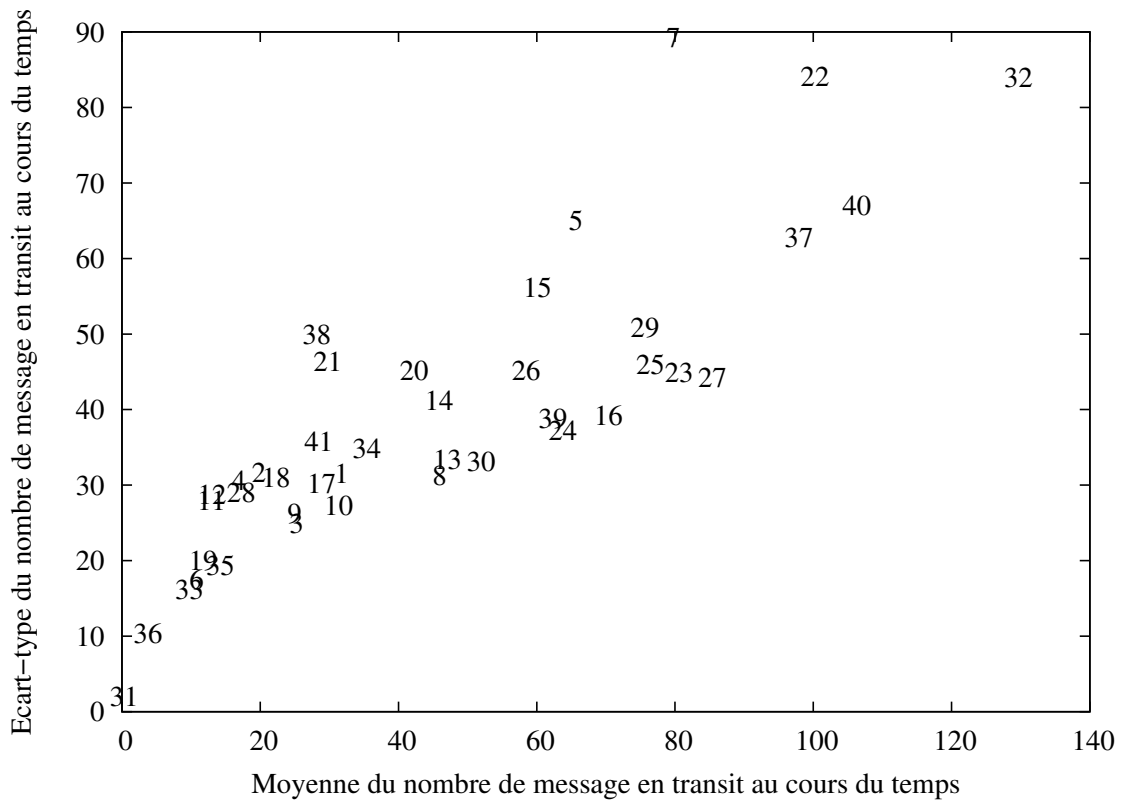


FIGURE 5.2 – Représentation des moyennes et écarts-types du nombre de messages en transit par nœud lors de la diffusion au cours de l’expérimentation Infocom. Le label indique le nœud correspondant à la mesure.

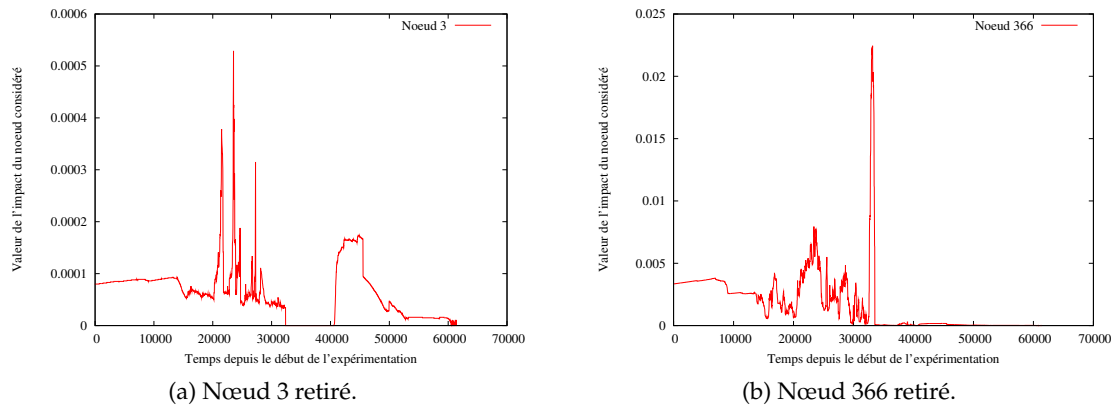


FIGURE 5.3 – Évolution de la valeur mesurée de l’impact du nœud retiré sur la diffusion selon la méthode proposée.

### 5.1.2 Pnas

Pour le jeu de donné Pnas, l’étude détaillée pour un temps de départ donné n’a pas pu être réalisée. Le jeu de donné présente un relativement grand nombre de nœuds, environ 800, et les méthodes utilisées pour générer les distributions était algorithmiquement trop coûteuses pour pouvoir le réaliser pour un grand nombre de nœuds. La mesure de l’impact au cours du temps du retrait d’un nœud a pu être réalisée pour la majorité des nœuds, le reste étant en cours de calcul. Malgré le grand nombre de nœuds, la durée relativement courte de l’expérience et le grand échelonnage temporel limitent le nombre de cas à étudier. La figure 5.3 montre l’évolution de cet impact pour 2 nœuds. Le nœud 366 est choisi pour sa grande moyenne et parce qu’il est représentatif de comportements extrêmes observés dus à de longues périodes peu actives au début et à la fin de l’expérience. Le nœud 3 est plus représentatif d’un comportement moyen, on observe quelques variations lentes de l’impact mesuré, des périodes de grande variabilité avec comme pour les autres jeux de données de grand pics soudains et aussi des périodes d’impact du retrait du nœud à peu près nul.

La figure 5.4 montre la valeur moyenne en fonction de l’écart-type pour les nœuds du réseau ayant fourni leur résultats. On observe une poignée de nœuds avec un impact bien plus important et un grand groupe avec une valeur en rapport très faible. Un regard rapproché montre une répartition qui semble plus ou moins logarithmique, contrairement à Rollernet, avec un nombre décroissant de nœuds au fur et à mesure que l’impact augmente. On observe bien toujours une forte corrélation entre moyenne et écart-type, bien que un cône semble se dessiner, avec une répartition bien plus importante sur les bordures plutôt que au centre de ce cône.

En ce qui concerne la mesure du nombre de messages en transit, les résultats sont disponibles pour l’ensemble des nœuds du réseau. La figure 5.5 présente les deux mêmes nœuds avec un comportement extrême conservé pour le nœud 366. On peut toutefois observer, et ce dans la majorité des cas, comme pour le nœud 3, une valeur généralement faible avec quelques pics très importants mais beaucoup moins nombreux que pour les autres jeux de données. La faible durée de l’expérimentation ne permet pas de pouvoir analyser plus en détail ce faible nombre relatif de pics vis à vis des autres jeux de données. La figure 5.6 représente la valeur moyenne et l’écart-type du nombre de messages au cours de l’expérimentation. On observe

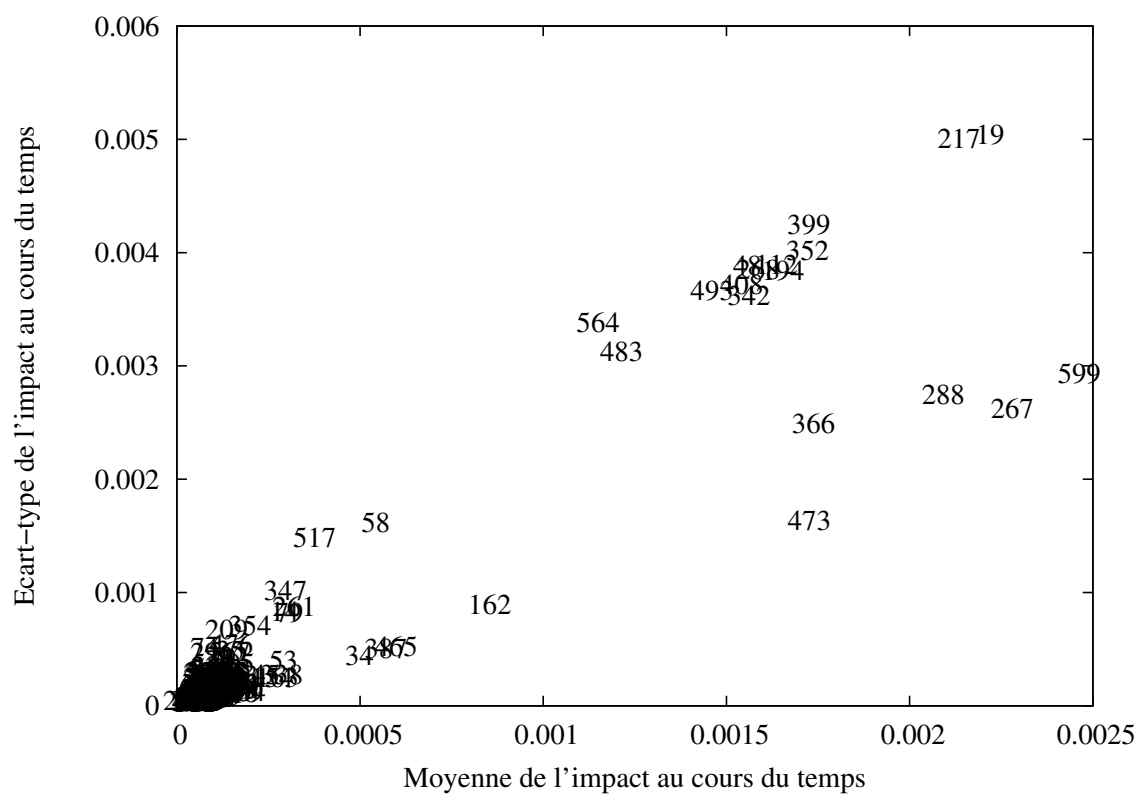


FIGURE 5.4 – Représentation des moyennes et écarts-types de la valeur de l'impact du retrait des nœuds au cours de l'expérimentation Pnas. Le label indique le nœud retiré correspondant à la mesure.

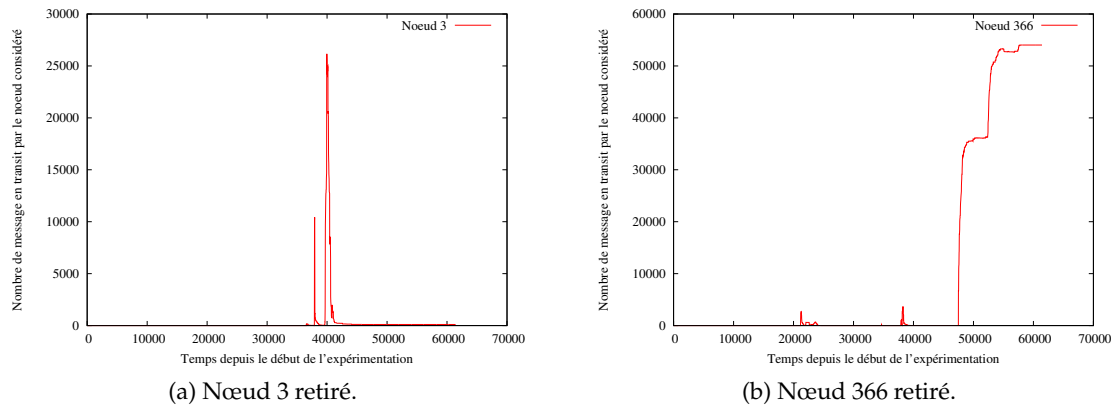


FIGURE 5.5 – Évolution du nombre de messages en transit par nœud lors d’une diffusion utilisant les plus courts chemins.

une structure globale très similaire au résultat obtenu pour l’impact du retrait d’un nœud. Une partie des nœuds avec une grande moyenne présentent aussi un comportement extrême, sans doute dû à des effets de début et fin d’expérimentation peu pertinents. La même répartition presque logarithmique est observable, bien que moins prononcée, et la répartition en cône est bien visible, montrant le même type de corrélation entre la moyenne et l’écart-type.

## 5.2 Comparaison entre différentes mesures de centralités dynamiques

Les deux notions proposées de centralité, mesurent chacune à leur manière l’importance d’un nœud donné pour la transmission de message *a posteriori*. Intuitivement, un nœud impactant fortement sur la diffusion par son absence laisse penser qu’un grand nombre de messages transite par lui. Réciproquement, si beaucoup de messages transitent par un nœud on peut penser que, si on l’enlève, les messages devront trouver un chemin plus long pour atteindre sa destination et que son retrait impactera fortement sur les temps de diffusion.

Il serait dommage de ne pas essayer de comparer les résultats obtenus par chacune de ces méthodes. Une juxtaposition de leur évolution pour certains nœuds pris pour exemple ne donne pas des résultats très satisfaisants. En effet, comme on a pu le voir précédemment, les deux mesures proposent des évolutions de même type pour les jeux de données dont les résultats sont disponibles à l’heure actuelle. Mais en comparant point à point les deux mesures on se rend compte que les pics ne se situent pas aux mêmes endroits, et qu’il ne semble pas y avoir de décalage plus ou moins régulier entre elles. Il faudrait tenter de comparer ces valeurs temporelles de manière plus automatisée pour d’une part quantifier la ressemblance et d’autre part généraliser la comparaison à tous les résultats obtenus. Il est possible de comparer visuellement quelques dizaines de graphiques, mais le faire pour plusieurs centaines n’est pas envisageable.

Toutefois, on avait pu remarquer, avec l’exemple donné pour Rollernet, que le nœud avec l’impact de plus grande amplitude possédait aussi le plus grand nombre de messages en transit. On peut donc essayer de comparer les valeurs moyennes de ces mesures afin de voir si une corrélation peut être observée. La figure 5.7 présente cette visualisation pour l’ensemble des nœuds du réseau. On observe pas de corrélation évidente entre les deux mesures. Les nœuds

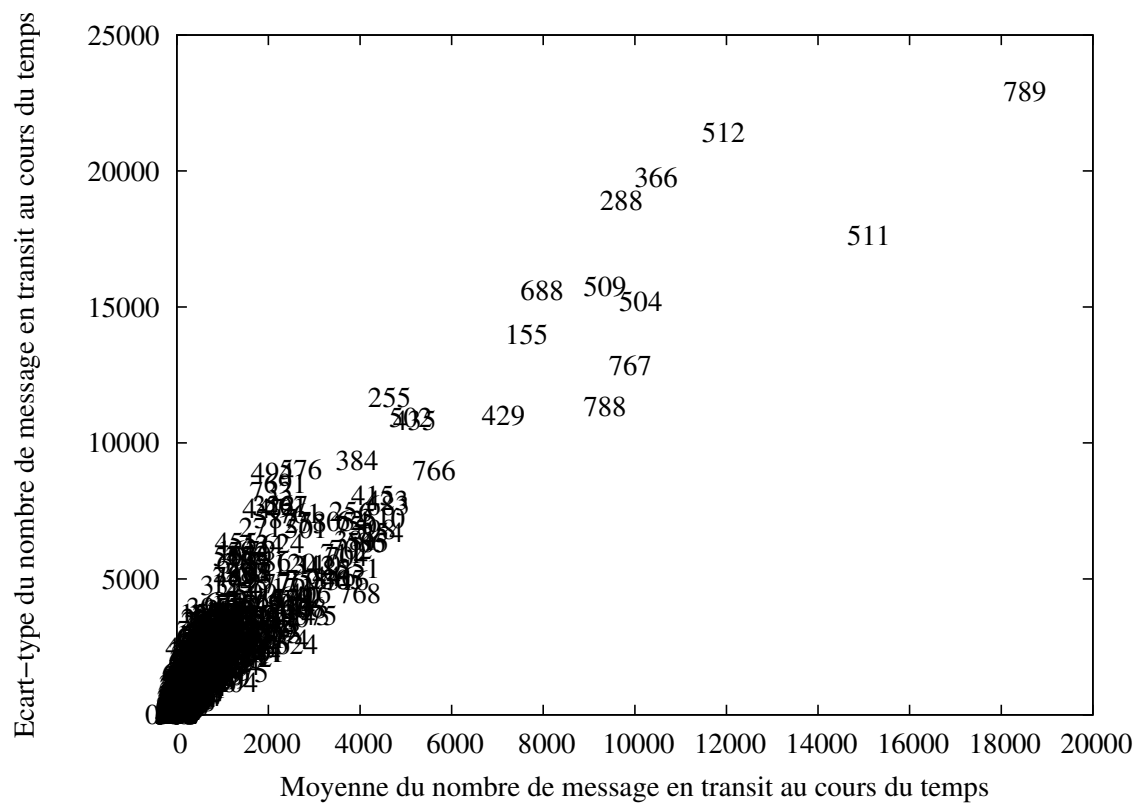


FIGURE 5.6 – Représentation des moyennes et écarts-types du nombre de messages en transit par nœud lors de la diffusion au cours de l'expérimentation Pnas. Le label indique le nœud correspondant à la mesure.

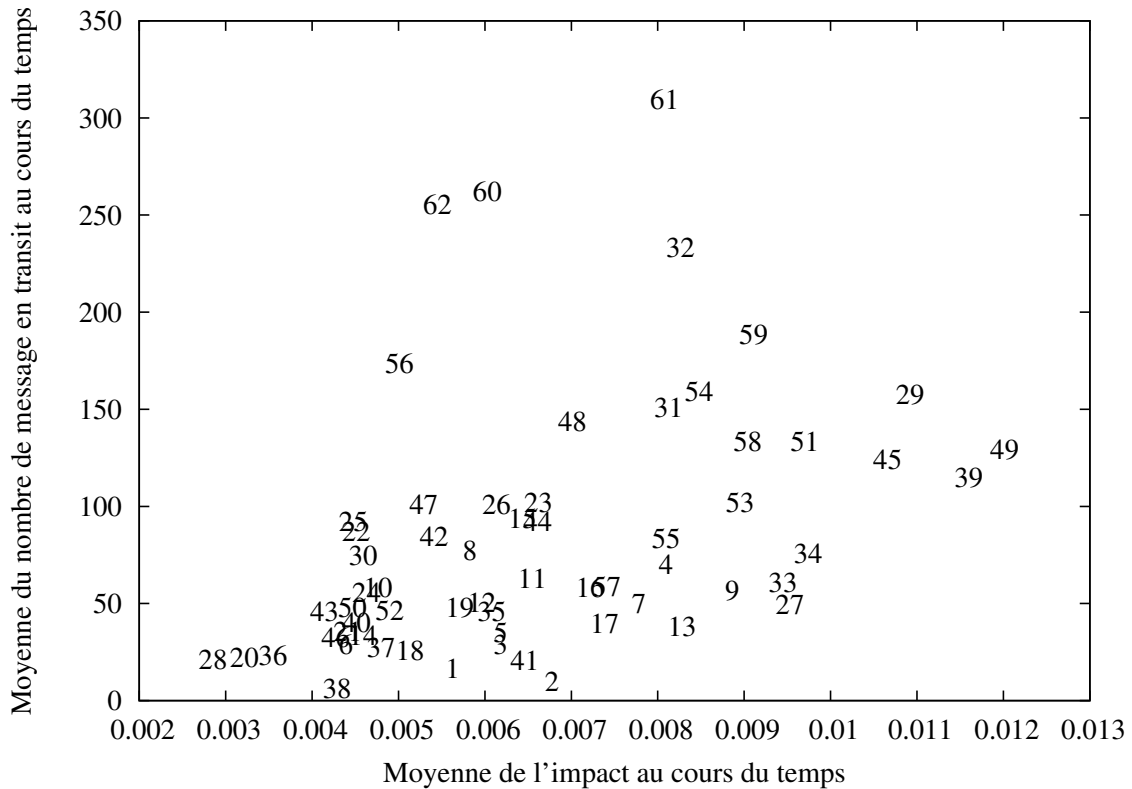


FIGURE 5.7 – Représentation croisée des moyennes de la valeur de l'impact du retrait d'un nœud et du nombre de messages en transit pour le jeu de données Rollernet. Le label indique le nœud correspondant à la mesure.

avec le plus fort impact lors de leur retrait du réseau ont un nombre moyen de messages qui transitent par eux. De même les nœuds avec le plus grand nombre de messages transitant par eux ont aussi un impact moyen sur la diffusion par leur retrait du réseau. On observe quand même que pour les nœuds avec des valeurs mesurées moyennes à faible, une certaine corrélation peut être observée, bien que faible.

La figure 5.8 représente elle la valeur moyenne de l'impact du retrait d'un nœud en fonction du nombre moyen de messages en transit par ce nœud pour les résultats disponibles. Avec les valeurs extrêmes observées par ces deux mesures, dont la pertinence reste toujours à vérifier, on pourrait presque croire à une corrélation inversée. En effet on peut observer des valeurs élevées pour les deux mesures de centralités pour seulement deux nœuds. La grande majorité des nœuds ayant une forte centralité pour l'une des deux mesures possède une valeur très faible pour l'autre. Il reste que la majorité des nœuds possèdent une valeur faible pour les deux mesures proposées, mais un zoom sur ces valeurs montre une répartition homogène dans l'espace. On ne peut donc pas, sur cette comparaison sommaire des deux mesures de centralités proposées, observer quelque corrélation que ce soit, et ce quels que soient les intervalles de valeurs comparés.

Les comparaisons proposées restent limitées, mais les premières observations faites sont plus que contre intuitives et aucune corrélation notable ne peut être observée à ce niveau-là. Les résultats pour d'autres jeux de données seraient intéressants pour essayer de généraliser ces premières observations.



## Chapitre 6

# Conclusion et perspectives

Nous avons recherché des moyens de caractérisation de réseaux de contacts tenant compte de leur dynamique qui ne permet pas d'utiliser les mesures de distance développées pour des réseaux statiques. J'ai participé au développement de nouveaux indicateurs au sein de mon équipe d'accueil. Les premiers résultats présentés dans ce mémoire sont plutôt encourageants car l'approche implémentée permet de détecter visuellement des noeuds "importants" pour la diffusion d'information à travers le réseau.

Je pense que ce premier travail pourrait être poursuivi sur plusieurs plans :

- en continuant l'analyse entreprise sur des jeux de données différents.
- en cherchant des indicateurs permettant de comparer différents types de réseaux de contacts entre eux, du point de vue de leurs caractéristiques de distribution temporelles de mesures de centralité dynamique pour chaque noeud, ce qui passerait par le développement de méthodes statistiques appropriées.
- en comparant les résultats obtenus par notre approche avec ceux obtenus par d'autres méthodes d'analyse de réseaux de contact.
- en cherchant à formaliser la méthodologie présentée, basée actuellement sur une démarche conceptuelle et algorithmique.
- en proposant de coupler des méthodes de visualisation de réseaux dynamiques et des méthodes de détection automatiques de noeuds centraux pour le routage dans des réseaux de contacts.



# Bibliographie

- [1] Lamia Benamara. *Dynamique des graphes de terrain : caractérisation et étude du biais lié à la mesure*. PhD thesis, UPMC, 2011.
- [2] B. Bui Xuan, Afonso Ferreira, and Aubin Jarry. Computing shortest, fastest, and foremost journeys in dynamic networks. Technical Report RR-4589, INRIA, October 2002.
- [3] Roberta Calegari, Mirco Musolesi, Franco Raimondi, and Cecilia Mascolo. CTG : a connectivity trace generator for testing the performance of opportunistic mobile systems. In *Proceedings of the the 6th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering, ESEC-FSE '07*, pages 415–424, New York, NY, USA, 2007. ACM.
- [4] A. Clauset and N. Eagle. Persistence and periodicity in a dynamic proximity network. In *DIMACS. Workshop*, 2007.
- [5] Nathan Eagle and Alex (Sandy) Pentland. Reality mining : sensing complex social systems. *Personal Ubiquitous Comput.*, 10 :255–268, March 2006.
- [6] Peter Grindrod, Mark C. Parsons, Desmond J. Higham, and Ernesto Estrada. Communicability across evolving networks. *Phys. Rev. E*, 83 :046120, Apr 2011.
- [7] Pan Hui, Augustin Chaintreau, James Scott, Richard Gass, Jon Crowcroft, and Christophe Diot. Pocket switched networks and human mobility in conference environments. In *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking, WDTN '05*, pages 244–251, New York, NY, USA, 2005. ACM.
- [8] David Kempe, Jon Kleinberg, and Amit Kumar. Connectivity and inference problems for temporal networks. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing, STOC '00*, pages 504–513, New York, NY, USA, 2000. ACM.
- [9] Hyounghick Kim and Ross Anderson. Temporal node centrality in complex networks. *Phys. Rev. E*, 85 :026107, Feb 2012.
- [10] Hyounghick Kim, John Tang, Ross Anderson, and Cecilia Mascolo. Centrality prediction in dynamic human contact networks. *Computer Networks*, 56(3) :983 – 996, 2012. <ce :title>(1) Complex Dynamic Networks (2) P2P Network Measurement</ce :title>.
- [11] D. Kotz and T. Henderson. Crawdad : A community resource for archiving wireless data at dartmouth. *Pervasive Computing, IEEE*, 4(4) :12–14, 2005.
- [12] M. Musolesi, S. Hailes, and C. Mascolo. Adaptive routing for intermittently connected mobile ad hoc networks. *World of Wireless Mobile and Multimedia Networks, 2005.*, pages 183–189, 2005.

- [13] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W. Feldman, and James H. Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51) :22020–22025, 2010.
- [14] Antoine Scherrer, Pierre Borgnat, Eric Fleury, Jean-Loup Guillaume, and Céline Robardet. Description and simulation of dynamic mobility networks. *Computer Networks (1976)*, 2008.
- [15] James Scott, Richard Gass, Jon Crowcroft, Pan Hui, Christophe Diot, and Augustin Chaintreau. CRAWDAD trace cambridge/haggle/imote/infocom2006 (v. 2009-05-29). Downloaded from <http://crawdad.cs.dartmouth.edu/cambridge/haggle/imote/infocom2006>, May 2009.
- [16] John Tang, Mirco Musolesi, Cecilia Mascolo, Vito Latora, and Vincenzo Nicosia. Analysing information flows and key mediators through temporal centrality metrics. In *Proceedings of the 3rd Workshop on Social Network Systems, SNS '10*, pages 3 :1–3 :6, New York, NY, USA, 2010. ACM.
- [17] Pierre Ugo Tournoux, Jérémie Leguay, Marcelo Dias de Amorim, Farid Benbadis, Vania Conan, and John Whitbeck. The accordion phenomenon : Analysis, characterization, and impact on DTN routing. In *In Proceedings of the 28rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, pages 1116–1124. IEEE, 2009.

# Table des figures

2.1	Un graphe $\mathcal{Z}$ avec un étiquetage invalidant le théorème de Menger . . . . .	6
3.1	Temps moyen requis pour transmettre un message dans le réseau, en utilisant les plus courts chemins. . . . .	13
4.1	Distribution cumulative normalisée des temps requis pour transmettre l'information de tout le monde vers tout le monde, avec tous les nœuds et sans les nœuds 3, 29, 32 et 53. . . . .	15
4.2	Représentation conjointe de la moyenne et de l'écart-type des différences entre la distribution cumulative du réseau complet et celle du réseau amputé d'un de ses nœuds. Le label correspond au numéro du nœud retiré du réseau. . . . .	16
4.3	Distributions cumulatives normalisées des temps requis pour transmettre l'information depuis des nœuds uniques vers tout le monde, avec et sans les nœuds 3, 29, 32 et 53. . . . .	17
4.4	Représentation conjointe de la moyenne et l'écart-type des différences entre la distribution cumulative du réseau complet et celle du réseau amputé d'un de ses nœuds pour une diffusion depuis une source unique. Le label correspond au numéro du nœud retiré du réseau, le nœud source de la diffusion n'est pas spécifié dans cette représentation. . . . .	18
4.5	Matrice représentant la catégorie associée pour tout couple donné de nœud retiré du réseau et nœud source de la diffusion. La catégorie correspondant à des variations très faibles vis à vis du réseau complet est représentée par un blanc. La ligne d'ordonnée 70 correspond à la catégorie obtenue pour une diffusion de tout le monde vers tout le monde. A comparer avec la figure 4.7 où les mêmes données sont représentées pour le numéro du nœud de destination de la diffusion. . . . .	19
4.6	Représentation conjointe de la moyenne et l'écart-type des différences entre la distribution cumulative du réseau complet et celle du réseau amputé d'un de ses nœuds pour un envoi de message de tout le monde vers un unique récepteur. Le label correspond au numéro du nœud retiré du réseau, le nœud de destination des messages n'est pas spécifié dans cette représentation. . . . .	21
4.7	Matrice représentant la catégorie associée pour tout couple donné de nœud retiré du réseau et nœud destinataire des messages. La catégorie correspondant à des variations très faibles vis à vis du réseau complet est représentée par un blanc. La ligne d'ordonnée 70 correspond à la catégorie obtenue pour une diffusion de tout le monde vers tout le monde. . . . .	22

4.8	Réseau utilisé pour l'exemple détaillé du déroulement de l'algorithme 1. Pour chaque arête sont indiqués les temps d'activité de celle-ci. . . . .	25
4.9	Valeurs en fonction du temps, dans l'ordre décroissant calculé par le programme, de la matrice <i>CommunicationTime</i> . À la ligne <i>i</i> et colonne <i>j</i> on peut lire le temps nécessaire pour transmettre un message du nœud <i>i</i> vers le nœud <i>j</i> . En légende sous chaque matrice est donné le temps correspondant et les composantes connexes pour ce temps donné (sauf singletons). . . . .	25
4.10	Évolution de la valeur mesurée de l'impact du nœud retiré sur la diffusion selon la méthode proposée. . . . .	26
4.11	Représentation des moyennes et écarts-types de la valeur de l'impact du retrait des nœuds au cours de l'expérimentation rollernet. Le label indique le nœud retiré correspondant à la mesure. . . . .	27
4.12	Évolution du nombre de messages en transit par nœud lors d'une diffusion utilisant les plus courts chemins. . . . .	29
4.13	Représentation des moyennes et écarts-types du nombre de messages en transit par nœud lors de la diffusion au cours de l'expérimentation rollernet. Le label indique le nœud correspondant à la mesure. . . . .	29
5.1	Évolution du nombre de messages en transit par nœud lors d'une diffusion utilisant les plus courts chemins. . . . .	31
5.2	Représentation des moyennes et écarts-types du nombre de messages en transit par nœud lors de la diffusion au cours de l'expérimentation Infocom. Le label indique le nœud correspondant à la mesure. . . . .	31
5.3	Évolution de la valeur mesurée de l'impact du nœud retiré sur la diffusion selon la méthode proposée. . . . .	32
5.4	Représentation des moyennes et écarts-types de la valeur de l'impact du retrait des nœuds au cours de l'expérimentation Pnas. Le label indique le nœud retiré correspondant à la mesure. . . . .	33
5.5	Évolution du nombre de messages en transit par nœud lors d'une diffusion utilisant les plus courts chemins. . . . .	34
5.6	Représentation des moyennes et écarts-types du nombre de messages en transit par nœud lors de la diffusion au cours de l'expérimentation Pnas. Le label indique le nœud correspondant à la mesure. . . . .	35
5.7	Représentation croisée des moyennes de la valeur de l'impact du retrait d'un nœud et du nombre de messages en transit pour le jeu de donné Rollernet. Le label indique le nœud correspondant à la mesure. . . . .	36
5.8	Représentation croisée des moyennes de la valeur de l'impact du retrait d'un nœud et du nombre de messages en transit pour le jeu de donné Pnas. Le label indique le nœud correspondant à la mesure. . . . .	37