



# Réseaux dynamiques de terrain : caractérisation et propriétés de diffusion en milieu hospitalier

Lucie Martinet

## ► To cite this version:

Lucie Martinet. Réseaux dynamiques de terrain : caractérisation et propriétés de diffusion en milieu hospitalier. Autre [cs.OH]. Ecole normale supérieure de lyon - ENS LYON, 2015. Français. NNT : 2015ENSL1010 . tel-01208023

**HAL Id: tel-01208023**

**<https://theses.hal.science/tel-01208023>**

Submitted on 1 Oct 2015

**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.

# THÈSE

en vue de l'obtention du grade de

**Docteur de l'Université de Lyon, délivré par l'École Normale Supérieure de Lyon**

**Discipline :** Informatique

**Laboratoire :** Laboratoire de l'Informatique du Parallélisme (LIP)  
UMR5668 : INRIA – ENS de Lyon – CNRS – UCBL Lyon I  
Université de Lyon – Labex Milyon

**École Doctorale :** École doctorale en informatique et mathématiques de Lyon (ED512)

présentée et soutenue publiquement le 18 septembre 2015

par Madame Lucie MARTINET

---

Réseaux dynamiques de terrain: caractérisation et propriétés de diffusion en milieu hospitalier.

---

**Directeur de thèse :** M. Éric FLEURY

**Co-encadrant de thèse :** M. Christophe CRESPELLE

**Après l'avis de :**

M. Jean-Loup GUILLAUME

Mme Clémence MAGNIEN

**Devant le jury composé de :**

M. Christophe CRESPELLE, maître de conférences à l'Université Claude Bernard Lyon 1, Co-encadrant

M. Eric FLEURY, professeur à l'ENS de LYON/ Chaire INRIA, Directeur

M. Jean-Loup GUILLAUME, professeur à l'Université de La Rochelle, Rapporteur

Mme Clémence MAGNIEN, directeur de recherche CNRS, Laboratoire d'Informatique de Paris 6, Rapporteur

M. Thomas NOEL, professeur à l'Université de Strasbourg, Examineur

M. Christophe PRIEUR, maître de conférences à Télécom ParisTech, Examineur

A ma formidable famille : ma mère, mon père et mon frère  
parce que sans eux, rien n'aurait été possible...



# Remerciements

Je ne pourrais pas commencer ce manuscrit sans remercier toutes les personnes qui ont contribué, de près ou de loin, à sa construction et sans qui cette thèse n'aurait jamais existé. J'espère n'oublier personne, mais vous êtes tellement nombreux ! Pardonnez-moi, ceux qui n'auront pas la joie de découvrir leur nom en ces quelques lignes.

Tout d'abord, je tiens à remercier sincèrement mes rapporteurs Jean-Loup Guillaume et Clémence Magnien pour leur relecture attentive de mon manuscrit. Je vous remercie véritablement pour votre patience et les remarques pertinentes que vous m'avez faites très rapidement et qui ont permis d'améliorer considérablement le manuscrit. J'aimerais aussi remercier mon encadrant Christophe Crespelle pour les nombreuses remarques judicieuses qu'il m'a faites tout au long de l'écriture du manuscrit ainsi que pour les discussions scientifiques que nous avons pu avoir au cours de ma thèse. Je le remercie aussi chaleureusement pour ses cours d'humours, diffusés tout au long de ma thèse. Je remercie également Eric Fleury, mon directeur de thèse, pour m'avoir permis de traiter ce sujet difficile et pourtant plein d'avenir. Je le remercie particulièrement pour son humanité et la très grande liberté d'organisation qu'il m'a laissée. Je remercie mes rapporteurs et examinateurs (Clémence Magnien, Jean-Loup Guillaume, Thomas Noel et Christophe Prieur) pour leurs questions concises pour la plupart et très pertinentes. Et je tiens à remercier les 6 membres de mon jury dont j'ai particulièrement apprécié la bienveillance tout au long de nos échanges ainsi que le jour de ma soutenance.

Durant ma thèse, j'ai vu se succéder un certain nombre de co-bureaux. Je voudrais tous les remercier pour leur bonne humeur et toutes les infos utiles que nous avons pu échanger. Je tiens particulièrement à remercier Benjamin qui m'a fait découvrir nombre de petits trucs utiles du point de vue technique mais aussi culturel. Je n'oublierai jamais les longues discussions passionnées et passionnantes que nous avons pu échanger pendant 3 ans. Et merci pour toutes tes aides matérielles, en particulier pour le jour de la soutenance ! Et aussi pour les debuggages réguliers, les infos sur LaTeX et toute l'aide sur les bug logiciels. Merci pour les remontages de moral et les quelques petits craquages qui nous ont échappés. Et merci pour ta tarte au citron.

Je n'oublierai pas non plus Yannick et Matteo, les pauvres co-bureaux qui m'ont subie pendant ma rédaction. Merci d'avoir participé à mes craquages réguliers, de m'avoir redonné la volonté d'aller jusqu'au bout et rendu ma bonne humeur ! Merci aussi à Brieuc qui a gentiment relu une partie de ma thèse. Tu ne peux pas savoir comme cela m'a rendu confiance !! Merci aussi à Qinna, qui a veillé à mon bien être durant la première année de thèse et a régulièrement pris de mes nouvelles. Merci à Guillaume A.G pour m'avoir aidée à régler les derniers détails peu avant la soutenance et avoir partagé son expérience de jeune docteur. Merci à Yongjun et Nghi pour leur soutien moral et leur nombreux conseils sur mes transparents. Merci de m'avoir permis de tester l'acoustique de la salle avant la soutenance !

Je tiens aussi particulièrement à remercier les permanents de mon équipe qui m'ont soutenue, conseillée et donné tout un tas de petits trucs pour me faire gagner du temps : Thomas,

---

Isabelle, Anthony, Paulo et Marton. Je commence par Thomas qui m'a fait confiance pour ses cours de réseaux et m'a fait participé à la mise en place de son nouveau cours. Il a aussi pris le temps de m'expliquer plein de notions, aussi en rapport avec sa recherche, pour que je sois à l'aise avec le sujet et qui m'a remonté le moral un certain nombre de fois. Merci Thomas ! Isabelle qui a été très présente à certains moments clefs et a aussi su partager son expérience et ses connaissances pour l'enseignement ET la recherche. Anthony qui m'a été d'une grande aide pour les dossiers, et avec qui j'ai été très heureuse de faire l'enseignement. Et qui m'a laissé gagner à la course en vélo. Je n'oublie pas Marton qui m'a conseillée scientifiquement et m'a obligée à parfaire mon anglais, ce que j'ai fait avec un immense plaisir. Et Paulo, qui m'a soutenue fortement sur la fin et m'a prodigué multitude de petits conseils hyper utiles que je n'oublierai pas !

Je voudrais aussi remercier d'autres personnes du LIP, avec qui j'ai eu la chance de travailler pour l'enseignement, et qui m'ont considérablement soutenue et conseillée. Laure (mention spéciale !) qui m'a repêchée par moments et a pris de son temps pour partager ses expériences en enseignement, recherche, relations humaines et musicales ! Qui m'a proposé de partager son bureau pour m'éviter des allers/retour inutiles entre Gerland et Villeurbanne. Qui m'a aussi proposé son aide pour la relecture du manuscrit. Laure : je te suis infiniment reconnaissante ! Eric T., avec qui j'ai eu un immense plaisir de travailler sur les graphes et des notions d'algorithmique peu classiques. Qui m'a souvent soutenue moralement et proposé son aide pour la relecture du manuscrit. J'ai appris beaucoup de choses à tes côtés Eric (T.) et j'ai vraiment apprécié de faire tes TDs, même si on manquait parfois de coordination. Damien Stehlé, avec son organisation sans faille, m'a permis d'enseigner efficacement sans perdre de temps pendant ma dernière année de monitorat et qui m'a fait confiance pour mener mon groupe en autonomie. Je tiens particulièrement à remercier Guillaume Hanrot pour le temps et l'énergie qu'il a consacrés à débayer le terrain glissant de ma situation administrative !

Et aussi des personnes du LIP et de l'IXXI qui m'ont soutenue gratuitement, par pur gentillesse : Natacha, Stéphan T., Nicolas S., Nicolas T., Pablo, Claire, Arnaud, Nathalie R, Nathalie A., Stephane et Stéphane (F. et G.). Et bien sûr tous les doctorants qui ont partagé avec moi leur joies et leurs souffrances, dans une ambiance joviale et dynamique. Je tiens aussi à remercier toute l'équipe d'assistants du labo, du département informatique et de l'IXXI pour leur aide, leur disponibilité et leur bonne humeur : Marie, Damien, Laetitia, Séverine, Sylvie, Evelyne, Nicole, Samantha, Catherine, Chiraz et Edwidge. Merci aussi au service des relations internationales pour leur soutien moral, leur qualité d'écoute et pour avoir stocké toutes les affaires dont j'avais besoin pour ma soutenance !

L'ambiance à l'IXXI n'aurait pas été la même sans la présence de l'équipe du CITI et certains stagiaires : en particulier, merci pour toutes les âneries et arguments de mauvaise foi qui ont égaillés les moments de midi grâce à Robert (champion), Briec (presque exaequo avec Robert), Adrien , Damien R., Billel, Etienne.... et aussi aux deux Stéphane qui se sont parfois joints à la partie. Je pense aussi à Elodie P. qui a partagé un temps les difficultés du travail de recherche en autonomie quasi-totale à l'IXXI. Merci aussi à Amaury pour m'avoir soutenue, conseillée scientifiquement depuis la L3 et pour les délires partagés par chat tout au long de ses longues années de thèse qui nous ont fait passer par tous nos états.

Et merci à Jean-Pierre C. qui m'a bien conseillée pendant la dernière ligne droite et permis de faire quelques pauses nécessaires, le temps de discuter un peu. Et Christophe Coupé, mon maître à penser et maître de stage de M2, qui me redonne toujours plein d'énergie à chaque fois que l'on se croise par hasard !

Et bien sûr tous les autres, co-bureau ou non, qui se sont succédés dans mon bureau ou mon équipe de recherche et ont contribué à me motiver, me rassurer, me redonner courage

---

et espoir : Imen, Awa, Marwa, Mohammed, Tudor, Adreea Chis, Emilie, Eleonora, Doreid, Thiago, Przemek, Aniela, Oana....et tous les autres que j'oublie.

Je voudrais aussi remercier une bonne partie de l'équipe d'enseignement du département informatique de l'UCBL qui m'a accueilli à bras ouverts pour mon ATER, m'ont appris plein de choses, laissé prendre des initiatives, remonté le moral, et avec qui j'ai passé d'excellents moments. Je pense en particulier à Marie L. (mention spéciale) qui m'a appris plein de choses et m'a permis de partager mes connaissances, idées et expériences. Tu m'a donné la pêche et confiance en moi (et espoir aussi)! Merci Marie! Je n'oublierai pas Nathalie G. (mention spéciale), sans qui je ne serai jamais allée aussi loin dans les études : merci pour ta bienveillance, ton attention toujours renouvelée et tes cours qui m'ont poussée vers la voie informatique. Merci Nicolas L. pour ton attention, tes conseils et le temps que tu m'as consacré dans de longues discussions qui m'ont redonné de l'énergie et de l'espoir. Merci à Julie et Nicolas B. d'avoir pris des nouvelles régulièrement et pour vos conseils et remontages de moral non négligeables. Merci à Olivier G. d'avoir pris le temps de m'expliquer les TDs pour que je sois à l'aise face aux étudiants et merci de m'avoir fait confiance. Merci pour tous tes conseils et ta bienveillance. Merci à Elodie D., Stéphanie, Thierry E., Brigitte et tous les autres pour m'avoir si bien accueillie et intégrée et les discussions au coin café qui m'ont permis de me détendre et de me sentir chez moi. Merci à Jacques B. qui a eu la gentillesse de me confier ses clefs de bureau pour que je puisse travailler au calme dans de bonnes conditions, un jour de forte chaleur. Merci à Samir A. qui m'a fait confiance et sans qui je n'aurais pas non plus entamé des études en informatique : merci pour tes TDs inoubliables, ton humour et ta bonne humeur.

Je remercie aussi le groupe d'Antoine, J&F pour leurs mails d'encouragements et surtout pour l'ambiance, leur humanité, leur sincérité et leur super musique, qui m'a accompagnée tout au long de l'écriture du manuscrit. Et merci à John! (musique très adaptée à l'exercice de la rédaction!)

Merci aux personnes qui comptent pour moi d'avoir fait l'effort de se déplacer pour me soutenir avec la crainte de ne rien comprendre.

Et bien sûr, un grand, méga, méga grand merci à ma famille (big, big mention spéciale) qui m'a portée et supportée pendant toutes ces années d'études, pas toujours faciles. Qui a subi mes sautes d'humeur, mon stress récurrent et mes gros craquages. Merci pour tous les bons petits plats, et pas seulement ceux du pot de thèse! Merci à Antoine qui m'a fait répéter et répéter encore malgré sa fatigue et a su me conseiller aussi efficacement que mes encadrants, si ce n'est plus! Un grand merci, donc, à ma mère, mon père et mon frère, sans qui je n'aurais jamais eu le courage d'aller aussi loin! Et je n'oublie pas qu'ils ont réussi la performance de corriger les multiples fautes de frappe, de grammaire, d'orthographe, de math...d'un manuscrit de plus de 150 pages en moins de 3 jours! C'est grâce à eux trois que ce manuscrit existe! Merci, merci, merci!!!!

---

## Résumé

Durant cette thèse, nous nous sommes intéressés aux outils permettant d'extraire les propriétés structurelles et temporelles de réseaux dynamiques ainsi que les caractéristiques de certains scénarios de diffusion pouvant s'opérer sur ces réseaux. Nous avons travaillé sur un jeu de données spécifiques, issu du projet MOSAR, qui comporte entre autre le réseau de proximité des personnes au cours du temps durant 6 mois à l'hôpital de Berk-sur-mer. Ce réseau est particulier dans le sens où il est constitué de trois dimensions : temporelle, structurelle par la répartition des personnes en services et fonctionnelle car chaque personne appartient à une catégorie socio-professionnelle. Pour chacune des dimensions, nous avons utilisé des outils existants en physique statistique ainsi qu'en théorie des graphes pour extraire des informations permettant de décrire certaines propriétés du réseau. Cela nous a permis de souligner le caractère très structuré de la répartition des contacts qui suit la répartition en services et mis en évidence les accointances entre certaines catégories professionnelles. Concernant la partie temporelle, nous avons mis en avant l'évolution périodique circadienne et hebdomadaire ainsi que les différences fondamentales entre l'évolution des interactions des patients et celle des personnels. Nous avons aussi présenté des outils permettant de comparer l'activité entre deux périodes données et de quantifier la similarité de ces périodes. Nous avons ensuite utilisé la technique de simulation pour extraire des propriétés de diffusion de ce réseau afin de donner quelques indices pour établir une politique de prévention.

## Abstract

In this thesis, we focus on tools whose aim is to extract structural and temporal properties of dynamic networks as well as diffusion characteristics which can occur on these networks. We work on specific data, from the European MOSAR project, including the network of individuals proximity from time to time during 6 months at the Brek-sur-Mer Hospital. The studied network is notable because of its three dimensions constitution : the structural one induced by the distribution of individuals into distinct services, the functional dimension due to the partition of individual into groups of socio-professional categories and the temporal dimension. For each dimension, we used tools well known from the areas of statistical physics as well as graphs theory in order to extract information which enable to describe the network properties. These methods underline the specific structure of the contacts distribution which follows the individuals distribution into services. We also highlight strong links within specific socio-professional categories. Regarding the temporal part, we extract circadian and weekly patterns and quantify the similarities of these activities. We also notice distinct behaviour within patients and staff evolution. In addition, we present tools to compare the network activity within two given periods. To finish, we use simulations techniques to extract diffusion properties of the network to find clues to establish a prevention policy.



# Table des matières

<b>Remerciements</b>	<b>1</b>
<b>Résumé</b>	<b>6</b>
<b>Abstract</b>	<b>6</b>
<b>Introduction</b>	<b>7</b>
<b>1 Contexte</b>	<b>11</b>
1.1 Analyse des réseaux . . . . .	11
1.2 Mesures et analyses de réseaux dynamiques par technologie sans fil. . . . .	13
1.3 Contexte expérimental : projet MOSAR . . . . .	13
<b>2 Structure spatiale et fonctionnelle.</b>	<b>31</b>
2.1 Répartition des contacts par service. . . . .	32
2.2 Différences patients/personnels . . . . .	37
2.3 Introversion des services . . . . .	42
2.4 Interaction entre les groupes d'un réseau de contacts. . . . .	51
2.5 Analyse de la dimension fonctionnelle . . . . .	68
<b>3 Une structure temporelle à plusieurs échelles</b>	<b>89</b>
3.1 Evolution journalière . . . . .	89
3.2 Analyse à l'heure . . . . .	110
<b>4 Structure de graphe temporel et diffusion</b>	<b>117</b>
4.1 Diffusion sur les données MOSAR : principe . . . . .	118
4.2 Transmission systématique . . . . .	118
4.3 Diffusions probabilistes et stratégies de vaccination . . . . .	136
<b>5 Conclusions et perspectives</b>	<b>147</b>
Bibliographie . . . . .	148



# Introduction

Le monde réel est façonné par la diversité et est intrinsèquement complexe. On peut observer cette complexité aussi bien dans la nature que dans les différentes sociétés ou technologies que l'on construit. Elle apparaît au sein même d'un individu, dans les interactions moléculaires et chimiques tout autant qu'entre les individus d'une société. Cette prise de conscience universelle de l'existence de phénomènes complexes, dans des domaines aussi variés que la biologie, la sociologie, l'économie, l'informatique ou la linguistique, ont fait émerger la notion de systèmes complexes, encore mal définie de nos jours [12]. On considère communément qu'un système complexe est un ensemble d'entités en interaction. Ces entités sont éventuellement hétérogènes et ont une structure interne elle-même complexe et non négligeable. Des propriétés émergent au niveau macroscopique comme au niveau microscopique et ces deux niveaux s'influencent l'un l'autre dans leur évolution au cours du temps. Plusieurs points de vue différents peuvent être adoptés pour l'étude de ces systèmes :

- Le point de vue *structurel*. Les caractéristiques quantifiables (nombre, taille, durée) du système aussi bien que des entités le constituant sont considérés.
- Le point de vue *fonctionnel*. La forme et les conditions des interactions ainsi que leurs conséquences sur l'état et le comportement global du système et des entités sont étudiées.
- Le point de vue temporel. L'aspect dynamique et l'évolution dans le temps sont observés. La direction de l'évolution du système est considérée. On cherche en particulier à déterminer si le système est stable ou non, s'il croît ou s'il peut s'adapter à certaines perturbations.

L'un des buts recherchés, lors de l'étude de ces systèmes, est de comprendre le fonctionnement du système afin d'être capable à terme de prédire les futurs états du système, ce qui permet l'anticipation. La compréhension de ces systèmes est aussi une source d'inspiration pour concevoir de nouveaux systèmes, artificiels, cette fois, en conservant certaines des propriétés des systèmes préalablement observés comme leur taille par exemple, afin de mieux généraliser les observations.

## Contexte

Prenons le cas des études épidémiologiques restreintes à des environnements fermés comme c'est le cas dans un hôpital conçu pour des hospitalisations de longues durées, cas auquel nous nous sommes intéressés dans cette thèse. Il est assez naturel de concevoir cet environnement comme étant un système complexe composé de différentes entités en interactions.

En effet, dans un tel contexte nous sommes en présence d'une population dont les personnes interagissent, ce qui signifie ici, qu'elles sont dans des situations leur permettant de se transmettre des bactéries par contact physique ou par voie orale. Cette population est composée de personnes ayant des rôles bien définis selon leur statut de patients, personnels soignants, personnels administratifs ou autres corps de métier représentés au sein des hôpitaux. L'hôpital est aussi souvent divisé en services ayant chacun une spécificité médicale bien définie. Le réseau des interactions humaines peut donc être en soi considéré comme un système complexe à plusieurs échelles : l'échelle de l'individu est l'échelle la plus fine, les services et les catégories socio-professionnelles sont deux échelles macroscopiques que l'on peut considérer en parallèle. On peut donc étudier ce système selon les trois points de vue complémentaires présentés plus haut, adaptés au contexte de l'étude d'un réseau d'interactions humaines, à savoir :

- Le point de vue structurel : nous nous intéressons particulièrement au nombre d'interactions et à leur durée de manière globale ainsi qu'à l'échelle de l'individu, des services et des catégories socio-professionnelles.
- Le point de vue fonctionnel : nous considérons les motifs des interactions entre les différentes entités telles que les services et les différents corps de métiers tout en considérant leur localisation.
- Le point de vue temporel : nous étudions l'évolution du nombre et de la durée des interactions au cours du temps, à différentes échelles temporelles telles que l'heure et la journée.

## Organisation du manuscrit

Dans cette thèse, nous nous intéressons à l'étude de la structure du réseau social d'une population hospitalière et les propriétés de diffusion de ce réseau. Par réseau social, nous entendons le réseau engendré par la succession d'interactions entre les personnes se trouvant face à face et à moins d'un mètre les unes des autres. Le but de ce travail est d'approfondir nos connaissances du milieu hospitalier afin de trouver des pistes pour lutter contre les maladies nosocomiales, qui sont des maladies contractées à la suite d'une admission en hôpital. Un grand nombre de facteurs intervient dans la transmission de ce type de maladies. Nous laissons aux scientifiques médicaux le soin d'identifier les facteurs biologiques et nous nous intéressons à la dynamique du réseau social qui joue aussi un grand rôle dans les disséminations épidémiologiques. Le manuscrit est organisé en cinq chapitres. Dans un premier chapitre, nous exposons le contexte dans lequel les données sur lesquelles nous appuyons, ont été collectées et donnons quelques ordres de grandeurs de l'ampleur du projet dans lequel s'inscrit cette thèse. Ensuite, nous présentons des outils d'analyse spatiale et fonctionnelle de l'hôpital destinés à extraire la structure du réseau social de l'hôpital. Nous nous intéressons, dans un troisième chapitre, à l'aspect dynamique des données pour découvrir le rythme de vie de l'hôpital et détecter les moments clefs d'une potentielle diffusion au sein du réseau. Ceci nous amène à identifier des cycles circadiens et hebdomadaires ainsi qu'à utiliser des outils permettant de quantifier le taux de similarité d'une période à une autre dans le but d'identifier des paramètres récurrents pouvant servir en modélisation. Dans un dernier chapitre, avant la conclusion générale, nous étudions le comportement du système face à des simulations simples afin de faire une première estimation des risques de diffusion.

# Chapitre 1

## Contexte

### 1.1 Analyse des réseaux

Les systèmes complexes sont généralement modélisés par des réseaux dont les nœuds représentent les entités du système et les liens représentent l'existence d'interactions entre ces entités. Le réseau est souvent lui-même représenté par un objet mathématique bien connu : un graphe. Les graphes construits à partir du système à modéliser peuvent être pondérés si les interactions entre les entités sont quantifiées. Ils peuvent aussi être orientés (un lien n'est valable que dans un sens) si les interactions représentées ne sont pas symétriques.

L'analyse des réseaux représentant des systèmes complexes intéresse un grand nombre de domaines [8]. La compréhension des propriétés des réseaux est en elle-même intéressante et amène un travail très vaste sur les structures topologiques. Les propriétés de diffusion des réseaux sont un autre sujet tout aussi visité et étudié que ce soit en économie [27], pour comprendre et définir les stratégies publicitaires ou dans le cadre d'études médicales, épidémiologiques [29, 6], sociales [69] ou bien même l'évolution du système prédateurs-proies [19]. Un très grand nombre de travaux ont d'abord été effectués sur des réseaux statiques, en laissant de côté l'aspect temporel. Ces travaux ont ensuite été complétés par l'étude des réseaux dynamiques dont les outils d'analyse sont encore loin d'être complets.

#### 1.1.1 Réseaux statiques

Les réseaux statiques sont des réseaux observés ou construits de toute pièce à partir de données réelles ou générés aléatoirement, sans conserver d'informations sur la dimension temporelle. Ils peuvent être de deux types : les réseaux issus de données réelles et les réseaux aléatoires générés automatiquement.

##### **Réseau de terrain**

Les réseaux de terrains sont les réseaux observés dans un contexte particulier. On peut par exemple, considérer qu'un réseau routier ou ferroviaire est un réseau statique en terme d'architecture. Un grand nombre de travaux de systèmes complexes ont été représentés jusqu'à présent par des réseaux statiques, afin d'identifier des caractéristiques communes de ces systèmes complexes, bien que les réseaux observés soient en fait intrinsèquement dy-

namiques. Un certain nombre de mesures sont régulièrement faites sur ces réseaux, comme les mesures de densité ou de voisinage des nœuds (cf. section 1.3.3). D'autres travaux sont plus algorithmiques pour, entre autres, faire des regroupements souvent hiérarchiques de nœuds interagissant plus entre eux qu'avec les autres nœuds du réseaux. On peut citer les algorithmes de recherche de communautés dans les réseaux statiques [13, 23, 61] et [49] qui comparent plusieurs d'entre eux.

Ces réseaux sont collectés par différentes techniques plus ou moins exhaustives, puis analysés afin d'en extraire des propriétés caractéristiques ainsi que répondre à des questions de domaines divers et variés. Parmi les techniques de collecte de données, on peut citer les référencements bibliographiques ou archéologiques dans le cadre d'études d'évolution de réseaux de co-auteurs [37] et co-acteurs [1] ou dans le cadre de l'étude de réseaux historiques politiques ou sociétaux [69]. On peut aussi citer les études de données numériques concernant les déplacements aériens qui permettent la transmission d'épidémies à grande échelle [24]. Un grand nombre d'expériences sociologiques s'appuie sur des réseaux sociaux non numériques pour lesquels les données ont été recueillies à l'aide de questionnaires. Par exemple, plusieurs expériences sont basées sur le protocole suivant : chaque personne participante fournit la liste des personnes auxquelles elle a parlé au cours de la journée et la durée approximative de chaque conversation [56]. Cela induit beaucoup d'oublis car la mémoire et même la perception humaine ne sont pas parfaites [73]. D'autres expériences impliquent une personne extérieure qui relève les interactions entre les participants [57]. Dans ce cas encore, des pertes d'information considérables sont possibles car la personne observant le système n'a pas toujours le temps de tout noter ou remarquer. Ces dernières années, les nouvelles technologies permettent de faire des relevés à très fine granularité du comportement d'une population, sans biais de perception humaine. Ce sont les technologies sans fil qui, distribuées au sein d'une population, permettent entre autre de relever les relations de proximité spatiale des personnes. Comme nous le verrons au cours de ce manuscrit, cette technique de relevés n'est pas parfaite et est aussi sujette à des pertes.

### Modèles de réseaux aléatoires

Les réseaux aléatoires sont des réseaux générés automatiquement. Les premiers graphes aléatoires ont été proposés par Erdős et Rényi fin des années 50 [87] afin de répondre aux questions d'existence de certains objets combinatoires. Erdős et Rényi ont proposé un modèle de base permettant de générer des graphes dont la distribution des degrés suit une loi de Poisson quand le nombre de nœuds du réseau tend vers l'infini. En effet, générer ces graphes aléatoires consiste, à partir d'un certain nombre de nœuds et d'une probabilité  $p$ , à construire les arêtes du graphe : chaque arête a une probabilité  $p$  indépendante d'exister. Les graphes générés ont donc une distribution homogène des degrés. Plus tard, dans les années 90, Watts et Strogatz ont proposé un autre modèle permettant d'introduire la notion de coefficient de Clustering afin de se rapprocher de la structure d'un graphe réel, tout en conservant une distribution homogène des degrés. Leur méthode consiste à construire un graphe dont les sommets sont répartis sur un cercle et  $k$ -régulier, c'est à dire dont tous les nœuds ont le même degré, et chaque sommet est relié à ses plus proches voisins. Certaines arêtes sont ensuite supprimées aléatoirement et immédiatement remplacées par une autre arête en conservant l'une des des deux extrémités. Cette méthode permet d'imposer le degré des nœuds, ce qui permet d'obtenir un graphe dit "petit monde", pour créer des graphes ayant des propriétés communes à la grande majorité des réseaux réels observés jusqu'à présent : une forte structure en communautés et des diamètres moyens faibles. Les modèles de réseau invariant d'échelle sont un troisième type de modèle de réseaux dont la distribution des degrés suit cette fois une loi de puissance. Ce modèle permet d'envisager la

croissance exponentielle d'un réseau dans le temps, par l'ajout de nœuds, qui seront reliés préférentiellement à d'autres nœuds ayant un degré élevé. Ces réseaux ont des propriétés "petit monde".

### 1.1.2 Réseaux dynamiques

En ce qui concerne les réseaux dynamiques, beaucoup d'efforts récents ont été faits pour l'analyse de ces réseaux. Il ne ressort pas, comme dans le cas statique, de propriétés incontournables dans les réseaux réels dynamiques, qui fassent intervenir l'aspect temporalité. Dans un cadre de modélisation, il est donc, pour l'instant, nécessaire de recueillir des données dans le contexte que l'on veut modéliser pour en extraire les propriétés caractéristiques et ensuite établir un modèle spécifique à l'étude du contexte. La génération est donc moins adaptée à la théorie des graphes que dans le cas statique car il faut rajouter la dimension temporelle et connaître les propriétés à conserver lors de la génération.

## 1.2 En pratique : mesure et analyse de réseaux dynamiques de terrain à l'aide de technologie sans fil.

Parmi les différents réseaux qui ont été collectés à l'aide de technologie sans fil, différents buts étaient poursuivis. L'objectif le plus souvent visé est l'étude des risques épidémiques au sein d'une population, mais pas uniquement. On peut relever aussi l'étude de la transmission d'informations ou de données au sein d'une population mobile, en utilisant les appareils modernes tels que les smartphones ou autres objets connectés.

Dans le cadre de l'étude de diffusions bactériologiques, différents contextes ont donné naissance à la collecte de réseaux de contacts : des écoles primaires, vivier de diffusions épidémiques [25, 75], des services d'hôpitaux [81, 83, 84, 64, 52, 42, 20], un hôpital dans son ensemble [63, 53], des conférences [74]. Dans le cadre de l'étude des sociétés animales telles que les bovins des données ont été collectées pour comprendre l'impact de l'évolution du réseau entre les fermes sur la diffusion d'épidémies [18]. Dans le cadre de l'étude de la mobilité humaine les données permettent de comprendre les possibilité de transmission d'informations ou données au sein d'un groupe.

## 1.3 Contexte expérimental : projet MOSAR

Afin de mieux comprendre l'expansion de la résistance aux antibiotiques de souches bactériologiques de certains types tels que les staphylocoques dorés, dans les hôpitaux, une expérience de longue durée a été mise en place, au sein du projet européen MOSAR (Mastering hOSpital Antimicrobial Resistance and its spread into the community). En effet, il a été constaté que le *staphylococcus aureus* est un organisme commensal (qui se nourrit des déchets d'un autre organisme) de la peau et des muqueuses qui colonise 30% à 50% des individus en bonne santé, dont 10% à 20% sont des porteurs chroniques [48, 30]. Ce portage "sain" peut entraîner l'infection de cet individu qui devient contaminé avec sa propre souche quand la barrière de la peau ou des muqueuses se rompt (pour cause de blessure, opérations, intervention médicale invasive, ...) [86]. Peu après l'introduction

de la méticilline dans les années 60, des souches de *staphylococcus aureus* résistant à cet antibiotique apparaissent et deviennent causes d'épidémies nosocomiales. Ces souches deviennent rapidement endémiques dans les établissements de soins en Europe comme aux Etats-Unis.

Un grand nombre de facteurs peut influencer la présence et la diffusion de ce type de bactéries. Un certain nombre de facteurs médicaux sont en jeu. En effet, du point de vue médical, il est important de comprendre le mécanisme qui régit les nombreuses mutations de ces bactéries et la sélection naturelle de certaines souches qui s'opère. Les interactions entre l'exposition aux antibiotiques et les transmissions croisées semblent prendre une part importante dans la diffusion des bactéries résistantes aux antibiotiques [16, 65, 38]. Depuis peu, il est admis que les bactéries communiquent entre elles et peuvent ainsi coopérer et adapter leur comportement en fonction de leur environnement "social" [46]. Plusieurs modèles mathématiques ont été publiés pour modéliser les interactions entre espèces microbiennes en concurrence pour un seul nutriment humain [66]. D'autres modèles ont aussi été créés afin de mieux cerner les facteurs susceptibles de permettre l'invasion de pathogènes sur un hôte humain exposé aux antibiotiques ainsi que pour reproduire la réponse immunitaire et l'évolution de la virulence des parasites [45, 51]. D'autres facteurs relèvent plus de l'étude des réseaux sociaux et des habitudes humaines. En effet, la capacité intrinsèque des souches bactériennes en jeu à se transmettre par contacts [7, 54], la dynamique des contacts entre les individus [15, 55, 68, 28] comme les interactions entre l'exposition aux antibiotiques et la transmission croisée semblent être les facteurs conduisant à la diffusion des bactéries résistantes aux antibiotiques [17, 71, 65, 38, 39]. La diffusion de ces bactéries est un phénomène complexe à plusieurs échelles, des écosystèmes d'individus et bactériologiques aux structures de larges populations. Ces dernières années, un grand nombre d'études de la structure du réseau humain ont été mises en place afin de mieux comprendre les risques de diffusions épidémiologiques, et ce à plusieurs échelles. En effet, un projet a été rapporté sur l'étude du transfert de patients entre les hôpitaux des Etats-Unis [36]. D'autres projets se sont focalisés sur l'étude du réseau humain au sein d'un service précis d'un hôpital [44, 82]. Comprendre le phénomène de diffusion des bactéries résistantes aux antibiotiques demande de cerner les différents aspects médicaux et sociologiques puis de comprendre comment ils interagissent, c'est pourquoi le projet MOSAR s'est attaché, pour la première fois, à collecter des données de réseaux d'individus et des relevés bactériologiques afin de suivre l'évolution des différentes souches de *staphylococcus aureus* au cours d'une expérience de 6 mois, pour pallier le manque de données dans ce domaine. En effet, combiner les données microbiologiques avec le réseau des individus semble une remarquable source d'informations pour mieux comprendre la dissémination des souches devenant résistantes aux antibiotiques [5, 16, 67, 66, 32]. Là encore, après avoir identifié certains facteurs clefs influençant le système, il est intéressant de produire des modèles permettant d'étudier différents scénarios et d'évaluer certaines mesures d'hygiène ou de vaccination. Souvent, les modèles utilisés sont des modèles multi-agents permettant de faire interagir des entités hétérogènes ayant des attributs, des réactions différentes les unes des autres [78, 77].

### 1.3.1 Contexte précis de collecte des données MOSAR

Afin d'avoir une vue d'ensemble sur les relations et risques de transmissions bactériologiques qui peuvent survenir dans un hôpital, une expérience inédite, impliquant un hôpital dans sa globalité, a été mise en œuvre pendant 6 mois. Le but est non seulement de comprendre ce qui se passe dans chacun des services de cet hôpital, comme il a déjà été proposé par



d'autres équipes de recherche telles que [82, 44] mais aussi de comprendre les interactions qui se font entre ces services et les risques en terme de diffusion épidémique que cela comporte.

### Description fonctionnelle de l'hôpital

L'hôpital Maritime de Berk-sur-Mer, un AP-HP (Assistance Publique-Hôpitaux de Paris), est d'une taille conséquente, prévu pour les séjours longue durée. De ce fait, il est composé de différents services de soins, chacun spécialisé pour un type de pathologie spécifique.

Nous distinguons les services de soins au nombre de cinq, composés de patients et de personnels, répartis selon leurs fonctions médicales. L'hôpital comporte aussi des services que nous appellerons *transversaux*, au nombre de quatre. Ces derniers sont constitués uniquement de personnels amenés à intervenir dans plusieurs services de soins, tels les personnels de rééducation comme les kinésithérapeutes ou les ergothérapeutes. Comme on peut le voir sur la figure 1.1, les services transversaux se situent au centre de l'hôpital, entre les deux ailes des bâtiments de service de soins. L'ensemble des services accueille en grande majorité des personnes dont la mobilité est relativement réduite. La table 1.1 donne le détail des activités de chacun des services et de la difficulté à se déplacer, plus ou moins avancée, des patients.

(a) Services de soins.				(b) Services transversaux	
	Nom	Lieu	Fonction	Nom	Fonction
Pavillon Sorrel	S1	RDC	Rééducation d'obésité	S6	Service de nuit
	S2	1er	Rééducation neurologique	S7	Kinésithérapie
	S3	2eme	Rééducation gériatrique	S8	Ergothérapie
Pavillon Ménard	S4	1er	Rééducation Neurologique	S9	Administration
	S5	2eme	Rééducation Neurologique		

TABLE 1.1 – Description des services

D'autre part, plusieurs corps de métiers sont réquisitionnés afin d'assurer le bon fonctionnement de l'hôpital et sont listés dans la table 1.2.

	Titres/Fonction	Description de la fonction
C1	Aide Soignant	Aide au soin, toilette des patients
C2	Cadre de santé	Responsable des soins
C3	Infirmier Diplômé d'Etat (IDE)	Donne les soins
C4	Médecins	Prescrit les soins, diagnostique les patients
C5	Élève stagiaire	
C6	Kinésithérapeute	Rééducation musculaire
C7	Ergothérapeute	Rééducation fonctionnelle : travail d'apprentissage de l'utilisation des membres et modification de l'environnement du patient pour plus de facilité dans la vie quotidienne
C9	Brancardier	Transport des patients
C10	Animateur / Coiffeur / ...	Participe à la vie et l'animation des patients longue durée
C11	Agent de Service Hospitalier (ASH)	Alimentation des patients, nettoyage du matériel hospitalier
C12	Logistique	Préparation du matériel
C13	Administration	Accueil et départ des patients
C14	Autre rééducateur	
C0	Patients	Pensionnaires à soigner

TABLE 1.2 – Description des catégories socioprofessionnelles présentes dans l'hôpital Berk-sur-Mer en 2009.

### Description géographique

L'hôpital de Berk-sur-Mer, dans lequel s'est déroulé l'expérience, est divisé en plusieurs bâtiments. Les différents services de soins (dénnotés S1, S2, S3, S4, S5) sont regroupés dans deux bâtiments (Sorrel et Ménard), chacun à un étage distinct, comme on peut le voir sur la figure 1.1. Les services de rééducation tels que kinésithérapie ou ergothérapie, composés uniquement de personnels, se situent principalement entre les deux ailes des bâtiments des services de soins.

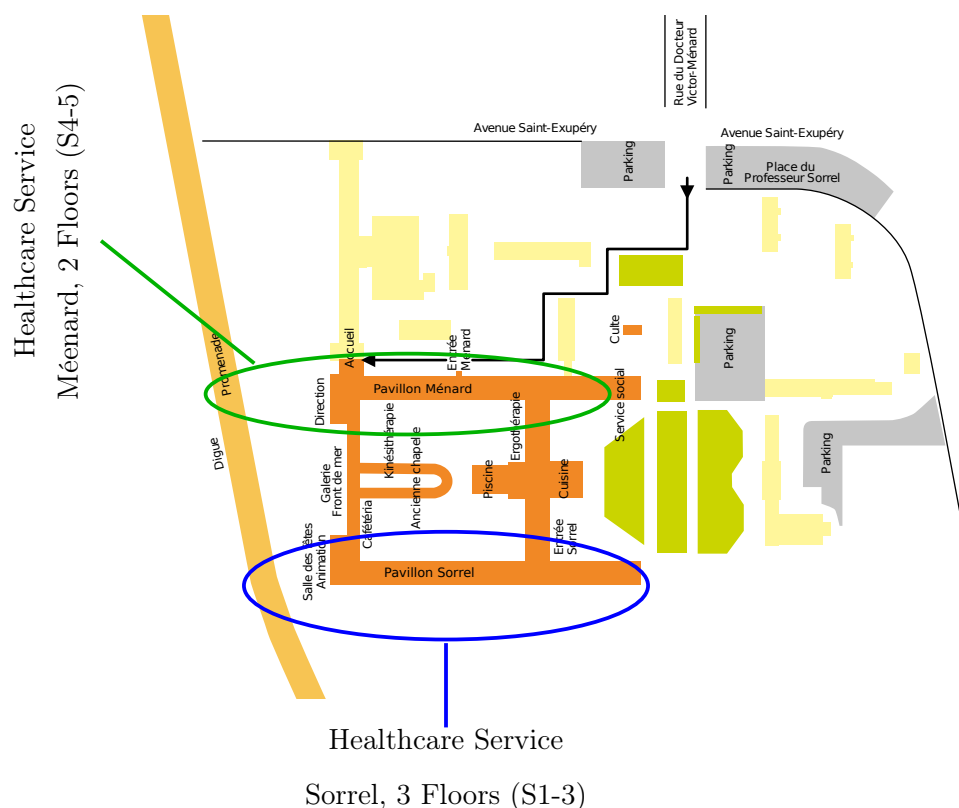


FIGURE 1.1 – Agencement de l'hôpital de Berk-sur-Mer.

### 1.3.2 Protocole de mesure

Afin de collecter des données qui permettent d'évaluer le lien entre diffusion bactériologique, relation de proximité humaine et résistance aux antibiotiques de souches bactériologiques, une expérience d'envergure a été mise en place à l'hôpital de Berk-sur-Mer. Plusieurs types d'informations ont été relevés durant cette expérience afin d'observer et de comprendre le fonctionnement d'un hôpital d'accueil de patients pour des séjours de longue durée. Ce type de structure assure une certaine stabilité en terme de présence, ce qui, ajouté à la longue durée de l'expérience nous permet de déterminer les cas d'anomalies dans les données et ainsi repérer les périodes en lesquelles nous pouvons avoir confiance. Les informations recueillies sont de trois types : les relevés bactériologiques, les relevés d'emploi du temps et la collecte des interactions humaines. Nous détaillons ci-dessous ces trois types de données.

#### Relevés bactériologiques

Des prélèvements bactériologiques ont été faits chaque semaine sur toutes les personnes participant à l'expérience afin de connaître l'évolution du réseau de souches bactériennes se développant dans l'hôpital. Le personnel médical et les chercheurs réalisaient chaque semaine, au cours des soins courants, des écouvillons nasaux, dont une image se trouve figure 1.2, qui étaient ensuite analysés pour tracer la présence chez les personnes de deux espèces bactériennes.

En particulier, l'analyse des prélèvements visait à détecter la présence de souches de staphylocoque doré, et en particulier les souches résistantes à la méticilline (SARM), qui est la plus fréquente des bactéries multirésistantes et responsable d'infections nosocomiales. Les chercheurs traçaient aussi les entérobactéries résistantes aux céphalosporines de 3ème génération (cefotaxime)



FIGURE 1.2 – Écouvillon nasal.

### Relevés des relations de proximité

Un autre défi de ce projet était d'enregistrer le réseau de proximité des personnes. On cherche donc à savoir, pour chaque instant, quelles sont les personnes qui se trouvent à proximité les unes des autres. Une relation de proximité entre deux personnes sera appelée contact dans la suite.

#### Definition 1.3.1 (Contacts)

Nous notons un *contact* entre une paire de personnes  $A$  et  $B$  par un quadruplet de la forme  $(A, B, t_s, l)$ .  $t_s$  est la date de début du contact et  $l$  la durée de celui-ci. Par convention, la durée minimale d'un contact est de 30 secondes,  $t_s < t_s + l$ .

Nous modélisons le réseau de proximité des personnes par un ensemble de contacts, notion définie ci-dessous.

#### Definition 1.3.2 (Ensemble de contacts)

Un *ensemble de contacts*  $C$  bien formé, pouvant donc modéliser un réseau de proximité des personnes sans ambiguïté, respecte les trois règles suivantes.

- Aucun contact n'est morcelé en plusieurs petits contacts séparés par des interruptions.
- Il n'existe pas de contacts distincts consécutifs en temps pour une même paire de personnes.
- Aucune paire de contacts d'une même paire de personnes ne peut se chevaucher, comme décrit par l'exemple 1.3.

Plus formellement, l'ensemble des contacts  $C$  représentant un réseau de proximité peut s'exprimer ainsi :

$$\forall (A, B, t_s, l) \in C, \nexists (A, B, t'_s, l') \text{ tel que } \{t_s \leq t'_s \leq (t_s + l) \text{ ou } t_s \leq t'_s + l' \leq (t_s + l)\}.$$

Nous considérons les relations de proximité comme vecteurs potentiels de transmission et donc comme étant des contacts tels que définis précédemment.

**Protocole :** Afin de relever des informations détaillées sur les relations de proximité humaines au cours des journées, chaque participant à l'expérience a été équipé d'un capteur sans fil capable d'émettre et de recevoir des messages toutes les 30 secondes environ à moins d'1,5 mètre. Cela nous permet d'avoir des informations sur la durée et la fréquence des relations de proximité face-à-face des personnes au cours de la journée. Afin de garantir l'anonymat des participants, un identifiant unique a été attribué à chaque personne.

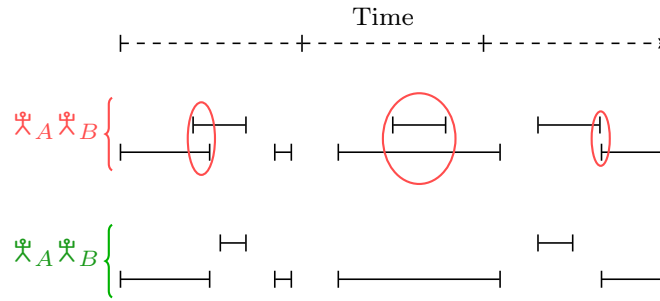


FIGURE 1.3 – Règle de modélisation des contacts d’un réseau de proximité : deux ensembles de contacts pour la paire de personnes  $(A, B)$  sont représentés ici. Seul celui précédé des petits bonhommes verts foncés est correct. Les défauts de construction des autres ensembles de contacts sont signalés par des ellipses rouge clair. Les contacts sont représentés par des segments délimités par des barres verticales.

Cet identifiant est aussi associé au numéro du badge que cette personne portait selon ses périodes de présence.

**Caractéristiques techniques des capteurs :** Les capteurs fusionnaient partiellement les contacts à la volée. En d’autres termes, toute suite de contacts consécutifs entre deux mêmes personnes étaient parfois enregistrés comme étant un seul contact de durée égale au produit du nombre de contacts consécutifs et de la durée d’un pas de temps (ici, 30 secondes). Les capteurs avaient la particularité de ne jamais s’éteindre et donc d’émettre même durant les périodes d’absence des participants qui laissaient leur badge sur panneaux prévus à cet effet et rassemblés dans une pièce spécifique. Nous expliquons dans la section 1.3.3 comment s’abstraire des informations supplémentaires qui en découlent. Un autre élément à prendre en compte, lié au traitement des données, est le fait que les capteurs ont une autonomie limitée en terme d’énergie ce qui a conduit à des périodes de changement de pile, pour les recharger. Durant ces périodes de trois jours consécutifs chacune, l’expérience était en suspens.

### Informations supplémentaires

Les admissions des patients et l’emploi du temps des personnels ont été enregistrés sur des cahiers, suivant les habitudes de fonctionnement de l’hôpital. Ainsi, nous connaissons les dates d’arrivée et de départ des patients ainsi que les causes de ces départs. Par ailleurs, nous bénéficions d’un emploi du temps détaillé des personnels dont certaines catégories socio-professionnelles font les 3/8. Ainsi, nous avons plusieurs types de personnels dont nous connaissons entre autres les heures d’arrivée et de départ chaque jour.

La figure 1.4 fait un résumé des présences et absences de participants à la journée. Les couleurs permettent de distinguer les présences prévues par l’emploi du temps de l’hôpital (vert) des présences non prévues (rouge) ainsi que les absences prévues (gris) des absences imprévues (noir). La sous-figure 1.4a ne rend compte que de la présence des patients. On peut constater de nouvelles arrivées constantes au cours du temps. Les départs sont moins fréquents. Les deux périodes d’absence de toutes les personnes simultanément correspondent aux périodes de changement de badges, durant lesquelles la collecte de données de contacts a été interrompue pendant 3 jours. Dans l’intervalle entre les deux périodes de

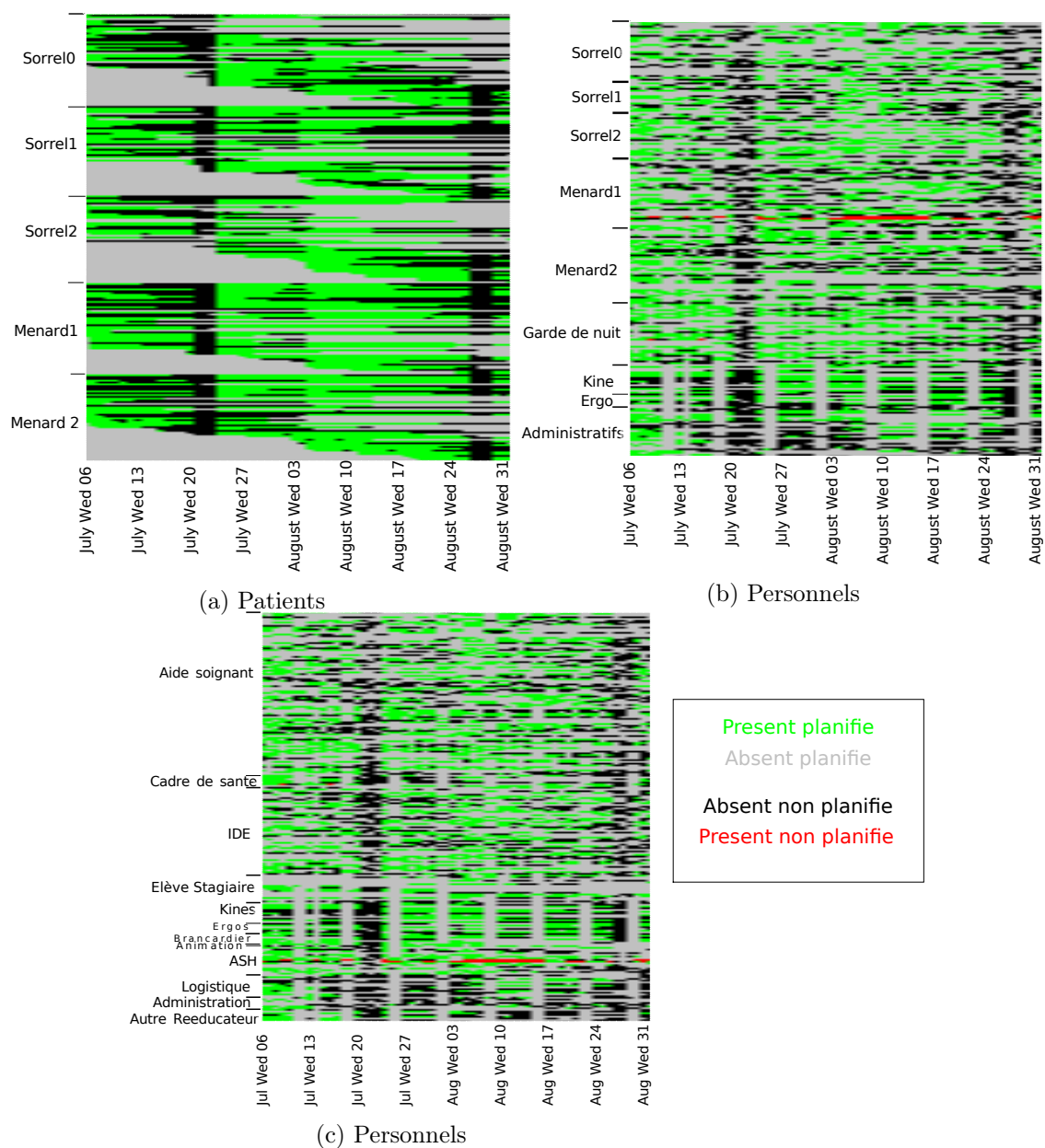


FIGURE 1.4 – Présence des personnes à la journée. En haut à gauche, la liste des patients regroupés par service. En haut à droite, la liste de personnels regroupés par service. En bas, les mêmes personnels, cette fois-ci regroupés en catégories professionnelles.

changement de badges, on peut observer deux périodes distinctes, dues à l'arrivée en masse de patients début août.

Nous bénéficions, par la même méthode, de la correspondance entre les identifiants et les numéros de badges.

Dans la section qui suit, nous expliquons comment, à partir de ces données brutes, nous pouvons obtenir des données plus faciles à manipuler et à interpréter.

### 1.3.3 Pré-traitements des données de mesures des capteurs

L'ensemble des données recueillies par les capteurs sans fil est biaisé par un certain nombre de facteurs. Il convient donc de s'abstraire au maximum de ces biais et pour cela les identifier. En premier lieu, l'échange de messages des capteurs restés sur les panneaux de la salle de récupération des capteurs sont des messages qui n'apportent aucune information sur les relations de proximité des personnes et il convient donc de s'en abstraire, ce que nous expliquons plus en détail ci-dessous. Ensuite, comme il a déjà été expliqué précédemment, les capteurs reconstituaient partiellement les contacts et il convient d'homogénéiser les données.

#### Nettoyages des données parasites

**Suppression des données les jours de changement de badge :** Les capteurs n'ayant pas une autonomie infinie, certains jours de l'expérience ont été consacrés à la recharge des piles des capteurs. Chaque période de recharge s'est faite sur trois jours, espacée d'environ un mois chacune, durant lesquelles aucune donnée de contact n'est conservée pour les analyses.

**Correction des données utilisant l'emploi du temps fourni :** Les capteurs stockés dans une même pièce lors de l'absence des personnels créent des données parasites qui ne correspondent pas à des relations de proximité de personnes. Il convient de les retirer. Nous nous sommes appuyés pour cela sur un emploi du temps d'abord noté sur les plannings des personnels qui ont ensuite été numérisés. Nous avons retiré tous les contacts impliquant deux badges dont l'emploi du temps des personnes portant ces badges attestait leur absence.

#### Reconstruction de contacts perdus

Lors de l'utilisation de technologie sans fil, un certains nombre de pertes peuvent se produire. En particulier, lorsque plusieurs messages sont envoyés en même temps par différents émetteurs, des collisions peuvent se produire et certains messages ne sont pas pris en compte. Une collision peut par exemple avoir lieu quand deux messages provenant de deux sources distinctes atteignent un unique destinataire en même temps. Il est possible de retrouver des messages perdus. En effet, les relations de proximité sont symétriques : quand une personne est à moins d'un mètre d'une autre, la réciproque est également vérifiée. Notre stratégie pour reconstruire les contacts a donc été de symétriser les relations.

Cette méthode ne permet pas de reconstituer tous les contacts. On peut penser à d'autres stratégies comme reconstruire les triangles. Cela consiste à estimer que lorsqu'un contact a lieu entre A et B, un autre entre B et C, alors A et C sont aussi en contact. Cela n'est pas toujours le cas, c'est pourquoi, nous avons choisi de ne pas utiliser cette méthode finalement.

### Vocabulaire et méthodes d'agrégation des données

Les données nettoyées issues des capteurs nous donnent une granularité très fine de 30 secondes. L'utilisation des données n'exige pas toujours une telle granularité d'information, c'est pourquoi ce type de données est souvent agrégé. Agréger les données permet une manipulation plus aisée et des calculs moins longs.

Avant d'agréger les données, il faut spécifier le type d'information qui nous intéresse ainsi que la représentation abstraite des données que l'on choisit. Durant cette thèse, nous nous appuyons principalement sur des outils et une méthodologie issus de travaux en physique statistique, que nous exposerons au cours des différentes parties de ce manuscrit. Pour décrire les phénomènes se déroulant dans le réseau des personnes nous nous appuyons sur du vocabulaire et des notions issus de la théorie des graphes [26, 87] que nous introduisons ici. Ce choix s'appuie sur le fait que les réseaux sont facilement représentables par un graphe [9].

#### **Definition 1.3.3** (Graphe)

On appelle graphe (ou réseau) la paire d'ensemble  $G=(V, E)$  où  $V$  est un ensemble de nœuds et  $E$  un ensemble d'arêtes notées  $e = (i, j)$ ,  $e \in E$ ,  $i, j \in V$ . Les arêtes peuvent être dirigées et dans ce cas, le graphe est dit orienté. Dans le cas contraire, le graphe est non-orienté et toute arête est considérée comme non-orientée. Les arêtes peuvent être pondérées. Dans ce cas, pour toute arête  $e = (i, j)$ ,  $i, j \in V$ , on associe une fonction de poids  $w_{ij} \in \mathbb{R}$ .

Dans le cas d'un réseau de proximité, les relations sont symétriques et peuvent par exemple être pondérées par la durée de la relation de proximité. On peut donc modéliser ce réseau par un graphe non-orienté pondéré résumant l'ensemble des interactions ayant eu lieu durant l'expérience. Ce faisant, la dimension temporelle est oubliée. Une technique utilisée, pour la prendre en compte est de découper le temps en périodes et de produire des séries de graphes statiques. Nous parlons alors de réseau dynamique et adaptons quelque peu les notions issues de la théorie des graphes statiques aux graphes dynamiques.

#### **Definition 1.3.4** (Graphe dynamique)

Soit  $G_t = (V, E)$  un graphe dynamique, dont les arêtes varient au cours du temps. Nous modélisons un graphe dynamique par une série temporelle de graphes statiques. Pour ce faire, nous conservons les ensembles  $V$  et  $E$  de la définition classique des graphes statiques et nous associons aux arêtes une fonction de poids et une fonction d'existence, qui renseigne sur les dates de présence de l'arête. Nous notons  $(u, v)_t$  l'existence d'une arête temporelle entre les nœuds  $u$  et  $v$  au temps  $t$ , avec  $u, v \in V$ ,  $(u, v) \in E$ . On notera  $V_t$  et  $E_t$  l'ensemble des nœuds et arêtes existant au temps  $t$ . Le poids des arêtes sera noté  $w_{uv}^t$ .



Les données de contacts de l'expérience MOSAR se prêtent bien à la représentation en séries temporelles de graphes car ces données sont discrètes et régulières. En effet, à partir des données de contact, telles que nous les avons décrites précédemment, créer le graphe dynamique revient à créer une série de graphes statiques par *pas de temps* et créer une arête entre tous les nœuds qui ont un contact en cours. L'expérience MOSAR nous permet de générer des graphes temporels avec des *pas de temps* de 30 secondes sans perdre d'informations par rapport aux données collectées.

### Exemple 1.3.1

Partons de la liste de contacts représentés figure 1.5. La technique pour créer la série de graphes temporels consiste à découper les contacts par *pas de temps*. Pour chaque pas de temps  $t$ , il existe une arête entre deux nœuds si et seulement si un contact entre ces deux nœuds est en cours à ce pas de temps  $t$ . Le graphe dynamique issu de la liste de contacts de la figure 1.5 est représenté figure 1.6.

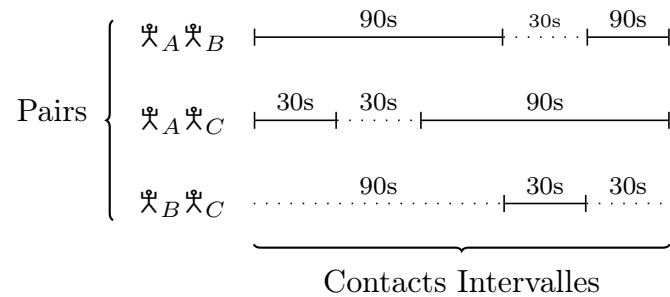


FIGURE 1.5 – Liste de contacts, base de la création d'un graphe dynamique.

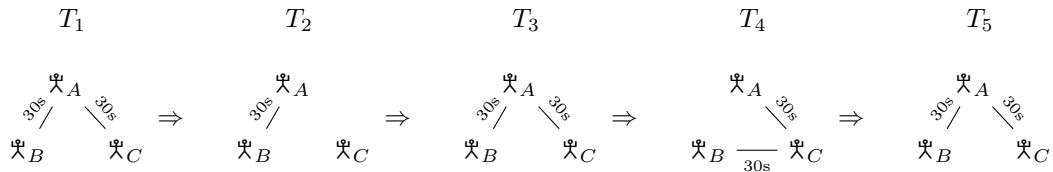


FIGURE 1.6 – Construction d'un réseau dynamique à partir de la liste de contacts de la figure 1.5.

#### Definition 1.3.5 (Voisin et degré.)

On dit qu'un nœud  $i$  est *voisin* d'un nœud  $j$  s'il existe une arête entre  $i$  et  $j$ , que le graphe soit pondéré ou non. Le nombre de voisins d'un nœud  $i$  est communément appelé *degré* du nœud et sera noté  $\mathfrak{d}(i)$ . Dans le cas de graphes dynamiques, nous parlerons de voisin au temps  $t$ .

#### Definition 1.3.6 (Chemin.)

Soit  $G=(V,E)$  un graphe statique. On dit qu'il existe un *chemin* allant du nœud  $i$  à  $j$  s'il existe une suite d'arêtes ou liens permettant de passer de  $i$  à  $j$  en suivant ces liens,

de la forme :  $\mathbf{p}(i, j) = \{(i, x), (x, w), \dots, (y, z), (z, j)\}$ . On note  $\mathcal{P}(i, j)$  l'ensemble des chemins allant de  $i$  à  $j$ .

**Definition 1.3.7** (Chemin temporel.)

Soit  $G=(V,E)$  un graphe ou réseau dynamique, dont les nœuds sont stables mais les liens changent au cours du temps. On dit qu'il existe un *chemin temporel* allant du nœud  $u$  à  $v$  à partir d'un temps  $t$  s'il existe une suite d'arêtes ou liens permettant de passer de  $u$  à  $v$  en suivant ces liens, de la forme :

$$\mathbf{p}_t(u, v) = \{(u, x)_{t_1}, (x, w)_{t_2}, \dots, (y, z)_{t_{n-1}}, (z, v)_{t_n}\}, \text{ avec } t_1 < t_2 < \dots < t_{n-1} < t_n.$$

**Definition 1.3.8** (Distance)

La distance  $\delta_{ij}$  entre deux nœuds  $i$  et  $j$  est définie par le nombre minimal d'arêtes qui les sépare dans le graphe.

$$\delta_{ij} = \min_{\mathbf{p} \in \mathcal{P}(i,j)} (|\mathbf{p}|)$$

**Definition 1.3.9** (Connexité)

On dit qu'un graphe est *connexe* si pour toute paire de nœuds du graphe, il existe un chemin reliant ces deux nœuds. Une composante connexe d'un graphe  $G$  est un sous-graphe connexe maximal (sous-graphe auquel on ne peut ajouter de nœud de  $G$  sans déconnecter le graphe qui en résulte).

La distance est définie uniquement dans le cas d'un graphe connexe. Dans le cas où le graphe n'est pas connexe, il est d'usage de définir les distances pour chaque composante séparément et de faire la plupart des analyses sur la plus grosse de ces composantes connexes.

**Definition 1.3.10** (Excentricité)

L'excentricité d'un sommet  $i$  est la distance maximale qu'il a avec tous les autres sommets du graphe.

$$\text{ex}(i) = \max_{k \in V} (\delta_{ik})$$

La notion de diamètre d'un graphe permet de qualifier en partie la connexité d'un graphe. Plusieurs mesures afin de qualifier la connexité d'un graphe existent. La mesure la plus traditionnelle est le diamètre.

**Definition 1.3.11** (Diamètre)

Le diamètre d'un graphe  $G = (V, E)$  est l'excentricité maximale de tous ses sommets.

$$D(G) = \max_{i \in V} (\text{ex}(i))$$

Dans le cadre de l'étude des réseaux réels, d'autres mesures sont aussi utilisées, comme l'excentricité moyenne [2, 79]. Donnons l'exemple de l'étude du réseau humain et de la distance à chaque personne lorsque les liens du réseau représente le fait que deux personnes se connaissent. Plusieurs expériences, menées depuis 1969, laissent penser que le graphe social terrestre aurait, entre deux personnes, une distance moyenne proche de 6. En effet, Milgram [79], le premier, a mené une expérience aux états-unis afin de mesurer la distance

sociale des personnes en envoyant des paquets à des personnes résidant en Omaha prises aléatoirement en leur demandant de le faire passer à une connaissance qu'ils pensaient plus proche d'une personne destinataire vivant à Boston. Cette expérience donne une moyenne de 5 intermédiaires et donc d'une distance moyenne de 6. D'autres expériences aboutissent à des moyennes similaires ou légèrement plus élevées [80, 60, 85, 47, 50]. Une étude plus récente du réseau social des utilisateurs actifs de facebook annonce même une distance moyenne de 4,74 [4].

Tout au long de ce manuscrit, nous nous intéressons à l'état et l'évolution des trois quantités distinctes suivantes :

- Les paires d'adjacence, qui sont des paires de personnes ayant eu un ou plusieurs contacts pendant une période donnée. On parlera de degré des personnes lorsque l'on se placera du point de vue des individus, ce qui correspond au nombre de voisins de chaque personne, généralement sur une journée.

**Definition 1.3.12** (Paire d'adjacence)

Soit un intervalle de temps  $T = [a, b]$  et un graphe dynamique  $G = (V, E)$  pour lesquels on associe des sous-ensembles temporels  $V_t$  et  $E_t$ . L'ensemble des paires d'adjacence de la période  $T$  est l'ensemble des arêtes de la forme  $(u, v)_t$ , avec  $(u, v) \in E$ , et  $t \in T$ . Notons  $N(u)_t$  l'ensemble des voisins de  $u$  au temps  $t$ , c'est à dire tous les nœuds  $v \in V_t | \exists (u, v) \in E_t$ . Le degré d'un nœud  $u$  sera alors  $|N(u)_t|$

- Le nombre de contacts, c'est à dire le cardinal d'un ensemble de contacts bien formé de la forme  $(A, B, t_s, l)$ .

**Definition 1.3.13** (Nombre de contacts)

Soit  $C$  un ensemble de contacts bien formé. Le nombre de contacts sera alors  $|C|$ .

- La durée cumulée des contacts appartenant à un ensemble de contacts bien formé de la forme  $(A, B, t_s, l)$  qui correspond à la somme des durées  $l$  des contacts de cet ensemble.

**Definition 1.3.14** (Durée cumulée de contacts)

Soit  $C$  un ensemble de contacts de la forme  $(A, B, t_s, l)$ , la durée cumulée des contacts sera  $L = \sum_{(A, B, t_s, l) \in C} l$ .

Nous parlerons d'*activité* lorsque nous décrirons des méthodes pouvant s'appliquer indifféremment à chacune de ces quantités.

## Agrégation des contacts

Nous décrivons ici deux types d'agrégations possibles dont une concerne les groupes de personnes, définis arbitrairement, l'autre concerne l'agrégation en temps.

### Agrégation par groupes

Afin d'avoir une vue plus globale du comportement d'un groupe assez conséquent de personnes, il paraît important de regrouper les personnes ayant un même rôle. Dans le cas de l'hôpital, nous avons décidé de construire des groupes de personnes, suivant les répartitions officielles en services et en catégories socio-professionnelles de l'hôpital hôte. Cela revient à faire une partition des sommets du réseau.

**Definition 1.3.15** (Partition des sommets d'un graphe)

Une *partition* des sommets d'un graphe  $G$  consiste à définir des ensembles de sommets disjoints et non vides dont l'union permet de reconstituer l'ensemble de sommets du graphe  $G$ . Formellement, soit  $G = (V, E)$  un graphe,  $P_k = \{V_1, \dots, V_k\}$  sera une partition de  $G$  en  $k$  parties si  $\cup_{i=1 \dots k} V_i$  et  $\forall (V_i, V_j) \in P_k, i \neq j \Rightarrow V_i \cap V_j = \emptyset$ .

Une fois les groupes de personnes définis, l'agrégation en groupe consiste à créer un nouveau réseau dont les nœuds sont les groupes prédéfinis et les arêtes résument l'ensemble des liens existants entre les personnes des groupes, ce que nous appellerons *contraction de nœuds*.

**Definition 1.3.16** (Fusion de nœuds)

Soit  $G = (V, E)$  un graphe. La fusion des nœuds  $i$  et  $j$  est l'opération qui consiste à transformer le graphe  $G$  en un graphe  $G' = (V', E')$ , où  $V' = V \setminus \{i, j\} \cup k$  où  $k$  est un sommet unique qui remplace  $i$  et  $j$  et  $E'$  est le même ensemble que  $E$  où toutes les occurrences de  $i$  et  $j$  sont remplacées par  $k$ . En fonction des domaines d'application, les arêtes de types  $(k, k)$  (boucle) ainsi créées, sont conservées ou non.

**Definition 1.3.17** (Contraction d'arête)

Soit  $G = (V, E)$  un graphe contenant l'arête  $(i, j)$  avec  $i \neq j$ . La contraction de l'arête  $(i, j)$  consiste à fusionner les nœuds  $i$  et  $j$ .

Nous procédons itérativement, pour agréger le réseau par groupe, à des contractions systématiques d'arêtes liant deux sommets du réseau appartenant à un même groupe. Les composantes connexes obtenues ainsi et appartenant à un même groupe sont ensuite fusionnées systématiquement.

Il existe une arête entre deux groupes  $U$  et  $V$  si et seulement si il existe au moins une paire de nœuds  $\{u, v\}$  tels que  $u \in U$  et  $v \in V$  reliés par une arête dans le réseau de départ. Le poids de l'arête créée correspond à la somme du poids des arêtes de type  $(u, v)$ , dans le cas d'un réseau pondéré. Dans ce cadre, nous conservons les boucles. Ainsi, contrairement au graphe de départ, le réseau résultant de l'agrégation possède des boucles : les contacts internes au groupe sont comptabilisés. Un exemple simple d'agrégation de groupe d'un réseau est représenté sur la figure 1.7.

**Agrégation en temps**

Le principe de l'agrégation en temps permet de réduire le nombre de pas de temps en résumant l'activité au cours du temps découpé en intervalles. Cela permet donc de réduire le nombre de graphes de la série temporelle représentant les données tout en conservant une certaine notion de causalité.

L'agrégation du graphe dynamique des contacts est en fait une série temporelle de plusieurs graphes séparés par un pas de temps  $\tau$  dont nous pouvons faire l'union.

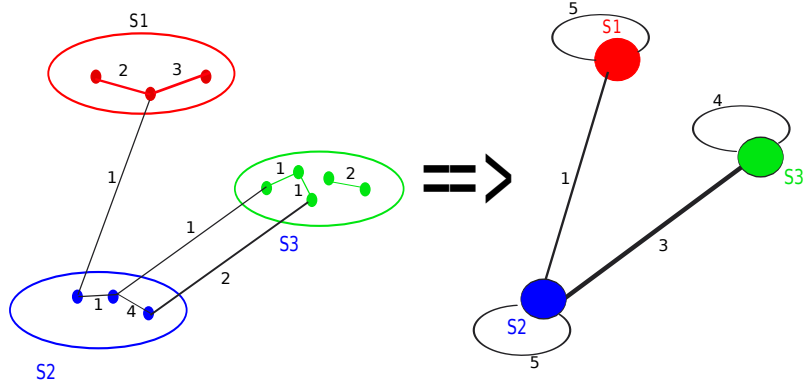


FIGURE 1.7 – Exemple d'agrégation par groupe d'un réseau simple.

**Definition 1.3.18** (Union de graphe)

Soit  $G = (V, E)$  un graphe non pondéré.  $G$  est l'*union* des graphes non pondérés  $G_1 = (V_1, E_1), \dots, G_k = (V_k, E_k)$  si  $V = \cup_{t=1}^k V_t$  et  $E = \cup_{t=1}^k E_t$ .

Dans notre cas, nous nous intéressons aux graphes pondérés, qui conservent deux informations : la durée cumulée des contacts et le nombre de contacts distincts, c'est à dire séparés par au moins un pas de temps d'inactivité. Nous appliquons donc l'union de graphes sur un intervalle de temps  $[d, f]$ , avec deux fonctions de poids supplémentaires :

- La fonction de poids informant sur le nombre de contacts, calculée à partir du graphe dynamique non pondéré de pas de temps  $\tau$ . On utilisera tout de même la notation de fonction de poids pour les arêtes de ce graphe afin de faciliter la lecture :

$$w_{uv}^t = \begin{cases} 1 & \text{si } (u, v)_t \in E_t \\ 0 & \text{sinon} \end{cases}$$

Les arêtes du graphe résultant sont le résultat de l'union de la série temporelle de graphes compris dans l'intervalle d'agrégation choisi issu du graphe temporel. A chaque arête  $e = (i, j)$  du graphe résultant, une fonction de poids est associée et compte le nombre d'intervalles de contacts non consécutifs :

$$w_{ij}^d = w_{uv}^{df} = \frac{w_{uv}^d * w_{uv}^f}{\tau^2} + \left\lceil \frac{\sum_{t=d}^{f-\tau} |w_{uv}^{t+\tau} - w_{uv}^t|}{2\tau} \right\rceil.$$

- La fonction de poids informant sur la durée cumulée des contacts. Nous travaillons à partir d'un graphe dynamique pondéré  $G = (V, E) | V = \cup_t V_t, E = \cup_t E_t$  pour obtenir un nouveau graphe dynamique pondéré. Les arêtes du graphe sont obtenues par union de graphes. Le poids d'une arête  $w_{ij}$  entre les sommets  $i$  et  $j$  correspond à la somme des poids de toutes les arêtes de type  $e_d = (u, v)_d \in \cup_t E_t$ .  $w_{ij}^d = w_{uv}^{df} = \sum_t w_{uv}^t$ .

Dans le contexte de la représentation de l'activité d'un hôpital, les échelles temporelles qui semblent les plus significatives sont celles de l'heure et de la journée car toutes les activités de cet environnement sont organisées selon ces deux grandeurs. Nous construisons donc des séries de graphes statiques qui sont le résultat d'agrégations à l'heure et à la journée.

**Definition 1.3.19** (Degré quotidien.)

Soit  $G_d$  le graphe dynamique agrégé au jour. A chaque nœud  $u$  du graphe est associé un ensemble de degrés quotidiens  $\mathcal{D}(u)$ , obtenu en calculant le degré de  $u$  pour

chaque graphe statique de la série temporelle du graphe dynamique  $G_d$ . L'ensemble de  $\mathcal{D}(u) = \cup_t \{\mathfrak{d}^t(u)\}$ , si on note  $\mathfrak{d}^t(u)$  le degré de  $u$  au temps  $t$ .

#### 1.3.4 Quelques ordres de grandeurs

Au total 592 personnes ont été prises en compte durant l'expérience. Nous avons un ensemble de 2671832 arêtes sur l'ensemble de la période. La durée des contacts est souvent très courte (inférieure à 2 minutes) comme l'atteste la distribution des durées de contacts sur l'ensemble du réseau (figure 1.8).

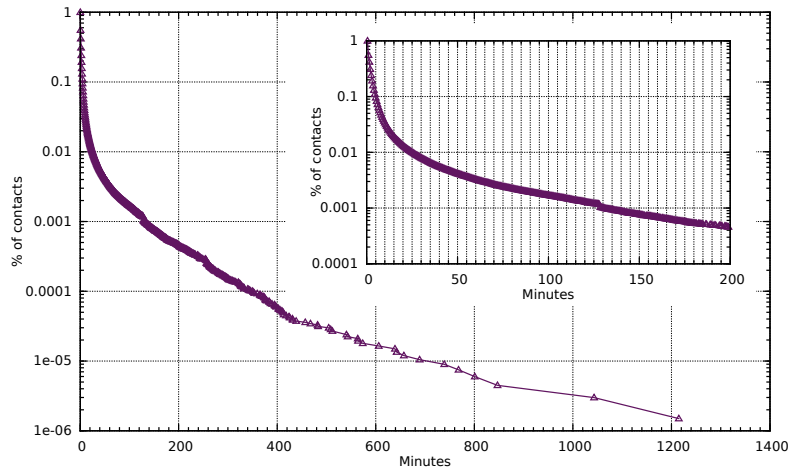


FIGURE 1.8 – Distribution des durées de contacts au cours de la période juillet/août

En effet, la figure 1.8 représente la distribution de chacun des contacts ayant eu lieu durant la période juillet et août 2009. Un zoom est fait sur la partie de la distribution des contacts durant jusqu'à 200 minutes, soit un peu plus de 3h. On constate que la majorité des contacts (90% d'entre eux) ont une durée inférieure ou égale à 5 minutes, ce qui est très court. Il existe tout de même quelques contacts extrêmement longs (presque une journée pour le plus long), ayant lieu entre patients d'une même chambre, qui peuvent être considérés comme étant de nature différente des contacts courts. En effet, dans un contexte d'étude de diffusion bactérienne, ces contacts n'auront sûrement pas le même rôle que des contacts courts.

Les quantités quotidiennes d'activité varient selon les personnes. Considérons la figure 1.9 qui représente la distribution des degrés quotidiens des personnes ainsi que le nombre quotidien de contacts et la durée cumulée de contact par personne. On constate que la majorité des valeurs obtenues sont contenues dans un intervalle assez restreint. Les valeurs très importantes sont tout de même nombreuses mais correspondent à une très faible proportion de cas. En effet, plus de la moitié des personnes ont un degré quotidien compris entre 6 et 16.

La figure 1.10 met en relation les différentes quantités d'activité de chaque personne entre juillet et septembre. Le nombre de contacts et la durée des contacts sont très corrélés comme l'atteste la figure 1.10c qui met en relation ces deux quantités. Pour chaque personne, la somme des degrés quotidiens des personnes est comparée à la somme de la durée de contacts sur les deux mois. On compare aussi la somme des longueurs avec le nombre de contacts.

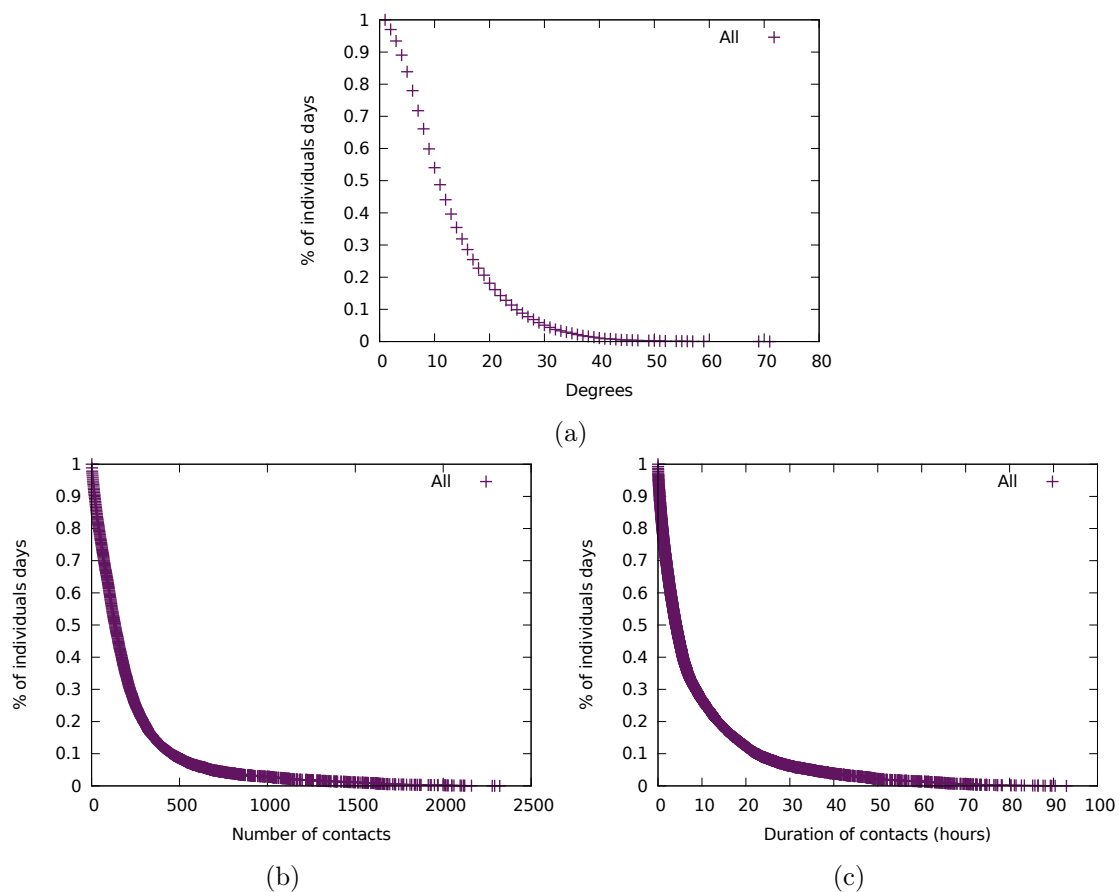
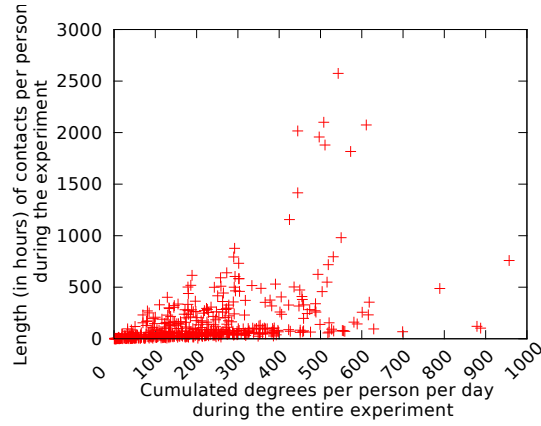
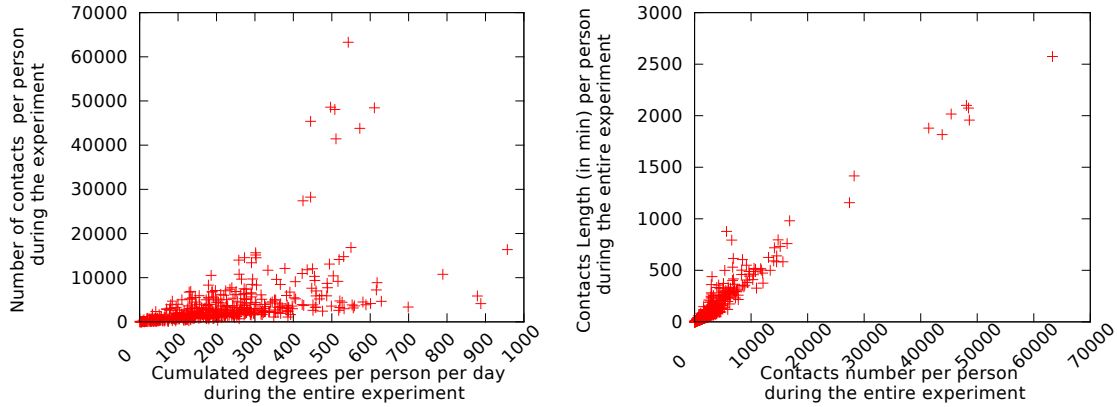


FIGURE 1.9 – Distribution des quantités par personne par jour. En haut : distribution des degrés, en bas à gauche distribution du nombre de contacts par jours. En bas à droite : distribution des durées de contacts par personnes

On peut remarquer que les durées de contact sont très corrélées avec le nombre de contacts pour la plupart des personnes, ce qui n'est pas le cas de ces deux quantités avec le nombre de voisins par jours. Nous verrons par la suite que le nombre de contacts et la durée de ces derniers nous apportent très souvent le même type d'informations.



(a) Degrés quotidiens cumulés des personnes (en abscisses) versus la durée cumulée de contacts des personnes (en ordonnées)



(b) Degrés quotidiens cumulés des personnes (en abscisses) versus le nombre cumulé de contacts des personnes (en ordonnées) (c) Nombre cumulé de contacts des personnes (en abscisses) versus la durée cumulée des personnes (en ordonnées)

FIGURE 1.10 – Corrélations entre les différentes quantités.



## Chapitre 2

# Structure spatiale et fonctionnelle.

L'analyse de la répartition et du volume de contacts ayant lieu entre les personnes, les services de l'hôpital ainsi que les catégories socio-professionnelles apporte un grand nombre d'informations au sujet du fonctionnement d'un hôpital constitué de services de soins de longue durée. Ce type d'analyse apporte aussi des pistes de réflexions sur les risques de propagation dans un environnement tel qu'un hôpital, qui engendre un réseau de relations de proximité des personnes ayant des caractéristiques très spécifiques. Dans notre cas, nous avons la chance de posséder des données précises qui nous donnent des informations non seulement sur les paires d'adjacence, pouvant être considérées comme les arêtes d'un graphe, mais aussi sur le nombre de contacts par paires d'adjacence ainsi que leurs durées respectives. Grâce à nos données, nous pouvons mesurer l'activité des personnes selon trois quantités que nous appellerons *quantité d'activité*. La première de ces quantités nous informe sur la *diversité topologique de l'activité*. Elle consiste à compter le nombre de paires d'adjacence, lorsque que nous considérons l'activité d'un point de vue global. Autrement dit, nous nous intéressons au nombre d'arêtes, ou la somme des degrés cumulés des personnes (égale au double du nombre de paires d'adjacence), lorsque nous prenons le point de vue individuel des personnes. Les deux autres mesures nous renseignent sur le *volume d'activité* ; nous nous intéressons au nombre de contacts, global ou par personne, ainsi qu'à la durée cumulée de ces contacts. Tout au long de ce chapitre, nous considérons régulièrement ces trois quantités qui nous apportent des informations complémentaires.

Afin de déterminer si l'activité des personnes est liée aux activités médicales variées des différents services, nous commençons par considérer l'activité globale de chacun des neuf services de l'hôpital. Nous étudions ensuite les différences d'activité des personnes en fonction de leur rôle dans l'hôpital, ce qui nous amène à distinguer les patients des personnels. Nous introduisons ensuite quelques outils permettant d'estimer si les caractéristiques du réseau sont proches ou éloignées d'un réseau aléatoire afin de mieux mesurer la spécificité d'un réseau hospitalier. Nous nous intéressons également à la répartition des quantités d'activité entre chaque paire de service ainsi qu'à la vraisemblance de cette répartition au regard d'une répartition aléatoire des liens selon un modèle de configuration que nous présentons préalablement. Pour terminer ce chapitre, nous appliquons une grande partie des outils présentés et développés pour l'étude de l'activité des services de l'hôpital cette fois-ci aux catégories socio-professionnelles. Cette dernière partie permet d'apporter des précisions sur le fonctionnement de l'hôpital et en particulier sur les relations entre les divers corps de métier, ainsi que de vérifier que les outils présentés peuvent être appliqués

et se révéler utiles pour l'étude structurelle de relations entre groupes de nœuds dans un réseau.

## 2.1 Analyse gros grains des contacts : répartition des contacts par service.

Nous allons, dans cette partie, étudier la répartition des contacts des personnes regroupées en services. Afin de mieux connaître chacun des services que nous considérons dans cette partie, la figure 2.1 nous informe sur le nombre moyen de personnes présentes par jour dans chaque service. On peut remarquer que les services de l'aile Sorrel (S1 à S3) sont tous composés d'un nombre similaire de personnes au quotidien. Le nombre de patients est presque deux fois supérieur au nombre de personnels. On remarque que la proportion de personnels, par rapport au nombre de patients, est un peu plus réduite dans le service S2, que pour les deux autres services de ce bâtiment. Il faut aussi garder à l'esprit, lors des analyses qui vont suivre, que les services S4 et S5, tous deux situés dans le bâtiment Ménard, ont une population plus importante que les services du bâtiment Sorrel et sont composés de beaucoup de patients et relativement moins de personnels.

Commençons par une approche simple : utilisons les résultats de l'agrégation de l'activité par groupe (cf page 25 pour la définition) sur la durée totale de l'expérience et considérons la moyenne de l'activité à la journée ayant lieu dans chacun des services. Pour plus de commodité, nous appellerons cette activité, l'activité des services. On calcule cette activité en faisant la somme des quantités des services sur l'ensemble de l'expérience normalisée par le nombre de jour. Une petite subtilité intervient pour le calcul des paires d'adjacence. En effet, pour connaître le nombre moyen de paires adjacentes par jour, nous sommes obligés d'utiliser les données déjà agrégées à la journée, pour ensuite appliquer l'agrégation de groupe. Nous décrivons le calcul des activités moyennes par jour d'un groupe à l'aide de l'équation 2.1. Nous notons  $D$  l'ensemble des jours sur lequel est appliqué le calcul et  $G_d$  l'ensemble des personnes d'un groupe  $G$  présentes un certain jour  $d \in D$ . Dans cette partie, les groupes considérés seront les services de l'hôpital. La fonction  $A(G)$  représente l'activité globale d'un groupe  $G$  qui peut-être tour à tour le degré total des personnes du groupe  $G$ , le nombre de contacts à la journée ou la durée des contacts. Nous notons aussi  $\mathcal{A}(D, G)$  l'activité du groupe  $G$  durant les jours appartenant à l'ensemble  $D$  et  $A(d, p)$  l'activité d'une personne un jour donné.

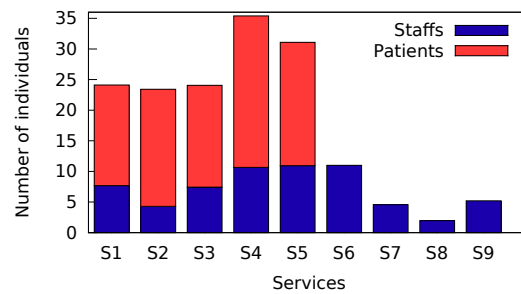


FIGURE 2.1 – Nombre quotidien de personnes présentes en moyenne dans chaque service. En rouge clair le nombre de patients et en bleu foncé le nombre de personnels.

$$\overline{A(G)} = \frac{\sum_{d \in D} \sum_{p \in G_d} A(d, p)}{|D|} \quad (2.1)$$

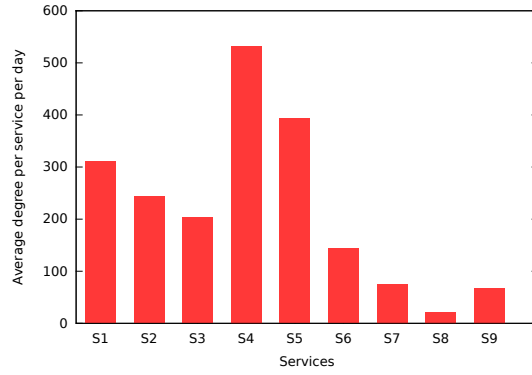
**Remarque:** Les contacts de l'équation 2.1 prennent en compte l'activité de chacune des personnes. Par conséquent, les contacts existant entre deux personnes d'un même groupe sont comptabilisés deux fois.

La répartition de l'activité de chacun des neufs services de l'hôpital est très différente en fonction des services comme le montre la figure 2.2. Cette figure nous informe sur les quantités quotidiennes moyennes d'activité de chacun des services : en haut, le degré quotidien moyen des services, en bas à gauche, leur nombre quotidien moyen de contacts et en bas à droite la durée quotidienne moyenne de contacts de ces services.

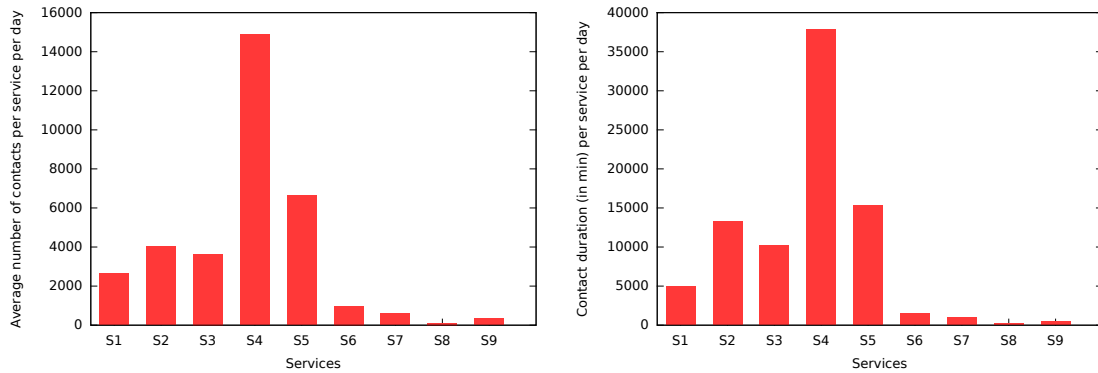
La figure 2.2 nous permet de distinguer nettement plusieurs niveaux d'activité, quelques soient les quantités observées. En effet, nous pouvons d'abord noter une très grande différence d'activité entre les services de soins, composés de patients et de personnels, qui sont très actifs comparés aux quatre services transversaux constitués uniquement par des personnels de l'hôpital et n'ayant pas de localisation propre. Il est à noter que les figures 2.2b et 2.2c ont relativement les mêmes patterns et apportent donc en grande partie les mêmes informations. Considérons maintenant les sous-figures plus en détails : on peut ainsi distinguer trois catégories de services, regroupés selon leur activité, en prêtant attention aux figures 2.2a et 2.2b. En effet, on peut repérer les services à très faible activité qui sont les services transversaux (de S6 à S9). Trois services ont une activité moyenne (S1 à S3). Ces trois services sont tous situés dans le même bâtiment, appelé Sorrel. Les deux derniers services ont une forte activité et sont eux aussi situés dans un même bâtiment nommé Ménard. Cette constatation n'est plus aussi évidente lorsque l'on considère les durées de contacts apparaissant sur la figure 2.2c où le service S5 a une activité beaucoup plus proche des services de l'aile Sorrel, ce qui isole le service S4 comme étant largement le service plus actif.

Pour continuer l'analyse, ordonnons les services par taux d'activité, selon chacune des trois mesures observées. On constate ainsi, d'après la figure 2.2a consacrée au degré cumulé des services, que plus les services sont près du rez-de-chaussée, plus leur degré journalier est important. En effet, les services S1 et S4 sont tous deux les plus actifs en terme de degrés et situés à l'étage du bas (rez-de-chaussée pour S1 et premier étage pour S4). On constate aussi que l'ordre des services est le même selon le nombre de contacts par jour et la durée cumulée quotidienne de ces contacts. Une grande différence apparaît par contre entre le degré quotidien des services et le nombre et la durée des contacts par jour, en particulier dans le bâtiment Sorrel dont les services S1 du rez-de-chaussée est de loin le plus actif en terme de degrés et semble très peu actif selon le nombre et la durée de contacts. Ce service semble donc privilégier la diversité de ses contacts plutôt que le volume contrairement aux deux autres services situés dans la même aile.

Ainsi, la figure 2.2 nous montre que chacune des quantités étudiées nous apporte des informations complémentaires. La figure concernant les degrés nous informe sur la diversité des contacts alors que les deux autres nous informent sur le volume de ces contacts. Nous avons déjà constaté que les services ayant une plus grande diversité de contacts ne sont pas nécessairement les plus actifs et donc ceux qui ont le volume d'activité le plus important.



(a) Degrés cumulés moyens par jour de chaque service.



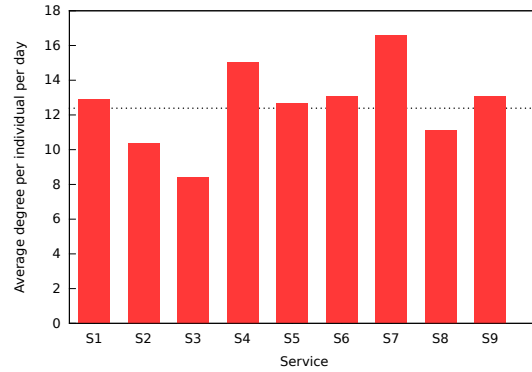
(b) Nombre de contacts cumulé moyen par service par jour. (c) Durée cumulée moyenne de contacts par service par jour.

FIGURE 2.2 – Moyenne des quantités d’activité de contacts entre les services de l’hôpital. En haut, le degré cumulé moyen par jour, en bas à gauche, le nombre de contacts moyen par jour et en bas à droite la durée cumulée par les personnes par jour.

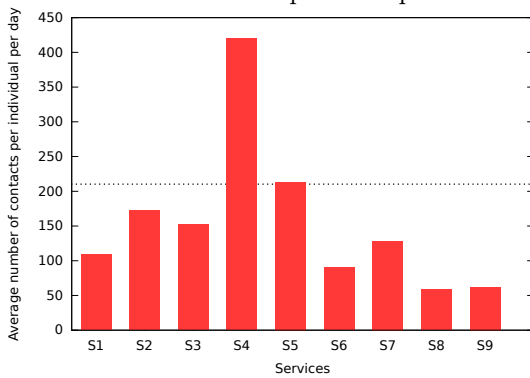
Il est nécessaire de s’assurer que les différences de comportement des services, observées sur la figure 2.2, ne sont pas uniquement dues à la taille de l’effectif de chacun des services. Une méthode simple pour s’abstraire du nombre de personnes consiste à considérer toutes les quantités par personne moyennées par jour. Ces quantités moyennes par individu sont obtenues simplement en sommant toutes les activités des personnes par jour. Ces sommes sont ensuite normalisées par la somme de personnes présentes par jour pour obtenir la moyenne individuelle de l’activité par jour, comme le décrit l’équation 2.2.

$$\langle A(G) \rangle = \frac{\sum_{d \in D} \sum_{p \in G_d} A(d, p)}{\sum_{d \in D} |G_d|} \quad (2.2)$$

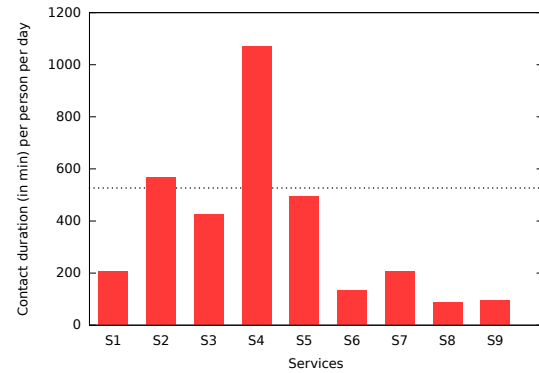
Les moyennes quotidiennes individuelles sont visibles sur la figure 2.3 : en haut, on peut lire le degré moyen quotidien par personne de chaque service ; en bas à gauche, on peut observer le nombre de contacts quotidiens moyens par personne de chaque service ; en bas à droite, la durée quotidienne moyenne par personne de chacun des services. Les disparités observées précédemment (figure 2.2) entre les services sont beaucoup moins marquées, quelques soient les quantités observées, bien qu’elles suivent les mêmes motifs.



(a) Degrés cumulé moyen par personne par jour pour chaque service.



(b) Nombre moyen cumulé de contacts par personne de chaque service par jour.



(c) Durée cumulée de contacts moyenne par personne par service par jour.

FIGURE 2.3 – Moyenne par personne des quantités d'activité de contacts entre les services de l'hôpital. En haut, le degré moyen par personne par jour, en bas à gauche, le nombre moyen de contacts par jour et en bas à droite la durée cumulée par les personnes par jour.

Les différences de patterns entre la figure 2.3a concernant les degrés moyens quotidiens par personnes des services avec les deux autres figures (2.3b et 2.3c) sont nettement plus marquées que précédemment (cf. figure 2.2a). On constate en effet des valeurs relativement homogènes à faible degré quotidien par personne quelque soit le service concerné, ce qui révèle une grande influence du nombre de personnes sur les valeurs de la figure précédente 2.2. Au contraire, le nombre quotidien moyen de contacts par personne et leurs durées restent important et très différents selon les services. On ne peut donc pas tenir pour seul responsable des disparités d'activité des services, en terme de volume, le nombre de personnes appartenant aux services.

Les activités moyennes des services transversaux (S6 à S9), en terme de nombre et de durées des contacts quotidiens par personnes, sont très faibles comparées à celles des autres services comprenant des patients. On constate aussi que les services les plus actifs sont les services S2, S4 et S5 qui contiennent tous trois une grande proportion de patients. L'activité individuelle des patients semble donc plus importante que l'activité des personnels. Cela n'explique pas toutes les disparités car les services S1 et S3 ont des quantités assez similaires de patients et de personnels mais des quantités d'activité très différentes. D'autres facteurs tels que les activités médicales et la localisation des services influencent aussi probablement ces disparités.

Jusqu'à présent, nous avons considéré puis comparé les résultats obtenus pour les trois quantités pour chacun des services. Nous avons remarqué que les résultats obtenus pour le nombre de contacts et leurs durées sont très similaires contrairement aux résultats concernant les degrés des services. Il est intéressant de savoir à quel point ces trois quantités sont corrélées ou non. La figure 2.4 vise à répondre à cette question. En effet, elle représente en haut la moyenne quotidienne par personne des degrés de chaque service en abscisses comparée au nombre quotidien de contacts par personne en ordonnées. En bas à gauche, la moyenne quotidienne par personne des degrés de chaque service en abscisses comparée à la durée quotidienne des contacts par personne en ordonnées. Enfin, en bas à droite, nous comparons le nombre de contacts quotidiens par personne en abscisses à la durée quotidienne des contacts par personne en ordonnées. Cette dernière figure nous permet d'obtenir une droite presque parfaite entre les points, ce qui signifie, comme nous l'avions déjà constaté pour les personnes, qu'il existe une très forte corrélation entre les durées de contacts et le nombre de contacts.

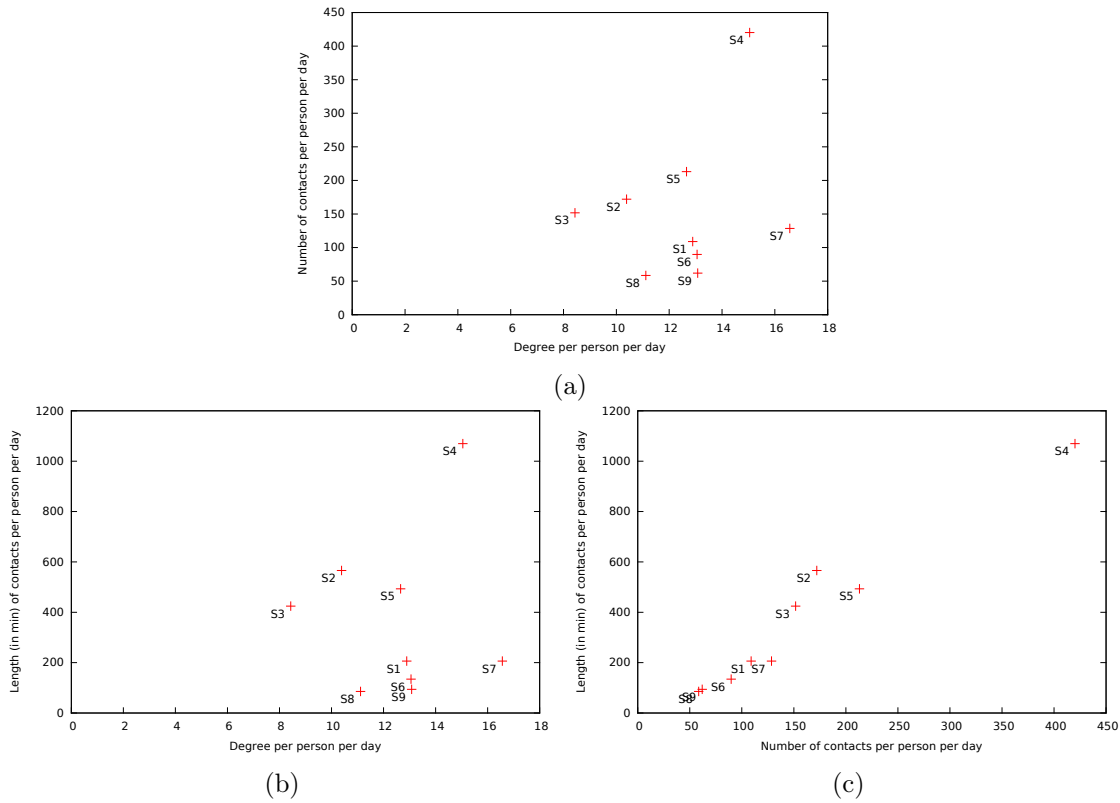


FIGURE 2.4 – Corrélation de la somme des valeurs des trois quantités obtenues pour chaque service par jour. En haut, le nombre de contacts cumulé comparé au nombre de voisins par jour et par personnes. En bas à gauche, la durée cumulée de contacts par services comparé au nombre de voisins par jour et par personne. En bas à droite, le nombre de contacts par service comparé avec la durée de contact par service.

En ce qui concerne les corrélations des figures en haut et en bas à gauche, aucune corrélation distincte n'apparaît. Le degré des personnes ne semble donc pas corrélé à la quantité d'activité, que ce soit en terme de nombre de contacts ou de durées. La diversité des contacts et la quantités de ces derniers sont donc deux informations complémentaires à prendre en compte.

Nous avons constaté jusqu'à présent que la quantité d'activité est répartie de manière plutôt déséquilibrée entre les services. Cela est en partie dû à une répartition inégale du nombre de personnes, mais aussi à deux autres facteurs moins quantifiables qui sont la localisation dans l'hôpital et leur fonction médicale. Nous avons aussi noté une différence considérable entre les services composés d'un grand nombre de patients et les autres, ce qui fait soupçonner une différence de comportement en fonction des rôles des personnes. Dans la partie suivante nous étudions donc le comportement des services, en distinguant l'activité des patients de celle des personnels pour comprendre plus en détails la répartition de l'activité des personnes à travers l'hôpital.

## 2.2 Différences patients/personnels

Afin de mieux comprendre l'activité des services, nous distinguons maintenant l'activité des personnels de celles des patients. Cela permet d'une part de comparer l'activité des services comportant des patients de ceux qui n'en ont pas en ne considérant que l'activité des personnels. D'autre part, cela permet de déterminer si l'activité précédemment observée des services n'est pas influencée par un comportement particulier des patients ou des personnels.

Commençons par une analyse générale du comportement global des patients et de celui des personnels en étudiant le tableau 2.1. Nous utilisons la notation PA pour désigner les patients et ST (abréviation de staff en anglais) pour désigner les personnels hospitaliers. Ce tableau est basé sur la répartition des quantités d'activité en trois types différents :

- Les interactions ayant lieu entre des patients uniquement.
- Les interactions impliquant deux personnes de groupes différents : un patient d'une part et un personnel d'autre part.
- Les interactions ayant lieu uniquement entre des personnels.

Le tableau 2.1 représente ainsi la répartition de l'activité entre les trois types définis dans cette partie. Nous considérons l'activité selon les trois quantités que nous observons depuis le début de ce chapitre (paire de contacts, nombre de contacts et durée de contacts). Deux point de vue sont donnés ici :

- Le point vue macroscopique global qui consiste à considérer chaque interaction faisant intervenir deux personnes une seule fois.
- Le point vue individuel, qui consiste à considérer les interactions de chaque personne, et revient à tenir compte de chaque interaction deux fois : une pour chacune des personnes interagissant.

Le tableau 2.1a résume les valeurs obtenues en considérant toutes les interactions entre les patients et les personnels. Les deux autres tableaux 2.1b et 2.1c résument ces valeurs du point de vue des patients puis du point de vue des personnels. En effet, le tableau 2.1b considère uniquement les interactions faisant intervenir au moins un patient tandis que le tableau 2.1c ne considère que les interactions impliquant au moins un personnel. On remarque que la majorité des paires de contact sont des paires impliquant un patient et un personnel. Le tableau 2.1a montre que le nombre de paires de patients (PA/PA) et de paires de personnels (ST/ST) est équitablement réparti ce qui met en évidence une répartition assez homogène des paires de contact entre personnes ayant un même rôle (patient ou personnel.). L'observation de la répartition de l'activité basée sur le nombre de contacts,

Valeurs	PA/PA	PA/ST	ST/ST	PA vs	PA	ST	ST vs	PA	ST
Paires	24%	56%	20%	Paires	46%	54%	Paires	58%	42%
Nombre	66%	21%	13%	Nombre	86%	14%	Nombre	44%	56%
Durée	80%	12%	8%	Durée	93%	7%	Durée	42%	58%

(a) Répartition globale pour les patients et les personnels.

(b) Répartition pour les patients

(c) Répartition pour les personnels.

TABLE 2.1 – Répartition des quantités de contact des patients et des personnels de l’ensemble de l’hôpital, présentée en pourcentages. Trois types d’interactions sont présentés ici : les interactions entre patients uniquement (indiquées par la colonne PA/PA), les interactions ayant lieu entre un patient et un personnel (représentées dans la colonne intitulée PA/ST) et enfin les interactions impliquant des personnels uniquement (apparaissant ici dans la colonne ayant pour titre ST/ST). Le tableau 2.1a fournit des valeurs globales, tandis que les tableaux 2.1b et 2.1c donnent un point de vue individuel. La première ligne (Paires) de chacun des tableaux présente la répartition des paires de contacts (égale au degré pour les tableaux du point de vue individuel 2.1b et 2.1c) pour chaque type d’interaction, en pourcentage. La deuxième ligne des tableaux (Nombre) concerne la répartition du nombre de contacts de chaque type. La troisième ligne (Durée) concerne la répartition de la durée cumulée des contacts.

puis sur la durée des contacts, souligne une répartition beaucoup plus hétérogène. En effet, en considérant ces quantités, on constate que les patients engendrent la grosse majorité de l’activité. Plus de 65% des contacts ont lieu entre deux patients et 20% entre un patient et un personnel. L’activité entre personnels étant de moins de 15% en terme de nombre de contacts quotidiens par personne et se réduit encore en terme de durée quotidienne de contacts par personne. Ces observations sont en grande partie dues au fait qu’il y ait plus de patients que de personnels présents sur le site au quotidien. En effet, si l’on considère le tableau 2.1c qui donne la répartition de l’activité des personnels, on constate que les quantités d’activité des personnels sont relativement équilibrées, avec une activité légèrement supérieure au sein du groupe des personnels. Du point de vue des patients 2.1b, l’activité des personnels semble très faible en terme de nombre et durée des contacts mais très équilibrée en terme de nombre de paires de contacts. La diversité des contacts est donc assez similaire chez les patients et les personnels. Les quantités d’activité observées sont par contre très différentes.

Les valeurs moyennes considérées précédemment nous permettent de soupçonner un comportement différent des patients et des personnels en terme de quantité. Qu’en est-il des valeurs particulières à chaque personne, chaque jour ? Sont-elles hétérogènes au sein des deux groupes ou plutôt réparties de manière homogènes ? La figure 2.6 permet de répondre à cette question en donnant la distribution des valeurs quotidiennes des personnes en terme de nombre de voisins (paires d’adjacence), de nombre de contacts et de durées cumulées de ces contacts. Cette figure est construite sur la base de résultats agrégés à la journée qui peuvent se concevoir comme étant une suite de réseaux statiques : un pour chaque jour. Il existe un lien entre deux personnes si ces deux personnes ont été en contact dans la journée représentée par le graphe. Ce lien est pondéré par le nombre de contacts ayant eu lieu entre ces deux personnes dans la journée ainsi que par la durée cumulée de leurs contacts ce jour-là. Une valeur est attribuée pour chaque jour de présence d’une personne, dont la situation est modélisée sur la figure 2.5 qui est un exemple de réseau temporel très simple de trois personnes. L’exemple 2.2.1 explique concrètement comment nous obtenons les quantités d’un réseau simple constitué de trois personnes, sur quatre jours consécutifs,



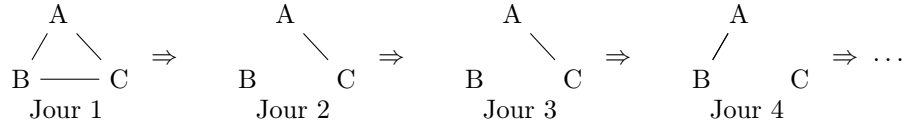


FIGURE 2.5 – Exemple simple de graphe dynamique.

Jour	1	2	3	4	Nombre de jours
A	2	1	1	1	4
B	2	X	X	1	2
C	2	1	1	X	3

TABLE 2.2 – Valeurs prises en compte pour la distribution des paires d’adjacence quotidiennes.

représenté par la figure 2.5. Le résumé des quantités apparaît sur le tableau 2.2. La figure 2.6 expose la distribution de ces trois quantités, obtenues à partir du réseau de proximité des personnes issu des données MOSAR.

### Exemple 2.2.1

Prenons le graphe de la figure 2.5 qui est formé par trois personnes étant présentes ou non sur quatre jours. Pour cette configurations, nous obtenons neuf valeurs dont nous cherchons à faire la distribution, comme présenté dans le tableau 2.2. Nous obtenons quatre valeurs pour la personnes A, trois pour la personnes C et deux pour la personne B. Dans cet exemple, nous considérons uniquement le nombre quotidien de paires d’adjacence.

On constate, en observant la figure 2.6 que les distributions des quantités quotidiennes des patients et des personnels ont des comportements assez différents. Comme nous l’avions déjà remarqué dans la partie 2.1, on peut constater que les personnels ont un degré journalier plus important que celui des patients. On peut noter un certain nombre de degrés quotidiens très importants pour les personnels, bien qu’ils ne représentent qu’un faible pourcentage. On peut noter les mêmes observations concernant les courbes des patients qui traitent du nombre et de la durée des contacts cumulés quotidiens. Malgré un grand nombre de valeurs extrêmes étendues sur des intervalles relativement grands et éloignés de la moyenne, la majorité des résultats se concentrent sur des intervalles très restreints, qui crédite les analyses basées sur des moyennes, qui sont significatives. En effet, plus de la moitié des personnes interagissent avec entre 6 à 16 personnes différentes au quotidien. Il y a environ 35% des personnels ayant un degré compris entre 16 et 30 alors que plus de 80% des patients n’atteignent pas 16 personnes différentes à la journée. Les patients ont donc une moins grande diversité de voisins à la journée mais ont une quantité d’activité beaucoup plus importante que celles de personnels. Les volumes d’activité plus faibles, observés chez les personnels, s’explique en partie par le fait que les personnels ne sont présents qu’une partie de la journée, contrairement aux patients qui ne quittent pas les lieux au cours de la journée. Le grand nombre de contacts, ainsi que les immenses durées cumulées des contacts relevées chez les patients comme chez les personnels suggèrent que beaucoup de contacts ont eu lieu en simultané.

Nous avons jusqu’à présent observé de très nettes différences entre le comportement des patients et celui des personnels. Nous savons aussi que les services ont une activité globale différente. Il est intéressant de savoir si le comportement des patients et celui des personnels de chaque service est identique, où si les activités des services restent différentes après

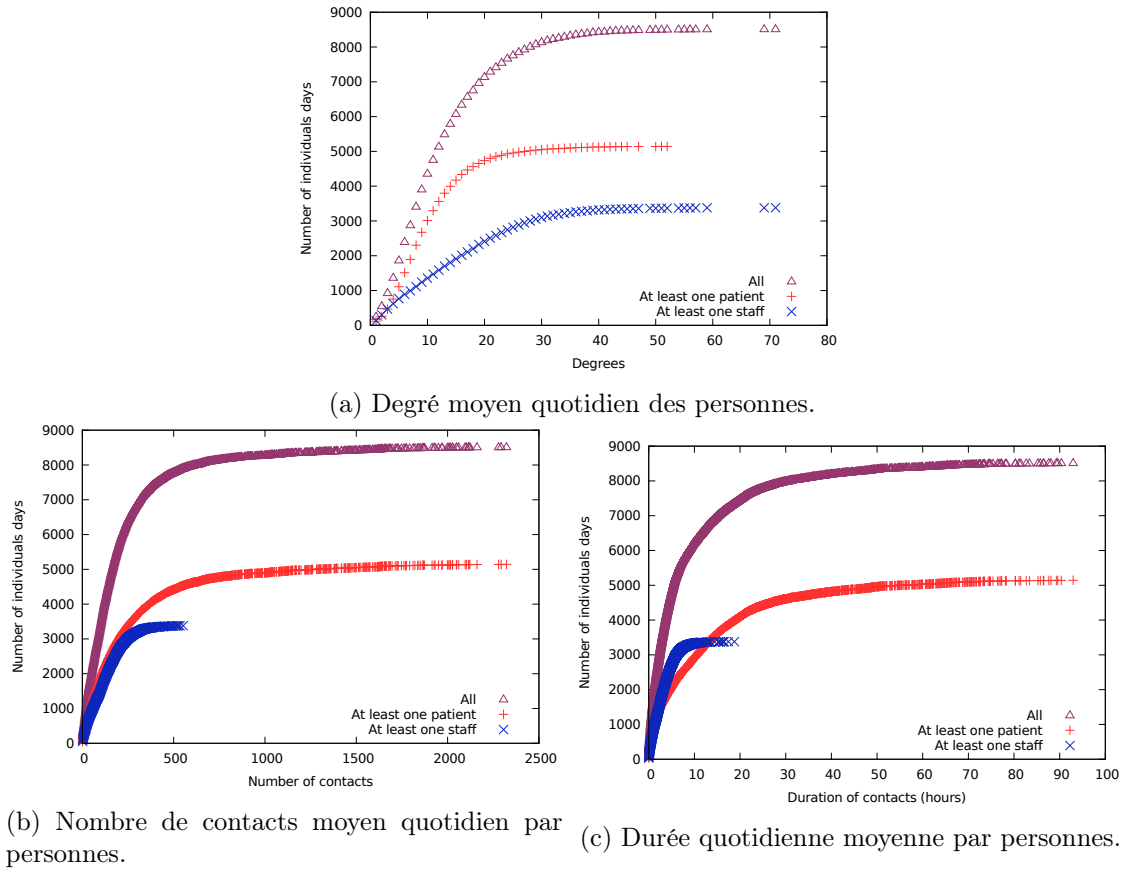


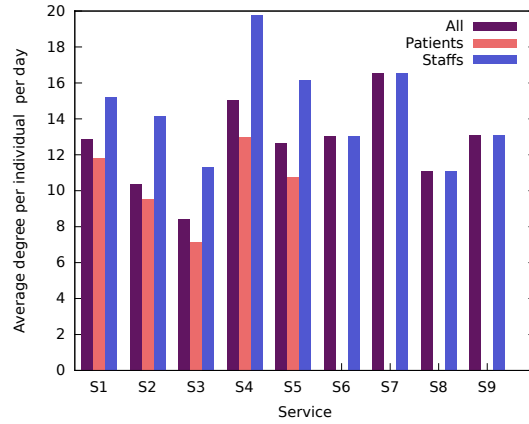
FIGURE 2.6 – Distribution des quantités quotidiennes par personne. En haut : distribution des degrés ; en bas à gauche distribution du nombre de contacts par jour ; en bas à droite : distribution des durées de contacts par personnes. Les courbes indigo sont les valeurs de l’ensemble des personnes, sans distinction de catégories ou de service. Les courbes bleues représentent les valeurs obtenues par les personnels de l’hôpital tandis que les courbes rouges représentent celles de patients.

avoir séparé les patients des personnels. Nous pouvons observer les résultats sur la figure 2.7 qui représente la moyenne quotidienne des quantités d’activité par personnes pour chacun des groupes. En indigo sont rappelés les valeurs obtenues pour chacun des services, sans distinguer les patients des personnels. En rouge sont représentées les valeurs obtenues pour patients et en bleu celles des personnels. Quelques soient les quantités observées, les valeurs des patients sont proches de celles de services globaux ce qui signifie que ce sont les valeurs des patients qui ont le plus d’influence sur les constatations générales que l’on peut faire. Il paraît donc encore une fois nécessaire de séparer l’étude des comportements des patients de ceux des personnels pour une meilleure compréhension du réseau que nous étudions. L’observation de la figure 2.7a montre que les patterns des patients et ceux obtenus pour les personnels sont très similaires bien que les valeurs soient éloignées. Les degrés moyens quotidiens de chacun des groupes restent très faibles. En ce qui concerne le nombre de contacts et la durée des contacts (cf figures 2.7b et 2.7c), les valeurs obtenues par les personnels sont très éloignées des courbes globales dont le pattern est relativement suivi par celles des patients. L’activité des personnels semble en effet relativement bien équilibrée et faible comparée à celle des patients qui se trouve être très déséquilibrée. Une partie de cette différence de concentration d’activité s’explique par le fait que les personnels sont présent dans la structure uniquement quelques heures par jour (8h environ)

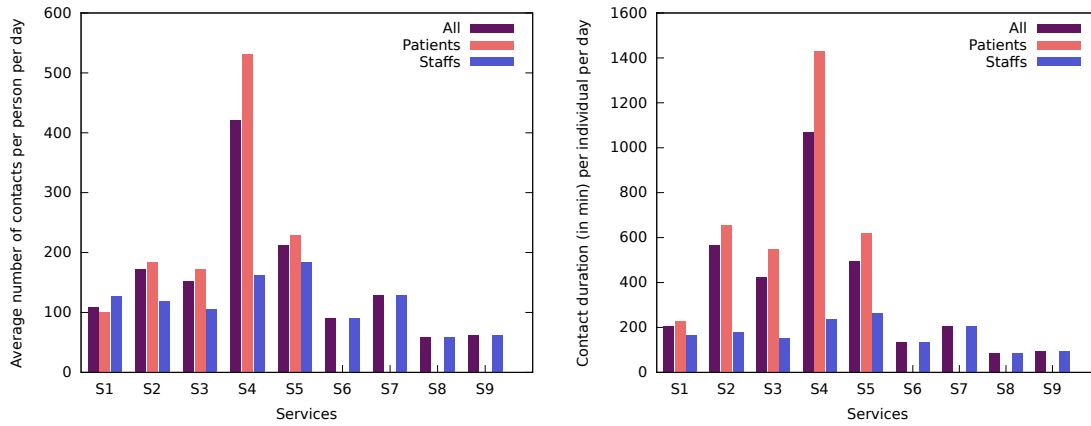
contrairement aux patients qui sont présents tout au long de la journée et de la nuit. Ce fait ne suffit pourtant pas à expliquer les différences considérables existant entre les patients des différents services.

Ainsi, les personnels peuvent être des vecteurs de transmission plus facilement que les patients car ils ont des contacts plus variés. Les patients ont par contre des temps de contacts très long qui peuvent eux aussi influencer considérablement la vitesse de transmission en cas de diffusion.

En résumé, nous observons une différence très nette entre le comportement des patients et de personnels que ce soit sur des constations globales ou par services. Nous savons quels sont les groupes qui ont tendance à être plus actifs, nous avons vérifié que les comportements les plus variés se trouvent parmi les groupes de personnels. Il nous reste à savoir si l'activité de chacun des services se fait plutôt en interne du service ou avec d'autres services de la structure pour avoir une meilleure vision des risques de transmission au sein de la globalité de l'hôpital. Il est aussi intéressant d'observer les préférences (accointances) de chacun des services en terme d'activité, ce qui fera l'objet de la section 2.4



(a) Degrés cumulés moyen par personne par jour pour chaque service.



(b) Nombre cumulé moyen personne par service (c) Durée cumulée de contacts moyenne par personne par service par jour.

FIGURE 2.7 – Moyenne des quantités d’activité de contacts entre des services. A droite, le degré moyen par jour, au centre, le nombre de contacts moyen par jour et à droite la durée cumulée par les personnes par jour. En indigo, tous les contacts sont pris en compte. En rouge, seuls les contacts impliquant des patients sont pris en compte. En bleu, seuls les contacts impliquant des personnels sont comptabilisés.

## 2.3 Introversion des services

Nous cherchons ici à caractériser les quantités d’activité échangées par les services ainsi que les quantités d’activité ayant lieu en interne de chaque service. Nous utilisons pour cela une notion d’introversion. Cette notion consiste à comparer le taux d’activité interne au taux d’activité externe au service.

### Definition 2.3.1 (Groupe)

Etant donné un ensemble  $H$  de personnes, nous appellerons *groupe* un sous-ensemble de  $H$ .

Nous noterons  $|H|$  l’effectif total de l’hôpital sur tout la durée de l’expérience, toute personne comptée une et une seule fois, qu’elle soit présente sur un seul ou plusieurs jours. Dans notre cas, l’ensemble principal  $H$  représente l’hôpital en entier. Les groupes de  $H$

peuvent être des services, des catégories socioprofessionnelles ou des sous-groupes encore plus spécifiques.

**Definition 2.3.2** (Contacts internes et externes)

Un contact est dit *interne* lorsque les deux personnes interagissant appartiennent au même groupe. Nous appellerons contact *externe*, tout contact faisant intervenir deux personnes appartenant à des groupes différents.

### 2.3.1 Notion d'introversion

Définissons maintenant plus formellement la quantité que nous appellerons introversion. Nous définissons l'introversion d'un groupe comme une valeur issue du rapport entre l'activité interne et l'activité externe d'un groupe. Les valeurs supérieures à 1, révèlent une activité interne des services plus forte que l'activité externe. Au contraire, les valeurs comprises entre 0 et 1 soulignent une activité externe plus importante que l'activité interne d'un service. Nous définissons ainsi l'introversion comme étant le rapport de l'activité interne d'un groupe  $G$  (notée  $Ia(G)$ ) et de l'activité externe de ce groupe  $G$  (notée  $Ea(G)$ ). La valeur 1 révèle une répartition parfaitement équilibrée des contacts internes et externes. Dans la suite, nous utiliserons les deux notations présentées par l'équation 2.3.

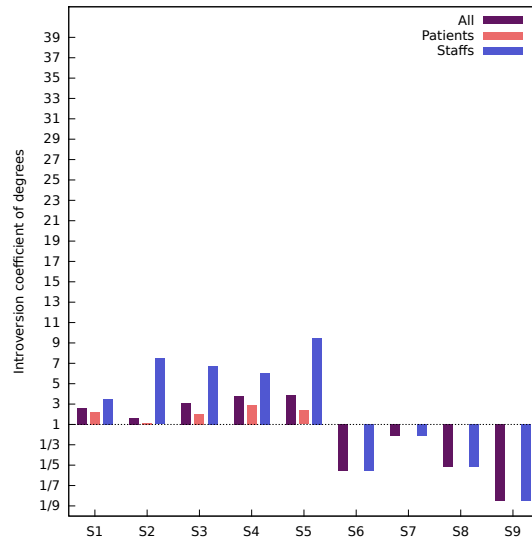
**Definition 2.3.3** (Introversion)

$$\mathcal{I}(G) = \mathcal{I}(Ia(G), Ea(G)) = \frac{Ia(G)}{Ea(G)} \quad (2.3)$$

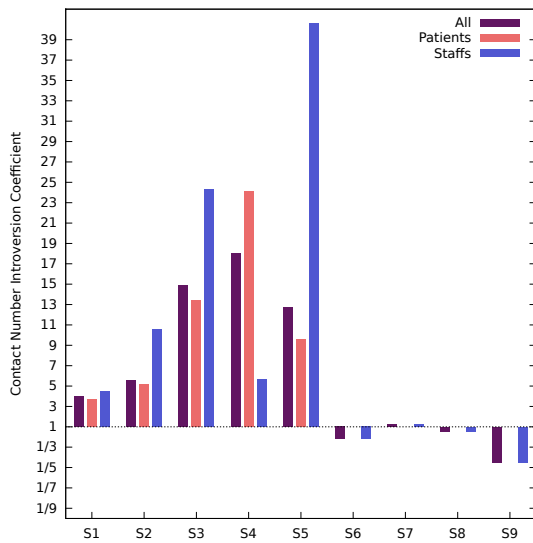
Nous nous intéressons à l'introversion des services en fonction des trois quantités habituelles : le nombre de voisins des personnes, leur nombre de contacts et la durée cumulée de ces contacts.

La figure 2.8 représente les valeurs d'introversion des services, calculées suivant l'équation 2.3. Nous distinguons encore une fois les valeurs globales, en indigo, des valeurs associées aux patients en rouge et des valeurs associées aux personnels décrites en bleu.

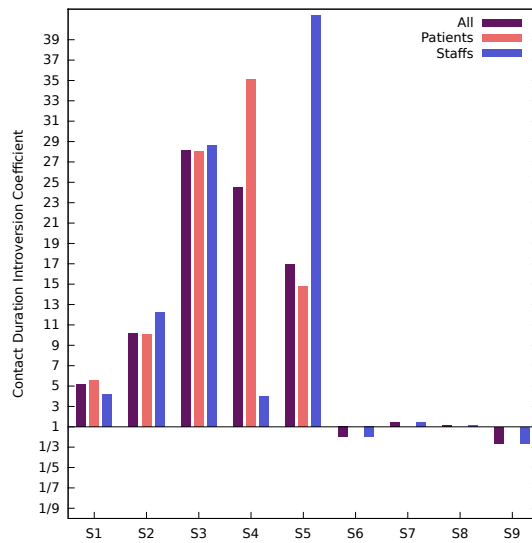
On peut constater, de par leurs valeurs toutes supérieures à 1, que tous les services de soins comportant des patients sont globalement introvertis quelques soient les mesures considérées. Au contraire les services transversaux semblent avoir une activité un peu plus équilibrée entre leurs activités intérieures et extérieures, lorsque l'on considère le nombre de contacts et leurs durées comme l'attestent les figures 2.8b et 2.8c. Ces mêmes services ont par contre plus de voisins à l'extérieur de leur service qu'à l'intérieur comme en témoigne la figure 2.8a qui résume l'introversion des services en terme de diversité d'activité (degré des personnes). Les services transversaux sont, pour au moins deux d'entre eux, extravertis, quelques soient les mesures sur lesquelles se basent les valeurs. Deux services transversaux de rééducation ont des tendances à l'introversion opposées en fonction des mesures considérées. Le service S8 a une activité très équilibrée dans la répartition de ses contacts internes et externes en terme de durée de contacts. Le service S7 a une répartition plus équilibrée en terme de nombre de contacts. Comparons maintenant le comportement des personnels et des patients des services de soins (de S1 à S5). On constate d'abord que l'introversion des patients, par rapport à celle des personnels, est très différente en terme de



(a) Introversion suivant le nombre de voisins quotidien.



(b) Introversion suivant le nombre de contacts



(c) Introversion suivant les durées de contacts

FIGURE 2.8 – Introversion des services en fonction des quantités considérées : en haut, les paires de contacts, en bas à gauche le nombre de contacts et en bas à droite la durée des contacts.

diversité (cf. figure 2.8a) ainsi qu'en terme de quantité (cf. figures 2.8b et 2.8c). La figure 2.8a fait ressortir une activité particulièrement équilibrée des patients du service S2, contrairement aux autres services dont les patients ont une introversion qui reste élevée, bien qu'elle soit inférieure à celle de leur personnels. En terme de nombre de contacts (cf. figure 2.8b) les introversions des patients et des personnels restent très différentes bien que l'on observe une augmentation générale des valeurs. On constate par contre que le service S4 a un comportement différent des autres services de soins. En effet, la valeur d'introversion des personnels est du même ordre que celle basée sur le degré des personnes tandis que la valeur d'introversion de l'activité des patients a considérablement augmenté et a largement dépassé celle des personnels du même service. Mis à part ce service, on constate une introversion toujours plus faible des patients que du personnel au sein d'un même service. Le

service 2, qui se distinguait pour son relatif équilibre en terme de degré des personnes agit comme les autres services en terme de nombre de contacts. Le nombre de contacts interne est donc très important, comparé à celui qui se fait à l'extérieur. La figure 2.8c qui traite de la durée des contacts fait ressortir encore une fois le service S4, qui, rappelons-le, est un service très actif en terme de diversité de contact des personnels. En effet, ce service obtient une valeur d'introversion beaucoup plus faible pour ses personnels en terme de durée de contacts que pour ses patients, ce qui n'est pas le cas des autres services. On remarque que les services du bâtiment Sorrel (S1-S3) l'introversion sur les durées de contact des personnels et des patients est relativement équilibrée, contrairement aux services du bâtiment Ménard (S4,S5).

Ainsi, la plupart des services sont introvertis, voire très introvertis, ce qui montre encore une fois encore la pertinence de considérer l'activité des différents services comme des groupes distincts ou des communautés bien définies. Les services transversaux partagent beaucoup plus équitablement leurs contacts entre l'intérieur et l'extérieur de leur service et sont extravertis en terme de diversité de contacts et presque introvertis en terme de quantité (nombre et durée des contacts).

### 2.3.2 Déviation d'introversion

Connaître l'introversion est une première information non négligeable, mais il est aussi important de savoir si le type d'introversion observé est caractéristique du réseau de terrain sur lequel on travaille ou s'il est commun avec des réseaux dont la répartition de l'activité entre les personnes est plus équitable ou aléatoire, en conservant certains paramètres fixes. Plusieurs méthodes peuvent être employées pour évaluer la structure moyenne du réseau. On peut citer les méthodes plutôt empiriques telles que les simulations [72] ou les méthodes analytiques telles que l'utilisation des modèles de configuration [10, 22, 59]. Nous avons privilégié ici une méthode analytique utilisant deux modèles de configurations différents.

**Definition 2.3.4** (Modèle de configuration)

Un modèle de configuration est un modèle de génération de graphe aléatoire qui à partir d'un nombre  $N$  de nœuds et d'une séquence de degrés  $(d_1, d_2, \dots, d_N)$  où  $d_1 \leq d_2 \leq \dots \leq d_N$  permet de définir l'espérance de l'existence d'une arête  $(i, j)$  comme  $(\frac{d_i * d_j}{2 \sum_{k=1}^N d_k})$ .

**Definition 2.3.5**

Nous appelons déviation d'introversion la distance entre l'introversion observée et l'introversion obtenue en utilisant un modèle de configuration choisi qui tient compte de certaines connaissances des données, à plus ou moins fort degré de détail.

Dans notre cas, deux modèles de configuration deux graphes à niveaux de détails différents sont pertinents.

- **Modèle simple** : Le modèle de configuration dont les nœuds sont les services du graphe. La séquence de degré est définie à partir des quantités (degré, nombre de

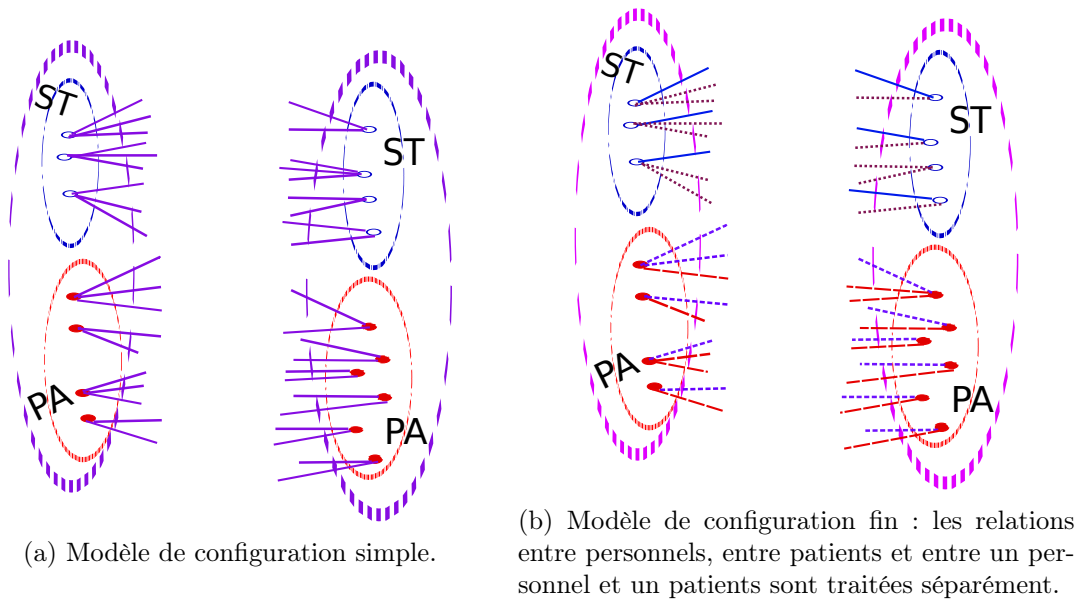


FIGURE 2.9 – Exemple simple de l'utilisation des modèles de configuration.

contact puis durées de contact) observées durant l'expérience. On autorise donc les arêtes multiples dans le cas du nombre et des durées de contact.

- **Modèle fin** : Le modèle de configuration est constitué d'un ensemble de nœuds correspondant aux services donnant chacun deux groupes quand ils existent : les patients et les personnels. Ce modèle de configuration contient donc théoriquement deux fois plus de nœuds que le premier. La séquence des degrés associés à ce modèle est toujours la liste des quantités observées associées à ces groupes durant l'expérience.

Une illustration de ces définitions se trouve sur la figure 2.9 qui montre la répartition des liens selon les deux modèles de configuration. Nous décrivons ici le phénomène de manière algorithmique pour des commodités visuelles. La figure 2.9a décrit la répartition moyenne des nœuds selon le modèle de configuration simple. Les demi-liens qui sont les traits associés à chacun des nœuds sur la figure sont ensuite associés deux à deux. La figure 2.9b donne une répartition des liens selon le modèle de configuration fin. Les traits pleins doivent ensuite être associés aux traits pleins (pour les entre relations personnels). Les traits rouges en pointillés très long doivent être associés à eux même (pour les relations entre patients). Les traits en pointillés courts bleus et pointillés en forme de point violets sont appareillés entre eux pour les relations personnels avec patients. En pratique seul le comportement moyen nous intéresse afin de calculer les déviations au modèle de configuration, ce qui nous permet de nous appuyer sur la définition mathématiques de l'existence de arêtes définie plus haut, sans passer par l'étape de simulation pour créer le réseau.

### Modèle de configuration simple

Le premier modèle de configuration, le plus simple et le plus généraliste, permet un calcul aisé des valeurs des introversions attendues d'un graphe aléatoire conservant les propriétés de quantités d'activité de chacun des services. En effet, nous calculons dans un premier temps les quantités attendues d'activité à l'intérieur des groupes et celles attendues avec l'extérieur en utilisant l'équation 2.4 qui peut s'appliquer à chacun des groupes définis au préalable. Nous notons ici  $Ta(G)$  l'activité totale observée pour le groupe  $G$  qu'elle soit



interne ou externe au groupe  $G$ . Nous dénotons l'ensemble de l'hôpital, privé du groupe  $G$  par  $H \setminus G$ . Ainsi, l'équation 2.4 répartit équitablement l'activité de  $G$  entre l'intérieur et l'extérieur de  $G$  (notées respectivement  $\overline{Ia_s(G)}$  et  $\overline{Ea_s(G)}$ ) en fonction de la proportion que l'activité de  $G$  représente face à celle de l'hôpital tout entier.

$$\begin{aligned} \begin{cases} \overline{Ia_s(G)} = Ta(G) * \frac{Ta(G)}{Ta(G) + Ta(H \setminus G)} & \text{si } |G| > 1 \\ \overline{Ia_s(G)} = 0 & \text{sinon} \end{cases} & \text{activité interne du groupe } G \\ \overline{Ea_s(G)} = Ta(G) * \frac{Ta(H \setminus G)}{Ta(G) + Ta(H \setminus G)} & \text{activité externe du groupe } G \end{aligned} \quad (2.4)$$

L'introversion de l'activité des services répartie selon le modèle simple de configuration, que nous noterons  $\mathcal{I}_s(G) = \mathcal{I}(\overline{Ia_s(G)}, \overline{Ea_s(G)})$ , s'obtient ensuite suivant l'équation 2.3 appliquée aux valeurs  $\overline{Ia_s(G)}$  et  $\overline{Ea_s(G)}$ .

L'évaluation de l'écart de l'introversion réelle comparée à l'introversion obtenue suivant un modèle de configuration simple s'obtient en faisant le ratio des deux valeurs, comme présenté dans l'équation 2.5.

$$Dev\mathcal{I}(G) = \frac{\mathcal{I}(G)}{\mathcal{I}_s(G)} \quad (2.5)$$

La figure 2.10 décrit la déviation d'introversion des différents services suivant la méthode simple décrite par l'équation 2.4 selon le nombre de paires des services, le nombre de contacts et la durée de ces contacts.

Une première observation montre que toutes les valeurs de déviation d'introversion sont considérablement supérieures à 1 ce qui signifie que tous les services, sans exception, ont une activité interne beaucoup plus conséquente que celle qui aurait lieu en suivant un modèle de configuration. A l'inverse, ils ont activité externe bien moins importante que dans un modèle de configuration ayant la même quantité d'activité, répartie aléatoirement. Nous constatons que seuls les services S6 (service de nuit) et S9 (service de rééducation) ont valeurs de déviation d'introversion inférieures à 10. En effet, ces deux services sont uniquement 2 à 3 fois plus introvertis que sur un réseau à répartition aléatoire de leur activité. Le service S4 se détache des autres avec le nombre et la durée des contacts. En effet, les valeurs de déviation d'introversion lui étant associées sont beaucoup plus faibles que pour les autres services de soin bien qu'elles indiquent toujours un fort éloignement par rapport au modèle de configuration simple. Ceci s'explique par la forte population constituant ce service : ainsi, le modèle de configuration attribut lui aussi une grande activité interne à ce service. Concernant les services transversaux n'ayant pas de localisation particulière (S6 et S9), nous constatons que la répartition des paires de contacts reste éloignée de ce que l'on obtient en suivant le modèle de configuration simple mais beaucoup plus faiblement que pour les autres services. Ce n'est plus le cas lorsque l'on considère le nombre et la durée des contacts : les services transversaux, bien qu'extravertis, ont eux aussi une activité interne beaucoup plus importante que celle à laquelle nous pourrions nous attendre.

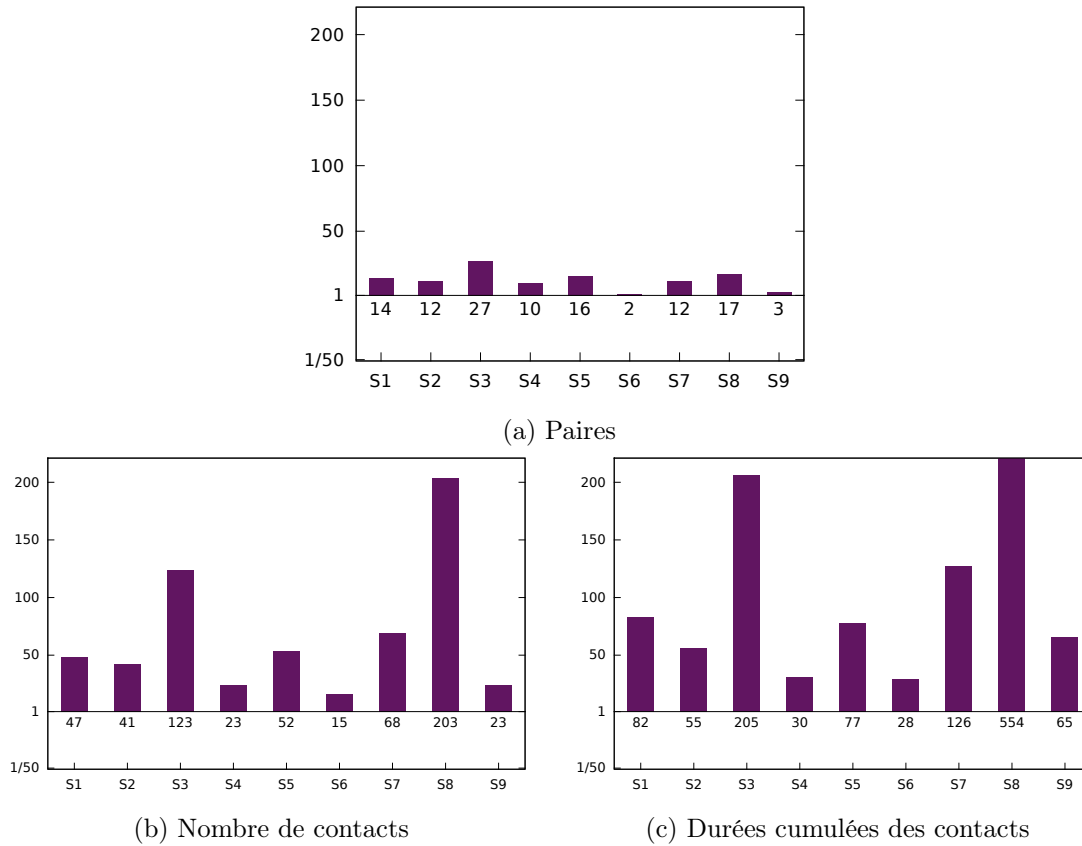


FIGURE 2.10 – Deviation d'introversion des services selon l'équation 2.4. En haut, selon les paires de contacts, en bas à gauche selon le nombre de contacts et en bas à droite selon la durée cumulée des contacts.

Les valeurs de déviation d'introversion par rapport au modèle de configuration qui répartit l'activité des services de manière aléatoire et équilibrée en fonction de la répartition de la population entre l'intérieur et l'extérieur des services confirme que le réseau de contact a une topologie particulière suivant la répartition de la population en services. Lors de la construction d'un réseau afin de modéliser une diffusion en milieu hospitalier, il est donc important de conserver ces informations topologiques caractéristiques en groupes bien formés.

### Modèle de configuration fin

La deuxième définition d'introversion est un peu plus fine et permet de se comparer au modèle de configuration qui redistribue équitablement l'activité en fonction de la proportion des éléments extérieurs et des éléments intérieurs tout en différenciant les relations patients et personnels. Cela nous permet d'examiner les activités des patients, puis des personnels que l'on pourrait observer en moyenne sur des réseaux conservant le même taux d'activité, mais en faisant varier la répartition des liens de manière aléatoire. Nous prenons en compte la quantité totale d'activité ayant eu lieu dans l'hôpital, et les quantités observées respectivement pour les patients et pour les personnels. Nous décrivons ensuite les quantités associées à chaque rôle que sont les patients et les personnels, à l'intérieur et à l'extérieur de chacun des services. Cela signifie qu'à chaque relation sont associés deux

paramètres : l'appartenance à un service et le rôle de l'individu (ici, patient ou personnel). Il existe donc quatre types des relations différentes à prendre en compte :

- Interaction entre deux personnes d'un même service  $S_i$  et d'un même rôle  $R_p$ , dont l'activité sera notée  $Ia((S_i, R_p, R_p))$ .
- Interaction entre deux personnes d'un même service  $S_i$  et d'un rôle différent  $R_p$  et  $R_s$ , dont l'activité sera notée  $Ia((S_i, R_p, R_s))$ .
- Interaction entre deux personnes de services différents dont l'une appartient au service  $S_i$  et d'un même rôle  $R_p$ , dont l'activité sera notée  $Ea((S_i, R_p, R_p))$ .
- Interaction entre deux personnes de services différents dont l'une appartient à  $S_i$  et d'un rôle différent  $R_p$  et  $R_s$ , dont l'activité sera notée  $Ea((S_i, R_p, R_s))$ .

Nous noterons  $Ta((S_i, R_p, R_s))$  la totalité de l'activité d'un groupe défini par son service  $S_i$  et son rôle  $R_p$  avec des personnes de rôle  $R_s$ , que les contacts soient internes ou externes au service.

Afin d'évaluer l'introversion attendue des patients et des personnels, nous devons tenir compte de ces quatre types de relations. Nous calculons l'introversion des services, en fonction des groupes de rôles. Pour chaque rôle, nous calculons le nombre d'interactions attendues à l'intérieur du service et le nombre d'interactions attendues avec l'extérieur. Nous calculons d'abord séparément le nombre d'interactions attendues pour les quatre types d'interactions décrits plus haut (conf. l'équation 2.6).

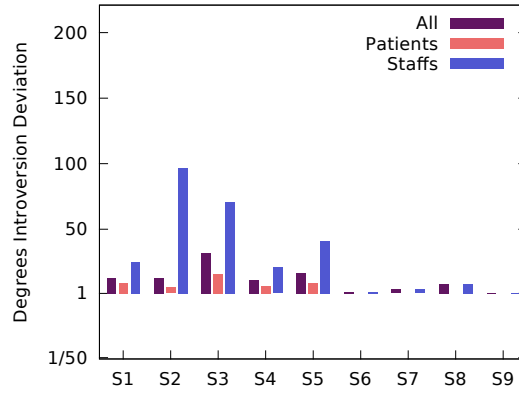
$$\begin{cases} \overline{Ia(S_i, R_p, R_p)} = Ta(S_i, R_p, R_p) * \frac{Ta(S_i, R_p, R_p)}{Ta(H, R_p, R_p)} \\ \overline{Ia(S_i, R_p, R_s)} = Ta(S_i, R_p, R_s) * \frac{Ta(S_i, R_p, R_s)}{Ta(H, R_p, R_s)} \\ \overline{Ea(S_i, R_p, R_p)} = Ta(S_i, R_p, R_p) * \frac{Ta(H \setminus S_i, R_p, R_p)}{Ta(H, R_p, R_p)} \\ \overline{Ea(S_i, R_p, R_s)} = Ta(S_i, R_p, R_s) * \frac{Ta(H \setminus S_i, R_p, R_s)}{Ta(H, R_p, R_s)} \end{cases} \quad (2.6)$$

Dans l'équation 2.6 nous dénotons les deux rôle différents que sont les patients et les personnels par  $R_p$  et  $R_s$ . De l'équation 2.6 nous déduisons facilement les quantités internes et externes attendues pour chacun des rôles. Ces quantités sont décrites par les deux premières lignes de l'équation 2.7. La dernière ligne décrit formellement l'introversion d'un sous-groupe d'un service.

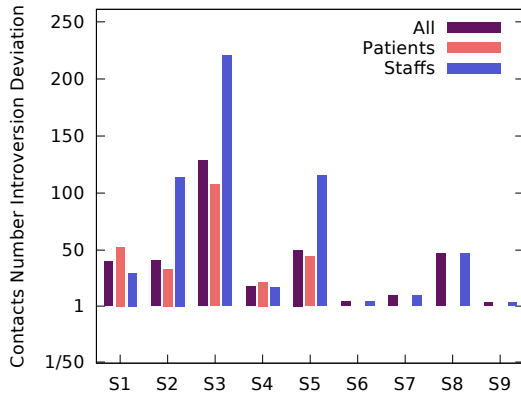
$$\begin{cases} \overline{Ia(S_i)} = \overline{Ia(S_i, R_p)} + \overline{Ia(S_i, R_s)} \\ \overline{Ea(S_i)} = \overline{Ea(S_i, R_p)} + \overline{Ea(S_i, R_s)} \\ \overline{\mathcal{I}(S_i, R_p)} = \mathcal{I}(\overline{Ia(S_i, R_p)}, \overline{Ea(S_i, R_p)}) \end{cases} \quad (2.7)$$

Connaissant les valeurs attendues pour chaque rôle (patient et personnel) et pour chaque service, en interne et en externe, il est aisé de reconstituer les valeurs attendues pour chacun des services en interne et en externe, suivant le modèle de configuration plus précis. L'équation 2.8 décrit comment obtenir ces valeurs.

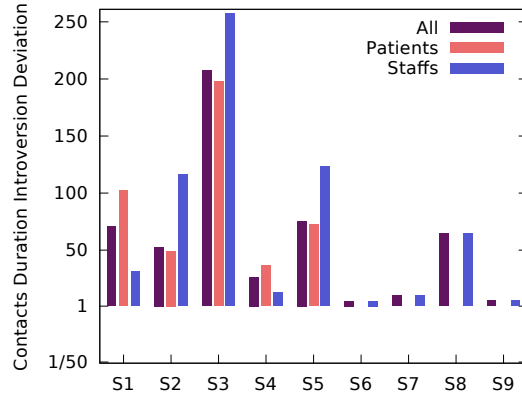
$$\begin{cases} \overline{Ia(S_i)} = Ia(S_i, R_p) + Ia(S_i, R_s) \\ \overline{Ea(S_i)} = Ea(S_i, R_p) + Ea(S_i, R_s) \end{cases} \quad (2.8)$$



(a) Déviation d'introversion basée sur les degrés quotidiens des personnes.



(b) Déviation d'introversion basée sur le nombre de contacts quotidiens des personnes.



(c) Déviation d'introversion basée sur la durée des contacts quotidiens des personnes.

FIGURE 2.11 – Déviation d'introversion selon l'équation 2.9 appliquée en haut aux paires de contacts, en bas à gauche au nombre de contacts et en bas à droite à la durée cumulée des contacts.

Les déviations d'introversion s'obtiennent, comme précédemment en calculant la distance entre l'introversion réelle du groupe considéré et l'introversion obtenu en suivant le modèle de configuration choisi (cf. équation 2.9).

$$Dev\mathcal{I}(S_i) = \frac{\mathcal{I}(S_i)}{\overline{\mathcal{I}(S_i)}} \quad (2.9)$$

La figure 2.11 représente les déviations d'introversion des patients, en rouge, et des personnels en bleu, obtenues suivant l'équation 2.7. En indigo sont représentées les valeurs de déviation d'introversion des services entiers calculées par rapport au modèle de configuration plus précis décrit par l'équation 2.8. On peut d'abord remarquer la forte similitude des courbes 2.11b 2.11c traitant des quantités d'activité qui sont très différentes de la courbe 2.11a traitant plutôt de la diversité des contacts. Nous analyserons donc les quantités d'activité séparément de la diversité de cette activité.

Commençons par quelques remarques concernant la diversité des contacts, représentée par la figure 2.11a. Les services transversaux ont toujours une très faible déviation d'introver-

sion et sont donc très proches du comportement moyen de modèle de configuration avec répartition aléatoire de la quantité d'activité bien qu'ils restent légèrement plus introvertis que lors d'une répartition aléatoire. Au contraire, les services de soins sont très éloignés des résultats que l'on obtiendrait en utilisant un modèle de configuration ce qui signifie qu'ils ont un comportement particulier, propre à l'environnement dans lequel les données ont été recueillies. La figure souligne aussi la différence de comportement des patients et des personnels. En effet, dans tous les services comportant des patients, les patients sont beaucoup moins éloignés du modèle de configuration que les personnels, bien qu'ils restent beaucoup plus introvertis que ce que prévoit le modèle de configuration. Les patients ont donc peu de diversité (cf. les conclusions de l'observation de la figure 2.7a) mais sont plus ouverts proportionnellement sur les autres services que les personnels. Ils peuvent donc aussi être des vecteurs de transmission non négligeables. Ceci sera à confirmer dans la partie 2.4. Concernant les valeurs de déviation d'introversion se basant sur le nombre et la durée des contacts (cf. figures 2.11b et 2.11c), on constate en premier lieu la grande similarité des valeurs des services de soin. On remarque aussi une différence entre les patients et les personnels très peu marquée à part pour le service S4. Encore une fois, ce service se détache des autres et les personnels de ce service sont plus introvertis que suivant un modèle de configuration, bien que beaucoup moins éloigné de ce modèle que les autres services. En terme de quantité des contacts les patients ont une plus nette tendance à l'introversion que les personnels.

Nous avons ainsi défini dans cette partie une mesure d'éloignement à la structure moyenne d'un réseau que nous appelons déviation d'introversion afin de quantifier la particularité du réseau avec lequel nous travaillons. Nous avons constaté que la structure de ce réseau est très centrée sur les services de soin, très introvertis. La structure de la répartition des contacts des patients et des personnels est distincte en terme de diversité des contacts mais relativement similaire en terme de nombre et de durée de contact bien que ces deux catégories (patients et personnels) n'aient pas la même quantité d'activité associée. Il est donc important de distinguer la quantité d'activité d'une part et la répartition de l'activité d'autre part car ce sont deux paramètres importants lorsque l'on étudie les capacités de diffusion d'un réseau.

Afin de mieux comprendre encore les relations et la répartition de l'activité des services, la partie 2.4 propose une autre étude plus détaillée de la répartition des contacts entre les différentes paires de services.

## **2.4 Etude des interactions de groupes prédéfinis d'un réseau de contacts.**

Cette partie est consacrée à l'étude des interactions entre les services pour estimer la répartition de l'activité entre les différents services. En particulier, nous cherchons à faire ressortir les accointances des services et à les expliquer.

### **2.4.1 Interactions entre les services**

La première évaluation des accointances des services passe par l'étude de la répartition de l'activité globale des services, toutes personnes confondues.

### Répartition des personnes

Pour évaluer les risques de diffusion du réseaux, il est très intéressant de connaître le nombre de personnes par jour établissant un lien avec chacun des groupes définis au préalable. En effet, plus il y a de personnes différentes interagissant avec des groupes autres que les leurs, plus il existe de "passerelles" entre les groupes ce qui augmente les possibilités de diffusion du réseau. La figure 2.12 représente la quantité de personnes de chaque service interagissant avec chacun des autres services. Dans cette partie, nous notons  $D$  l'ensemble des jours de l'expérience et  $d \in D$  un jour donné. En pratique, nous calculons, pour chaque service  $S$ , le nombre de personnes présentes par jour, noté  $|S(d)|$  puis nous sommons toutes ces valeurs quotidiennes pour avoir une valeur globale dans une unité que nous nommerons personnes-jours et noterons  $|S^g(D)| = \sum_{d \in D} |S(d)|$  pour le service  $S$ . Nous calculons de même le nombre de personnes-jours d'un groupe  $G_1$  qui interagissent avec un groupe donné  $G_2$ , ce que nous notons  $|G_1^g(D)|_{\mapsto G_2}$ . La moyenne du nombre de personnes d'un groupe  $G_1$ , interagissant avec un autre groupe  $G_2$  est alors le quotient du nombre de personnes-jours de  $G_1$  interagissant avec le groupe  $G_2$  et du nombre de jours  $|D|$ . Cette quantité, que l'on peut aussi décrire comme étant le nombre quotidien moyen de personnes du groupe  $G_1$  interagissant avec  $G_2$  est décrite formellement par l'équation 2.10.

$$\overline{|G_1^g(D)|}_{\mapsto G_2} = \frac{\sum_{d \in D} |G_1(d)|_{\mapsto G_2}}{|D|} \quad (2.10)$$

Le pourcentage du nombre de personnes d'un groupe  $G_1$ , interagissant avec un autre groupe  $G_2$  est le quotient du nombre de personnes-jours de  $G_1$  interagissant avec le groupe  $G_2$  sur le nombre réel de personnes-jours présentes dans le groupe  $G_1$ , ce qui peut s'écrire comme proposé dans l'équation 2.11.

$$P(G_1^g(D))_{\mapsto G_2} = \frac{\sum_{d \in D} |G_1(d)|_{\mapsto G_2}}{\sum_{d \in D} |G_1(d)|} \quad (2.11)$$

Les matrices représentées sont asymétriques. Les valeurs de chacun des services se lisent en ligne. Ainsi, la case  $M[i][j]$ , avec  $i$  l'indice des lignes et  $j$  l'indice des colonnes indique la quantité de personnes du groupe  $i$  interagissant quotidiennement avec une ou des personnes du groupe  $j$  (cf. l'exemple 2.4.1). Une interprétation visuelle des nombres apparaissant dans les cases de la matrices 2.12a est donnée par l'exemple 2.13.

#### Exemple 2.4.1

Prenons la matrice  $M$  de la figure 2.12a en guise d'exemple de lecture. La case  $M[S6][S2]$  contenant la valeur 4.58 indiquent que 4 à 5 personnes par jour en moyenne, appartenant au service S6, interagissent avec au moins une personne du service S2. La figure 2.12b indique que cela représente au total un tiers (0,33) des personnes du service S6. A l'inverse, la case  $M[S2][S6]$  contenant la valeur 15.30 indique qu'environ 15 à 16 personnes interagissent quotidiennement avec une personnes du service S6. Cela représente deux tiers (0,65) des personnes du service S2, d'après la figure 2.12b

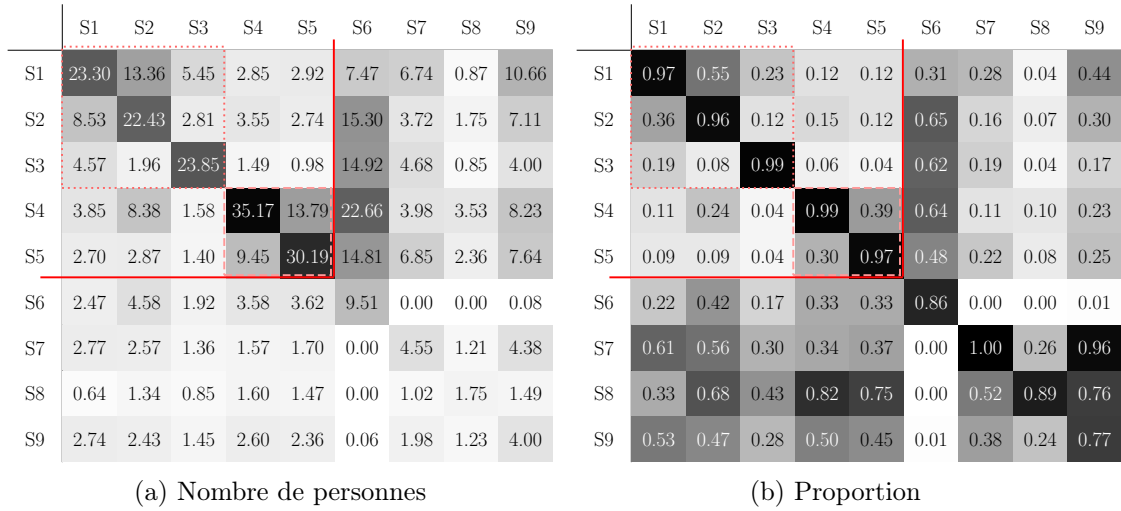


FIGURE 2.12 – Quantité de personnes par jour de chaque service (en ligne) interagissant avec les autres services (en colonne). La figure de gauche donne la moyenne du nombre de personnes quotidien des groupes interagissant avec chacun des services. La figure de droite donne la proportion de personnes des services interagissant avec les différents services.

#### Exemple 2.4.2

Cet exemple se base sur la figure 2.13. Prenons les deux cas  $M[S1][S7]$  et  $M[S7][S1]$  de la figure 2.12a et interprétons la en terme de réseau. On commence par arrondir le nombre de personnes concernées par les échanges entre les services S1 et S7. Cela nous donne 3 personnes en moyenne appartenant à S7 interagissent avec 7 personnes de S1. La figure 2.13a résume cette information à l'aide d'un réseau dont les arêtes relient les deux groupes de personnes des deux services concernés par les échanges. Il faut au minimum 3 arêtes pour relier toutes les personnes de S7 interagissant avec S1 à ce dernier service. La figure 2.13b donne trois scénarios différents de répartition des paires de contacts pouvant exister entre S1 et S7. Les deux premiers scénarios utilisent 7 arêtes qui est le nombre minimum pour satisfaire les contraintes, à savoir sept personnes de S1 ayant au moins une arête avec quelqu'un de S7 et 3 personnes de S7 partageant au moins une arête avec une personne de S1. La dernière figure, en bas, donne un scénario possible avec plus d'arêtes que nécessaire pour satisfaire les contraintes. Ainsi, avec uniquement l'information du nombre de personnes concernées par les échange avec chacun des services, cela permet de savoir si l'interface entre les différents services est large ou non. Nous ne savons par contre par si cette interface permet de multiples échanges ou non.

La figure 2.12a, donnant le nombre de personnes quotidien interagissant avec les services de l'hôpital fait ressortir les services de soins, qui ont une population plus nombreuse, ayant des interactions internes très prononcées. En effet, une diagonale apparaît sur la partie supérieure gauche de la figure, partie qui est consacrée aux interactions entre les services de soins. On peut aussi distinguer nettement les deux ailes de services de soins. La première est l'aile du bâtiment Sorrel, constituée des trois premiers services (S1-S3). Le deuxième bâtiment, nommé Ménard, accueille les services S4 et S5. Ces deux derniers services forment un carré plus foncé car le nombre de personnes impliquées dans leurs interactions, de part et d'autres est plus important que le nombre de personnes interagissant avec les autres services de l'hôpital. On peut aussi remarquer un plus grand nombre de personnes interagissant entre les services S1 et S2 qui sont situés respectivement au rez-de-chaussée et au premier étage du bâtiment Sorrel. Connaissant la topologie de l'hôpital, et la répartition

- (a) Donnée : 3 personnes de S7 interagissent avec S1 (b) Trois scénarios différents correspondant à l'information donnée et représentée par la figure 2.13a

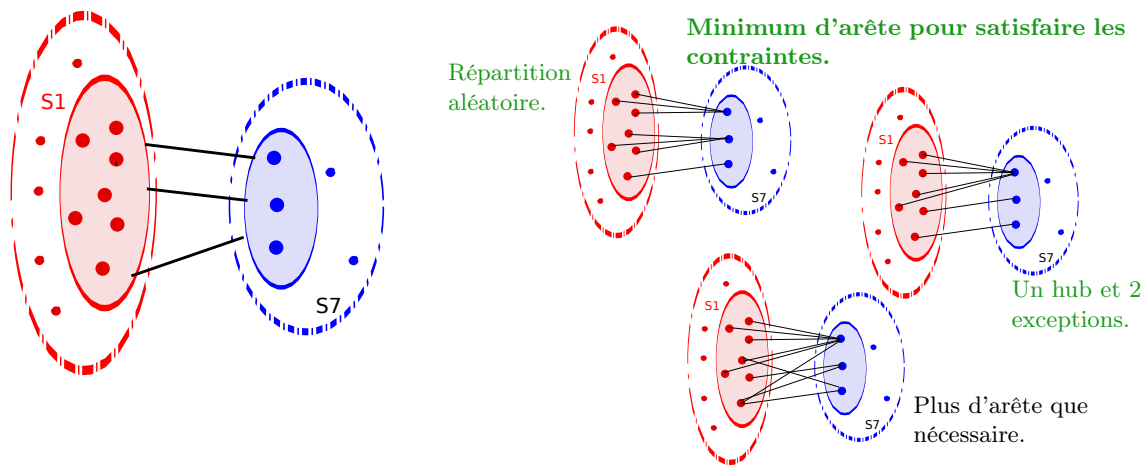


FIGURE 2.13 – Répartition possible des paires de contacts, vues comme des arêtes, entre les services S1 et S7 qui ont en moyenne arrondie à l'entier 3 personnes de S7 qui interagissent avec 7 personnes de S1.

des cinq services de soins dans l'espace, on peut remarquer que plus ils sont physiquement proches, plus ils semblent avoir d'interactions impliquant un grand nombre de personnes. On retrouve donc relativement bien la topologie de l'hôpital et la séparation des bâtiments en observant le nombre de personnes interagissant entre les différents services géolocalisés que sont les services de soins. Les services transversaux, quant à eux, ont une population très réduite qui semble se répartir de manière relativement équilibrée entre tous les services. Les membres de services transversaux interagissent aussi beaucoup les uns avec les autres au sein d'un même service, ce qui se remarque très aisément sur la figure 2.12b, nous informant sur le pourcentage de personnes du services impliquées par les interactions. Nous pouvons remarquer une diagonale parfaite qui confirme que la grosse majorité des personnes de chacun des services interagissent avec des personnes de leur propre service. La valeur la plus basse étant 77%, concernant le service S9 qui est composé de personnels administratif, pouvant éventuellement être représenté par une seule personne, certains jours, ce qui empêche toute interaction avec son propre service. Les services transversaux de jour (S7-9) ont une forte proportion (plus de 25%) de leur population qui interagit avec les services de soins, quels qu'ils soient. Le service de nuit S6 semble intervenir beaucoup plus dans l'aile Ménard (S4-5) ainsi qu'au niveau du service S2 que dans les deux autres services de soins (S1 et S3). Le service S3 semble d'ailleurs avoir un comportement différent des autres services de soins. En effet, la proportion de population interagissant avec des services extérieurs est très faible (inférieure à 20%) sauf avec le service de nuit qui semble sollicité par une majorité de la population.

Ainsi, les différents services de soins échangent avec d'autant plus de personnes qu'ils sont géo-localement proches, exception faite du service S3 n'interagissant que très peu avec l'extérieur, à part avec le service de nuit. Le service S3 semble donc être l'un des plus isolés en terme de nombre de diversité de population interagissant avec d'autres services. Au contraire, les services transversaux semblent faire des passerelles entre tous les autres services, bien qu'ils soient peu nombreux.



## Quantité moyenne d'activité échangée par les services

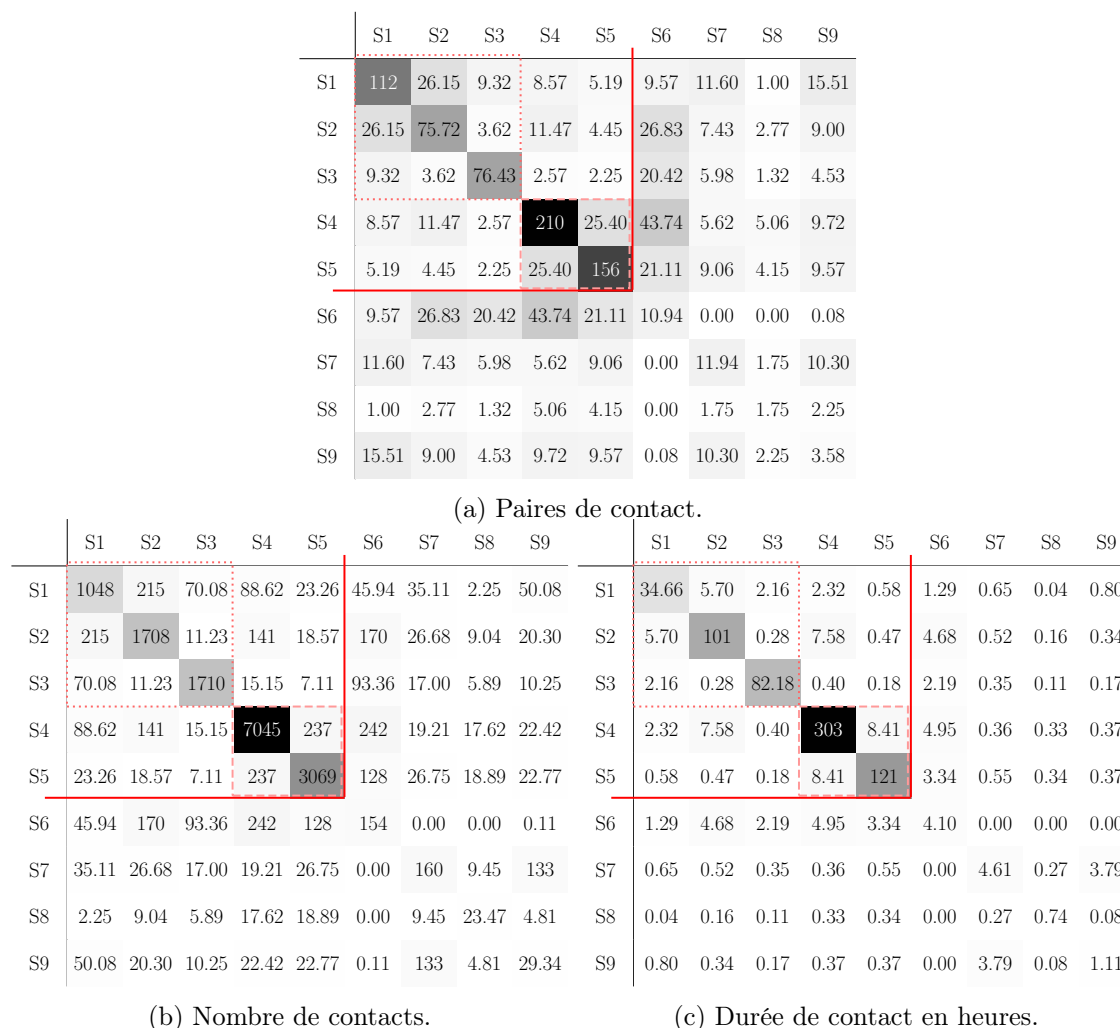


FIGURE 2.14 – Matrices d'activité entre les services par jour toutes personnes confondues, selon les degrés quotidiens, le nombre de contacts et la durée des contacts.

Nous nous intéressons ici à la répartition de l'activité quotidienne de chacun des services. La figure 2.14 résume la répartition du nombre de paires de contacts impliquées, du nombre de contacts et de la durée de ces contacts, exprimée en heure entre les différents services. Cette figure est symétrique, tout comme les interactions qu'elle représente. Cette figure fait ressortir encore une fois une activité interne prépondérante des services de soins face à l'activité externe entre les services, ce qui se voit à diagonale formée entre les services de soins qui ressortent grâce à l'importance de leurs valeurs, et ce sur les trois figures. La figure 2.14a concernant les paires de contact fait aussi ressortir les deux bâtiments de services de soins. En effet, nous pouvons distinguer le bâtiment Ménard, constitué des services S4 et S5, entouré sur la figure par un ligne en pointillés long. Ces deux services répartissent la grosse majorité de leur voisins entre eux. De même dans l'aile Sorrel constituée des services S1, S2 et S3, la grosse majorité des voisins de ces trois services reste concentrée dans le bâtiment. On peut remarquer une nette préférence entre S1 et S2 ; S3 étant plus introverti.

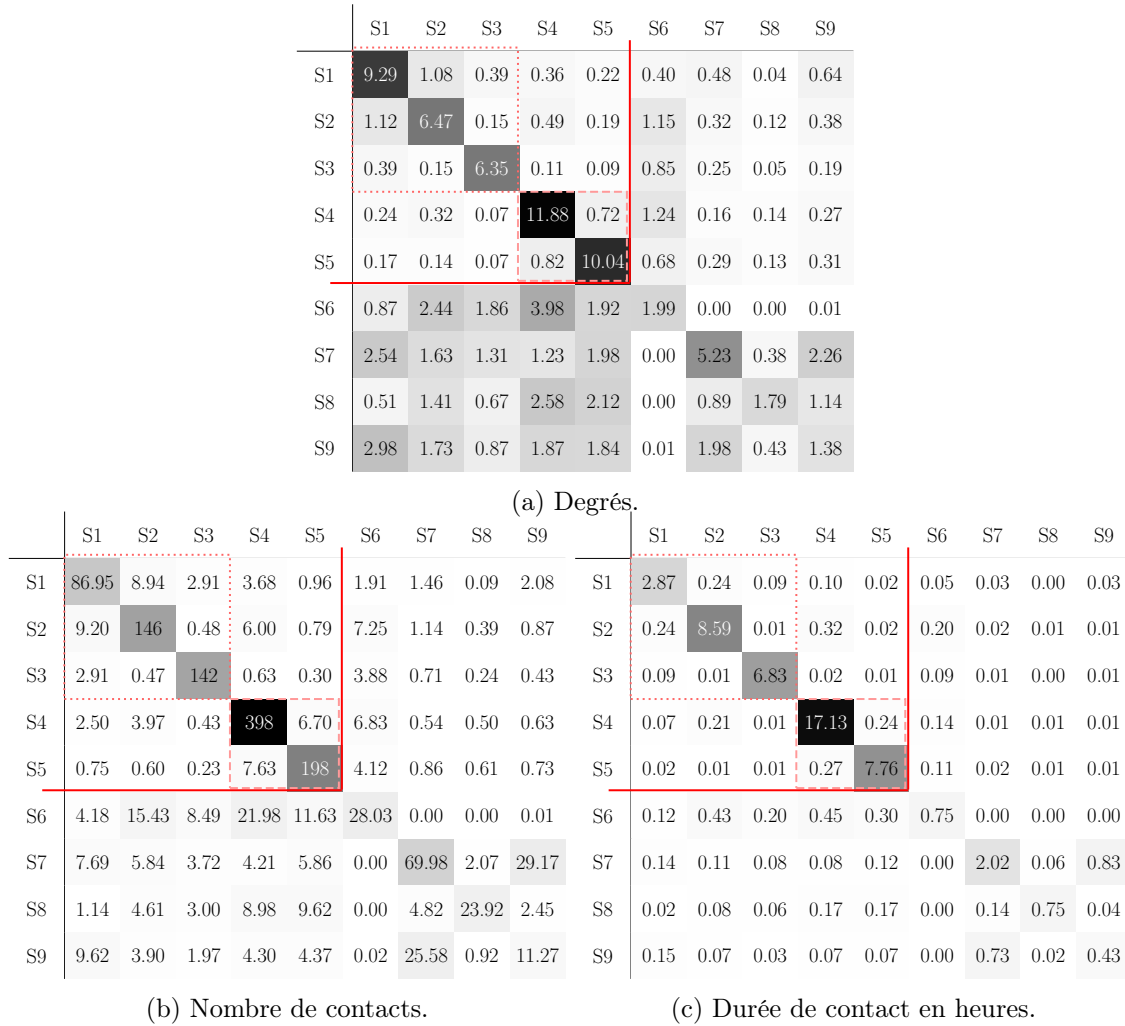


FIGURE 2.15 – Matrices d'activité entre les services par personne jour, selon les degrés quotidiens, le nombre de contacts et la durée des contacts.

La figure 2.15 représente l'activité ayant lieu entre chacun des services par personnes par jour, suivant le nombre de paires, le nombre de contacts et la durée des contacts par jour. Ces trois matrices sont asymétriques du fait de la différence de population des services et se lisent en ligne. Chaque ligne représente un service et donne l'activité quotidienne par personne que ce service partage avec les services lus en colonne (cf. l'exemple 2.4.3).

### Exemple 2.4.3

Prenons la matrice  $M$  de la figure 2.15a en guise d'exemple de lecture. Nous dénotons la case située ligne 2 colonne 6 par  $M[S2][S6]$ . Cette case comporte la valeur 1.15 ce qui signifie qu'environ chaque personne du service S2 interagit avec une seule personne de S6. Au contraire, la case  $M[S6][S2]$  comportant la valeur 2.44 signifie chaque personne appartenant au service S6 interagit avec un peu plus de deux personnes de S2 par jour en moyenne.

La figure 2.15 fait ressortir encore une fois une activité interne prépondérante des services de soins face à l'activité externe entre les services, ce qui se voit aux trois diagonales qui ressortent comme précédemment du fait de l'importance de leur valeur. On peut aussi très nettement distinguer les services de soins des services transversaux. En effet, les services

transversaux semblent répartir leur activité de manière plus homogène entre les différents services, dont le leur, que les services de soins. La figure 2.15a concernant le degré des personnes renforce cette idée d’homogénéité de répartition des contacts des services transversaux. En effet, les services transversaux répartissent de manière équilibrée le nombre de leur voisins entre les services de soins. L’exception à cette règle est le service S6 qui est un service particulier fonctionnant uniquement de nuit, ce qui explique sa quasi-inactivité avec les autres services transversaux qui sont eux, des services travaillant uniquement de jour. Les services de soins S2, S3 et S4 sollicitent beaucoup le service de nuit S6 car on peut remarquer que chaque personnes de ces services de soins interagissent en moyenne chaque jour avec une personne de S6. Des soins sont donc plus régulièrement donnés la nuit dans ces trois services que dans les deux autres services de soins que sont S1 et S5 qui ont peu de contacts avec le service S6. Les figures 2.15b et 2.15c nous indiquent que la plus grosse majorité des contacts extérieurs représentent une très petite quantité d’activité comparée à l’activité des personnes à l’intérieur de leur services. Le nombre de contacts par personnes, avec chacun des autres services est tout de même plus élevé entre des services géo-localement proches, de même que la durée totale de leurs contacts, toujours inférieure à une heure par personne par jour.

### Répartition de l’activité entre les services

Nous avons jusqu’à présent considéré les valeurs brutes des quantités d’activité ayant lieu entre les services. Une autre question importante est de savoir ce que représente les quantités d’activité du point de vue des services : est-ce la majorité de leurs contacts ou bien une faible proportion. Nous appellerons cette notion le *taux d’occupation* des groupes et nous l’évaluerons toujours selon les trois mesures d’activité (le degré des personnes, le nombre de contacts et la durée cumulée des contacts). Ce taux d’occupation est calculé de manière très naturelle pour chaque groupe  $G$  en normalisant la quantité d’activité que ce groupe  $G$  partage avec un autre groupe  $G'$  par la somme totale d’activité du groupe  $G$ . En d’autres termes, si la matrice M1 contient les valeurs des quantités d’activité de chaque service, comme les matrices de la figure 2.14, la matrice M2 contenant les taux d’occupation des services sera calculée comme suit :

$$M2[i][j] = \frac{M1[i][j]}{\sum_k M1[i][k]}$$

Les taux d’occupation des services sont représentés sur la figure 2.16 qui exprime la répartition de l’activité du point de vue de chacun des services. Cette fois encore, cette figure se lit en ligne, ce qui signifie que la case M[S6][S4] de la figure 2.16a indique que 33% des voisins quotidiens du service transversal S6 appartiennent au service S4. Au contraire, la case M[S4][S6] signifie que seulement 14% des voisins du service S4 sont du service S6.

La figure 2.16 confirme la relative homogénéité de la répartition de l’activité des services transversaux, comparée à celle des services de soins. En effet, les services transversaux partagent moins de 40% de leur activité au sein de leur propre service, quelque soit le type d’activité considéré. Au contraire les services de soins partagent beaucoup plus de 50% de leur activité au sein de leur services, mise à part une exception : le service S2 ne forme que 45% de ses paires de contacts à l’intérieur de son service et interagit plus avec le service S1, dans la même aile que lui, et le service de nuit. Cette importante différence de

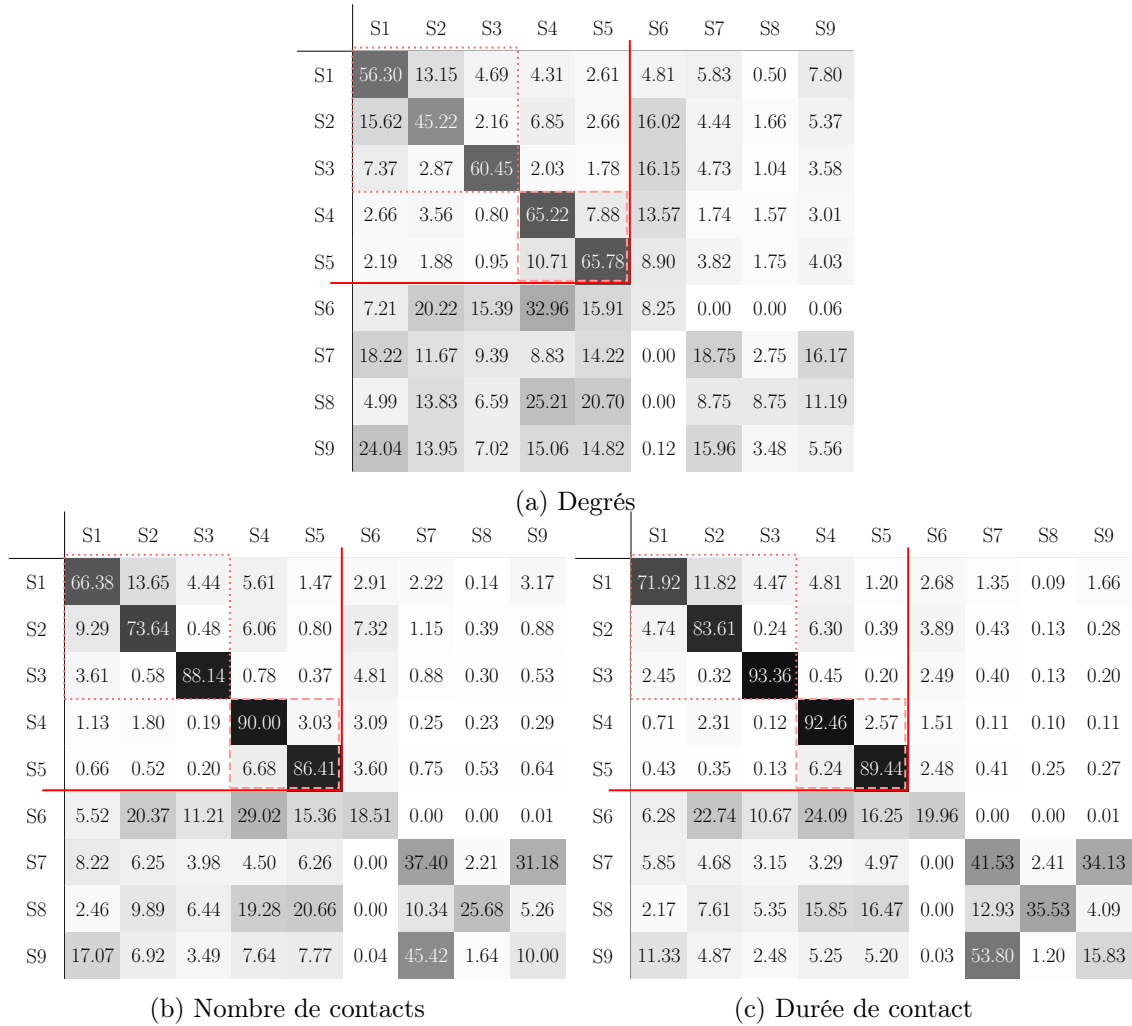


FIGURE 2.16 – Matrices d'activité entre les services, selon les degrés quotidiens, le nombre de contacts et la durée des contacts, affiché en pourcentage d'occupation.

comportement, visible au niveau des paires est nettement moins importante lorsque l'on considère le nombre de contacts ou les durées cumulées des contacts. Nous pouvons aussi constater que les services S7 et S9 partagent une forte proportion de leurs interactions, en particulier lorsque l'on considère le nombre de contacts et la durée de ces contacts. Ces deux services qui sont respectivement un service de kinésithérapie (S7) et un service administratif (S9) semblent avoir des rôles qui les amènent à se rencontrer plus souvent.

## 2.4.2 Déviation de l'activité entre les services.

Nous savons maintenant les quantités d'activité échangées entre les services. Ces quantités sont-elles particulières à notre environnement ou sont-elles relativement proche de la répartition aléatoire et homogène des quantités d'activité sur un réseau de même topologie ? La comparaison des quantités obtenues avec celles que l'on obtient suivant un modèle de configuration y répondra. Nous utiliserons dans cette partie uniquement le modèle de configuration simple que nous avons déjà défini dans la partie 2.3. Nous nous appuyons donc sur un modèle qui répartit uniformément la quantité d'activité associée à un service entre tous les services de l'hôpital proportionnellement à leur taux d'activité global. L'équation



2.12 décrit formellement le calcul de l'activité obtenue entre chacun des services  $S_i$  et  $S_j$  suivant ce modèle de configuration. Nous notons  $Ta(H)$  l'activité totale de l'hôpital et  $(S_i, S_j)$  l'activité existant entre les services  $S_i$  et  $S_j$ .

$$\overline{(S_i, S_j)} = Ta(S_i) * \frac{Ta(S_j)}{Ta(H)} \quad (2.12)$$

Nous nous intéressons particulièrement à la comparaison des valeurs obtenues par le modèle de configuration aux valeurs réelles, ce que nous appellerons déviation au modèle de configuration. Le calcul de déviation est décrit par la formule 2.13 qui est fortement inspirée de la partie 2.3.

$$DevM(S) = \begin{cases} \frac{(S_i, S_j)}{\overline{(S_i, S_j)}} & \text{si } (S_i, S_j) > \overline{(S_i, S_j)} \\ -\frac{(S_i, S_j)}{\overline{(S_i, S_j)}} & \text{sinon} \end{cases} \quad (2.13)$$

De cette façon, les valeurs seront positives lorsque la quantité d'activité réelle est supérieure à la quantité d'activité obtenue suivant un modèle de configuration. Elles seront négatives dans le cas contraire.

La figure 2.17 donne le résultat de la déviation des valeurs réelles et des valeurs annoncées par le modèle de configuration. Les matrices sont symétriques car nous travaillons sur les quantités échangées par les services deux à deux. Les cases des matrices nous renseignent sur l'éloignement au modèle de configuration. La valeur de déviation est inscrite dans chaque case dont la couleur est plus foncée à mesure que la valeur absolue de la valeur est grande et donc que les valeurs réelles sont éloignées des valeurs obtenues en utilisant le modèle de configuration. Les valeurs négatives indiquent que les valeurs réelles sont inférieures à celles que prévoit le modèle de configuration et sont surlignées par une forme hexagonale bleue claire. La case  signifie que l'activité entre les deux services choisis est 46 fois inférieure à celle prévue par le modèle de configuration, ce qui est une valeur moyenne par rapport aux autres valeurs obtenues pour la matrice considérée car la case est gris clair. Les valeurs positives signifient que les valeurs réelles sont supérieures aux valeurs du modèle de configuration. Elles sont surlignées par une forme ovale rouge claire. Ce code de couleur sera conservé dans la suite. Par exemple, la case  signifie que l'activité entre les deux services choisis est 5 fois supérieure à celle prévue par le modèle de configuration, ce qui est une valeur faible par rapport aux autres valeurs obtenues pour la matrice considérée car la case est presque blanche.

Au premier regard, les trois mesures d'activité donnent des résultats assez similaires. L'activité des services de soins avec les services transversaux est assez proche d'une répartition suivant la répartition des contacts en utilisant un modèle de configuration des contacts car les valeurs de déviation obtenues sont assez proches de 1. Au contraire l'activité entre services de soins ou entre service transversaux est très éloignée de ce que le modèle de configuration propose. Les services de soins interagissent beaucoup moins entre eux que ce que le modèle de configuration prévoit alors que les services transversaux interagissent beaucoup plus entre eux que dans le modèle de configuration.

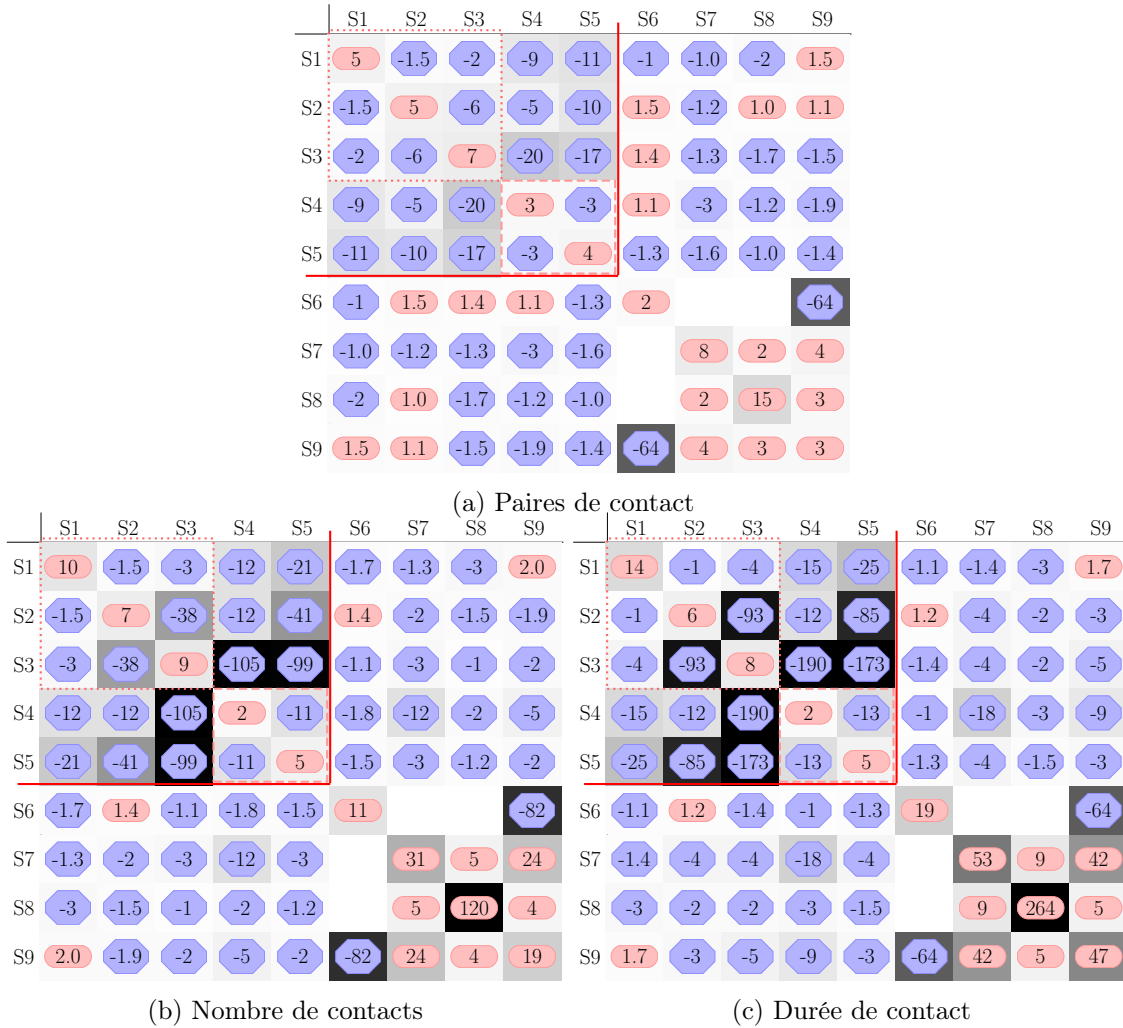


FIGURE 2.17 – Matrices de déviation de l'activité entre les services par rapport à un modèle de répartition de l'activité aléatoire, selon les degrés quotidiens, le nombre de contacts et la durée des contacts en prenant en compte la grande quantité interne des services. Le fond gris des cases est d'autant plus foncé que la valeur de déviation est forte. Les valeurs négatives en bleu ( $-64$ ) dénotent une activité plus faible que prévue par le modèle de configuration. Les valeurs positives en rouge ( $15$ ) dénotent une activité plus importante que ne le prévoit le modèle.

Cette représentation nous apporte peu d'informations sur les relations entre les services. Pour avoir une vue plus précise des interactions entre les services, nous faisons les calculs, sans tenir compte de l'activité interne des services.

Nous répartissons donc entre les services uniquement l'activité externe des ces services. La comparaison entre les valeurs obtenues par le modèle de configuration et les valeurs réelles est résumé par la figure 2.18. La diagonale n'a plus aucune signification dans ce cas. Dans ce cas, la quantité de valeurs réparties est très inférieures au premier modèle de configuration car la majorité des interactions sont des interactions internes aux services. Le rapport des valeurs du modèle de configuration et des valeurs réelles sera donc plus élevé que dans le cas précédent.

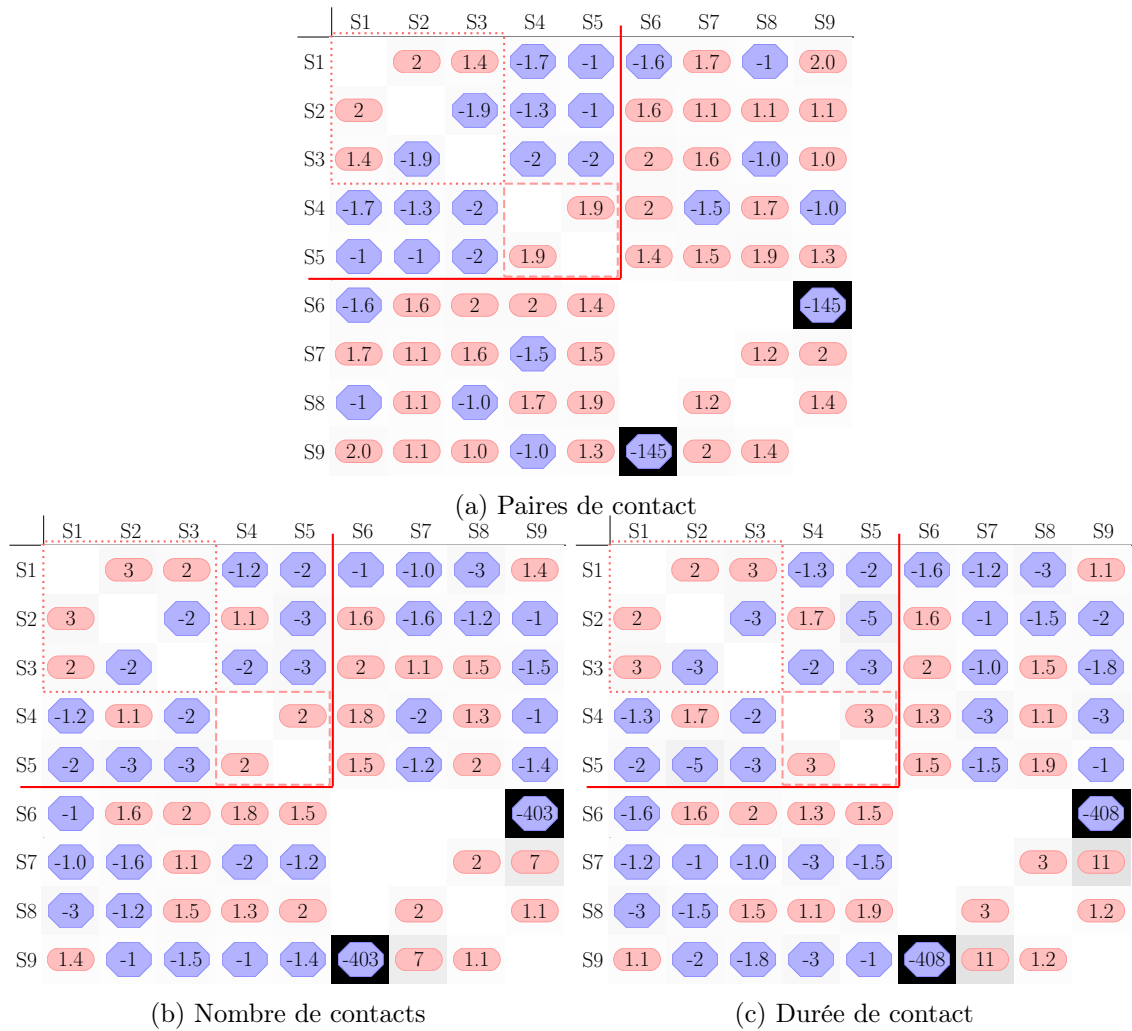


FIGURE 2.18 – Matrices de déviation de l'activité entre les services par rapport à un modèle de répartition de l'activité aléatoire, selon les degrés quotidiens, le nombre de contacts et la durée des contacts en ne travaillant avec les interactions externes uniquement. Les valeurs négatives en bleu ( $-3$ ) dénotent une activité plus faible que prévue par le modèle de configuration. Les valeurs positives en rouge ( $11$ ) dénotent une activité plus importante que ne le prévoit le modèle.

La figure 2.18 montre une répartition des contacts moins éloignée qu'auparavant dans la grosse majorité des cas. Nous pouvons remarquer que certains services ont une activité plus importante entre eux que le modèle de configuration ne le prévoit. Au contraire, d'autres ont une activité toujours plus faible que celle donnée par le modèle de configuration. La figure 2.18a montre que le service S1 interagit 2 à 3 fois plus que ne le prévoit le modèle de configuration avec les autres services situés dans le même bâtiment que lui (S2 et S3). Ce service interagit aussi beaucoup avec le service S7 des kinésithérapeutes ayant une location particulière et S9 (service logistique), se déplaçant dans tout l'hôpital. Les services S2 et S3 ont un taux d'interactions entre eux plus faible que ne le prévoit le modèle de configuration alors qu'il se situent dans un même bâtiment. Au contraire, ces deux services, comme nous l'avons déjà souligné, interagissent plus que prévu avec le service S6 de nuit. Les interactions qu'ils ont avec les autres services transversaux sont très proches de ce que nous annonce le modèle de configuration. On peut remarquer que le service des kinésithérapeutes intervient

principalement dans les services S1, S3 et S5 en terme de diversité des paires, ce qui représente par contre une très faible activité en terme de nombre et de durée de contacts qui sont beaucoup plus faibles en réalité que suivant le modèle de configuration.

L'éloignement au modèle de configuration de l'activité des services S2 et S3 s'accroît lorsque nous considérons les quantité d'activité (figure 2.18b et 2.18c)

### 2.4.3 Interactions patients/personnels entre les différents services.

Dans cette partie, nous nous intéressons toujours aux interactions entre les différents services, mais nous considérons séparément l'activité des personnels et celles de patients.

La figure 2.19 représente le nombre de patients et le nombre de personnels par jour impliqués dans au moins une interaction avec respectivement des personnels ou des patients de chacun des services suivant l'équation 2.10 présenté dans la sous-section précédente.

Conservons nos conventions et présentons les figures donnant des valeurs par personnes de manière à ce que ces matrices se lisent en ligne (cf 2.4.4).

La figure 2.19a donne en ligne le nombre de patients des cinq services de soin interagissant avec les patients de chacun des services de l'hôpital. La figure 2.19b donne le nombre quotidien de patients des services de soins lus en ligne interagissant avec les personnels des services différents services de l'hôpital. La figure 2.19c donne le nombre quotidien de personnels des services lus en ligne interagissant avec des patients des services de l'hôpital lus en colonne. La figure 2.19d indique le nombre quotidien de personnels des services lus en ligne interagissant avec des personnel des services de l'hôpital lus en colonne.

#### Exemple 2.4.4

La case  $M[PAS1][PAS2]$  de la figure 2.19a avec PAS1 lue en ligne et PAS2 lu en colonne signifie que 9 patients du service S1 interagissent quotidiennement avec au moins un patient du service S2. la case de  $M[PAS1][PES2]$  de la figure 2.19b signifie que moins d'un patient en moyenne par jour interagit avec un personnel appartenant au service S2.

Toutes les matrices représentant le nombre de personnes interagissant avec les différents services de l'hôpital de la figure 2.19b sont formées par une diagonale très visible qui montre encore une fois que beaucoup plus de personnes interagissent avec d'autres personnes de leur propre service qu'avec des personnes extérieures à leur service. Les services transversaux ne dérogent pas à la règle, bien que le nombre de personnes concernées soit plus faible. Le service de soins isolé S3 mis à part, les deux bâtiments Sorrel et Ménard sont toujours apparents par leur activité plus importante au sein même des bâtiments qu'à l'extérieur. Les figures 2.19a et 2.19b concernant le nombre de patients interagissant avec les différents groupes de personnes montrent que beaucoup des échanges avec l'extérieur des services dans un même bâtiment sont dus aux échanges entre patients. Quelques patients du bâtiment Sorrel (S1-3) interagissent avec des personnels du service S1 se situant au rez-de-chaussée du bâtiment. De la figure 2.19b on apprend que beaucoup de patients sollicitent le service de nuit, mis à part le service S1 qui semble agir différemment des autres services de soins.

#### Remarque 2.4.1

La moyenne du nombre de personnes par jour interagissant avec des personnes d'un groupe suit la formule 2.10. Suivant cette formule, les groupes n'ayant certains jours



aucun contact entre eux auront une moyenne plus basse que ce que l'on peut s'attendre d'une telle moyenne au sein d'un groupe. En effet, lorsque nous étudions le nombre quotidien moyen de personnes interagissant au sein d'un même groupe  $\overline{G_1(D)}_{\leftrightarrow G_1}$ , nous attendons des valeurs supérieures ou égales à 2 ou la valeur 0 car une personne ne peut interagir seule. Or, prenons la matrice de la figure 2.19d. La valeur moyenne du nombre de personnes interagissant au sein service S8 est inférieure à 2 : en effet, une interaction implique au moins deux personnes, donc il y a au moins deux personnes du même groupe qui interagissent avec des personnes de leur groupes respectifs. Cela signifie qu'il n'y a pas d'interaction entre les membres de S8 tous les jours.

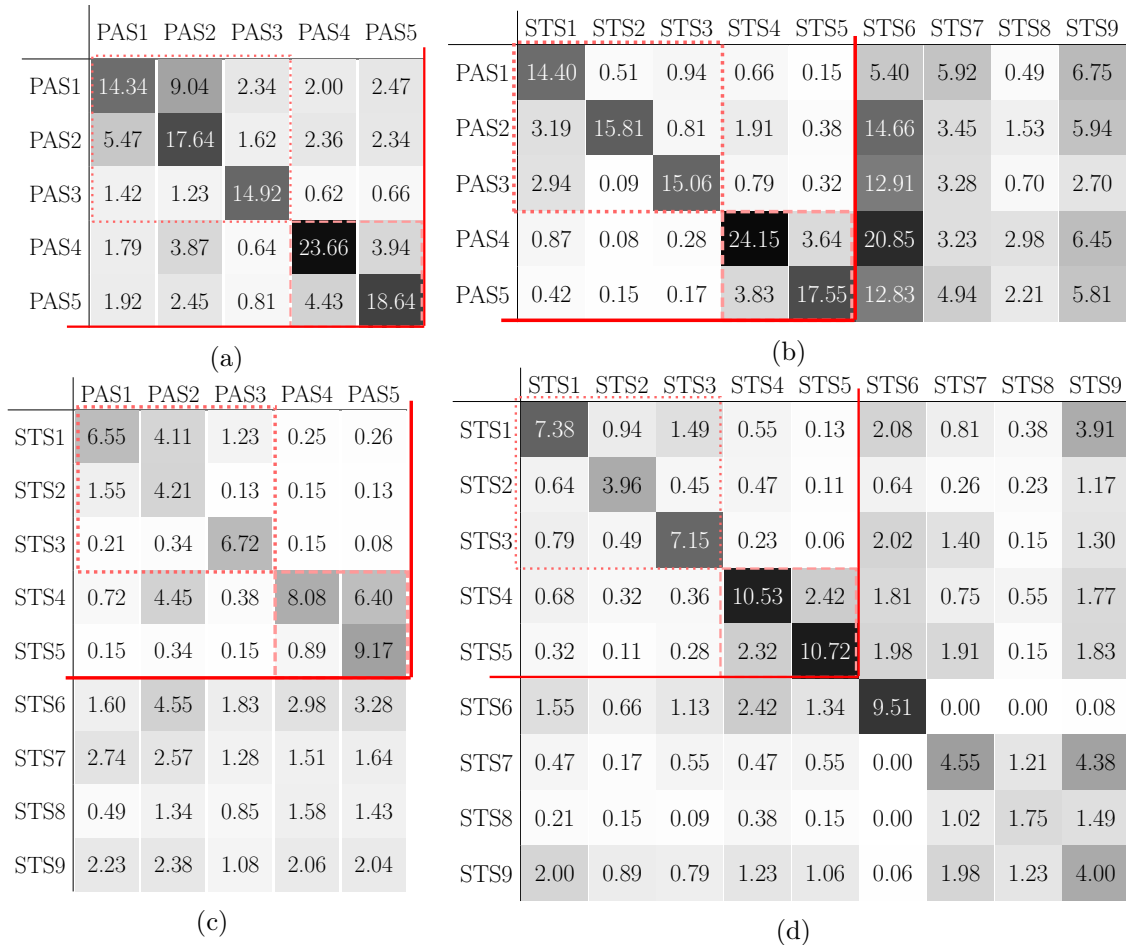


FIGURE 2.19 – Nombre de personnes par jour concernées par les interactions : la figure est à lire en ligne. La case  $M[i][j]$ , avec  $i$  numéro de la ligne et  $j$  numéro de la colonne, alors la case  $M[i][j]$  se lit : il y a  $M[i][j]$  personnes du service  $i$  qui interagissant avec le service  $j$  (cf. l'exemple 2.4.4).

La figure 2.20 représente le pourcentage de personnes d'un groupe  $G_1$  interagissant avec une personne d'un autre groupe  $G_2$ , suivant l'équation 2.11.

Au sein des services, on observe un très fort pourcentage de personnes impliquées dans les interactions, que ce soient des patients entre eux, des personnels entre eux ou des patients avec des personnels et inversement. Nous observons tout de même une différence assez marquée des personnels du service S4, service comportant le plus de personnes. En effet, seulement 76% des personnels de ce service sont impliqués dans des interactions avec des

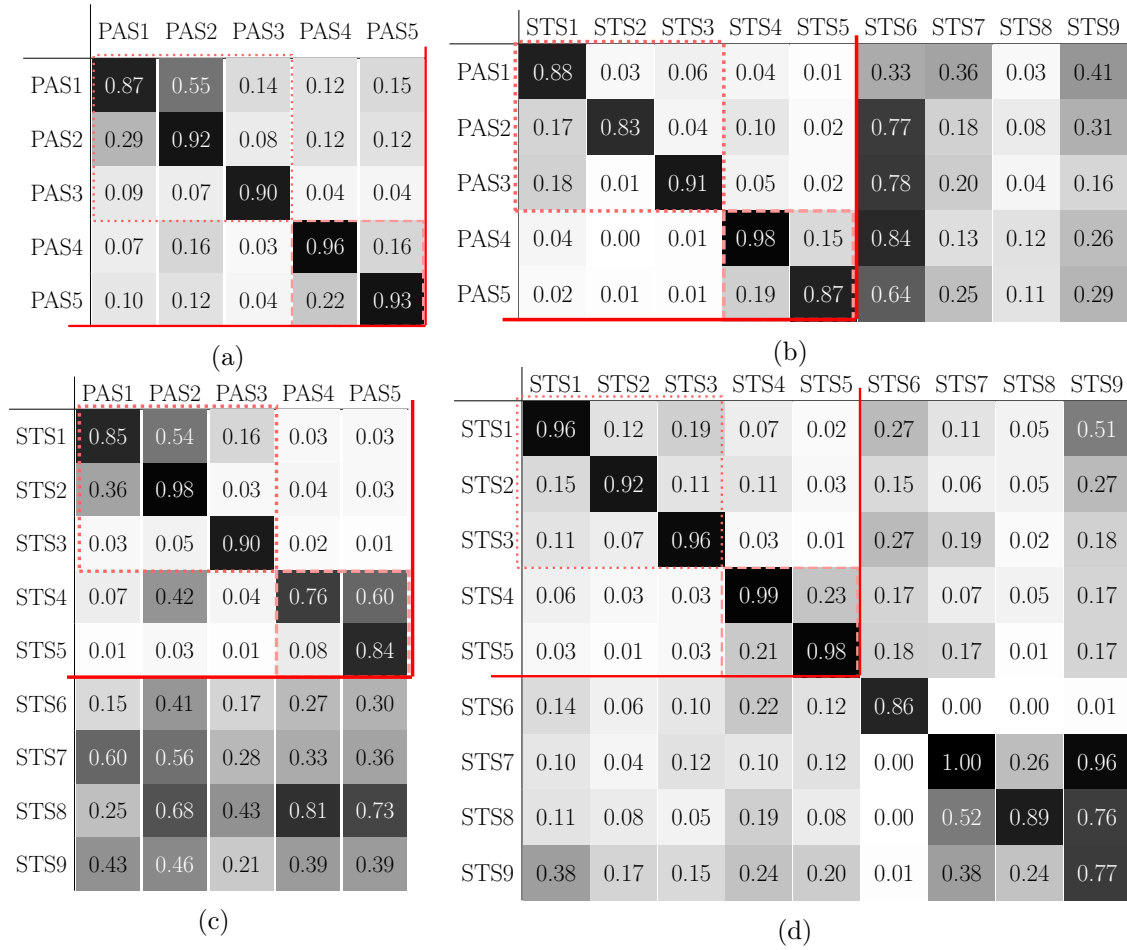


FIGURE 2.20 – Pourcentage de personne de chaque groupe en ligne interagissant avec des personnes des groupes en colonne.

patients de ce service ce qui signifie que certains personnels ne rencontrent de près qu'uniquement des personnels de leur service, comme le montre la figure 2.20b. Nous remarquons ensuite qu'une grosse majorité des patients sollicitent le service de nuit car tous les services, mis à part le service S1, comportent un ensemble de patients dont plus de 60% interagissent avec le service de nuit. Ce service de nuit semble lui répartir relativement équitablement (cf. figure 2.20c) entre les services de soins. À noter que le service S2 interagit avec presque la moitié du service S6. Les services transversaux, quant à eux, interagissent principalement avec les services de soins par le biais des patients, ce qui se comprend en comparant les figures 2.20c et 2.20d. On peut remarquer aussi que la majorité des personnels de S4 (60%, sur la figure 2.20c) interagissent avec une infime partie des patients de S5 (19%, figure 2.20b). Il y a donc une plus grande interface d'échange entre service pour le service S4 que pour le service S5, dans cette situation. On remarque que cet échange entre service n'est pas réciproque dans le sens où très peu de personnels du service S5 interagissent avec des patients de S4 (seulement 8%).

Les trois figures 2.21, 2.22 et 2.23 qui suivent résument les quantités quotidiennes d'activité par personne selon les trois critères habituels que sont le degré moyen des personnes, le nombre de contact et la durée cumulée des contacts. Nous distinguons comme précédemment, les contacts entre patients, ceux entre personnels et ceux qui ont lieu entre patients

et personnels. Les résultats étant donnés en quantité par personne, cela nous amène à considérer 4 matrices pour chacune des quantités.

La figure 2.21a donne le nombre quotidien de patients avec qui les patients interagissent dans chaque service et la figure 2.21b résume le nombre de personnels avec lesquels les patients de chacun des services de soins interagissent chaque jour.

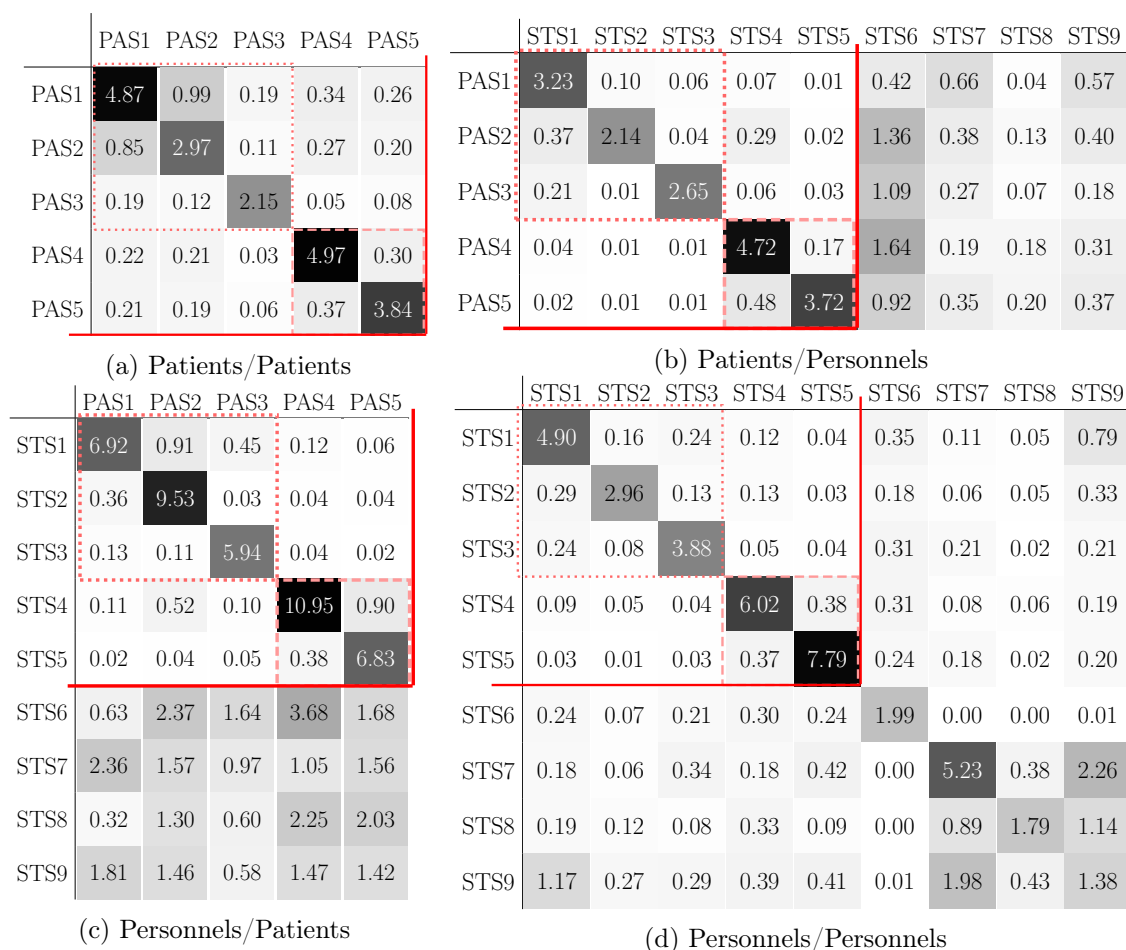


FIGURE 2.21 – Matrices d’activité entre les services selon les degrés quotidiens par personnes.

On remarque tout d’abord que le nombre de voisins des patients à l’intérieur de leur services est assez équilibré entre patients et personnels. Les patients interagissent en moyenne avec légèrement plus de patients qu’avec des personnels mais la différence est très faible. Ainsi, les patients ont des contacts avec en moyenne 2 à 5 patients et 2 à 5 personnels de leur service par jour. Le nombre de voisins quotidien à l’extérieur des services est pour tous inférieur à un. Les patients interagissent donc avec moins d’une personne par jour dans chacun des autres services. La seule exception est le service de nuit. En effet, les services que nous avons déjà remarqué comme sollicitant beaucoup le service S6 de nuit, comme les services S2, S3 et S4 sont composés de patients qui interagissent avec entre 1 et 2 personnels par jour du service de nuit. Les patients de chacun des services de soins n’interagissent que peu souvent avec les personnels des services transversaux car ils ont moins d’un contact par jour en moyenne avec les personnels de chacun des ces services.

De même, la figure 2.21d donne le nombre quotidien de personnels avec qui les personnels interagissent dans chaque service et la figure 2.21c, le nombre de patients avec lesquels les personnels de chacun des services interagissent chaque jour.

Les personnels des services de soins, mis à part le service S5 qui agit à l'inverse des autres services, rencontrent quotidiennement dans leur service environ 2 à 4 patients de plus que de personnels, également dans leur services. Les personnels interagissent donc avec plus de patients qu'avec leurs collègues dans une journée et cette remarque s'applique aussi aux personnels et patients appartenant aux autres services, bien que le nombre de personnes impliquées soit beaucoup plus négligeable : entre 0 et 2 personnes pour la plus part des services. Le service S6 de nuit est composé de personnes qui interagissent avec 1 à deux personnes par service. Une exception peut-être relevée concernant les interactions avec le service S4 car les personnes du service S6 interagissent avec entre 3 et 4 personnes de ce service de soins, ce qui est nettement au-dessus de la moyenne des autres services et s'explique par le grand nombre de patients appartenant au service S4. Les personnels des services transversaux interagissent peu entre eux du fait de leur faible nombre et ont aussi peu de voisins à l'intérieur de leur propre service.

La figure 2.22 donne le nombre de contact quotidien moyen par personne avec chacun des groupes définis comme étant les personnels d'un côté et les patients de l'autre de chacun des services.

Le pattern général observé ressemble beaucoup à la figure précédente 2.21, mais les différences sont beaucoup plus marquées. En effet, le nombre de contacts entre patients d'un même service est beaucoup plus important que le nombre moyen de contacts quotidiens par patient avec des personnels de leur service. Il existe donc plus de chemin de transmission au sein du groupe des patients dans les services de soins. Le nombre de contact ayant lieu entre les patients des services S1 et S2 situés dans le même bâtiment est peu différent du nombre de contacts ayant lieu entre les personnels et les patients du service S1 ce qui signifie que ces contacts sont à prendre avec égale considération. Les personnels du service de nuit, quant à eux, ont un nombre assez important de contacts avec 4 des services de soins. Les patients de ces services de soins ont également un nombre non négligeable de contacts avec le service de nuit ce qui leur permet, en terme de diffusion, de servir de "passerelle" entre les personnels du service de nuit et les personnels de leur service.

La figure 2.23 donne la durée cumulée des contacts quotidiens moyens par personne exprimée en heures avec chacun des services subdivisés en deux groupes patients et personnels lorsque cela est possible.

Encore une fois, nous remarquons de beaucoup plus fortes valeurs situées en sein des services, puis dans une moindre mesure entre patients des services situés dans un même bâtiment. La durée cumulée des contacts entre personnels des services transversaux est peu importante mis à part ceux des services de kinésithérapie S7, S8 et administratif S9 qui, comme nous l'avions déjà remarqué, interagissent beaucoup ensemble. On remarque aussi qu'en moyenne, les personnels des services transversaux ont moins d'une demi-heure de contact avec les patients des services de soins. Au sein même des services de soins, le nombre d'heure des personnels passé avec les patients dépasse l'heure pour la plupart des services. Les personnels des services S2 et S4 atteignent en particulier presque 2 heures de durée cumulée de contacts. On remarque aussi que les personnels des services S4 et S1 interagissant plus de 20 minutes par jour avec des patients appartenant au service situé un étage au-dessus du leur (resp. S2 et S5). La durée des contacts entre patients et personnels

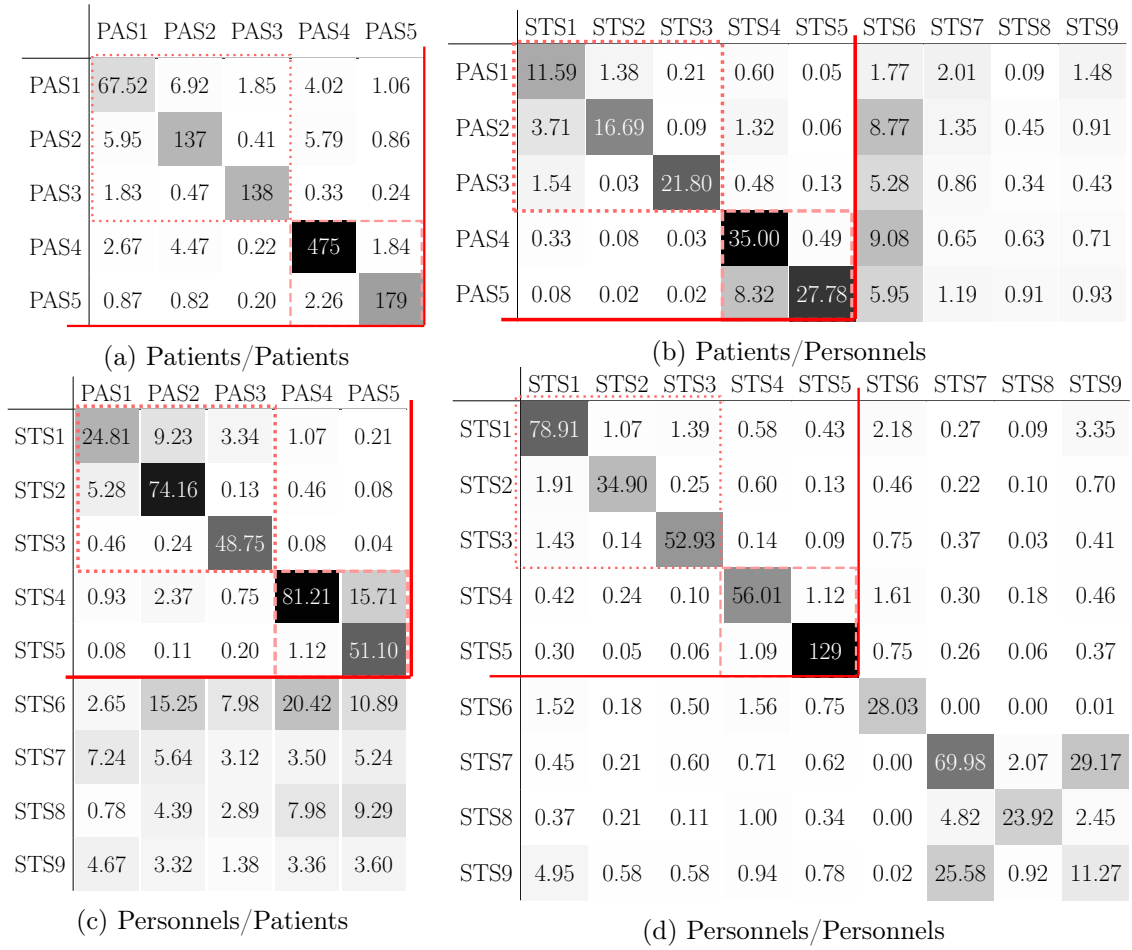


FIGURE 2.22 – Matrices d’activité entre les services selon le nombre de contacts quotidiens par personnes.

s’élève à moins d’une heure par personnes par jour. La durée des contacts par personnes impliquant deux patients de services différents est inférieure à 15 minutes par jour alors qu’en interne elle varie d’environ 3 heures à environ 10 heure, sauf pour le service S4 qui est un service particulier. En effet, le service S4 comprend des personnes partageant leur chambre en état végétatif chronique ce qui explique une importante durée des contacts. Les patients de tous les services, à part le service S1, passent environ un quart d’heure par jour avec le service de nuit.

### En résumé

Nous avons, dans cette partie, considéré les interactions entre les services de l’hôpital. Pour cela nous avons agrégé les données par groupe de service et utilisé la méthode de représentation par des matrices de contact souvent utilisé dans le domaine [76, 44]. Nous avons constaté que le réseau des interactions est très structurés en "communautés de services" ainsi que communauté de bâtiments. En effet, bien que la majorité des interactions ait lieu au sein des services, un nombre non négligeable d’interactions entre les services existe, ce qui est ignoré lors d’études épidémiologiques s’appuyant sur des données considérant un unique service[44]. Nous avons comparé le réseau réel à deux modèles de configuration afin de confirmer que la structure observée est bien une structure spécifique, à prendre en compte lors de la conception de modèle de diffusion pour étudier l’environnement hospitalier. Nous avons aussi séparé l’activité des patients et des personnels ce qui nous a permis

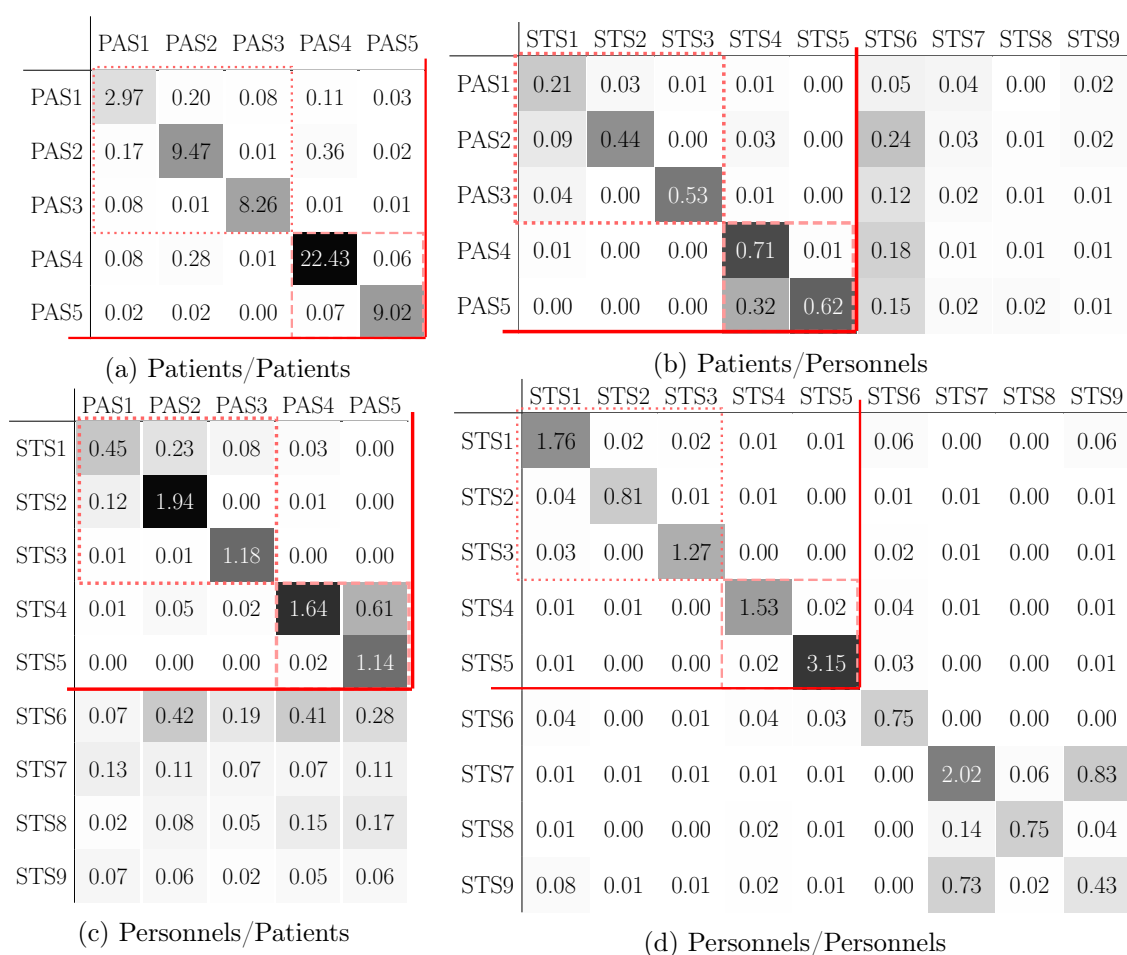


FIGURE 2.23 – Matrices d’activité entre les services selon les durées de contact quotidiennes par personnes.

de constater que les échanges entre les différents services de soin d’un même bâtiment sont principalement initiés par les patients, qui sont donc une passerelle pour une diffusion potentielle. Le service de nuit, constitué uniquement de personnels, est une autre passerelle potentielle car les personnels de ce services interagissent entre eux comme avec les patients des différents services de soins.

## 2.5 Analyse de la dimension fonctionnelle

Dans cette partie, nous nous intéressons au fonctionnement des différentes catégories socioprofessionnelles et à leurs interactions. Nous allons appliquer à ce nouvel ensemble de groupes de personnes les outils d’analyse décrits et appliqués jusqu’à présent dans ce chapitre 2 aux services de l’hôpital. Ceci permet d’affiner notre perception des interactions des personnels avec leurs collègues ainsi qu’avec les patients. Rappelons d’abord les 14 catégories socioprofessionnelles différentes présentes dans l’hôpital de Berk-sur-Mer en 2009 (cf. table 2.3).

Le numéro des catégories ayant été assignés arbitrairement, nous les avons réordonnées par rapport à leur fonction et leur type d’intervention afin de rapprocher les catégories supposées travailler ensemble telles que les cadres de santé (C2) avec les infirmiers (C3)

	Titres/Fonction	Description de la fonction
C1	Aide Soignant	Aide au soin, toilette des patients
C2	Cadre de santé	Responsable des soins
C3	Infirmier Diplômé d'Etat (IDE)	Donne les soins
C4	Médecins	Prescrit les soins, diagnostiques les patients
C5	Élève stagiaire	Rééducation musculaire
C6	Kinésithérapeute	Rééducation fonctionnelle : travail d'apprentissage de l'utilisation des membres et modification de l'environnement du patient pour plus de facilité dans la vie quotidienne
C7	Ergothérapeute	Rééducation fonctionnelle : travail d'apprentissage de l'utilisation des membres et modification de l'environnement du patient pour plus de facilité dans la vie quotidienne
C9	Brancardier	Transport des patients
C10	Animateur / Coiffeur / ...	Participe à la vie et l'animation des patients longue durée
C11	Agent de Service Hospitalier (ASH)	Alimentation des patients, nettoyages du matériel hospitalier
C12	Logistique	Accueil et départ des patients
C13	Administration	Accueil et départ des patients
C14	Autre rééducateur	
C0	Patients	Pensionnaires à soigner

TABLE 2.3 – Description des catégories socioprofessionnelles présentes dans l'hôpital de Berk-sur-Mer en 2009.

ou les aides soignants (C1). La figure 2.24 donne le nombre quotidien moyen de personnes présentes dans chaque catégories socioprofessionnelles des personnels de l'hôpital.

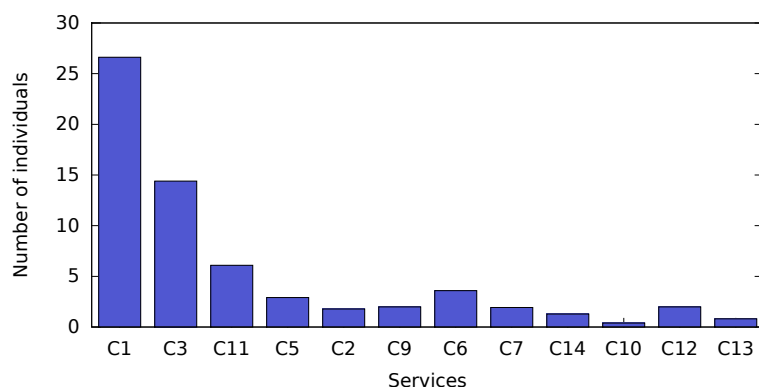


FIGURE 2.24 – Nombre moyen quotidien de personnes présentes par catégories.

Certaines catégories sont beaucoup plus nombreuses que d'autres, en particulier les catégories amenées à s'occuper directement des patients tels que les aides soignants (C1) ou les infirmiers (C3). La majorité des catégories socio-professionnelles n'ont pas plus de cinq membres par jour, mis à part les ASH qui sont moins d'une dizaine, les aides soignants (C1) et les infirmiers (C3). Ces deux dernières catégories sont amenées à se déplacer auprès des patients pour leur prodiguer les soins quotidiens et sont respectivement composées de 27 et 14 personnes par jour. A noter que durant l'analyse, nous ne travaillerons pas avec les médecins, peu nombreux au cours de la période étudiée et peu présents.

Chaque personne n'appartient qu'à un service et une seule catégorie socioprofessionnelle au cours de l'expérience. On peut donc facilement étudier la répartition des catégories au sein

des services. Le tableau 2.4 nous informe sur le nombre exact de personnes appartenant à chaque paire (service, catégorie).

Cat. \ Serv.	S1	S2	S3	S4	S5	S6	S7	S8	S9	Total
C1	17	8	10	20	19	20			1	95
C3	8	6	7	11	7	12				51
C11	3	2	3	5	4					17
C5	4		2		7		2		1	16
C2	1	1	1	1	1	1	1			7
C9							1		5	6
C6							12			12
C7								6		6
C14			2	1	3				2	8
C10									1	1
C12								1	12	13
C13						1			6	7
C0	53	50	50	53	51					257

TABLE 2.4 – Répartition des personnes sur toute la période étudiée entre les services et catégories. Le nombre inscrit dans une case correspond au nombre de personnes ayant appartenu au service lu en colonne et à la catégorie socioprofessionnelle lue en ligne durant l'été 2009.

Ainsi, d'après le tableau 2.4, certaines des catégories socioprofessionnelles voient leurs membres répartis dans les différents services de soins. D'autres catégories sont toutes assignées à un seul et même service, parfois un service de soins, parfois un service transversal. Les catégories n'auront donc pas toutes le même rôle dans la diffusion à travers les services de l'hôpital.

On peut ainsi distinguer quatre types de catégories : les catégories intervenant uniquement dans les services de soins, celles qui se partageant entre services de soins et services transversaux, celles dont les membres sont repartis uniquement entre les services transversaux et celles associées à un seul et unique service transversal. Le tableau 2.5 répartit les différentes catégories socioprofessionnelles selon les 4 types de catégories annoncés.

Services	Soins	Soins et Transversaux	Transversaux	Transversal unique
	C11 ASH	C1 Aide Soignant	C12 Logistique	C6 Kiné
	C4 Médecin	C2 Cadre de Santé	C13 Administration	C7 Ergo
		C3 IDE	C9 Brancardier	C10 animateurs
		C5 Stagiaire		
		C14 Rééducateur		

TABLE 2.5 – Répartition des catégories socioprofessionnelles en fonction de leur activité au sein des services.



### 2.5.1 Analyse gros grain des catégories socioprofessionnelles

#### Valeurs moyennes

La figure 2.25 représente la quantité d'activité quotidienne des différentes catégories socio-professionnelles selon les différentes mesures d'activité que nous avons identifiées au début de ce chapitre. Pour rappel, nous étudions la diversité de l'activité en s'appuyant sur la répartition des degrés des personnes et la quantité d'activité en s'appuyant sur le nombre d'interactions et la durée des interactions. Comme nous l'avons fait pour les services, nous travaillons sur l'activité cumulée de toutes les personnes constituant chacune de ces catégories. En observant la figure 2.25 on remarque immédiatement la grande similarité des

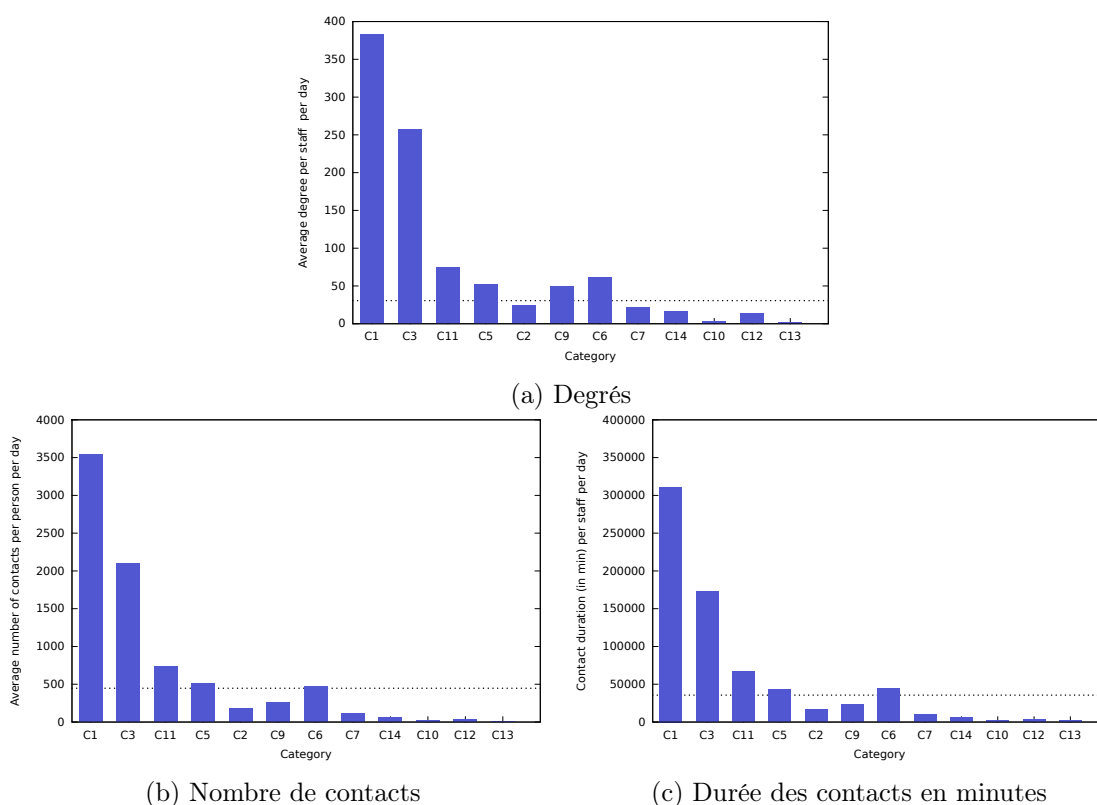


FIGURE 2.25 – Valeurs moyennes des catégories socioprofessionnelles par jour des différentes quantités.

patterns obtenus pour chacun des trois mesures d'activité ce qui suggère que la diversité des contacts (degrés) et la quantité de ces contacts (nombre de contacts et durée) sont proportionnels. Cette remarque reste à confirmer par l'observation de l'activité individuelle qui sera décrite par la figure 2.26. Comme nous pouvions nous y attendre en connaissant le nombre de personnes appartenant à chaque catégories, les deux groupes les plus actifs sont les catégories avec une très forte population C1 (aides soignants) et C3 (infirmiers).

La figure 2.26 représente les valeurs quotidiennes moyennes par personnes de chaque catégorie socioprofessionnelle selon les trois mesures d'activité : leur degré, le nombre de contacts et la durée cumulée de ces contacts. Comme précédemment, ces valeurs sont obtenues en sommant l'ensemble de l'activité des groupes sur toute le période et en divisant par la

somme du nombre de personnes présentes chaque jours dans ces groupes (cf. l'équation 2.2)

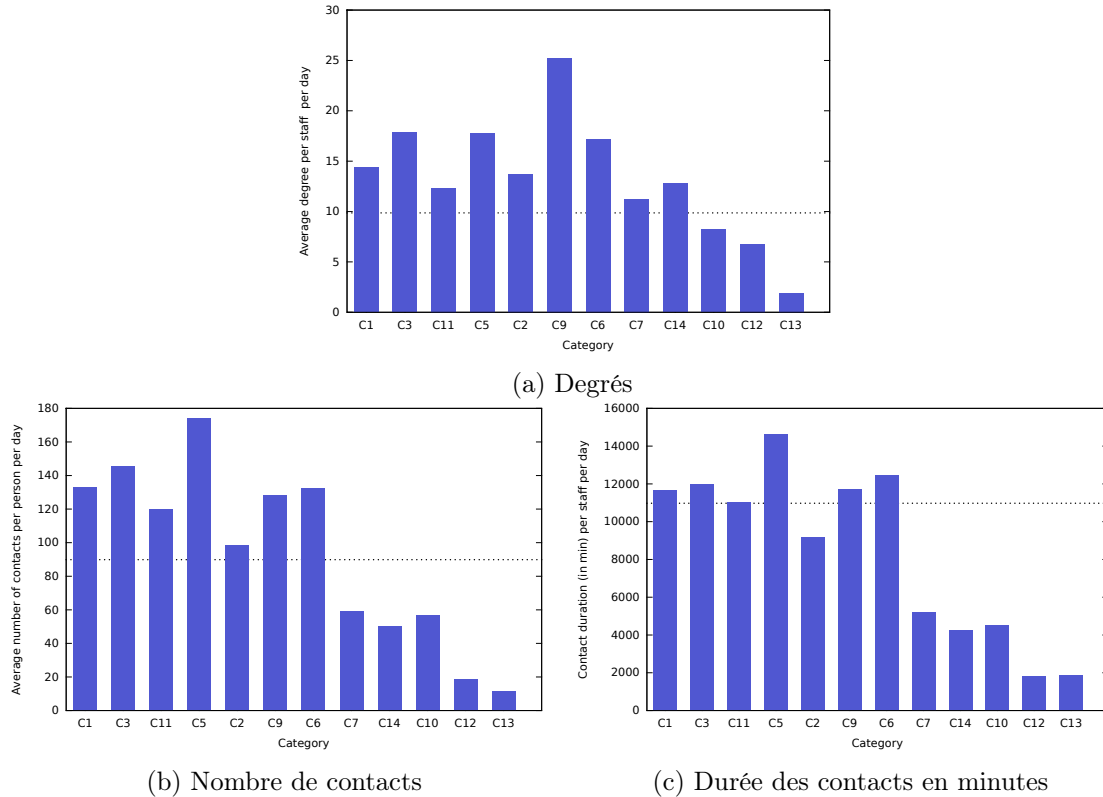


FIGURE 2.26 – Valeurs moyennes par personnes organisées en catégories socioprofessionnelles par jour des différentes quantités.

Contrairement aux mesures quotidiennes des catégories socioprofessionnelles, les mesures de quantités d'activité individuelles par jour n'ont pas le même pattern selon les différentes mesures d'activité observées. En particulier, les valeurs maximales ne sont pas atteintes par les mêmes catégories en fonction des mesures d'activité. En terme de diversité d'activité, identifiée par le degré des personnes, le service C9 représentant les brancardiers a beaucoup plus d'activité que les autres catégories socioprofessionnelles. Ce n'est plus le cas en terme de quantité d'activité, donné par le nombre de contacts et la durée de ces contacts. En terme de quantité, la catégorie dominante est la catégorie C5 des élèves stagiaires. De même, nous n'obtenons pas les mêmes groupes de catégories lorsqu'on regroupe les catégories ayant les mêmes valeurs individuelles quotidiennes.

La figure 2.27 établit les corrélations entre les différentes mesures d'activité individuelles moyennes quotidiennes. Ainsi, la figure 2.27b compare les degrés moyens individuels quotidiens des catégories à la durée individuelle moyenne quotidienne des contacts. La figure 2.27a compare les degrés quotidiens individuels des catégories au nombre individuel moyen de contacts quotidiens. La figure 2.27c compare le nombre individuel moyen de contacts quotidiens à la durée individuelle moyenne quotidienne des contacts.

Les deux premières sous-figures, qui comparent le degré moyen individuel quotidien des catégories respectivement aux durées et aux nombres moyens individuels quotidiens, donnent des motifs assez similaires. Cela s'explique par la très forte corrélation des deux

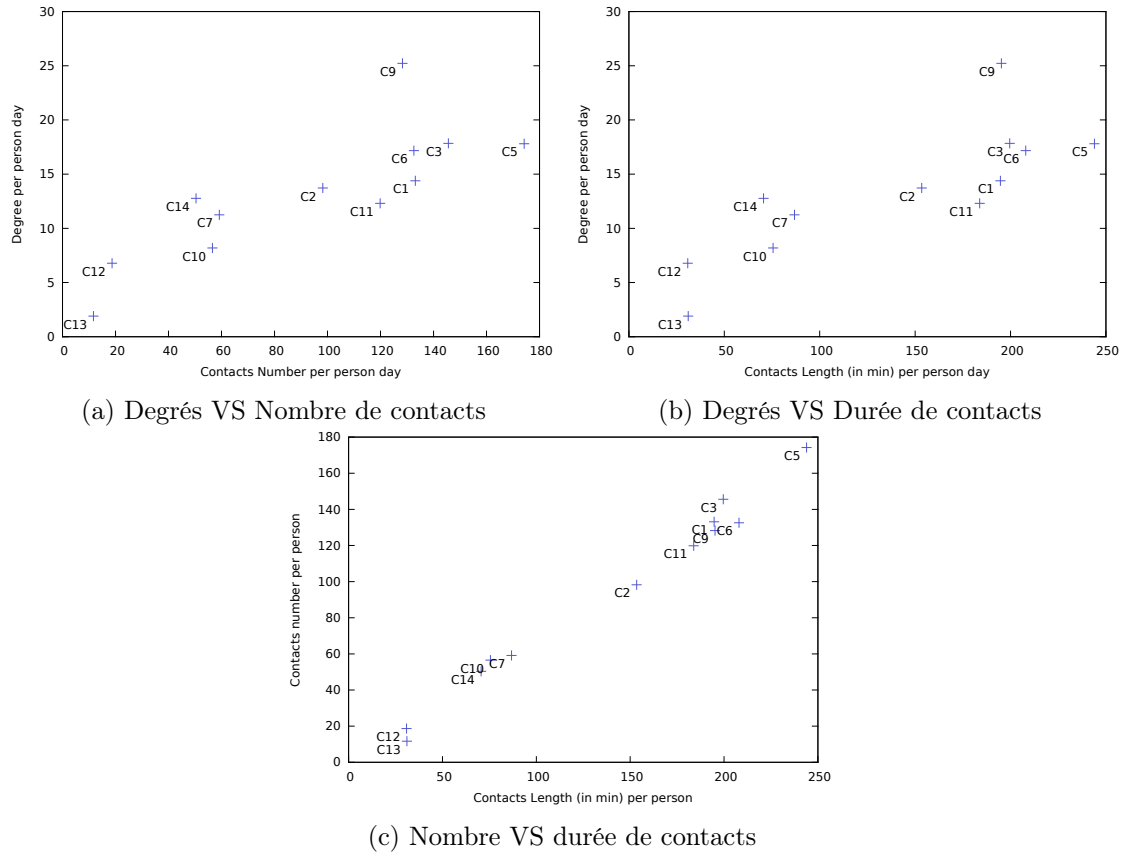


FIGURE 2.27 – Corrélation des différents types d’activité par catégories socio-professionnelles, n’impliquant que les personnels de l’hôpital.

dernières mesures d’activité comme le montre la figure 2.27c. Ces deux quantités sont en effet presque parfaitement corrélées et permettent de distinguer 3 à 4 groupes de catégories socioprofessionnelles ayant des caractéristiques d’activité communes. Les deux catégories dédiées à l’administratif et à la logistique ont en effet des durées cumulées quotidiennes de contacts très courtes et un nombre de contacts quotidiens assez réduit comparé à celui des autres catégories. On remarque que la catégorie administrative a plus de contacts et donc des contacts plus courts que celle des logisticiens. Un autre groupe de catégories est constitué par les rééducateurs intervenant dans les différents services, les ergothérapeutes et les animateurs qui ont une activité un peu plus importante que les deux catégories précédentes. Ensuite vient la catégorie de cadres de santé, qui se distingue légèrement des autres catégories intervenant principalement au près des patients sur la figure 2.27c par une activité plus faible. La différence est moins marquée sur les deux premières figures qui comparent le degré aux durées et nombres de contacts. Nous pouvons aussi considérer comme un groupe de catégories ayant les mêmes caractéristiques d’activité les catégories des aides soignants (C1), des ASH (C11), des infirmiers (C3) et contre toute attente des kinésithérapeutes, qui n’ont apparemment pas le même type de comportement que les autres rééducateurs. Nous dirons que ce groupe est proche des patients dans la suite. Les brancardiers changent de groupe en fonction des mesures d’activité considérées. En effet, les mesures d’activité très corrélées comme le nombre de contacts et leur durée cumulée placent les brancardiers dans les même groupes que le groupe proche des patients. On constate que les brancardiers sont la catégorie ayant le plus fort degré. En revanche, les deux autres figures (fig. 2.27a et 2.27b) révèle un comportement différent des catégories proches de patients. En

effet, les brancardiers ont des degrés plus élevés que les autres catégories pour un nombre et une durée de contacts équivalents. Les stagiaires, quant à eux, ont beaucoup plus de contacts que les autres catégories, mais ont un degré similaire aux catégories proches de patients.

### 2.5.2 Introversion des catégories socioprofessionnelles.

Etudions maintenant l'introversion des catégories socioprofessionnelles afin de mieux comprendre l'organisation de leur activité.

La figure 2.28 donne les valeurs d'introversion des catégories socioprofessionnelles, suivant les définitions qui ont été décrites dans la partie 2.3. Nous utilisons en particulier l'équation 2.3 pour obtenir ces valeurs, appliquée à l'ensemble des catégories.

L'introversion a été calculée sur deux ensembles différents : le premier, représenté en clair sur les figures, inclut l'ensemble de l'activité de l'hôpital. Le deuxième ensemble de catégories, apparaissant en couleur foncée sur les figures qui suivent, est constitué de toutes les catégories socioprofessionnelles, sans les patients (C0).

On remarque que la majorité des catégories socioprofessionnelles sont fortement extraverties et donc interagissent en grande majorité avec des personnes n'appartenant pas à leur propre catégorie, et ce, quelque soit le type d'activité observé. Les catégories de rééducation comme les kinésithérapeutes (C6) et les ergothérapeutes (C7) ainsi que la catégorie de logisticiens (C12) sont peu extraverties par rapport aux autres catégories, que l'on prenne en compte ou non l'activité des patients. La catégorie des animateurs (C10) ne comportant qu'une seule personne pour la période étudiée n'a pas de contacts intérieures et donc a une introversion nulle, comme attendu par le modèle de configuration. On remarque que la valeur d'introversion des catégories est très touchée par la prise en compte de l'activité des patients car une grande partie de l'activité des personnels est partagée avec les patients. Cependant, lors de l'absence des patients dans le calcul de l'introversion, nous obtenons toujours des valeurs de forte extraversion des catégories, ce qui met en évidence que les différentes catégories professionnelles de l'hôpital communiquent entre elles. En comparant les résultats d'introversion obtenus en fonction des différentes quantités, on remarque que la diversité des contacts met en avant une extraversion beaucoup plus importante que les résultats basés sur les quantités d'activité (nombre et durée des contacts). Même si les catégories ont une plus grande diversité de contacts avec l'extérieur de leur catégorie qu'à l'intérieur, le déséquilibre tant à se rétablir en terme de quantité de contact (en particulier les rééducateurs – C6, C7, – et logisticiens – C12 –) qui deviennent même légèrement introvertis.

La figure 2.29 nous informe sur la déviation d'introversion des catégories comparées à celle obtenue suivant le modèle de configuration simple décrit dans la partie 2.3, qui s'appuie sur une répartition homogène de l'activité entre les personnes des différentes catégories. Nous utilisons ici l'équation 2.5 décrite précédemment au sujet des services de l'hôpital. Encore un fois, nous travaillons avec deux ensembles différents de catégories socioprofessionnelles : l'un complet, l'autre ne prenant pas en compte l'activité des personnels avec les patients.

Les valeurs de déviation d'introversion sont presque toutes supérieures à 1 ce qui signifie que la plupart des catégories ont moins d'activité avec l'extérieur de leur catégorie qu'elles ne le devraient en suivant un modèle de configuration répartissant équitablement l'activité des

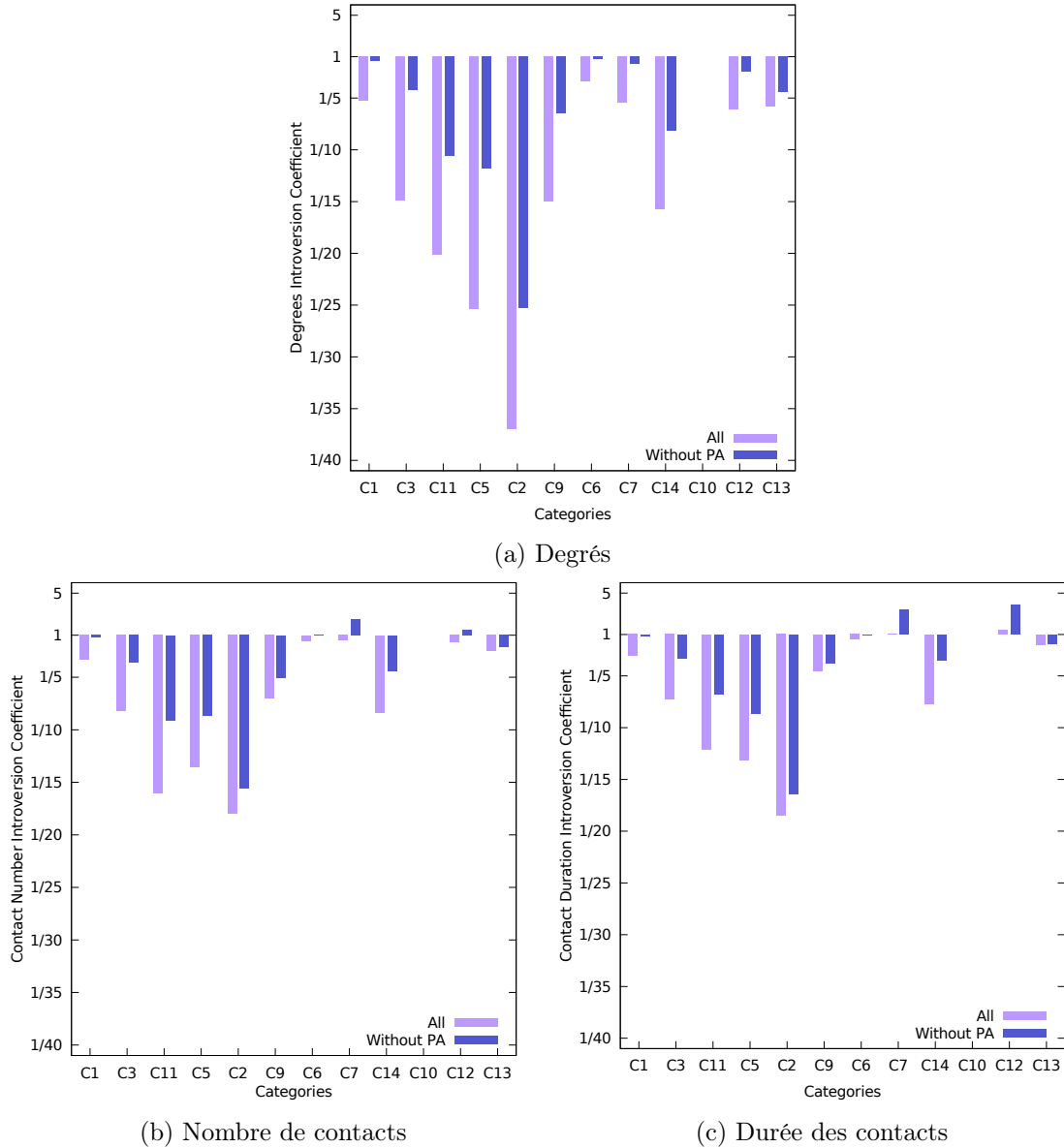


FIGURE 2.28 – Introversion des différentes catégories selon les 3 quantités habituelles.

catégories entre les personnes : les catégories socio-professionnelles, bien qu'elles interagissent beaucoup avec des personnes appartenant à d'autres catégories ont un comportement introverti.

On remarque plus précisément que les catégories socioprofessionnelles dispersées dans les différents services de soins (C1, C3, C11, C5, C2, C14) et pour certaines aussi dans les services transversaux (C9, C6, C7), sont très peu déviées du modèle de configuration répartissant leur activité de manière aléatoire et équilibrée, en particulier lorsque l'on considère uniquement l'activité des personnels. En observant la figure 2.29a basée sur le degré des personnes, les catégories de personnels des services de soins (sauf C14 qui est une catégorie de rééducation) se distinguent des autres catégories par leur valeurs de déviation d'introversion très faibles. En effet, ces valeurs sont très proches de 1 et montre que la répartition de l'activité de ces catégories, en terme de diversité, est très proche de celle d'une répartition aléatoire équilibrée, en particulier lorsque l'on ne considère que l'activité des personnels.

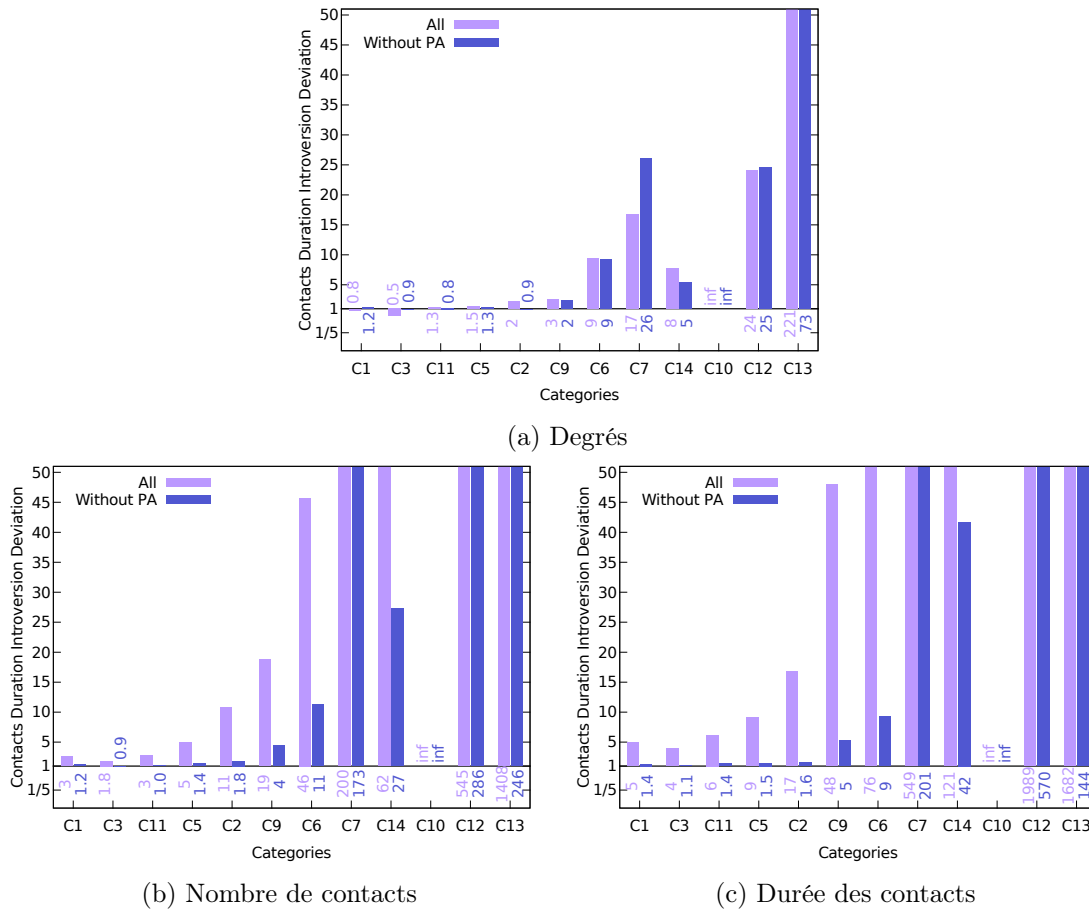


FIGURE 2.29 – Déviation d’introversion des différentes catégories selon les 3 quantités habituelles.

Cela n’est pas le cas des autres catégories de personnels hospitaliers qui ont une activité interne beaucoup plus importante que dans un modèle de configuration de répartition aléatoire. Parmi les catégories de rééducateurs (C6,C7,C14) ou transporteurs (C9), pourtant, certaines sont réparties dans les services de soins et pourrait avoir les mêmes caractéristiques que les personnels des services de soins tels que les infirmiers ou ASH (C1, C3, C11, C5, C2). Etre réparti à travers les différents services de soins n’est pas un critère suffisant pour permettre de savoir si la diversité de l’activité des catégories suivra ou non le modèle de configuration. Le rôle de chacune des catégories, indépendamment de leur rattachement à un service spécifique semble dicter la répartition de l’activité de ces catégories. Il faut donc faire ce travail d’étude pour toutes les catégories socio-professionnelles afin faire des modèles réalistes car chacune a ses caractéristiques propres. En effet, lorsque l’on s’intéresse aux valeurs de déviation d’introversion basées sur le nombre ou la durée des contacts, on constate des valeurs assez différentes pour chacune des catégories. Bien que beaucoup de valeurs restent faibles pour les catégories réparties dans les services de soins, elles ne sont pas non plus négligeables.

Ainsi, les catégories socio-professionnelles sont très extraverties et interagissent donc avec beaucoup de corps de métiers différents. Pour autant, les affinités à l’intérieur du corps de métier sont très fortes car les catégories socio-professionnelles ont des déviations d’introversion pour la plupart supérieure à 1 qui montrent qu’elles ont une activité interne supérieure à ce que l’on obtiendrait avec un modèle de configuration. Les personnels de

soins généraux (C1, C3, C11, C5, C2) sont tout même comparables a la répartition du modèle de configuration ce qui n'est pas le cas des autres catégories sont très éloignées du modèle.

### 2.5.3 Interaction entre les différentes catégories

De même que nous avons observé les interactions entre les différentes paires de services, nous nous intéressons aux interactions et échanges qui se font entre les différentes catégories. Comme précédemment, nous nous intéressons dans un premier temps à la répartition du nombre de personnes concernées par les échanges entre chacune des catégories socioprofessionnelles pour établir l'étendue de la surface d'échange entre les différentes catégories. Nous nous intéressons ensuite aux différentes quantités d'interactions existant sur ces surfaces d'échange déjà observées. Nous comparons pour finir la répartition de l'activité entre les catégories à un modèle de configuration pour évaluer encore une fois la spécificité de structure du réseau étudié.

#### Personnes impliquées dans les interactions entre les catégories socio-professionnelles.

Commençons par étudier la répartition du nombre de personnes concernées par les échanges entre les différentes catégories socioprofessionnelles.

La figure 2.30 nous indique, pour chaque catégorie, la quantité quotidienne de personnes ayant une activité commune avec au moins une personnes d'une autre catégorie de l'hôpital. Cette figure contient deux sous-figures qui nous informent respectivement du nombre de personnes, puis du pourcentage de personnes concernées des catégories. Comme dans la partie concernant les services, ces deux matrices se lisent en ligne de telle sorte que la case  $M[C3][C5]=3,34$  de la figure 2.30a donne le nombre quotidien moyen de personnes appartenant à la catégorie C3 interagissant avec au moins une personne appartenant à la catégorie C5.

D'après la figure 2.30a le nombre de personnes impliquées dans les interactions au quotidien est très faible dans chacune des catégories, mises à part trois d'entre elle qui sont beaucoup plus nombreuses (C1, C3 et C11). Pour beaucoup des catégories intervenant uniquement dans les services transversaux, le nombre quotidien moyen de personnes interagissant avec les différentes catégories est inférieur à 1, ce qui signifie qu'elles n'ont pas d'interaction du tout certains jours de l'expérience avec certaines des catégories. Il y a donc moins de risques de transmission pour les catégories socioprofessionnelles réparties au sein des différents services transversaux. Au contraire, il existe plus d'échanges entre les catégories associées à au moins un service de soins, qu'elles soient constituées d'un grand nombre de personnes ou non. En terme de proportion de personnes des catégories impliquées dans les échanges, donné par la figure 2.30b, on peut constater que la majorité des personnels interagissent avec certains patients. Les catégories interagissant les moins avec les patients sont celles dispersées dans les services transversaux comme les personnels administratifs (C13) et logistiques (C12). Toutes les catégories socioprofessionnelles réparties dans les différents services de soins ont plus de 60% de leur membre qui interagissent avec une personne appartenant aux catégories C1, C3 et dans une moindre mesure C11, dont les rôles les amènent à interagir beaucoup avec les patients. Ainsi, les personnels permettant la diffusion à travers l'hôpital sont les aides soignants (C1) et les infirmiers (C3). Parmi

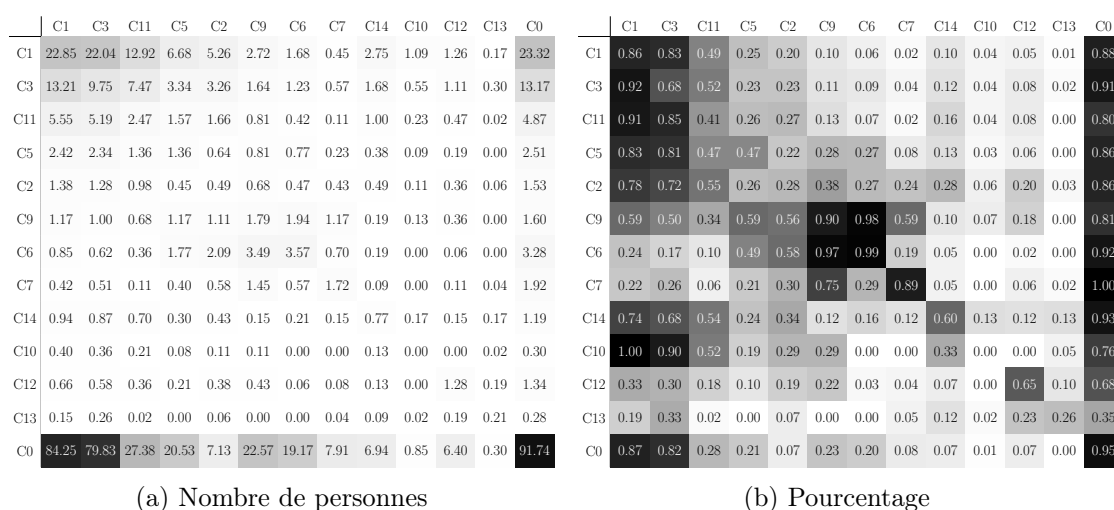


FIGURE 2.30 – Quantité de personnes de chaque catégorie (en ligne) interagissant par jour avec les autres catégories (en colonne). La figure de gauche donne la moyenne du nombre de personnes quotidien des groupes interagissant avec chacun des catégories. La figure de droite donne le pourcentage de personnes des catégories interagissant avec les différents services.

les catégories concernées par l'activité des patients, on peut aussi compter les catégories de rééducation de kinésithérapie et d'ergothérapies, susceptibles d'avoir un impact sur la diffusion de bactéries car beaucoup de patients sont concernés par leur activité et ce quelque soit le service auquel ils sont rattachés.

### Quantités moyennes d'activité échangée entre les catégories

La figure 2.31 donne les quantités d'activité quotidiennes individuelles moyennes des personnes de chaque catégories, encore une fois lues en ligne.

La majorité de l'activité des catégories socioprofessionnelles, comme nous pouvions nous y attendre, est concentrée autour des patients. La figure 2.31a concernant le degré des personnes confirme ce qui a été observé précédemment en terme de quantité de personnes concernées par les échanges. En effet, les catégories avec lesquelles il y a le plus d'échanges sont la catégorie des patients, celles des aides soignants (C1), des infirmiers (C3) ainsi que des ASH (C11). La figure 2.31b restreint le nombre de catégories très actives avec les autres à 2 (les aides soignants (C1) et les infirmiers (C3)) car le nombre de contacts par personnes avec les autres catégories est beaucoup plus élevé que les autres, ce qui s'explique par leur nombre plus important. Dans la suite (cf. figures 2.36, 2.37 et 2.38), nous verrons si ces valeurs sont plus ou moins élevées que ce que l'on obtiendrait en répartissant équitablement l'activité à l'aide d'un modèle de configuration. Le nombre de contacts entre les catégories ainsi que la durée cumulée quotidienne de ces contacts par personne sont très élevés avec les patients pour les aides soignants (C1), les infirmiers (C3), les ASH (C11) et les élèves stagiaires. La durées cumulée de ces contacts, bien qu'ils soient tous très courts, s'élève à plus d'une heure par jour. On remarque des taux de contacts aussi très élevés et dépassant l'heure par jour entre la catégorie de rééducation des kinésithérapeutes et des brancardiers.



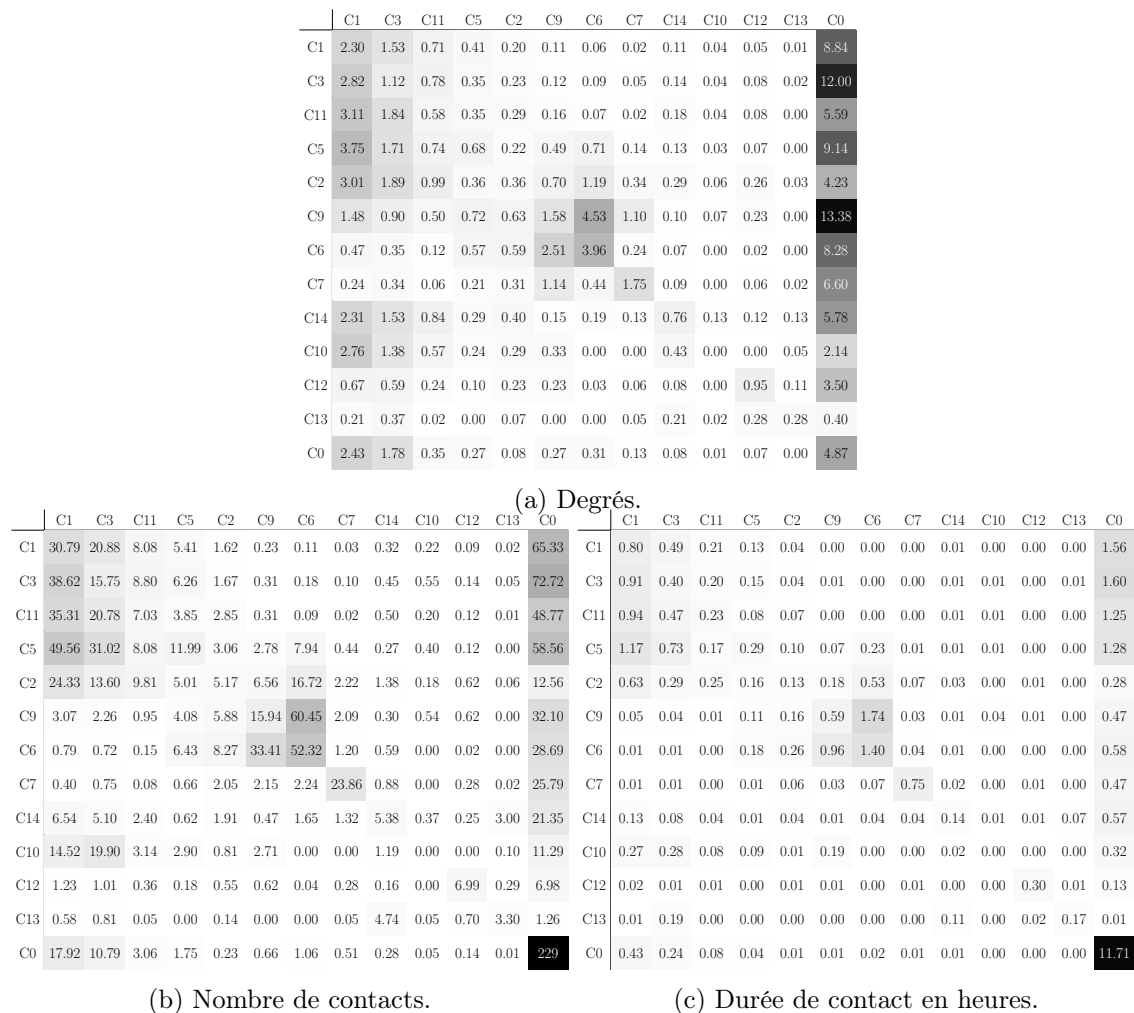


FIGURE 2.31 – Matrices d’activité entre les catégories par personne jour, selon les degrés quotidiens, le nombre de contacts et la durée des contacts.

## Répartition de l’activité entre les catégories

Les trois figures qui suivent sont relatives au pourcentage d’occupation des catégories, c’est à dire à la répartition de l’activité de ces catégories entre les différentes catégories socioprofessionnelles de l’hôpital.

La première de ces figures (fig. 2.33) s’appuie sur le degré des personnes tandis que les deux suivantes s’appuient sur la quantité d’activité en terme de nombre de contacts quotidiens (fig. 2.34) et durée de contacts quotidiens 2.35. Chacune de ces figures est constituée de deux sous-figures. La première de ces sous-figures prend en compte l’activité des patients contrairement à la deuxième qui s’appuie uniquement sur l’activité des personnels et qui permet de mieux évaluer les accointances de certaines catégories travaillant en binômes.

La figure 2.33a, nous informant sur la répartition de l’activité des degrés des catégories, montre encore une fois que la plupart des catégories partagent plus de 40% de leurs interactions avec des patients, que nous nommerons paires d’adjacence pour la description de cette courbe. Nous rappelons que les paires d’adjacences sont des paires de personnes ayant eu au moins une interaction durant la période observée et la somme des degrés de

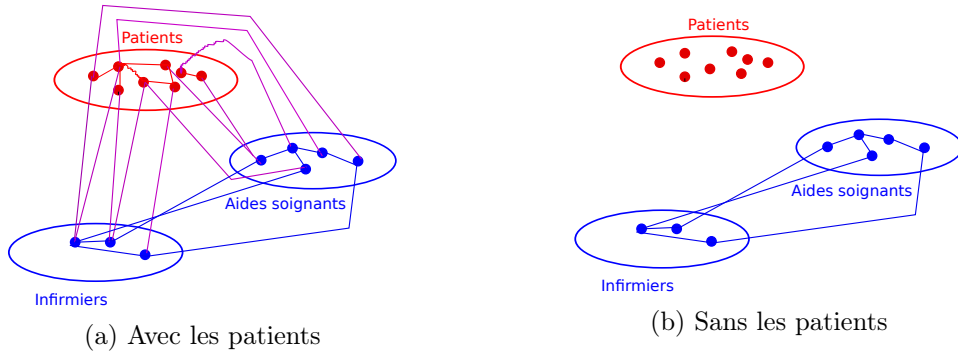


FIGURE 2.32 – Exemple de sélection d’activité pour le calcul de répartition de l’activité des catégories. Toutes les arêtes présentes sur les graphes sont prises en comptes. Les arêtes des patients sont représentées en rouges et en violet. Ces deux types d’arêtes sont prise en compte dans les premier cas et supprimées dans le deuxième.

toutes les personnes est égal au double du nombre de paires d’adjacence. Seules les cadres de santé, les animateurs, les personnels administratifs et logistiques ne partagent que entre 21 et 32 % de leur activité avec les patients et donc plus de 65 % avec leurs collègues. Comme il était prévisible, la deuxième catégorie partageant le plus d’activité avec toutes les autres catégories est celle des aides soignants (C1), les plus nombreux. En effet, pour beaucoup de catégories, la quantité de paires d’adjacences partagées avec les aides soignants (C1) représente au moins 8 à 10 % de leur activité. Les exceptions à cette règle sont les brancardiers et rééducateurs kinésithérapeutes et ergothérapeutes qui ont par contre une activité dépassant les 50 % avec les patients. Les cadres de santé et les logisticiens mis à part, toutes les catégories partagent plus de 60 % de leur paires d’adjacences avec ou bien la catégorie des patients ou bien des aides soignants (C1).

	C1	C3	C11	C5	C2	C9	C6	C7	C14	C10	C12	C13	C0		C1	C3	C11	C5	C2	C9	C6	C7	C14	C10	C12	C13
C1	8.70	11.53	5.38	3.09	1.52	0.83	0.48	0.13	0.84	0.31	0.37	0.05	66.78	C1	26.19	34.70	16.18	9.30	4.56	2.50	1.45	0.39	2.53	0.93	1.13	0.15
C3	16.33	3.25	4.51	2.00	1.35	0.71	0.50	0.27	0.79	0.22	0.47	0.12	69.47	C3	53.50	10.64	14.79	6.56	4.42	2.34	1.64	0.87	2.58	0.72	1.54	0.40
C11	25.87	15.33	2.42	2.94	2.40	1.37	0.57	0.15	1.47	0.31	0.64	0.03	46.51	C11	48.36	28.66	4.53	5.49	4.48	2.55	1.06	0.29	2.75	0.58	1.20	0.05
C5	21.45	9.81	4.24	1.93	1.26	2.83	4.05	0.78	0.74	0.19	0.41	0.00	52.30	C5	44.97	20.58	8.89	4.05	2.65	5.92	8.50	1.64	1.56	0.39	0.86	0.00
C2	22.23	13.98	7.31	2.67	1.34	5.18	8.80	2.51	2.12	0.47	1.89	0.24	31.26	C2	32.34	20.34	10.63	3.89	1.94	7.54	12.80	3.66	3.09	0.69	2.74	0.34
C9	6.04	3.66	2.07	2.96	2.57	3.24	18.56	4.52	0.39	0.27	0.94	0.00	54.78	C9	13.36	8.10	4.57	6.55	5.69	7.16	41.03	10.00	0.86	0.60	2.07	0.00
C6	3.12	2.29	0.76	3.78	3.88	16.49	13.03	1.56	0.45	0.00	0.10	0.00	54.54	C6	6.86	5.03	1.68	8.31	8.54	36.28	28.66	3.43	0.99	0.00	0.23	0.00
C7	2.27	3.31	0.57	1.98	3.02	10.96	4.25	8.41	0.85	0.00	0.57	0.19	63.61	C7	6.23	9.09	1.56	5.45	8.31	30.13	11.69	23.12	2.34	0.00	1.56	0.52
C14	18.65	12.35	6.77	2.38	3.21	1.19	1.54	1.07	3.09	1.07	0.95	1.07	46.67	C14	34.97	23.16	12.69	4.45	6.01	2.23	2.90	2.00	5.79	2.00	1.78	2.00
C10	33.72	16.86	6.98	2.91	3.49	4.07	0.00	0.00	5.23	0.00	0.00	0.58	26.16	C10	45.67	22.83	9.45	3.94	4.72	5.51	0.00	0.00	7.09	0.00	0.00	0.79
C12	10.57	9.37	3.78	1.66	3.63	3.63	0.45	0.91	1.21	0.00	7.55	1.81	55.44	C12	23.73	21.02	8.47	3.73	8.14	8.14	1.02	2.03	2.71	0.00	16.95	4.07
C13	11.84	21.05	1.32	0.00	3.95	0.00	0.00	2.63	11.84	1.32	15.79	7.89	22.37	C13	15.25	27.12	1.69	0.00	5.08	0.00	0.00	3.39	15.25	1.69	20.34	10.17
C0	29.53	21.68	4.27	3.33	0.94	3.33	3.73	1.59	0.93	0.11	0.87	0.04	29.66													

(a) Toutes les catégories
(b) Sans les patients

FIGURE 2.33 – Pourcentage d’occupation des catégories (en ligne) selon le degré des catégories. A gauche, le pourcentage en tenant compte de l’activité de toutes les catégories. A droite, le pourcentage d’occupation des catégories sans tenir compte de l’activité des patients, prépondérante.

La figure 2.33b donne la répartition des paires d’adjacences des catégories en oubliant toute l’activité liée aux patients. Cette figure souligne la forte influence des aides soignants (C1) sur l’activité des autres catégories, ce qui n’est pas surprenant étant donné leur nombre.

On remarque que les infirmiers (C3) partagent aussi un grand nombre de paires d'adjacence des différentes catégories de l'hôpital. On peut par contre remarquer les aides soignants (C1) partagent plus de paires d'adjacence avec les infirmier (C3) qu'avec leur propre catégories, malgré le fait qu'ils soient beaucoup plus nombreux. Cette observation s'explique par l'organisation même du travail à l'hôpital : en effet, les aides soignants (C1) travaillent en binôme avec les infirmiers (C3). De même, on remarque que les kinésithérapeutes ainsi que les ergothérapeutes partagent une bonne proportion de leurs paires de contacts. Les personnels administratifs et logistiques semblent aussi interagir beaucoup ensemble. Seulement 30% des paires impliquant un patient impliquent un deuxième patient. Les patients partagent donc leurs paires d'adjacence équitablement entre avec la catégorie des patients, celle des aides soignants (C1) et celle des infirmiers (C3).

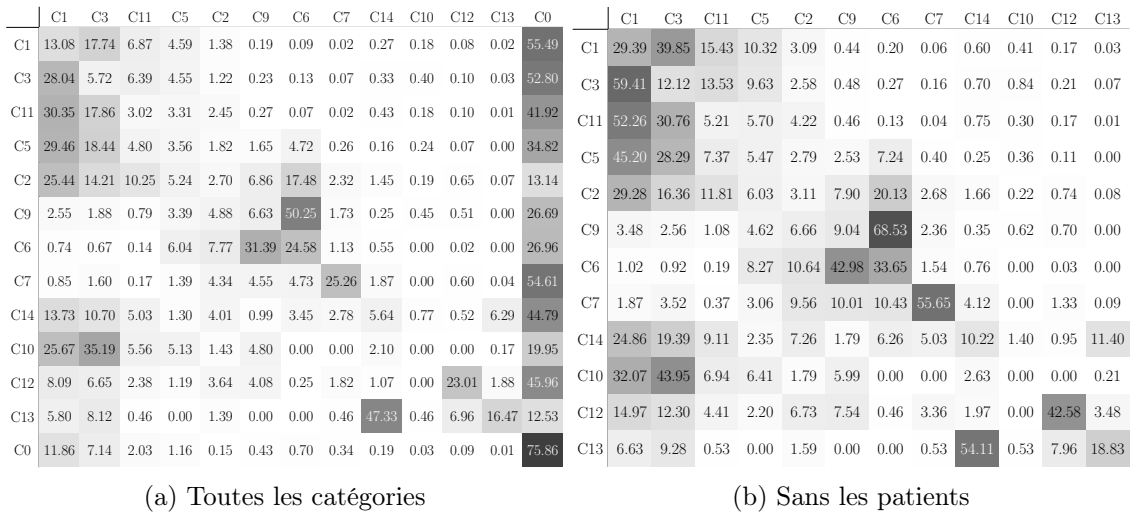


FIGURE 2.34 – Pourcentage d'occupation des catégories (en ligne) selon le nombre de contacts des catégories. A gauche, le pourcentage en tenant compte de l'activité de toutes les catégories. A droite, le pourcentage d'occupation des catégories sans tenir compte de l'activité des patients, prépondérante.

La figure 2.34a donne la répartition du nombre de contacts partagés par les différentes catégories de l'hôpital. Les résultats obtenus sont plus marqués que ceux observés précédemment au sujet des paires d'adjacence. En effet, la catégorie des patients partage proportionnellement beaucoup moins de contacts avec les personnels et partagent par contre 75% de ses contacts avec des patients.

Les catégories de personnels partageant une importante proportion de leur paires d'adjacence partagent une plus grosse proportion de contact encore, ce qui est déjà visible sur le figure prenant en compte l'activité des patients.

Les rééducateurs affiliés à des services de soins particuliers partagent plus de 15% de leur contacts avec les personnels en logistique. Ce fort taux d'interactions entre les deux services n'était pas mise en avant jusqu'à présent.

La figure 2.34b renforce les observations faites précédemment et souligne plus fortement les appariement des différentes catégories.

A titre informatif, la figure 2.35 donne la répartition des durées de contacts partagées entre les différentes catégories. Les observations qu'on peut en faire sont les mêmes que la figure précédente 2.34 concernant le nombre de contacts.

	C1	C3	C11	C5	C2	C9	C6	C7	C14	C10	C12	C13	C0		C1	C3	C11	C5	C2	C9	C6	C7	C14	C10	C12	C13
C1	13.99	17.24	7.52	4.50	1.48	0.13	0.06	0.02	0.22	0.14	0.05	0.01	54.64	C1	30.85	38.00	16.58	9.93	3.27	0.28	0.14	0.04	0.49	0.31	0.10	0.02
C3	29.03	6.43	6.35	4.69	1.12	0.17	0.11	0.06	0.23	0.24	0.06	0.35	51.15	C3	59.44	13.17	12.99	9.60	2.30	0.34	0.22	0.12	0.48	0.50	0.12	0.71
C11	31.75	15.90	3.96	2.70	2.48	0.12	0.05	0.01	0.30	0.19	0.06	0.01	42.48	C11	55.19	27.64	6.89	4.69	4.32	0.21	0.08	0.01	0.53	0.32	0.11	0.01
C5	29.95	18.52	4.25	3.67	2.52	1.90	5.81	0.15	0.13	0.32	0.04	0.00	32.76	C5	44.54	27.54	6.32	5.46	3.74	2.82	8.64	0.22	0.20	0.47	0.05	0.00
C2	25.40	11.44	10.09	6.49	2.63	7.23	21.21	2.70	1.05	0.11	0.38	0.08	11.19	C2	28.60	12.89	11.36	7.31	2.96	8.14	23.89	3.04	1.18	0.12	0.43	0.09
C9	1.63	1.29	0.37	3.68	5.45	9.91	58.87	0.99	0.24	1.28	0.47	0.00	15.82	C9	1.93	1.53	0.44	4.38	6.47	11.77	69.93	1.18	0.28	1.53	0.56	0.00
C6	0.49	0.50	0.08	6.68	9.45	34.80	25.29	1.33	0.49	0.00	0.03	0.00	20.86	C6	0.61	0.63	0.11	8.43	11.94	43.97	31.96	1.68	0.62	0.00	0.04	0.00
C7	0.67	1.24	0.07	0.82	5.80	2.83	6.41	35.15	2.33	0.00	0.56	0.02	44.09	C7	1.20	2.23	0.12	1.46	10.38	5.07	11.46	62.87	4.16	0.00	1.01	0.04
C14	11.76	7.42	3.82	1.07	3.25	0.98	3.43	3.37	6.09	0.48	0.61	6.13	51.59	C14	24.29	15.32	7.90	2.21	6.71	2.03	7.08	6.96	12.58	0.98	1.26	12.67
C10	21.40	22.03	6.68	7.25	0.95	15.10	0.00	0.00	1.36	0.00	0.00	0.06	25.18	C10	28.60	29.44	8.93	9.69	1.26	20.18	0.00	0.00	1.81	0.00	0.00	0.08
C12	4.79	3.68	1.53	0.57	2.36	3.83	0.44	1.62	1.20	0.00	41.80	1.75	36.42	C12	7.54	5.78	2.41	0.90	3.72	6.02	0.69	2.55	1.89	0.00	65.75	2.75
C13	2.07	44.59	0.40	0.00	0.99	0.00	0.00	0.13	24.92	0.09	3.59	20.03	3.19	C13	2.13	46.06	0.42	0.00	1.02	0.00	0.00	0.14	25.74	0.09	3.71	20.69
C0	6.38	3.54	1.18	0.57	0.08	0.14	0.32	0.14	0.11	0.02	0.04	0.00	87.48													

(a) Toutes les catégories

(b) Sans les patients

FIGURE 2.35 – Pourcentage d’occupation des catégories (en ligne) selon la durée cumulée des contacts des catégories. A gauche, le pourcentage en tenant compte de l’activité de toutes les catégories. A droite, le pourcentage d’occupation des catégories sans tenir compte de l’activité des patients, prépondérante.

### Déviations de l’activité entre les catégories.

Nous allons maintenant comparer les valeurs décrites précédemment à celles obtenues en répartissant de manière équilibrée la quantité d’activité des catégories. Nous appliquons, pour ce faire, le modèle de configuration simple que nous avons décrit dans la partie 2.3.2. Nous employons la formule 2.13 que nous avons utilisée sur l’ensemble des services, cette fois-ci appliquée à l’ensemble des catégories socioprofessionnelles. Les résultats vont être présentés comme précédemment sous forme de matrices de contacts dont chaque case  $M[i][j]$  donnera la valeur de déviation d’activité entre les catégories  $i$  et  $j$ . Ces matrices sont symétriques car il s’agit de la comparaison de mesures directes, qui sont symétriques dans les données. L’activité des patients étant très importante par rapport à l’activité des personnels, nous avons aussi comparé la répartition des quantités d’activité en prenant en compte uniquement l’activité ayant lieu entre les catégories professionnelles, sans plus tenir compte des activités impliquant au moins un patient. Nous utiliserons dans la suite l’expression *modèle complet* pour désigner le modèle tenant compte de l’activité totale de l’hôpital, y compris celle des patients. Nous appellerons *modèle partiel* le modèle qui s’appuie uniquement sur l’activité des personnels, en ne tenant aucun compte de l’activité des patients.

La figure 2.36 donne le résultat du calcul des déviations d’activité des catégories socioprofessionnelles en s’appuyant sur le degré des personnes. La figure 2.36a prend en compte l’activité de toutes les catégories tandis que la figure 2.36b ne tient pas compte de l’activité des patients. La couleur des cases des matrices est plus foncée lorsque le modèle est très éloigné des valeurs réelles. Les valeurs négatives indiquent que les valeurs réelles sont inférieures à celles que prévoit le modèle de configuration. Elles sont surlignées par une forme

hexagonale bleu clair. Les valeurs positives signifient que les valeurs réelles sont supérieures aux valeurs obtenues en utilisant le modèle de configuration. Elles sont surlignées par une forme ovale rouge clair.

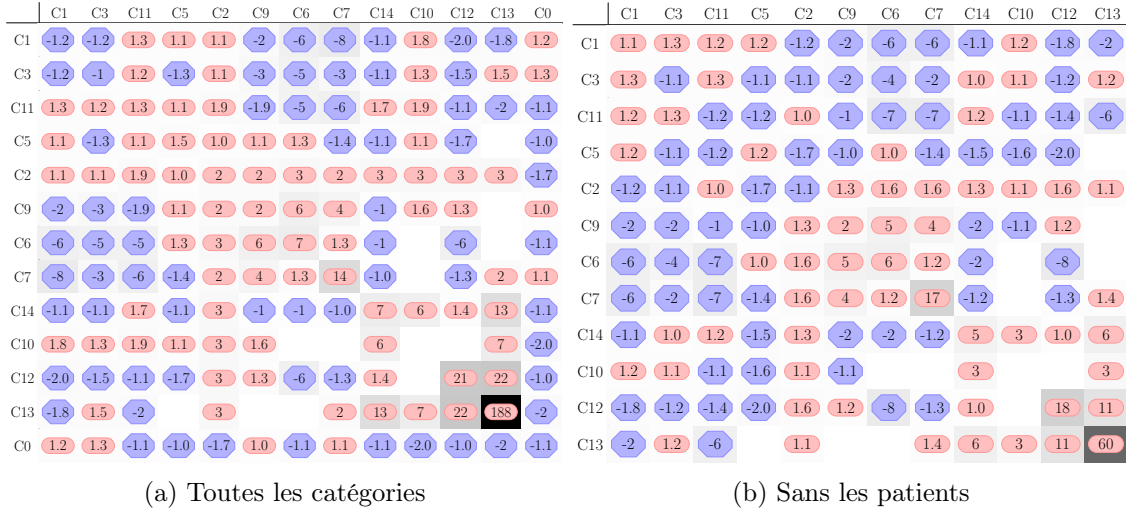


FIGURE 2.36 – Matrices de déviation de l'activité entre les catégories par rapport à un modèle de répartition de l'activité aléatoire, selon les degrés quotidiens des différentes catégories. A gauche, l'activité de toutes les catégories est prise en compte. A droite, nous ne prenons pas en compte l'activité de la catégorie C0 des patients. Le fond des cases est gris et d'autant plus foncé que la valeur absolue de déviation au modèle de configuration est grande. Les valeurs négatives en bleu ( -3 ) dénotent une activité plus faible que prévue par le modèle de configuration. Les valeurs positives en rouge ( 11 ) dénotent une activité plus importante que ne le prévoit le modèle.

Commençons par étudier la figure 2.36a qui s'appuie sur l'activité de l'ensemble des catégories socioprofessionnelles, y compris les patients. On remarque immédiatement que certaines paires de catégories sont très proches du modèle de configuration, qui sert de modèle de référence, car les valeurs de déviation sont très proches de 1. Au contraire, certaines paires sont très éloignées du modèle de configuration. Parmi ces paires, certaines sont beaucoup plus actives en réalité qu'en suivant le modèle alors que d'autres sont à l'inverse beaucoup moins actives que lors d'une répartition équilibrée de l'activité. Parmi les catégories ayant un comportement très proche de ce que prévoit le modèle de configuration simple, nous pouvons citer la catégorie des stagiaires, qui n'est pas plus de 1,5 fois plus ou moins active que ne le prévoit le modèle, mises à part deux exceptions : son activité avec les personnels administratifs (C13) et celle avec les personnels logistiques (C12). La catégorie des patients est aussi assez proche du modèle de configuration, bien que leur activité avec les cadres de santé (C2), les animateurs (C10), les personnels administratifs (C12) soit 2 fois inférieure à ce que prévoit le modèle de configuration.

La catégorie des cadres de santé (C2) ressort par les résultats particuliers qu'elle obtient. En effet, c'est la seule catégorie à avoir plus d'activité que prévu par le modèle avec toutes les catégories professionnelles sans exception. La seule catégorie avec laquelle elle interagit moins selon les données expérimentales que selon le modèle est celle des patients. L'éloignement de l'activité des cadres de santé au modèle est assez marqué (facteurs multiplicatifs 2 à 3) avec toutes les catégories à part 3 exceptions que sont les stagiaires (C5) ainsi que les personnels soignants infirmiers (C3) et aides soignants (C1).

Toutes les autres catégories montrent des accointances ou des répulsion particulières pour différentes catégories. Ainsi, les personnels soignants au quotidien (C1,C3) et les ASH (C11) ont une activité très peu éloignée du modèle de configuration au sein de leur propre catégorie et très faible avec les personnels de rééducation (C6, C7) et les brancardiers (C9). Les brancardiers (C9) et les personnels de rééducation kinésithérapeutes (C6) et ergothérapeutes (C7) interagissent au contraire beaucoup plus entre eux que ce qui est prévu par le modèle, ainsi qu'avec les cadres de santé.

La figure 2.36b, qui contient les valeurs de déviation d'activité des catégories au modèle de configuration en utilisant uniquement l'activité des personnels, apporte le même type d'informations que la figure 2.36a : la plupart des valeurs proches de 1 le restent, même si certaines changent de signe, globalement, les déviations au modèle de configuration sont moins importantes lorsque l'on ne prend pas en compte l'activité des patients. Ainsi, l'activité des cadres de santé (C2) paraît beaucoup plus proche du modèle de configuration que sur la figure 2.36a. En effet, toutes les valeurs absolues de déviation sont inférieures à 2. Au contraire, la déviation d'activité des ASH (C11) augmente un peu avec les personnels de rééducation (C6 et C7).

Considérons maintenant les déviations au modèle de configuration basé sur les quantité de contacts afin de vérifier s'il est possible d'extraire d'autres informations. La figure 2.37 nous informe sur la déviation au modèle de configuration basé sur le nombre de contacts des personnes. Comme précédemment, la première figure 2.37a donne les résultats obtenus en appliquant le modèle de configuration à toutes les personnes de l'hôpital. La deuxième figure 2.37b concerne les valeurs obtenues en utilisant le modèle de configuration appliqué uniquement au personnel de l'hôpital, sans tenir compte des patients.

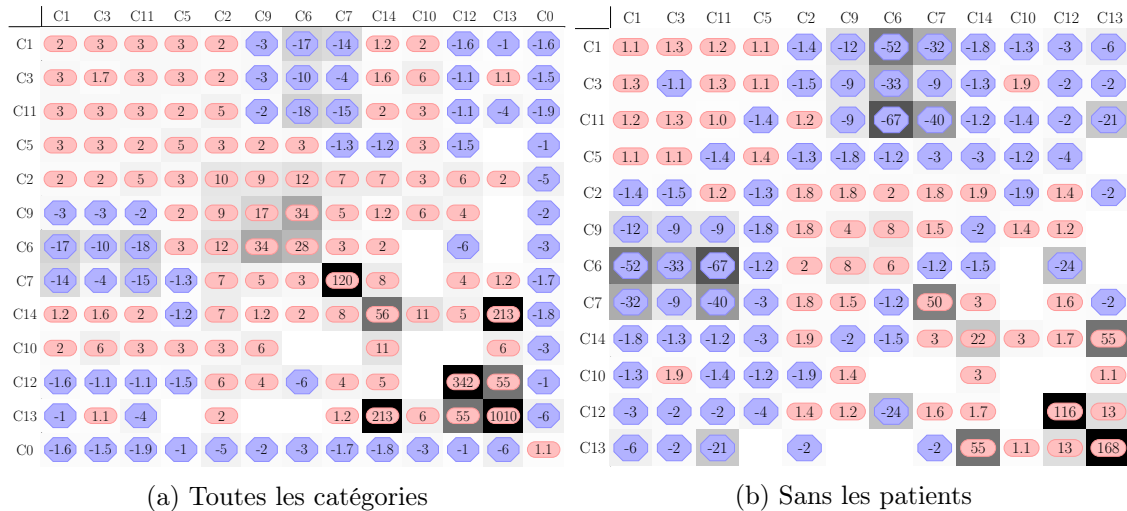


FIGURE 2.37 – Matrices de déviation de l'activité entre les catégories par rapport à un modèle de répartition de l'activité aléatoire, selon le nombre de contacts quotidiens des différentes catégories. A gauche, l'activité de toutes les catégories est prise en compte. A droite, nous ne prenons pas en compte l'activité de la catégorie C0 des patients. Le fond des cases est gris et d'autant plus foncé que la valeur absolue de déviation au modèle de configuration est grande. Les valeurs négatives en bleu ( -3 ) dénotent une activité plus faible que prévue par le modèle de configuration. Les valeurs positives en rouge ( 13 ) dénotent une activité plus importante que ne le prévoit le modèle.

On peut remarquer, en premier lieu, que les valeurs obtenues sont beaucoup plus élevées que les valeurs de déviations basées sur les degrés. En particulier, les valeurs tenant compte de l'activité des patients (cf. figure 2.38a) sont beaucoup plus éloignées de 1 que précédemment. Nous devons aussi noter la grande différence entre les deux types de déviation toutes deux basées sur le nombre de contacts, lorsque l'on prend ou non en compte l'activité des patients. En effet, la première figure tenant compte de l'activité des patients donne des déviations majoritairement positives contrairement aux déviations d'activité calculées sur les personnels uniquement, qui sont, elles, en grande majorité négatives. Toutes les valeurs de déviation diminuent lorsque que l'on ignore l'activité des patients, ce qui rapproche du modèle de configuration les catégories plus actives que prévu par ce modèle. Les paires de catégories concernées sont principalement l'ensemble des catégories de personnels de vie quotidienne (C1, C3, C5, C11) qui ont entre elles une activité proche de celle du modèle.

Les résultats obtenus par comparaison des données expérimentales au modèle de configuration partiel, basé sur le nombre de contacts, et représentés par la figure 2.37b confirment les informations obtenues en utilisant le modèle partiel basé sur le degré des personnes. En effet, en s'appuyant sur le modèle ne prenant en compte que l'activité des personnels, l'activité des cadres de santé (C2) reste très proche du modèle de configuration avec toutes les autres catégories. L'activité des personnels de santé de vie quotidienne C1, C3, C11 et C5 est peu éloignée du modèle. Nous pouvons donc donner des règles de répartition de l'activité simple pour construire un réseau entre ces catégories qui ait les mêmes propriétés que le réseau réel que nous observons. Certaines paires de catégories ne suivent absolument pas le modèle de configuration, comme les personnels de rééducation (C6 et C7) avec les personnels de santé de vie quotidienne, ce qui ne nous permet pas de construire un modèle simple de réseau de répartition des liens qui reproduise leur comportement.

Vérifions si le même phénomène est observable lorsque l'on considère la répartition des durées de contacts entre les différentes catégories. La figure 2.38 nous informe de la déviation au modèle de configuration basé sur la durée cumulée de contacts des personnes. La première figure 2.38a donne les résultats obtenus en appliquant le modèle de configuration complet, appliqué à toutes les personnes de l'hôpital. La deuxième figure 2.38b concerne les valeurs obtenues en utilisant le modèle de configuration partiel, appliqué uniquement au personnel de l'hôpital, sans tenir compte des patients.

La figure 2.38 donne des résultats similaires à la figure 2.37, bien qu'avec des valeurs plus importantes, ce qui était attendu car le nombre des contacts et leur durée cumulée sont très corrélés comme nous l'avons déjà constaté précédemment. Les valeurs de déviation de l'activité du modèle complet ont toutes, mises à part quelques exceptions, des valeurs extrêmement élevées, supérieures à 2, quelles soient positives ou négatives. Cela signifie que l'activité des personnes est répartie selon une autre loi que la répartition équilibrée des contacts et donc ont une structure particulière à l'environnement étudié. Le modèle de configuration s'appuyant uniquement sur l'activité des personnels donne des valeurs beaucoup moins importantes et fait mieux ressortir les catégories socio-professionnelles interagissant beaucoup ensemble de celles qui ont une quantité d'interactions faible. Il ressort par exemple que les aides soignants (C1), les infirmiers (C3) et les ASH (C11) forment une "communauté" interagissant entre eux, légèrement plus que ne le prévoirait une répartition équitable des contacts. Ces trois catégories interagissent par contre beaucoup moins avec les brancardiers (C9), les kinésithérapeutes (C6), les ergothérapeutes et les

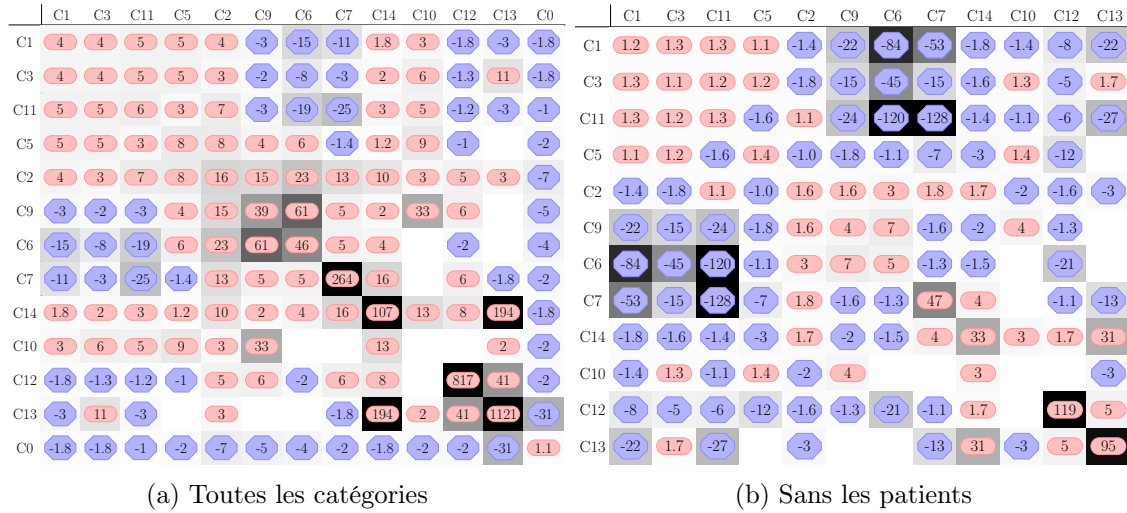


FIGURE 2.38 – Matrices de déviation de l'activité entre les catégories par rapport à un modèle de répartition de l'activité aléatoire, selon la durée des contacts quotidiens des différentes catégories. A gauche, l'activité de toutes les catégories est prise en compte (modèle complet). A droite, nous ne prenons pas en compte la catégorie C0 des patients (modèle partiel). Le fond des cases est gris et d'autant plus foncé que la valeur absolue de déviation au modèle de configuration est grande. Les valeurs négatives en bleu ( $-3$ ) dénotent une activité plus faible que prévue par le modèle de configuration. Les valeurs positives en rouge ( $15$ ) dénotent une activité plus importante que ne le prévoit le modèle.

logisticiens que suivant un modèle de configuration. Ces groupes étaient déjà perceptibles sur la figure 2.38a, mais moins repérables.

Ainsi, quelque soit les quantités sur lesquelles nous appliquons le modèle de configuration, qu'il soit complet ou partiel, nous obtenons des informations relativement similaire et pouvons donc utiliser au choix chacune de ces quantité pour modéliser la répartition de l'activité. Pour que le modèle soit proche des données de l'expérience, il est préférable d'utiliser dans un premier le modèle complet pour assigner la quantité d'activité ayant lieu avec les patients. Pour avoir un modèle plus fin de la répartition de l'activité des personnels, le modèle partiel apporte plus d'informations quant à la structure du réseau attendu.

Nous avons appris que les personnels soignants (infirmiers (C3), aides soignants (C1), stagiaires (C5), cadres de santé (C2)) ainsi que les ASH (C11) interagissent relativement peu avec les personnels de leur propre catégorie en regard de leurs échanges avec les autres catégories parmi les personnels soignants et ASH (C11). Les autres catégories professionnelles interagissent encore moins avec les infirmiers (C3), aides soignants (C1) et ASH (C11), mis à part les animateurs et les rééducateurs quelconques (sans les kinésithérapeutes et les ergothérapeutes).

### En résumé

Dans ce chapitre, nous avons traité de la structure du réseau de contacts suivant deux types de répartition des personnes. Nous avons dans un premier temps étudié les interactions des personnes regroupées en services de soins et services transversaux. Ces services sont définis par le fonctionnement même de l'hôpital et chaque personne est rattachée à l'un d'entre eux uniquement. Dans un second temps, nous avons considéré le réseau des contacts en répartissant les personnes selon les catégories socioprofessionnelles, également établie par l'hôpital, cette fois selon le statut professionnel des personnes et en particulier du métier



qu'elles exercent. Encore une fois, chaque personne n'appartient qu'à une et une seule catégorie socioprofessionnelle.

Concernant l'organisation en services, nous en avons identifié de deux types : les premiers ont des tailles importantes et sont composés par des patients et des personnels. Les seconds sont composés uniquement de personnels et ont des populations en faible nombre. Nous avons introduit plusieurs outils afin de caractériser la structure du réseau des services. Nous avons en particulier présenté une mesure d'introversion qui permet d'évaluer la quantité d'échanges entre les différents services. Nous en avons conclu que les services de soins sont très introvertis, contrairement aux services transversaux qui le sont beaucoup moins. Nous avons ensuite introduit un modèle de configuration qui permet de répartir équitablement l'activité des différents services afin d'évaluer la spécificité du réseau hospitalier, comparé à un modèle aléatoire, et donc de comportement moyen. Nous constatons, dans le cas de services, que la majorité de leur activité était interne, ou avait lieu avec des services localement proches (dans le même bâtiment) ainsi qu'avec le service de nuit. Nous avons aussi constaté que la quantité d'activité des personnels était très inférieure à celle des patients et réparties plus équitablement en terme de durée de contacts.

En ce qui concerne l'étude de l'organisation de l'activité des catégories socioprofessionnelles, nous avons d'abord constaté qu'elles sont très hétérogènes en terme de population. Nous avons identifié quatre types de catégories : les catégories intervenant uniquement dans les services de soins, celles se partageant entre services de soins et services transversaux, celles dont les membres sont repartis uniquement entre les services transversaux et celles associées à un seul et unique service transversal. Les outils de mesure d'introversion ont mis en évidence que ces catégories socioprofessionnelles sont très extraverties mises à part quelques unes d'entre elles, telle que la catégorie particulière des patients.

Pour les services comme pour les catégories, la répartition de l'activité est très éloignée des modèles de configuration qui permettent de répartir équitablement l'activité entre les groupes définis, que l'on considère le volume d'activité ou sa diversité. L'hôpital a donc une organisation propre très centrée autour des services.

Les différentes catégories sont ensuite organisées à l'intérieur même des services pour celles qui sont réparties dans plusieurs services. Les catégories à service unique sont plus centrées sur elles-mêmes.

Toutes ces informations confirment l'idée que le réseau de contact humain en milieu hospitalier est régi par les activités médicales de chacune des catégories socio-professionnelles et de l'organisation des services de soins. Lors de la construction de modèle de diffusion ou de l'étude des risques de diffusion, cette organisation hiérarchique en service puis en catégories socioprofessionnelles spécifiques aux réseaux hospitaliers est à prendre en compte.



## Chapitre 3

# Une structure temporelle à plusieurs échelles

Nous avons jusqu'à présent considéré la dimension structurelle voire spatiale des interactions ayant lieu dans un hôpital constitué de plusieurs services et faisant intervenir différentes catégories d'individus. Dans ce chapitre, nous nous intéressons à l'évolution temporelle des interactions. Nous disposons de données constituant un réseau de contacts à granularité très fine, puisque nous avons un relevé des interactions toutes les 30 secondes. Considérer directement comme un réseau à part entière chaque période de 30 seconde, de manière indépendante, engendre une multitude de réseaux de très faible densité. Hors, travailler sur des réseaux de très faibles densités n'apporte que peu d'informations utiles et exploitables en l'état. Nous avons donc choisi de travailler sur des réseaux obtenus en agrégeant les données selon plusieurs échelles de temps. Trouver des échelles de temps d'agrégation pertinentes et exploitables n'est pas une tâche aisée. Nous travaillons sur des réseaux agrégés au jour, puis à l'heure car ce sont des échelles pertinentes et faciles à analyser et à comprendre car nous sommes habitués à découper le temps de cette façon. Nous travaillons ainsi sur des réseaux agrégés sur des périodes d'un jour dans un premier temps. Nous nous intéressons à l'évolution du nombre de personnes actives au cours du temps, au nombre de paires de contacts actives ainsi qu'au nombre et à la durée de ces contacts au cours du temps. Nous considérons ensuite plusieurs mesures permettant de quantifier le taux de similarité entre deux réseaux, agrégés à la journée, construits sur des périodes de temps différentes. Enfin, nous considérons l'évolution temporelle des différentes quantités d'activité à l'heure pour extraire les patterns journaliers et hebdomadaires de l'activité de l'hôpital de Berk-sur-Mer.

### 3.1 Evolution journalière

Nous nous intéressons dans cette partie à l'évolution de l'activité de la globalité de l'hôpital au cours des jours. Considérons dans un premier temps simplement l'évolution de la population active au cours des jours. Nous rappelons qu'un individu est dit actif quand il partage au moins un contact avec un autre individu durant la période observée (ici, un jour).

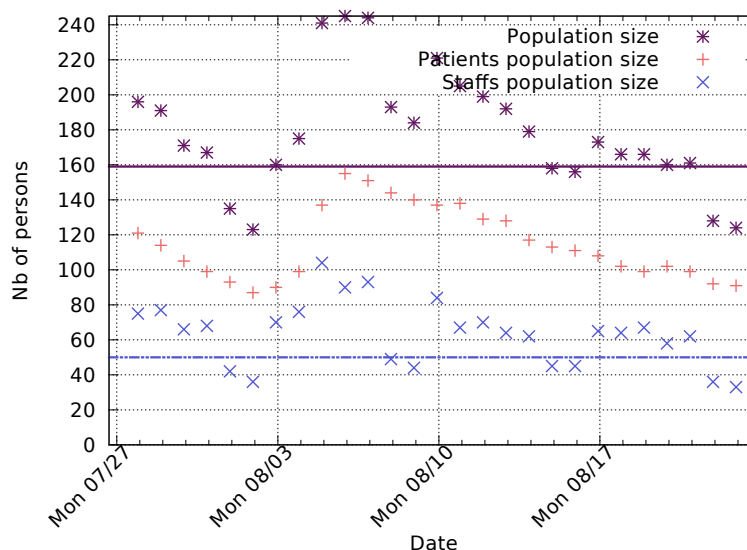


FIGURE 3.1 – Évolution de la population au cours du temps. En ordonnées se trouve le nombre de personnes présentes dans l'hôpital. L'échelle temporelle est représentée par période d'un jour.

La figure 3.1 représente l'évolution de la taille de la population chaque jour sur un mois, du 27 juillet au 23 août. Les étoiles en violet foncé représentent la population totale tandis que les plus (+) rouges indiquent le nombre de patients par jour. Les croix bleues indiquent le nombre journalier de personnels. Comme il est prévu par l'emploi du temps des personnels de l'hôpital, la population totale est toujours moins importante durant les week-ends que pendant les jours de la semaine, de même que la taille de la population des personnels. En effet, la population de l'hôpital est toujours inférieure à 160 personnes différentes, présentes dans la journée durant les week-ends. À l'inverse, les jours de semaine, l'hôpital compte toujours plus de 160 personnes présentes. Il est à noter que la taille de la population est très variable d'un jour à l'autre et que la forme de la courbe représentant la variation de la population totale suit celle de la population des personnels. Les seules exceptions notables sont les 5, 6 et 7 août, dont la variation de population est aussi fortement influencée par la population des patients qui augmente considérablement.

La taille de la population, au cours des jours, est donc assez instable, ce qui devra être pris en compte lors de l'étude de l'évolution de l'activité quotidienne de l'hôpital, en particulier lors du choix des mesures destinées à comparer l'évolution de cette activité au cours des jours.

### 3.1.1 Connexité et diamètres journaliers

Les mesures importantes lors de l'étude de la structure d'un réseau, en particulier dans le but de comprendre les risques de diffusion pouvant avoir lieu sur un réseau sont les mesures de connexité et de diamètre. Nous nous intéressons particulièrement à l'évolution de ces deux quantités au cours du temps.

On appellera dans la suite **réseau journalier**, un réseau construit en agrégeant les contacts des données à la journée. Nous avons calculé le nombre de composantes connexes par réseau

journalier ainsi que le diamètre de chacune des composantes connexes pour chaque réseau journalier durant un mois.

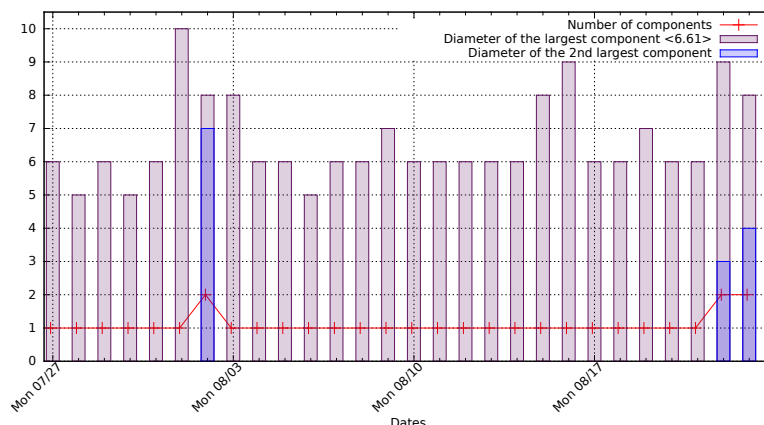


FIGURE 3.2 – Diamètres journaliers des composantes connexes formées par les personnes de la globalité de l’hôpital. Les diamètres sont représentés par des barres. Une couleur est attribuée par composante connexe. La ligne rouge indique le nombre de composantes connexes par jour.

La figure 3.2 indique le nombre de composantes connexes distinctes existant chaque jour dans l’hôpital, sur une période d’un mois (du lundi 27 juillet au dimanche 23 août). On constate que le plupart du temps, le réseau hospitalier est entièrement connecté à la journée et n’est donc composé que d’une seule et unique composante connexe. Cela signifie qu’il n’y a généralement pas de groupes de l’hôpital entièrement isolés à l’échelle de la journée. On peut donc supposer qu’une transmission totale à travers le réseau est possible. Il faut tout de même rester prudent car l’agrégation à la journée fait perdre un grand nombre d’informations sur la temporalité des interactions qui joue un rôle important lorsque l’on étudie les possibilités de diffusion dans un réseau. On constate que les seuls jours pour lesquels le réseau est composé de deux composantes connexes distinctes s’inscrivent durant les week-ends : en effet, deux dimanches sont concernés par cette anomalie ainsi qu’un samedi. Il nous reste à comprendre si la division en deux composantes connexes distinctes est équilibrée ou non, c’est à dire si elle répartie la population en deux groupes de taille équilibrée et si la connexion de ces composantes est plus ou moins forte. La figure 3.4 permet de confirmer que la séparation des composantes n’est pas équilibrée lorsqu’il y a deux composantes connexes. En effet cette figure 3.4 fait la corrélation entre la valeur du diamètre et le nombre de nœuds constituant chacune des composantes connexes. Lorsqu’il y a deux composantes connexes pour une même journée, la plus grosse est représentée en rouge avec le symbole (+) tandis que la deuxième est représentée en vert avec le symbole (X). Lorsque plusieurs composantes coexistent, la première contient toujours une grosse majorité de la population du jour tandis que l’autre n’est formée que par un tout petit groupe de personnes (inférieur à 25 personnes, ce qui représente au maximum 1/5 des effectifs de l’hôpital).

Revenons à la figure 3.2 et considérons la taille des diamètres des composantes connexes journalières. On remarque que sur l’échantillon représenté ici, les diamètres journaliers sont plus élevés durant les week-ends, y compris lorsque le réseau est réparti selon deux composantes connexes distinctes. On constate aussi que les diamètres journaliers sont généralement égaux à six. La moyenne de ces diamètres est légèrement plus élevée et se rapproche

de 7, ce qui paraît élevé pour une si petite structure. La structure du réseau de l'hôpital est donc assez peu connectée et laisse un espoir pour trouver des stratégies efficaces afin de déconnecter une partie du réseau dans une optique de ralentir ou d'enrayer une diffusion.

Pour ce type d'expérience, il serait intéressant de pouvoir prédire, la taille de la population donnée, le diamètre journalier ainsi que le nombre de composantes connexes. Nous considérons donc la taille de la population de l'hôpital et essayons d'établir si elle a une incidence sur le nombre de composantes connexes du réseaux de proximité des personnes. La figure 3.3 met en relation ces deux quantités. Le nombre de personnes présentes dans l'hôpital est représenté en abscisses de cette figure. Le nombre de composantes connexes du réseau est placé en ordonnées.

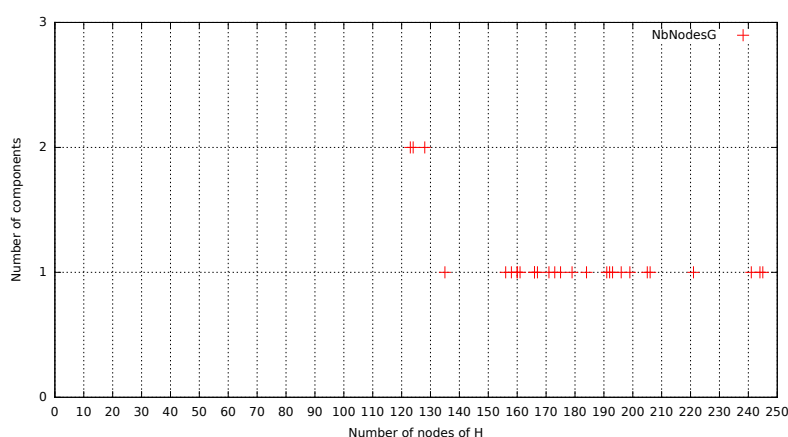


FIGURE 3.3 – Corrélation entre le nombre de composantes connexes du réseau de personnes (en abscisses) et la taille journalière de la population de l'hôpital (en ordonnées).

Nous savons, d'après la figure 3.2, que seuls trois jours ont un réseau constitué de deux composantes connexes. Les trois jours "déconnectés" sont les jours dont la population est la plus faible, comme le souligne la figure 3.3. En effet, les seuls jours dont le réseau est formé par deux composantes connexes, la taille de la population était inférieure à 130 personnes. Ces jours correspondent à deux dimanches et un samedi d'après la figure précédente 3.2. Il semble donc, sur le petit échantillon sur lequel nous travaillons, que le réseau est déconnecté quand il est composé d'une faible population et qu'il est au contraire plus connecté lorsque la population est plus importante.

Le diamètre du réseau est-il lui aussi dépendant de la taille de la population ? La figure précédente, comparant le nombre de composantes connexes, tend à nous faire penser que plus la taille de la population est importante, plus le réseau de proximité des personnes sera connecté et certainement avec de faibles distances entre les paires. Vérifions ces hypothèses en considérant les corrélations pouvant exister entre la taille de la population des composantes connexes et leur diamètre (cf. figure 3.4).

Les différentes composantes connexes sont considérées séparément, comme des graphes indépendants. Considérons d'abord uniquement les composantes connexes constituées de la majorité des effectifs de l'hôpital, représentées par les plus (+) en rouge sur la figure. On constate alors que lorsque les effectifs sont plus réduits, les diamètres, eux, sont plus grands (compris entre 8 et 10). Au contraire, lorsque les effectifs dépassent 160, mises à

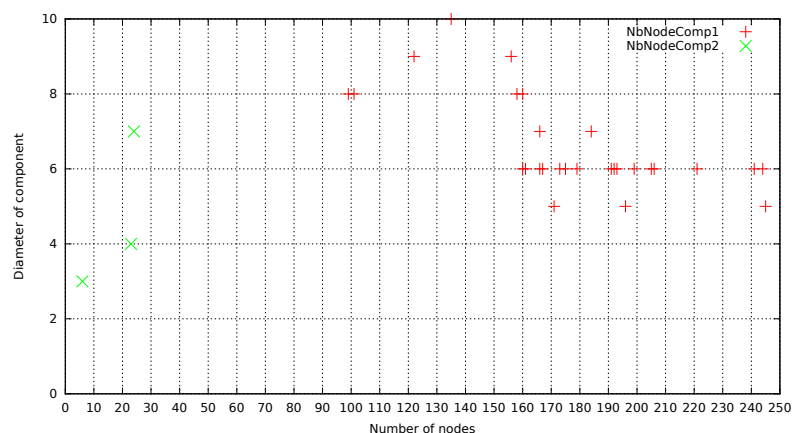


FIGURE 3.4 – Corrélation entre le diamètre journalier des composantes connexes du réseau de personnes de l’hôpital (en abscisses) et le nombre de nœuds qui les composent (en ordonnées.).

part deux exceptions, les diamètres observés sont toujours inférieurs ou égaux à 6. Les deux exceptions à cette règle sont des sous-réseaux d’un réseau global non connexe avec un nombre de nœuds faible (dans le sous-réseau comme dans le réseau global) car les seuls réseaux déconnectés ne contiennent pas plus de 130 nœuds, l’une des tailles de population les plus faibles observées. Suivant ces observations, plus le nombre de nœuds est important, plus le réseau est bien connecté en terme de composantes connexes et le diamètre est faible. A l’inverse, moins le réseau contient de nœuds, plus les liens sont clairsemés. Or, d’après la figure 3.1, nous avons remarqué que les seuls jours durant lesquels le nombre de personnes est plus faible que 160 correspondent à des jours de week-end du fait de la restriction de personnels. Cela implique encore une fois que les personnels jouent un grand rôle dans la connexion du réseau, et donc facilitent la diffusion au sein du réseau.

Ainsi, les diamètres étant grands pour une population durant les week-ends, les risques de transmission semblent plus réduits que les autres jours de la semaine, et ce, d’autant plus que la population est réduite. Non seulement, le nombre de personnes pouvant être infectées est plus faible, mais le réseau des personnes semble peu connecté ce qui réduit la vitesse d’une potentielle propagation. Au contraire, les jours de semaines, le réseau des personnes semble mieux connecté, avec des diamètres plus courts et une population plus importante ce qui présage une propagation plus rapide à travers le réseau.

### 3.1.2 Evolution de l’activité au cours des jours

Nous savons à présent que les réseaux journaliers sont pour la plupart connexes avec des diamètres variant de 5 à 10, ce qui ne prend pas en compte la quantité d’interactions qui se produit sur le réseau. Commençons donc par examiner les quantités d’interactions de l’ensemble des personnes de l’hôpital, sans se préoccuper de leur fonctions respectives. La figure 3.5 représente à gauche (3.5a) l’évolution du nombre de personnes dans l’hôpital, courbe qui a déjà été détaillée précédemment. La deuxième courbe (3.5b) représente les quantités d’activité sur la globalité de l’hôpital au cours des jours. Ainsi, on peut lire le nombre de paires différentes de contacts existant chaque jour (en rouge), le nombre de contacts exprimés en dizaines ainsi que la durée cumulée de ces contacts exprimée en heures.

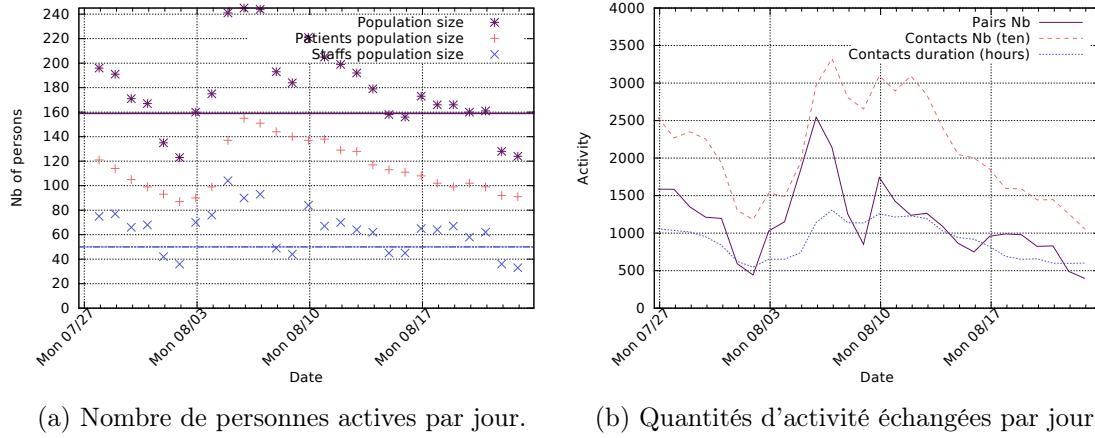


FIGURE 3.5 – Évolution à la journée du nombre de personnes en activité le jour (3.5a) et des quantités d'activité observées (3.5b). La figure 3.5b donne en violet le nombre de paires d'adjacence observées chaque jour, en rouge le nombre de contacts et en bleu la durée cumulée de tous les contacts.

On remarque que globalement, ces trois quantités suivent l'évolution de la taille de la population. En effet, plus il y a de personnes présentes, plus il existe de paires différentes, plus le nombre de contacts est important et plus leur durée cumulée est élevée. On a donc l'impression que l'activité est, à part quelques exceptions, uniquement liée au nombre de personnes présentes et que, par conséquent, du point de vue individuel, l'activité journalière est assez semblable.

Vérifions ces impressions, en observant ce qui se passe en moyenne par individu chaque jour. Nous nous intéressons à l'évolution des quantités d'activité au cours du temps pour la globalité de l'hôpital ainsi que l'évolution de ces quantités pour les deux grands types de personnes que sont les patients et les personnels. La figure 3.6 est ainsi composée de trois sous-figures qui représentent respectivement la moyenne par individu et par jour du degré des individus (3.6a), la moyenne du nombre de contacts par individu (3.6b) ainsi que la moyenne quotidienne des durées cumulées par individu (3.6c). Les quantités calculées pour le groupe des patients (resp. personnels) prennent en compte tous les contacts qui impliquent au moins un patient (resp. personnel). Chacune de ces sous-figures contient trois courbes : en violet, l'hôpital dans sa globalité ; en rouge, les patients et en bleu pointillé les personnels.

On observe que, pour chacune des quantités, les trois courbes (population de l'hôpital, patients, personnels) semblent avoir globalement une même évolution temporelle car leurs patterns sont proches, bien que suivant des échelles différentes. On remarquera que le nombre de paires impliquant au moins un personnel est beaucoup plus élevé que le nombre de paires impliquant des patients et ce quelque soit le jour observé. Au contraire, les quantités de contacts et durées de contacts journaliers des personnels sont beaucoup plus faibles que celles des patients. Les courbes concernant les patients ont toujours un pattern assez semblable, quelque soient les quantités d'activités dont il s'agisse, et suivent l'évolution de la taille de la population. Les courbes des personnels, mise à part celles concernant le degré moyen des personnes, sont beaucoup moins corrélées à la taille de la population, en particulier en terme de durées de contact. En effet, la durée cumulée moyenne des contacts d'un personnel varie peu au cours du temps et varie entre 2 à 4 heures de contacts par jour.



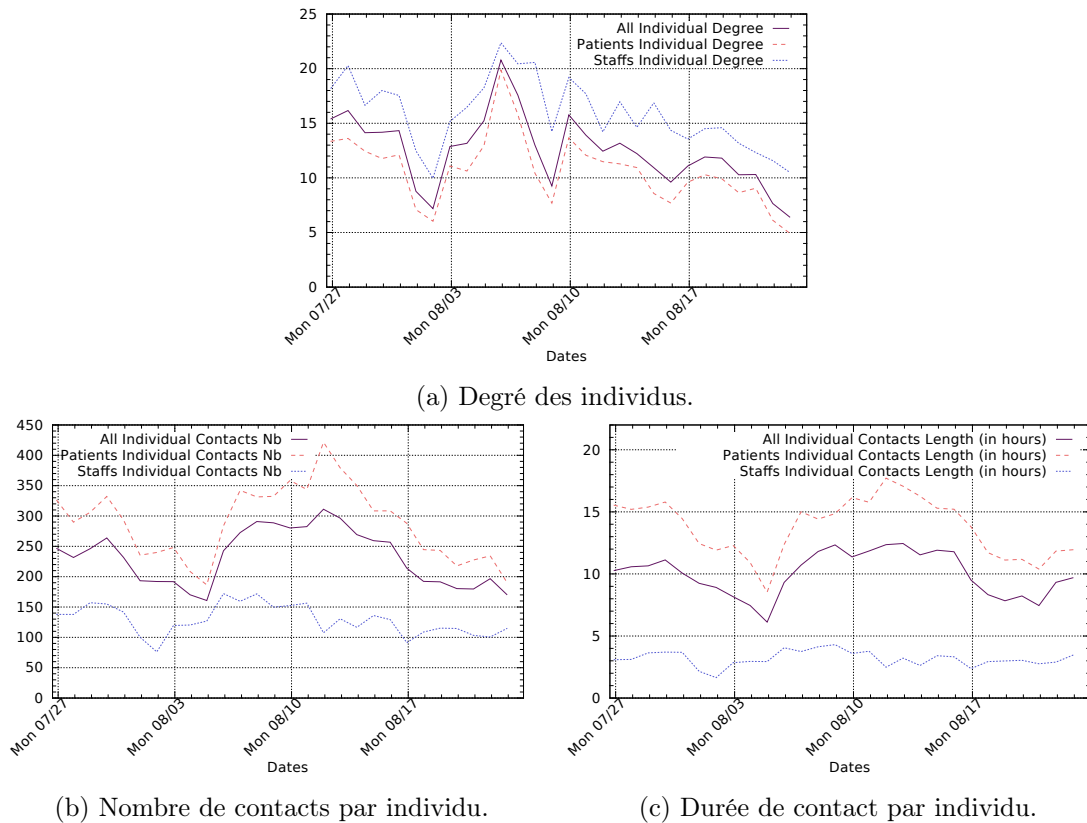


FIGURE 3.6 – Évolution des degrés quotidiens des personnes (3.6a), du nombre de contacts par individu (3.6b) ainsi que la durée individuelle de contacts quotidiens (3.6c) au cours d'un mois. La figure 3.6a représente le degré moyen par individu, sur l'ensemble de l'hôpital, pour les patients et pour les personnels.

On peut donc penser que l'activité des patients est beaucoup plus affectée par un changement d'effectif global que celle des personnels.

Si le volume d'activité individuel varie au cours du temps, qu'en est-il du volume d'activité par paires. La figure 3.7 donne le nombre moyen de paires par individu (degré des individus, en violet foncé) au cours du temps ainsi que le nombre moyen de contacts par paire (en rouge clair) et la durée cumulée moyenne par paire (en bleu pointillé).

Les quantités par paires impliquant au moins un personnel sont très stables tandis que celles des patients et par conséquent de l'hôpital dans sa globalité sont plus variables au cours du temps. La moyenne des durées cumulée des personnels ne dépasse jamais 20 minutes par jour alors que celle des patients varie de plus d'une heure en fonction des jours considérés. Cette variation est parfois inversement proportionnelle au degré moyen individuel (figure 3.7b, courbe violet foncé) ainsi qu'à la taille de la population (figure 3.5a). En particulier, le 2 août ainsi que tous les autres dimanches, on observe un pic passant de 100 à 120 minutes de contacts par paire alors que le degré moyen passe de 11 à 8 et la population se réduit de 92 patients à 87. Inversement, le lundi 3 août, la durée de contact par paire chute de près de 40 minutes par rapport au jour précédent alors que le degré des patients augmente tout comme la taille de la population. Donc plus la population est petite, plus la durée cumulée des contacts entre paires sera importante.

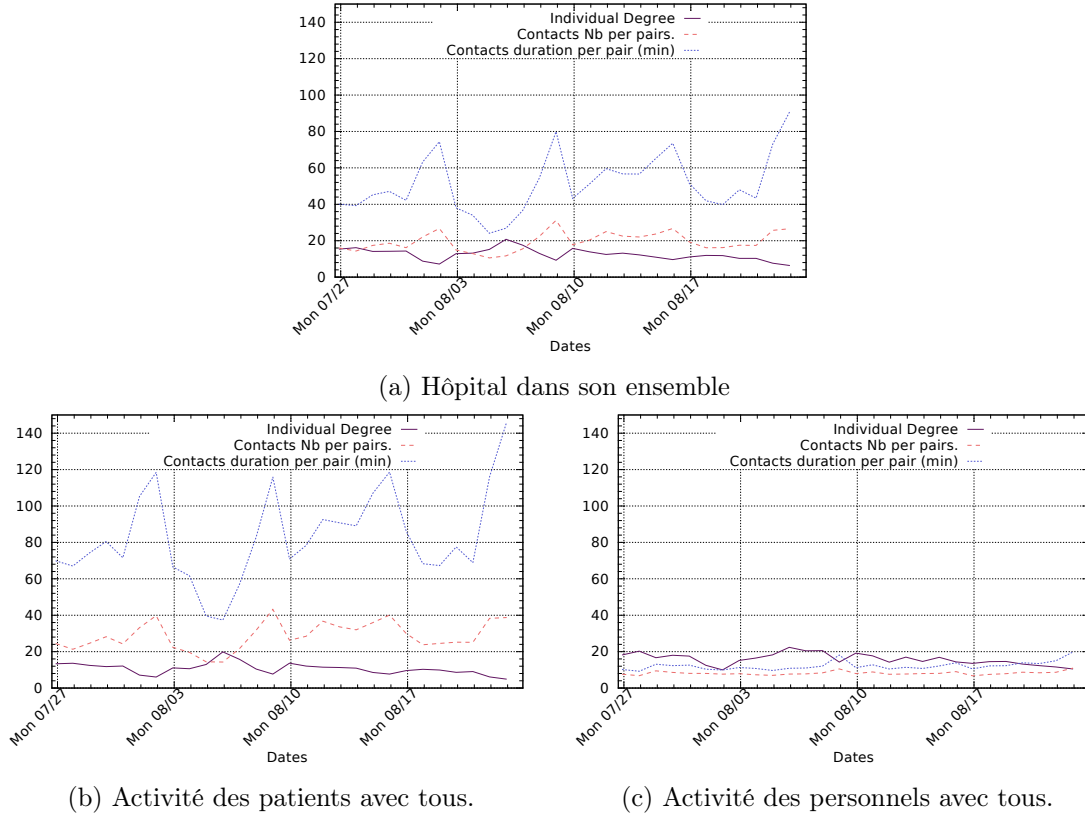


FIGURE 3.7 – Évolution et corrélation entre le degré des personnes et la durée de contacts par individu. La figure 3.7a donne le nombre degré moyen des personnes de tout l’hôpital (ou aussi paires par individu), la durée individuelle moyenne et la durée moyenne par paire par individu. La figure 3.7b donne ces trois quantités pour les patients. La figure 3.7c représente ces mêmes trois quantités pour les personnels de l’hôpital.

Nous avons remarqué jusqu’à présent que la taille de la population varie au cours du temps et que les quantités d’activité échangées suivent cette évolution. Pour autant, les quantités d’activité individuelles des personnels semblent très stables d’un jour à l’autre alors que celles des patients semblent assez variables.

Jusqu’à présent, nous nous sommes uniquement intéressés à des mesures assez basiques de quantification d’activité moyennées par individu. Les moyennes ne sont pourtant pas toujours représentatives et nous masquent un certain nombre d’informations : ainsi, nous avons par exemple l’impression que l’activité des personnels est très semblable d’un jour à l’autre. Nous allons voir, dans la partie suivante, que ce n’est pas tout à fait le cas, en introduisant des mesures de similarité quotidiennes d’activité individuelles.

### 3.1.3 Comment capturer les similarités ou les différences d’un jour à l’autre

Dans cette partie, nous considérons plusieurs mesures afin de connaître, d’un point de vue individuel, le taux de changement d’un jour à l’autre, ou inversement le taux de similarité d’un jour à l’autre. Étudier la similarité d’activité d’une période à l’autre est important pour décrire l’évolution d’un réseau temporel, afin d’être capable de détecter des caractéristiques

spécifiques au réseau [76]. Il existe un grand nombre de mesures destinées à quantifier la similarité de deux jeux de données [21]. Dans notre cas, nous nous intéressons à la mesure de similarité d'activité de deux jours consécutifs, d'abord calculée pour chaque personne, puis nous étudions la moyenne sur l'ensemble du premier jour des similarités individuelles obtenues par la comparaison de l'activité du jour suivant. Ceci est un moyen de quantifier les changements qui s'opèrent d'un jour à l'autre.

Dans la suite, nous noterons  $w_{ij}(t)$  le poids du lien entre les personnes  $i$  et  $j$ , au temps  $t$  du réseau agrégé au jour. Ce poids sera 1 (lien existant) ou 0 (aucun lien) quand il s'agira des paires de contacts, le nombre cumulé de contacts entre les deux individus pour jour considéré ou la durée cumulée entre les deux protagonistes au temps  $t$  (un jour donné). Le tableau 3.1 résume les différentes mesures de similarités que nous avons appliquées aux données MOSAR pour extraire des informations de similarités temporelles des données. Nous précisons, pour chacune, sur quel type de réseau, construit à partir des données MOSAR, nous les avons appliquées.

Types de similarité \ Types de réseaux	Paires	Durée de contact	# de contacts
Taux de conservation du jour 1	(3.1)		
Différence symétrique	(3.3) ou (3.2)	(3.3)	(3.3)
Indice de Jaccard	(3.4) ou (3.5)	(3.5)	(3.5)
Similarité cosinus traditionnelle	(3.7)	(3.7)	(3.7)
Activité	(3.8)	(3.8)	(3.8)
Similarité cosinus avec coefficient	(3.11)	(3.11)	(3.11)

TABLE 3.1 – Tableau récapitulant les différentes mesures de similarité et sur quels types de réseaux les applications sont justifiées.

Voici maintenant comment fonctionne chacune des mesures de similarités que nous avons considérées afin de trouver des caractéristiques temporelles stables et donc caractérisant le réseau d'activité contenu dans les données MOSAR.

### Taux de conservation

Commençons par le taux de conservation de l'activité de chacune des personnes entre un jour et un autre. En d'autres termes, nous nous intéressons ici à la proportion de voisins d'une personne  $i$  qui reste en contact avec  $i$  les deux jours considérés par la mesure. Cette mesure permet d'évaluer si la diversité de contact des personnes est constante d'un jour à l'autre. Or, connaître la diversité des contacts est important lors de l'étude des risques de diffusion au sein d'un réseau car elle correspond aux différentes sources susceptibles de transmettre.

Le taux de conservation du voisinage d'un nœud s'exprime par le rapport entre le nombre de voisins communs d'un nœud durant les deux jours comparés et le nombre de voisins de ce même nœud durant le premier jour considéré (cf. équation 3.1).

$$S_C(i, t_1, t_2) = \frac{|N(i, t_1) \cap N(i, t_2)|}{|N(i, t_1)|} \quad (3.1)$$

### Différence symétrique

La différence symétrique révèle le taux de changement d'un jour à l'autre entre les ensembles de voisins, c'est à dire le taux d'apparitions ou de disparitions de voisins d'un nœud. Ce taux de changement peut s'exprimer de manière ensembliste, lorsque le réseau n'est pas pondéré comme la différence symétrique entre les ensembles des voisins d'un nœud pour chacun des jours comparés (cf. équation 3.2). Nous avons choisi de pondérer cette différence par la taille de l'ensemble des voisins du premier jour afin d'avoir une proportion de différence et ainsi s'abstraire du nombre réel de personnes présentes pour avoir des valeurs comparables au cours du temps.

$$S_{DP}(i, t_1, t_2) = \frac{|N(i, t_1) \Delta N(i, t_2)|}{|N(i, t_1)|} \quad (3.2)$$

Nous pouvons aussi exprimer cette mesure de similarité, que nous appelons différence symétrique, appliquée sur des réseaux pondérés (cf. équation 3.3).

$$S_D(i, t_1, t_2) = \frac{\sum_{j \in N(i, t_1) \cup N(i, t_2)} |w_{ij}(t_2) - w_{ij}(t_1)|}{\sum_{j \in N(i, t_1)} w_{ij}(t_1)} \quad (3.3)$$

L'inconvénient de cette mesure est qu'elle donne des valeurs positives non bornées ce qui peut donner des résultats plus difficiles à interpréter que les mesures bornées entre 0 et 1.

### Indice de Jaccard

L'indice de Jaccard est une autre mesure destinée à évaluer la similarité de deux réseaux et est traditionnellement utilisée pour des réseaux pondérés. Cet indice évalue le taux d'activité commune des voisins d'un nœud durant les deux jours comparés. Cette mesure permet de détecter simultanément si le nombre de voisins entre les deux périodes est identique ainsi que si ces voisins sont différents ou non. La mesure de Jaccard appliquée aux réseaux non pondérés peut s'exprimer comme le rapport de l'intersection des ensembles des voisins des deux jours d'un nœud sur leur union (cf. équation 3.4).

$$S_{JP}(i, t_1, t_2) = \frac{|N(i, t_1) \cap N(i, t_2)|}{|N(i, t_1) \cup N(i, t_2)|} \quad (3.4)$$

Traditionnellement utilisée sur des réseaux pondérés, l'indice de Jaccard peut s'exprimer comme dans l'équation 3.5, en suivant les notations utilisées jusqu'à présent. Soit deux ensembles de valeurs de liens de même taille. A chaque voisin du nœud dont on veut connaître le taux de similarité d'activité entre deux jours, on associe deux valeurs : la valeur du lien le premier jour et celle associée au deuxième jour. Le premier ensemble va contenir, pour chaque voisin existant au moins un des deux jours, la valeur minimale obtenue parmi les deux jours comparés. Le deuxième ensemble de valeurs contient, quant à lui, les valeurs maximales. L'indice de Jaccard s'exprime simplement par le rapport de la somme des valeurs minimales (du premier ensemble) sur la somme des valeurs maximales

regroupées dans le deuxième ensemble (cf. équation 3.5).

$$\begin{aligned}
 S_J(i, t_1, t_2) &= \frac{\sum_{j \in N(i, t_1) \cup N(i, t_2)} \min(w_{ij}(t_1), w_{ij}(t_2))}{\sum_{j \in N(i, t_1) \cup N(i, t_2)} \max(w_{ij}(t_1), w_{ij}(t_2))} \\
 &= 1 - \frac{\sum_{j \in N(i, t_1) \cup N(i, t_2)} |w_{ij}(t_2) - w_{ij}(t_1)|}{\sum_{j \in N(i, t_1) \cup N(i, t_2)} \max(w_{ij}(t_1), w_{ij}(t_2))} \quad (3.5)
 \end{aligned}$$

### Similarité cosinus

Nous avons jusqu'à présent présenté des mesures qui mettent en avant des modifications de structure du voisinage ainsi que des changements globaux de volume en terme de répartition des contacts par voisin et de quantité de voisins. La mesure qui suit (3.7), souvent appelée similarité cosinus, n'indiquera que des modifications de structure des voisins ainsi que de leur nombre. Cela signifie que si l'activité du voisinage d'un nœud est exactement la même à un facteur multiplicatif près, alors cette mesure indiquera une similarité exacte entre les deux jours comparés. Cette mesure est donc intéressante car elle permet de s'abstraire du taux de volume global réparti entre les différents voisins pour s'intéresser uniquement aux changements de forme du réseau.

La similarité cosinus peut s'exprimer, dans un cas non pondéré, en utilisant une notation ensembliste, comme dans l'équation 3.6.

$$S_{CosP}(i, t_1, t_2) = \frac{|N(i, t_1) \cap N(i, t_2)|}{\sqrt{|N(i, t_1)| * |N(i, t_2)|}} \quad (3.6)$$

Généralisée à un réseau pondéré, la mesure de similarité cosinus peut s'écrire comme dans l'équation 3.7.

$$S_{Cos}(i, t_1, t_2) = \frac{\sum_{j \in N(i, t_1) \cap N(i, t_2)} w_{ij}(t_1) * w_{ij}(t_2)}{\sqrt{\sum_{j \in N(i, t_1)} w_{ij}(t_1)^2 * \sum_{j \in N(i, t_2)} w_{ij}(t_2)^2}} \quad (3.7)$$

Nous disposons ainsi d'une mesure permettant de détecter uniquement un changement dans la répartition relative des contacts entre les individus indépendamment du volume total des contacts.

### Différence d'activité

La mesure de différence d'activité d'une personne d'un jour à l'autre permet, elle, de s'abstraire de la différence de structure pour ne plus s'intéresser qu'à la différence d'activité globale de cette personne avec son voisinage. Dans le cas pondéré, nous exprimons traditionnellement cette différence comme la racine du quotient des normes euclidiennes des

deux vecteurs représentant l'activité d'une personne  $i$ .

$$S_A(i, t_1, t_2) = \frac{\sqrt{\sum_{j \in N(i, t_2)} w_{ij}(t_2)^2}}{\sqrt{\sum_{j \in N(i, t_1)} w_{ij}(t_1)^2}} \quad (3.8)$$

Dans le cas non pondéré, seule l'existence du lien est exprimée, dans une seule dimension. Nous préférons de ce fait utiliser la norme 1, ce qui donne l'équation 3.9 qui suit .

$$S_{AP}(i, t_1, t_2) = \frac{|N(i, t_2)|}{|N(i, t_1)|} \quad (3.9)$$

L'inconvénient de cette mesure est qu'elle donne des valeurs positives non bornées ce qui peut donner des résultats plus difficiles à interpréter que les mesures bornées entre 0 et 1.

Enfin, nous proposons une mesure dont le but est de détecter à égalité des changements de structure et de répartition proportionnelle du voisinage d'un côté et de changements globaux de volume d'activité de l'autre. Cette mesure est construite à partir de la similarité cosinus et d'un coefficient compris dans l'intervalle  $[0 : 1]$  permettant de détecter un changement global de volume dans l'activité du voisinage d'un nœud.

### Séparation des termes de changements structurels et volumiques.

La notation ensembliste de cette nouvelle mesure est décrite par l'équation 3.10.

$$\begin{aligned} S_{CRP}(i, t_1, t_2) &= \frac{|N(i, t_1) \cap N(i, t_2)|}{\sqrt{|N(i, t_1)| * |N(i, t_2)|}} * \frac{(|N(i, t_1)| * |N(i, t_2)|)}{|N(i, t_1) \cap N(i, t_2)| + \frac{N(i, t_1) \Delta N(i, t_2)^2}{2}} \\ &= |N(i, t_1) \cap N(i, t_2)| * \frac{\sqrt{|N(i, t_1)| * |N(i, t_2)|}}{\left(|N(i, t_1) \cap N(i, t_2)| + \frac{N(i, t_1) \Delta N(i, t_2)^2}{2}\right)^2} \end{aligned} \quad (3.10)$$

L'expression de cette mesure sur des réseaux pondérés se trouve dans l'équation 3.11.

$$S_{CR}(i, t_1, t_2) = \frac{\sum_{j \in N(i, t_1) \cap N(i, t_2)} w_{ij}(t_1) * w_{ij}(t_2)}{\sqrt{\sum_{j \in N(i, t_1)} w_{ij}(t_1)^2 * \sum_{j \in N(i, t_2)} w_{ij}(t_2)^2}} * \frac{\sum_{j \in N(i, t_1)} w_{ij}(t_1) * \sum_{j \in N(i, t_2)} w_{ij}(t_2)}{\left(\sum_{j \in N(i, t_1) \cup N(i, t_2)} \frac{w_{ij}(t_1) + w_{ij}(t_2)}{2}\right)^2} \quad (3.11)$$

### Etude des mesures de similarité sur le réseau journalier des paires de contact.

Voici quelques exemples simples illustrant les différences entre les différentes mesures de similarité, regroupés dans le tableau 3.2.

Les trois exemples résumés dans ce tableau s'intéressent aux valeurs de similarité d'une personne A ayant B et C pour voisin le premier jour  $t_1$  et un autre ensemble de voisins le deuxième jour  $t_2$ . Les trois ensembles choisis pour le deuxième jour sont les suivants :

- les mêmes voisins que le premier jour plus deux nouveaux :  $\{B, C, D, E\}$ . Le volume d'activité est donc doublé.
- deux voisins différents du premier jour :  $\{D, E\}$ . Il y a donc conservation du volume d'activité mais pas du pattern des contacts.
- un voisin est conservé, un autre est nouveau :  $\{B, D\}$ . Il y a encore une fois conservation du volume, mais le pattern est modifié de moitié.

On rappelle que toutes les mesures valent 1 quand il n'y a aucun changement, mis à part les mesures non bornées et le taux d'apparition et de disparition (3.3) qui vaut 0 quand il n'y a aucune modification entre les ensembles des deux jours. On remarque que la seule mesure de similarité qui ne permette pas de détecter un changement entre les deux jours dans la premier cas est celle du taux de conservation. La mesure d'activité ne détecte elle, aucun changement lorsque les voisins de A sont aussi nombreux d'un jour à l'autre mais sont différents, comme nous l'avons annoncé lors de la présentation des différentes mesures.

Toutes les figures qui suivent dans cette partie sont obtenues en calculant les similarités entre l'activité d'un jour  $t_1$  et d'un jour  $t_2$  pour chaque individu présent le jour  $t_1$ . L'ensemble des valeurs obtenues est ensuite moyenné par la taille de l'ensemble des individus présents le jour  $t_1$ , en suivant la formule 3.12. Nous appelons  $S_X$  la variable qui nous indique de quelle mesure de similarité il s'agit.

$$S_{X_{D1}}(t_1, t_2) = \frac{\sum_{i \in V(t_1)} S_X(i, t_1, t_2)}{|V(t_1)|} \quad (3.12)$$

Dans certains cas, à titre de comparaison, nous utiliserons aussi la moyenne s'appuyant sur la taille de l'ensemble des individus présents obligatoirement les deux jours comparés.

$$sim_{X_{D1 \cap D2}}(t_1, t_2) = \frac{\sum_{i \in V(t_1) \cap V(t_2)} sim_x(i, t_1, t_2)}{|V(t_1) \cap V(t_2)|} \quad (3.13)$$

Types de similarité \ Type de contact	$N(A, t_1) = \{B, C\}$ $N(A, t_2) = \{B, C, D, E\}$	$N(A, t_1) = \{B, C\}$ $N(A, t_2) = \{D, E\}$	$N(A, t_1) = \{B, C\}$ $N(A, t_2) = \{B, D\}$
Taux de conservation 3.1	$\frac{ \{B, C\} \cap \{B, C, D, E\} }{ \{B, C\} } = 1$	$\frac{ \{B, C\} \cap \{D, E\} }{ \{B, C\} } = 0$	$\frac{ \{B, C\} \cap \{B, D\} }{ \{B, C\} } = 0.5$
Différence symétrique 3.2	$\frac{ \{B, C\} \Delta \{B, C, D, E\} }{ \{B, C\} } = 1$	$\frac{ \{B, C\} \Delta \{D, E\} }{ \{B, C\} } = 2$	$\frac{ \{B, C\} \Delta \{B, D\} }{ \{B, C\} } = 1$
Indice de Jaccard (3.4)	$\frac{ \{B, C\} \cap \{B, C, D, E\} }{ \{B, C\} \cup \{B, C, D, E\} } = 0.5$	$\frac{ \{B, C\} \cap \{D, E\} }{ \{B, C\} \cup \{D, E\} } = 0$	$\frac{ \{B, C\} \cap \{B, D\} }{ \{B, C\} \cup \{B, D\} } = 1/3$
Similarité cosinus (3.7)	$\frac{ \{B, C\} \cap \{B, C, D, E\} }{\sqrt{ \{B, C\}  *  \{B, C, D, E\} }} = \frac{1}{\sqrt{2}} = 0.7$	$\frac{ \{B, C\} \cap \{D, E\} }{\sqrt{ \{B, C\}  *  \{D, E\} }} = 0$	$\frac{ \{B, C\} \cap \{B, D\} }{\sqrt{ \{B, C\}  *  \{B, D\} }} = 0.5$
Activité (3.8)	$\frac{ \{B, C, D, E\} }{ \{B, C\} } = \frac{4}{2} = 2$	$\frac{ \{D, E\} }{ \{B, C\} } = 1$	$\frac{ \{B, D\} }{ \{B, C\} } = 1$
Similarité cosinus renormée (3.11)	$S_{CRP}(A, t_1, t_2) = \frac{4\sqrt{2}}{9} = 0.6$	$S_{CRP}(A, t_1, t_2) = 0$	$S_{CRP}(A, t_1, t_2) = 1/2$

TABLE 3.2 – Trois exemples illustrant le fonctionnement des différentes mesures de similarités appliquées aux paires de contacts. Les calculs sont faits pour la personne A.

Nous appliquons maintenant les différentes mesures de similarité moyennées par individu au cours d'un mois sur les données MOSAR, collectées en milieu hospitalier, en utilisant d'abord les réseaux non pondérés.

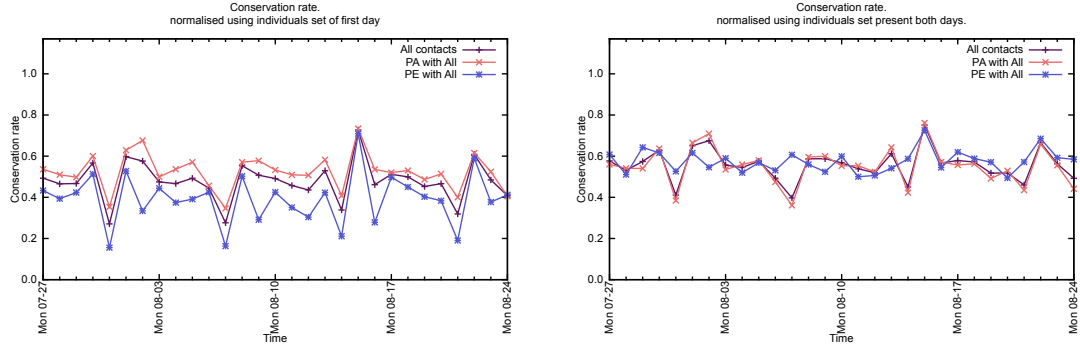
Les courbes violettes (+) s'appuient toujours sur l'ensemble des contacts. Les courbes rouges (X) utilisent uniquement les contacts impliquant au moins un patient. Les courbes bleues (\*) sont basées uniquement sur les contacts impliquant au moins un personnel de l'hôpital.

### Changement de pattern et volume.

La figure 3.8a présente trois courbes donnant les valeurs moyennes de similarité des contacts des personnes en terme de voisinage, en utilisant la mesure de taux de conservation.

On repère immédiatement le début des week-ends qui donne, pour les vendredis une valeur de similarité avec le samedi très faible comparée aux autres jours de la semaine. Au contraire, le samedi est très semblable au dimanche. On remarque que la courbe des personnels est toujours inférieure à la courbe des patients ce qui révèle une plus forte instabilité de leur activité. Cette forte variabilité est due au fort taux d'absence d'un jour à l'autre des personnels dont l'emploi du temps n'est pas régulier et ne requiert pas leur présence systématique tous les jours. En effet, considérons la courbe 3.8b, qui présente les moyennes de taux de conservation par individu établies sur l'ensemble des individus présents les deux jours uniquement (cf. eq. 3.13).

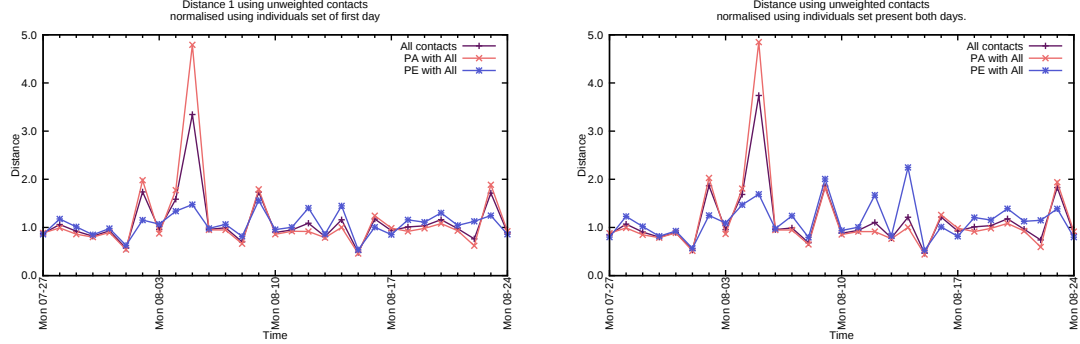




(a) Taux de conservation des paires entre jours consécutifs (équation 3.1) moyenné sur le nombre de personnes présentes le jour 1. (b) Taux de conservation des paires entre jours consécutifs (équation 3.1) moyenné sur l'ensemble des personnes présentes les deux jours comparés.

FIGURE 3.8 – Taux de conservation des paires entre jours consécutifs (équation 3.1)

On remarque que les taux moyens de conservation sont plus élevés, supérieurs à 0.5 pour la plupart. Dans ce cas, les personnels, moins nombreux obtiennent des taux moyens de conservation beaucoup plus stables que ceux des patients dont les taux de conservation sont beaucoup plus sensibles aux absences des personnels. Finalement, les personnels, au regard de la mesure du taux de conservation quotidien, ont une variété de contacts moins importante que les patients, ce qui n'était pas évident en s'appuyant sur la figure 3.8a.



(a) Différence symétrique sur les paires (eq. 3.2), (b) Différence symétrique sur les paires moyennée sur l'ensemble des personnes présentes le jour 1. (c) Différence symétrique sur les paires moyennée sur l'ensemble des personnes présentes le jour 1 et le jour 2 (intersection des personnes présentes les deux jours) (eq. 3.3).

FIGURE 3.9 – Différence symétrique sur les paires (eq. 3.3).

Intéressons nous maintenant aux valeurs moyennes des similarités symétriques individuelles au cours d'un mois (figure 3.9). On remarque quelques pics particuliers, affectant certains jours les similarités moyennes individuelles, quels que soient les ensembles d'individus considérés (globalité de l'hôpital, patients ou personnels). Le plus gros pic, le 5 août, correspond à une arrivée massive de patients ainsi qu'à une augmentation du personnel et donc à un gros changement de taille de population. Le cycle de la semaine n'est pas particulièrement remarquable. Aussi, contrairement au taux de conservation, les différences entre les moyennes des similarités symétriques des patients et celles des personnels sont beaucoup moins marquées et non constantes. En effet, parfois les personnels sont considérés comme ayant une activité moins semblable d'un jour à l'autre à la différence des patients

qui ont une activité similaire entre jours consécutifs. Nous observons parfois l'inverse, pour d'autres jours. La différence symétrique appliquée uniquement aux personnes présentes les deux jours (figure 3.9) met un peu plus en avant quelques jours particuliers en ce qui concerne l'activité des personnels, autour du week-end du 15 août.

L'évolution des moyennes des similarités selon l'indice de Jaccard, appliquée sur les paires de contacts (figure 3.10), fait apparaître un pattern hebdomadaire très régulier. En effet, on constate que les week-ends ressortent nettement par rapport aux autres jours de la semaine par leur forme caractéristique. Ainsi, l'indice moyen de Jaccard le vendredi chute considérablement alors que celui du samedi est très haut, quelque soit la population considérée (la globalité de l'hôpital, uniquement l'activité des personnels ou seulement celle des patients.) L'indice moyen de Jaccard est toujours plus faible pour les personnels que pour les patients, toujours à cause de leurs absences plus fréquentes.

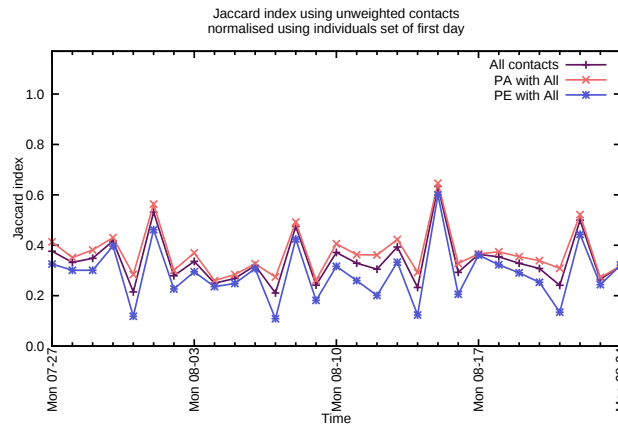


FIGURE 3.10 – Indice de Jaccard sur les paires (eq. 3.4).

### Pattern uniquement

La figure 3.11 nous informe sur les variations des moyennes de similarité cosinus au cours du mois. Nous remarquons avec étonnement que l'évolution de ces valeurs est exactement la même que lors de l'utilisation de l'indice de Jaccard pourtant sensible aux variations globales de volume. Les valeurs sont légèrement plus élevées que celles obtenues avec l'indice de Jaccard. Leur amplitude est par contre presque identique et le pattern global est conservé. Ici encore, les week-ends apparaissent donc nettement.

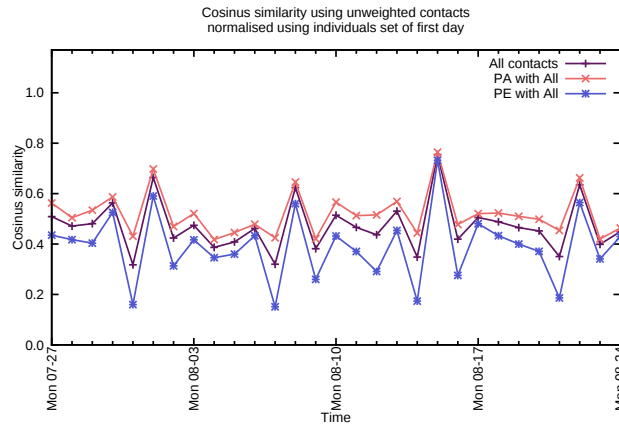


FIGURE 3.11 – Similarité cosinus sur les paires (eq. 3.6).

### Changement de volume global.

La figure 3.12 présente les variations de volume moyen au cours du mois. Plus la moyenne est proche de 1, plus les deux jours considérés sont similaires en terme de degré individuel. Les valeurs positives indiquent que le jour suivant sera plus actif que le jour que l'on considère. Au contraire, les valeurs négatives dénotent une activité moins importante pour le lendemain. On constate encore une fois que les valeurs obtenues indiquent une plus grande différence d'activité pour les personnels que pour les patients. En effet, les valeurs des personnels sont souvent fortement éloignées de 1 contrairement aux patients dont seulement quelques valeurs, à intervalle régulier, s'écartent de 1. Les vendredis, les valeurs des patients sont, pour 3 d'entre elles nettement en-dessous de 1, ce qui signifie que le samedi s'annonce moins actif que le vendredi. Au contraire, les dimanches ont des valeurs supérieures à 1, ce qui signifie que le lundi sera plus actif. On remarque tout de même des valeurs moins éloignées de 1 pour le week-end du 15 août qui semble affecté par le jour férié. En effet, cela se retrouve aussi sur la courbe des personnels qui subit chaque vendredi une forte décroissance, invisible le week-end du 15 août, mais apparaissant plutôt le dimanche, ce qui peut s'expliquer par un fort renouvellement de l'équipe du personnel. Le comportement particulier autour du 5 août, dû à la très forte augmentation de population, reste visible avec cette mesure de similarité.

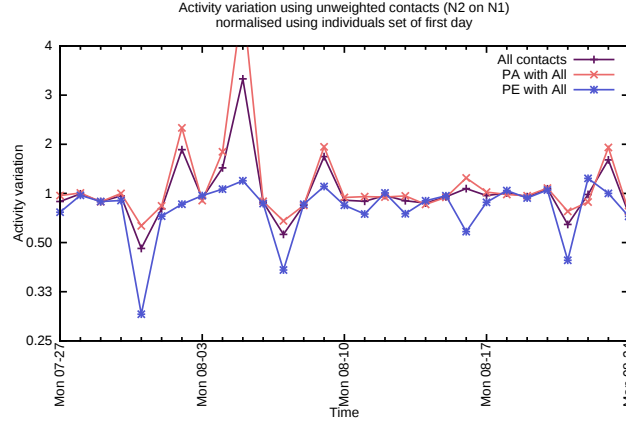


FIGURE 3.12 – Différence d'activité des paires moyennées sur l'ensemble des personnes présentes le jour 1 (eq. 3.9).

### Séparation des termes de changements structuraux et volumiques.

La figure 3.13 montre l'évolution des valeurs moyennes de similarité cosinus dans laquelle nous avons ajouté un terme pour y intégrer une information sur l'évolution globale de volume. On remarque que le résultat obtenu est très proche de ceux obtenus avec la similarité cosinus, ce qui était voulu, par construction, pour le cas non pondéré.

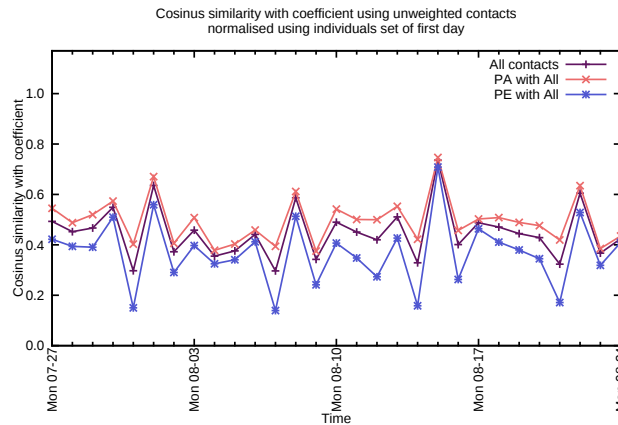


FIGURE 3.13 – Similarité cosinus renormé sur les paires (eq. 3.10).

L'étude des différentes mesures de similarité pour des périodes de temps consécutives, appliquée au réseau non pondéré journalier a permis de révéler, pour certaines mesures, un pattern hebdomadaire très prononcé. D'autres mesures de similarité, telle que la différence symétrique, ne nous ont pas apporté beaucoup plus d'informations et nous laissent supposer que l'activité des personnes est toujours assez similaires d'un jour à l'autre, sans distinguer de jour particulier. En effet, pour des réseaux aussi peu denses, les valeurs proches de 50% de similarité sont élevées par comparaisons avec ce que l'on obtiendrait en répartissant les liens aléatoirement.

### 3.1.4 Similarité appliquée au réseau pondéré.

Dans la suite, nous nous intéressons maintenant à l'évolution de valeurs des mesures de similarité sur les réseaux journaliers pondérés, soit en terme de nombre de contacts, soit en terme de durées cumulées des contacts entre paires d'individus. Comme précédemment, nous commençons par nous intéresser à quelques exemples simples afin de mieux appréhender les résultats sur les données MOSAR par la suite. Voici donc trois exemples simples pour illustrer le fonctionnement des différentes mesures de similarité appliquées à une personne et son voisinage. Les trois exemples sont regroupés dans le tableau 3.3. La première ligne de ce tableau décrit l'ensemble des voisins de A au temps  $t_1$ , avec le poids de leur interaction. La deuxième ligne de ce tableau nous décrit l'ensemble des voisins de A au temps  $t_2$ , toujours accompagnés du poids des interactions.

Types de similarité \ Type de contact	$N(A, t_1) = \{B : 2, C : 2\}$ $N(A, t_2) = \{B : 1, C : 1, D : 1, E : 1\}$	$N(A, t_1) = \{B : 2, C : 2\}$ $N(A, t_2) = \{D : 2, E : 2\}$	$N(A, t_1) = \{B : 2, C : 2\}$ $N(A, t_2) = \{B : 1, C : 1\}$
Différence symétrique 3.2	1	4	$\frac{1}{2}$
Indice de Jaccard (3.4)	$\frac{1}{3}$	0	$\frac{1}{2}$
Similarité cosinus (3.7)	$\frac{1}{\sqrt{2}}$	0	1
Activité (3.8)	1	1	$\frac{1}{2}$
Similarité cosinus renormé (3.11)	$\frac{1}{\sqrt{2}} = 0.7$	0	$8/9 = 0.88$

TABLE 3.3 – Trois exemples illustrant le fonctionnement des différentes mesures de similarités appliquées aux réseaux pondérés. Les calculs sont faits pour la personne A.

Le tableau 3.3 présente un premier exemple (colonne 2) qui conserve la quantité d'activité mais la répartie sur le double de personnes au deuxième temps. Dans ce cas, il est évident que la mesure d'activité ne permettra de détecter aucun changement. Nous savons aussi que les mesures de similarité cosinus et la similarité que nous proposons donneront la même valeur car le seul coefficient qui les différencie vaut 1 dans ce cas précis. Ces deux mesures mettent en avant la différence de structure du premier réseau. Les autres mesures (indice de Jaccard et différence symétrique) soulignent aussi une différence entre les deux pas de temps. Concernant le deuxième exemple, qui conserve exactement la quantité d'activité mais la répartie sur des nœuds totalement différents, quatre des mesures indiquent que la structure du réseau est totalement différente entre les deux pas de temps (la différence symétrique, l'indice de Jaccard, la similarité cosinus et la similarité que nous proposons.). Notre nouvelle mesure permet, en effet, d'être sensible au changement de structure et de quantité d'activité à égalité uniquement dans le cas où les deux éléments ne sont pas totalement différents. Dans le troisième cas, nous conservons le même réseau d'un temps

à l'autre, mais nous divisons par 2 la quantité d'activité. Dans ce cas, seule la similarité cosinus ne détecte aucun changement.

L'ensemble des mesures de similarité journalière appliquées aux réseaux pondérés des données MOSAR est présenté dans la suite (figure 3.14 à 3.18). Les résultats obtenus concernant les personnels sont très similaires à ceux obtenus pour les réseaux non pondérés, en particulier en terme de pattern, car les amplitudes s'accroissent. Ceci n'est pas très surprenant car nous avons remarqué que les contacts des personnels sont très limités et très courts. En revanche, les courbes obtenues pour l'activité des patients sont moins marquées, plus lisses avec de plus faibles amplitudes entre jours consécutifs et les valeurs obtenues sont nettement plus élevées que celles des personnels. Cela signifie que, contrairement au cas non pondéré, les patients ont une activité moins différente d'un jour à l'autre que les personnels. En outre, ces différences d'activité sont presque constantes et très peu influencées par les variations de taille de population hebdomadaire. Le comportement des patients est donc certainement plus facile à modéliser que celui des personnels qui paraît très chaotique et qui marque un cycle hebdomadaire. On remarque aussi que les similarités cosinus et celles basées sur l'indice de Jaccard ont les mêmes patterns mais que, contrairement au cas non pondéré, les amplitudes obtenues passent du simple au double. En effet, l'indice de Jaccard montre une moins grande variation de similarité journalière que la similarité cosinus, principalement sensible aux changements de structure du réseau. Cela signifie que les modifications structurelles sont plus ou moins compensées par des paires de contacts identiques d'un jour à l'autre et qui ont une durée de contact très conséquente.

### Changement structurel et volume global.

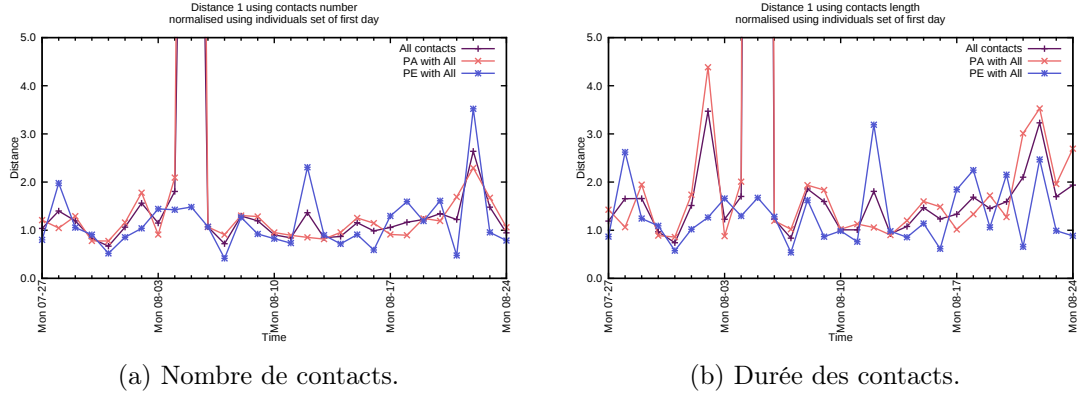


FIGURE 3.14 – Différence asymétrique sur le réseau pondéré.

### Changement de voisins et du volume relatif entre voisins.

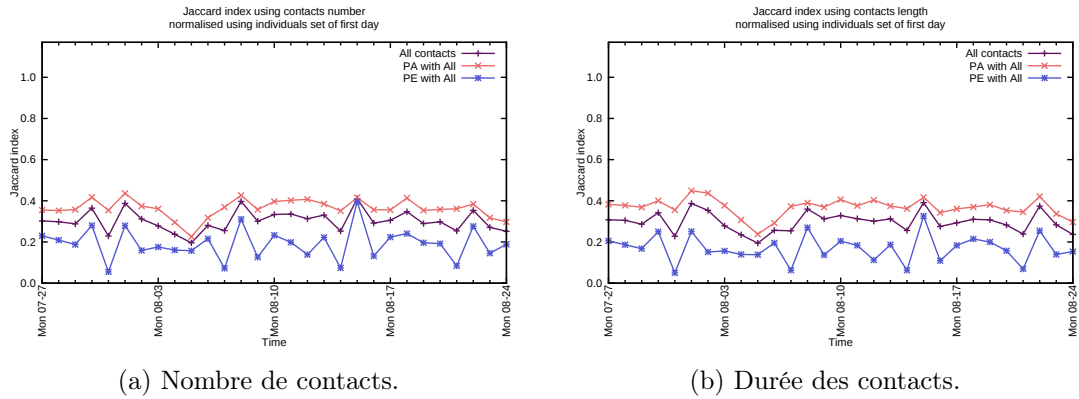


FIGURE 3.15 – Indice de Jaccard sur le réseau pondéré.

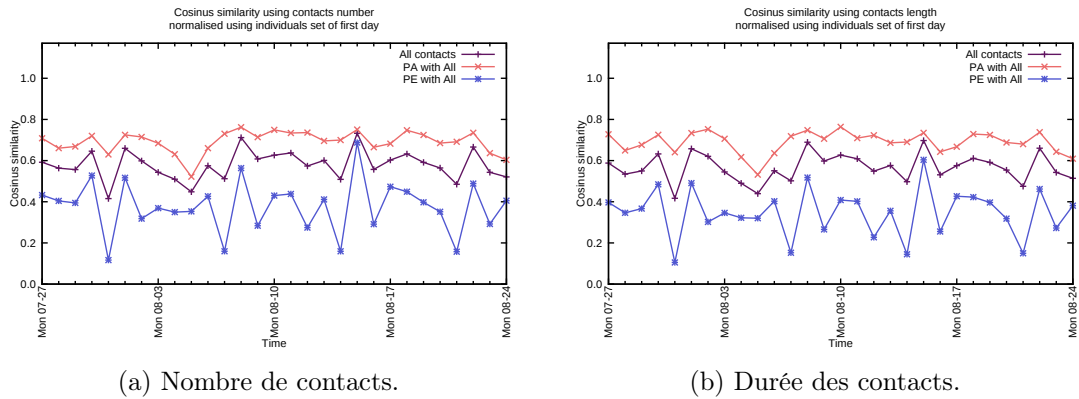


FIGURE 3.16 – Similarité cosinus sur le réseau pondéré.

### Changement de volume global.

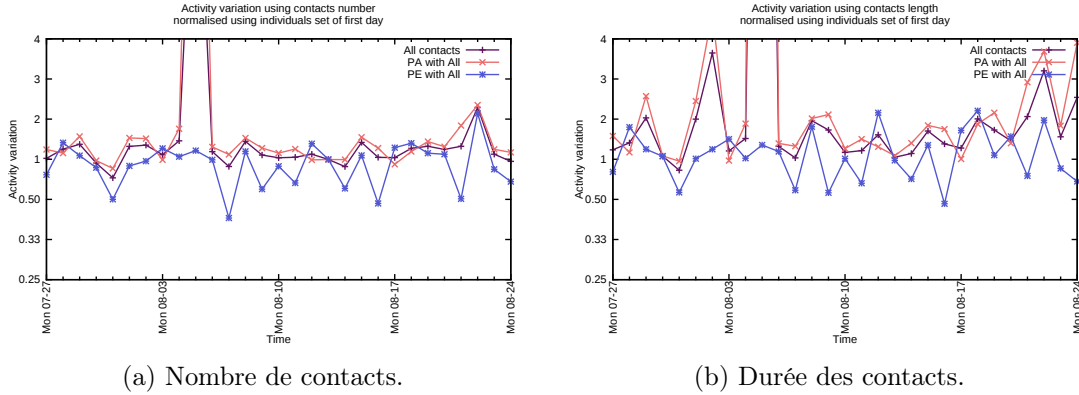


FIGURE 3.17 – Différence d'activité sur le réseau pondéré.

### Séparation des termes de changements structuraux et volumiques.

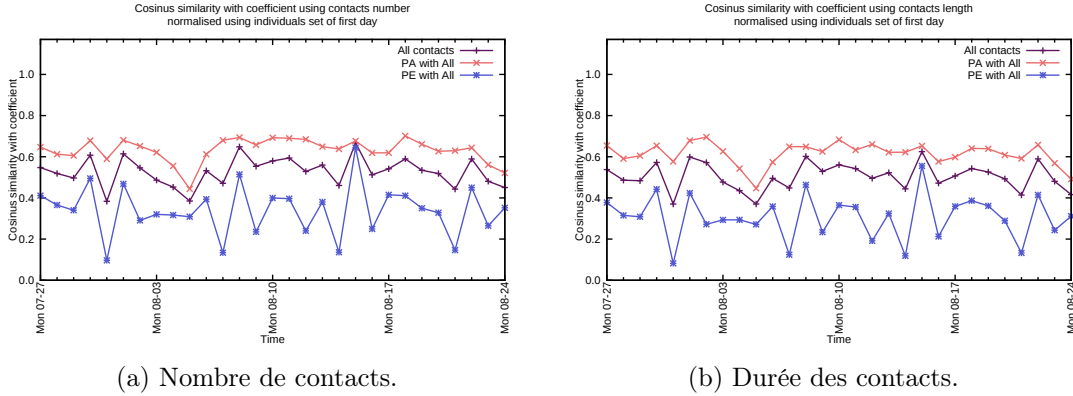


FIGURE 3.18 – Similarité cosinus renormé sur les paires.

Jusqu'à présent, nous nous sommes principalement intéressés à l'évolution quotidienne de l'activité de l'hôpital. Nous avons pu remarquer des comportements différents durant les week-ends, en terme de quantité d'activité. Nous avons aussi retenu que l'activité des personnels est constitué de beaucoup de contacts avec des personnes différentes, mais très courts et peu répétés, contrairement aux patients qui ont plutôt des degrés faibles, mais des contacts plutôt long. Nous savons aussi que les modifications de comportement d'un jour à l'autre sont assez importantes que ce soit en terme de paires ou de quantité d'activité. Ces différences ne sont pas toujours de même intensité pour les personnels, d'un jour à l'autre alors qu'elles sont presque constantes pour les patients en terme de quantité d'activité. Qu'en est-il à une échelle plus détaillée ? La section 3.2 traite de l'évolution des quantités d'activité ayant lieu au cours de heures.

## 3.2 Analyse à l'heure

Nous nous intéressons, dans cette partie, à l'évolution de la quantité d'activité par heure sur l'ensemble de l'hôpital. Dans toute cette partie, nous divisons le temps en intervalles d'une heure. Tout contact chevauchant deux intervalles distincts est coupé en deux (cf. figure 3.19).



(b) Quantité d'activité observée.

On retrouve le rythme circadien de nombreuses activités humaines, sur la figure 3.20a. En effet, le nombre d'individus actifs est beaucoup plus élevé entre 6h00 et 24h00, avec un pic remarquable autour de midi. Ce nombre d'individus actifs la nuit entre minuit et 6h00 est très faible. On peut aussi remarquer, encore une fois, un pattern hebdomadaire évident, indiqué par un nombre plus restreint d'individus actifs durant les week-ends, tout au cours de la journée. Cela signifie que l'ensemble de la journée est affectée en même proportion par une réduction de la population.

On distingue, sur la figure 3.20a les patients actifs des personnels actifs. Ces deux sous-groupes sont tous les deux régis par les mêmes rythmes circadien et hebdomadaire.

Néanmoins, la variation du nombre de personnels actifs est beaucoup plus importante que la variation du nombre de patients actifs au même moment de la journée. L'une des raisons qui peut être responsable de ces observations est le fait que les patients restent la journée ainsi que la nuit entière à l'intérieur de l'hôpital ce qui n'est pas le cas des personnels qui sont beaucoup moins nombreux la nuit (seul le service de nuit S6 est présent). Nous verrons plus loin qu'une autre raison explique cette différence entre le pattern des patients et celui des personnels.

Considérons l'évolution des quantités d'activité (nombre de paires, nombre de contacts et durée cumulée des contacts) au cours des heures pour la population globale (figure 3.20b). Ces trois quantités suivent le même pattern signalé précédemment concernant le nombre de personnes actives, ce qui indique une solide structure temporelle de l'activité de l'hôpital.

La figure 3.21 donne l'évolution au cours des heures du nombre de contacts par individu, de la durée cumulée moyenne de contacts par individu ainsi que la durée de chacun des contacts.

Les trois sous-figures de la figure 3.21 permettent d'avoir une vue d'ensemble de l'hôpital (figure 3.21a) ainsi que l'évolution de l'activité des patients (figure 3.21b) et celle des personnels (figure 3.21c). On rappelle que l'activité des patients (resp. personnels) prend en compte tous les contacts qui impliquent au moins un patient sur les deux individus en contact (resp. personnel).

Sur les trois sous-figures de la figure 3.21, on retrouve le rythme circadien déjà remarqué précédemment, mais le cycle hebdomadaire n'est plus identifiable. En effet, durant les week-ends, les valeurs moyennes par individu sont assez équivalentes à celles observées durant la semaine de travail.

Considérons d'abord la figure 3.21a, concernant l'ensemble de l'hôpital. On remarque que le nombre de contacts moyens par individu varie majoritairement entre 10 et 15 contacts par individu par heure. Les variations du nombre de contacts par individu sont très spécifiques en fonction du moment de la journée. En effet, on observe une hausse brutale du nombre de contacts moyens par individu entre 8h00 et 14h00, chaque jour, et des valeurs plus faibles le reste de la journée. On observe au contraire que la durée moyenne de contacts par individu est très importante la nuit entre minuit et 6h00 (située entre 40 à 50 minutes par heure) alors qu'elle est beaucoup plus faible dans la journée (entre 20 et 35 minutes par heure) ce qui est surprenant à première vue. Pour l'expliquer, considérons non pas la durée moyenne des contacts individuels, mais la durée moyenne de chacun des contacts, non agrégés, ni en temps, ni en paires. On constate alors que la durée moyenne d'un contact est plus longue la nuit de quelques minutes. On notera que la moyenne de la durée de ces contacts se

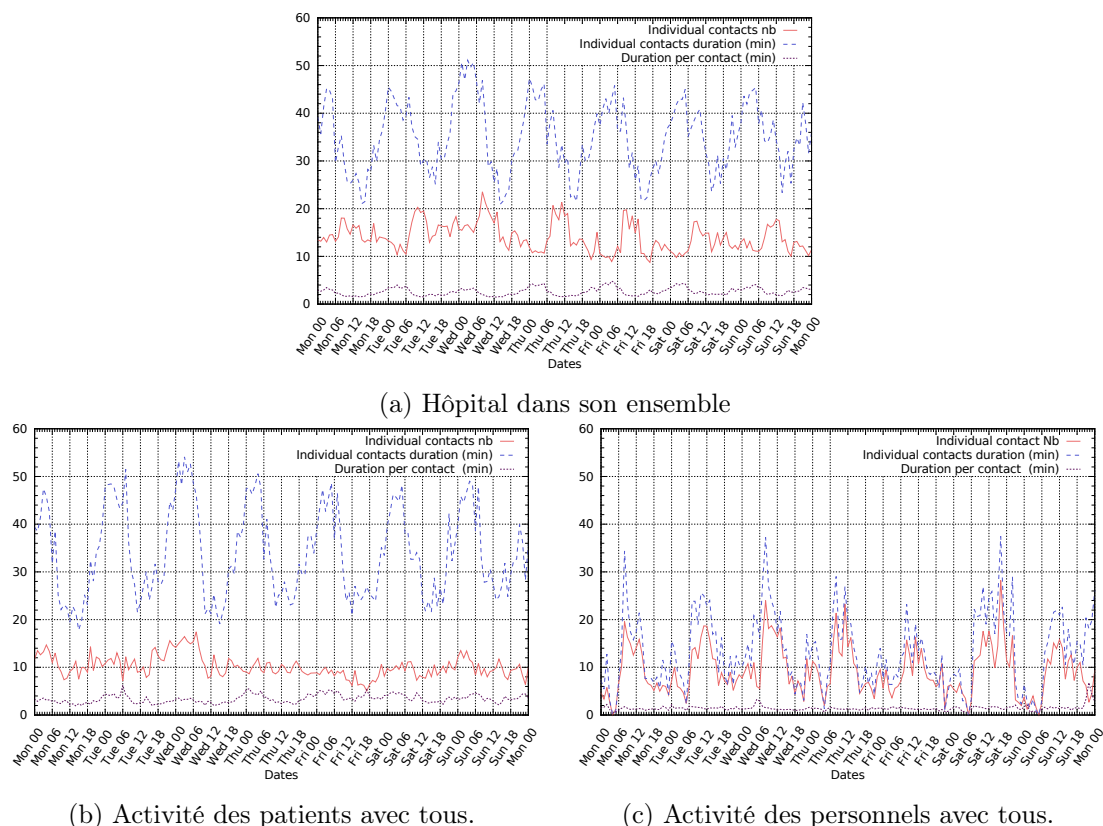


FIGURE 3.21 – Évolution du nombre de contact et de leur durée au cours d’une semaine. La figure 3.21a donne le nombre moyen de contacts des personnes de tout l’hôpital, durée cumulée moyenne de contacts par individu et la durée moyenne de chacun des contacts. La figure 3.21b donne ces trois quantités pour les patients. La figure 3.21c représente ces mêmes trois quantités pour les personnels de l’hôpital.

situe rarement au-dessus de 5 minutes. La figure 3.21b, traitant des contacts impliquant des patients, nous amène à faire des constats similaires. Les durées cumulées moyennes des patients sont un peu plus élevées que les moyennes obtenues pour la globalité de l’hôpital alors que le nombre des contacts par patient est lui plus bas. La durée moyenne des contacts des patients est donc plus élevée tout au cours de la journée que celle de l’ensemble de l’hôpital. Nous avons déjà pointé ce trait dans la partie structurale, et ceci se retrouve indépendamment de l’heure de la journée ou de la nuit.

La figure 3.21c donnant le nombre et la durée moyenne des contacts par personnel ainsi que la durée moyenne de chacun des contacts impliquant un personnel met en évidence que la plupart des contacts sont très courts comme tend à le faire penser la moyenne des durées de chacun des contacts, inférieure à 2 minutes dans la majorité des cas. On constate que le nombre de contacts ainsi que la durée cumulée de ces contacts par personnel sont très corrélées. Ce n’était pas le cas des courbes s’appuyant sur les données globales de l’hôpital ainsi que sur les courbes concernant l’activité des patients. La durée des contacts des personnels ne semble pas varier considérablement non plus, contrairement à celle des patients.

La figure 3.22 représente l’évolution conjointe du degré des individus et de la durée de contacts par individu. La sous-figure 3.22a s’appuie sur le degré moyen des personnes au cours des heures ainsi que la durée de contacts par individu. Nous constatons encore une

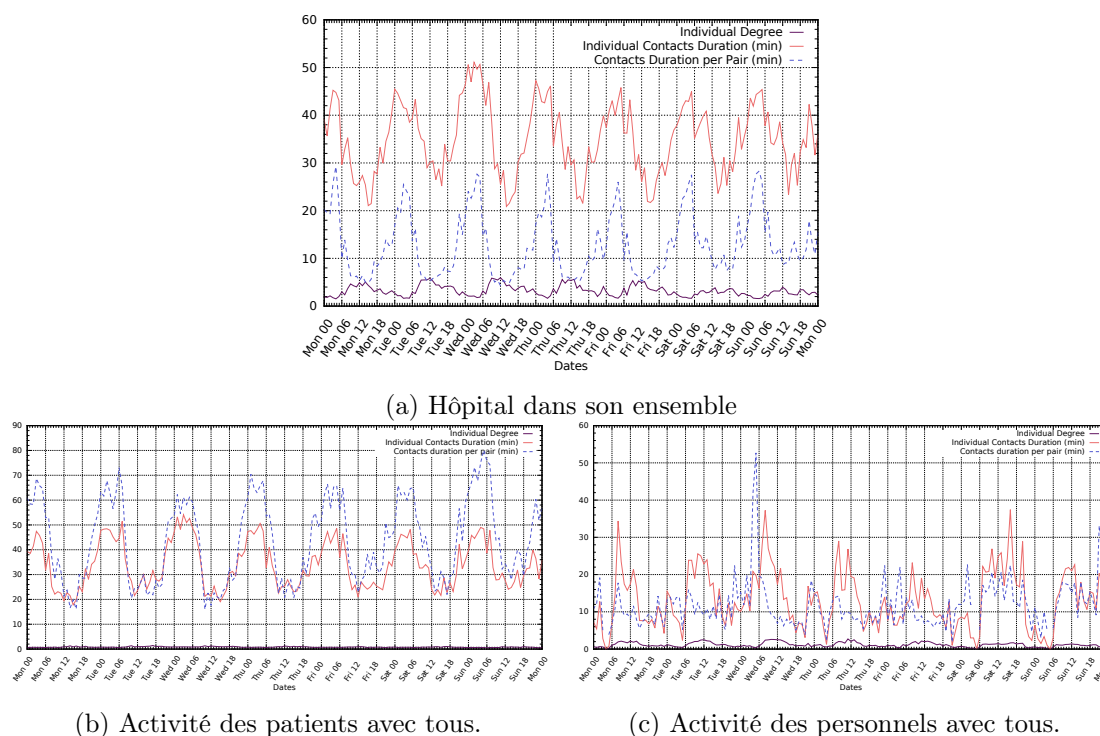


FIGURE 3.22 – Comparaison de l'évolution à l'heure du degré des personnes et de la durée de contacts par individu. La figure 3.22a donne le degré moyen des personnes de tout l'hôpital (ou aussi paires par individu), la durée moyenne par individu et la durée moyenne par paire par individu. La figure 3.22b donne ces trois quantités pour les patients. La figure 3.22c représente ces mêmes trois quantités pour les personnels de l'hôpital.

fois l'apparition d'un cycle circadien. Le week-end n'est par contre pas différent des autres jours en terme d'activité par individu. Nous remarquons avec étonnement que la durée cumulée des contacts d'une personne est beaucoup plus élevée la nuit que le jour. Nous observons de même que pour une même paire d'individus, la durée de leurs interactions est plus élevée la nuit que le jour. Au contraire, bien que le degré moyen par individu au cours des heures soit très faible (toujours inférieur à 5) nous distinguons une différence entre le jour et la nuit : le degré moyen des personnes actives est plus élevé durant la journée que durant la nuit.

Ainsi, il semblerait que les contacts entre les personnes soient plus longs la nuit que le jour, mais soient moins diversifiés. Il est possible de mieux comprendre ce phénomène en étudiant séparément l'activité des patients de celle des personnels. Les résultats sont résumés par les sous-figures 3.22b pour les patients et 3.22c pour les personnels. Le rythme circadien, très visible pour l'activité des patients (figure 3.22b) est un peu moins marqué sur l'activité des personnels, en particulier sur la courbe des durées de contacts par paire. On remarque que la durée des contacts des personnels varie beaucoup plus dans le temps que celle des patients. Et regardant plus attentivement les heures de pics d'activité, on remarque que la durée des contacts par patient est très importante la nuit (entre 22h00 et 8h00). Pour les personnels, nous constatons le contraire. En effet, la durée de contact pour les personnels encore actifs la nuit (entre minuit et 6 :00), la durée de leurs contacts est très faible. On constate aussi une diminution considérable au milieu de la nuit (aux alentours de 4h du matin). Les pics d'activité des personnels se produisent généralement en journée. Certains

changement brusques d'activité surviennent en début de journée (aux alentours de 8h00), en milieu de journée (vers midi), et en fin de journée (20h00 et 23h00).

On peut expliquer cette différence de comportement en reportant notre attention sur la comparaison des courbes représentant le nombre de paires par individu et la durée moyenne par paire. En ce qui concerne les patients, le degré individuel est très faible, compris entre 1 et 3. On constate souvent une légère augmentation du degré moyen lorsque la durée cumulée des contacts par individu diminue. Les variations de durée moyenne de contacts par paire suivent celles de la durée cumulée. On en déduit que durant le jour, les patients sont en contact avec plusieurs personnes, durant un temps très court. Ils ont donc des contacts courts et un peu plus variés, contrairement à la nuit, durant laquelle la durée des contacts par paire est très élevée, mais seules quelques paires existent. Ce phénomène peut s'expliquer par le fait que certains patients partagent leur chambre et sont donc à proximité quasi-permanente la nuit. Pour les personnels, la situation est opposée. La durée des contacts par individu suit l'évolution du degré moyen des personnels. Ainsi, la durée cumulée de contacts des personnels est d'autant plus importante que ces mêmes personnels rencontrent plus d'individus et est plus faible durant la nuit qui n'implique que quelques paires d'adjacence isolées. Ainsi, pour les patients, la durée cumulée des contacts par individu dépend principalement de la durée des contacts par paire et non pas du nombre de paires impliquées à chaque instant de la journée. En revanche, la durée cumulée par personnel est dépendante du nombre de paires actives car la durée des contacts par paire d'adjacence varie très peu au cours des heures : la durée de contact au sein d'une paire est rarement supérieure à 10 minutes en moyenne par heure. Ces différences profondes d'activité des personnels et des patients ont certainement une très grande influence sur la manière dont les patients et les personnels peuvent propager une diffusion dans ce réseau.

### **En résumé**

Dans ce chapitre, nous avons étudié l'évolution temporelle des interactions ayant lieu dans l'hôpital. Nous avons commencé par étudier l'évolution des quantités d'activité sur des données agrégées au jour et constaté que le réseau des personnes est très souvent connecté dans son ensemble mais est peu dense et formé d'un grand diamètre pour une population si réduite. Nous avons aussi constaté que la taille de la population est assez variable au cours des jours et corrélée à la quantité d'activité se produisant dans l'hôpital. Au contraire, d'un point de vue individuel, les quantités d'activités par jour sont comparables. Afin de savoir si l'activité d'un jour à l'autre est vraiment stable, nous avons travaillé sur des mesures de similarité de données. Nous avons présenté plusieurs mesures existantes et les avons appliquées au réseau tiré des données MOSAR. Nous avons aussi proposé une autre mesure de similarité inspirée d'une mesure déjà connue. Nous avons constaté que les interactions sont assez différentes d'un jour à l'autre et que le taux de différence varie énormément pour les personnels, faisant ressortir les week-ends. Au contraire, les différences de structure de réseau du point de vue des patients sont assez constantes ce qui laisse un espoir pour modéliser leur comportement. Nous avons enfin étudié le comportement de l'activité à l'heure sur une semaine et avons encore constaté une nette différence entre le comportement des patients et des personnels. Les personnels ont toujours à peu près la même quantité de contacts au cours du temps, bien que nous puissions distinguer des pics d'activité aux heures des repas et des relèves de personnels. Les patients, eux, ont deux types de comportements : un le jour, avec des contacts assez nombreux et courts, et un la nuit, avec un degré très faible, mais des durées de contacts très longues.



## Chapitre 4

# Structure de graphe temporel et diffusion

Les modèles de diffusion épidémique sont très largement utilisés pour décrire des données empiriques et générer des diffusions artificielles ainsi que pour extraire des caractéristiques du réseau [11]. C'est seulement récemment que ces méthodes ont été appliquées aux études épidémiologiques [3, 31, 34, 33, 35]. Par exemple, certains modèles multi-agents ont été utiles pour étudier la diffusion de la variole ou des pandémies de grippe dans de larges communautés [33, 70, 43, 40]. Plusieurs travaux ont été conduits sur des réseaux dynamiques pour évaluer l'impact de la structure sur la vitesse de diffusion possible dans les réseaux [41, 20, 62]. Les différents modèles de diffusion existants, tels les modèles SI, SIR ou SEIR, sont relativement simples et dans certains cas particuliers analysables mathématiquement sur des réseaux statiques. Ces trois modèles s'appuient sur le fait que les réseaux sont constitués de nœuds étant dans un état qui peut changer au cours du temps, au gré des interactions jouées entre les nœuds. Ces états sont au nombre de 4 et leurs initiales permettent de donner leur nom aux modèles :

- S sain : nœud qui ne porte pas la bactérie mais est susceptible d'être infecté par la suite
- E exposé : touché par l'infection mais n'est pas encore infectieux
- I infectieux : nœud portant l'infection et qui peut la transmettre
- R remis : n'est plus infectieux, ni susceptible de le devenir (guéri et immunisé ou mort).

Chacun des trois modèles permet de révéler les propriétés de diffusion clefs du réseau. Ils permettent en particulier de mettre en avant les caractéristiques topologiques suivantes : le taux d'infection à la fin de la diffusion, la profondeur de cette diffusion, ou, pour des réseaux dynamiques, la durée de cette diffusion ainsi que le nombre de liens utilisés. Nous avons par contre peu de connaissances sur le comportement de ces modèles sur des réseaux dynamiques dont les nœuds et les arêtes changent au cours du temps. Pourtant, ces modifications, comme la structure de voisinage des nœuds des réseaux statiques, ont une forte influence sur l'efficacité de la diffusion.

Dans la littérature, certains auteurs tentent de calibrer leurs modèles avec des paramètres appropriés pour essayer de reproduire des résultats proches d'observations obtenues sur des données empiriques. Ici, nous nous intéressons plutôt à l'influence de la structure du réseau

sur les caractéristiques de la diffusion. Nous comparons le comportement de diffusion selon quelques configurations des paramètres et considérons l'influence que cela peut avoir sur les caractéristiques principales de la diffusion. En particulier, nous nous intéressons au taux d'infection à la fin de la période de simulation, à la durée et à la vitesse de propagation des diffusions jouées. Nous considérons le nombre de liens partagés et qualifions leur importance dans la diffusion.

## 4.1 Diffusion sur les données MOSAR : principe

Pour faire une première évaluation des possibilités de diffusion du réseau de proximité humaine obtenu avec les données MOSAR, nous jouons des diffusions simples suivant le modèle SI. Pour mettre en place ce modèle, nous affectons aux nœuds du réseau l'un des deux statuts suivants, au cours du temps :

- Sain
- Infecté

Le premier des statuts, dénommé "Sain", indique que le nœud n'est pas porteur de l'information ou du virus et qu'il ne peut par conséquent le transmettre dans l'immédiat. Ce nœud est par contre susceptible de recevoir l'information par un autre nœud qui, lui, porte le statut "Infecté". Les nœuds ayant le statut "Infecté" le sont pour toute la durée de la simulation et peuvent infecter un nœud "Sain" avec une probabilité choisie avant de lancer la simulation.

Toutes les simulations qui ont été conduites dans la suite sont initialisées à partir d'un temps donné et d'une seule personne infectée.

## 4.2 Transmission systématique

Nous avons lancé une diffusion systématique sur 8 semaines des données MOSAR (du 6 juillet au 2 septembre), à partir de toutes les sources possibles, à toutes les heures possibles, ce qui produit plus de 100000 simulations. Nous étudions ensuite le comportement de ces diffusions systématiques pour connaître la propension globale de ce réseau dynamique à diffuser une information. Nous travaillons sur le réseau non pondéré agrégé à l'heure, obtenu comme expliqué page 26.

De cette manière, nous obtenons une première évaluation des chemins de diffusion possibles. Un chemin est, dans ce contexte, le parcours à faire pour passer d'une personne à l'autre en suivant les contacts que nous avons enregistrés (cf définition 1.3.7).

La transmission systématique, sur un graphe non pondéré, permet de trouver les chemins les plus courts en temps, entre toutes les paires de nœuds.

### Lecture des axes temporels

\* Dans la suite, tous les axes représentant le temps sur une période de 8 semaines sont exprimés en pas de temps, chaque pas de temps correspondant à une heure



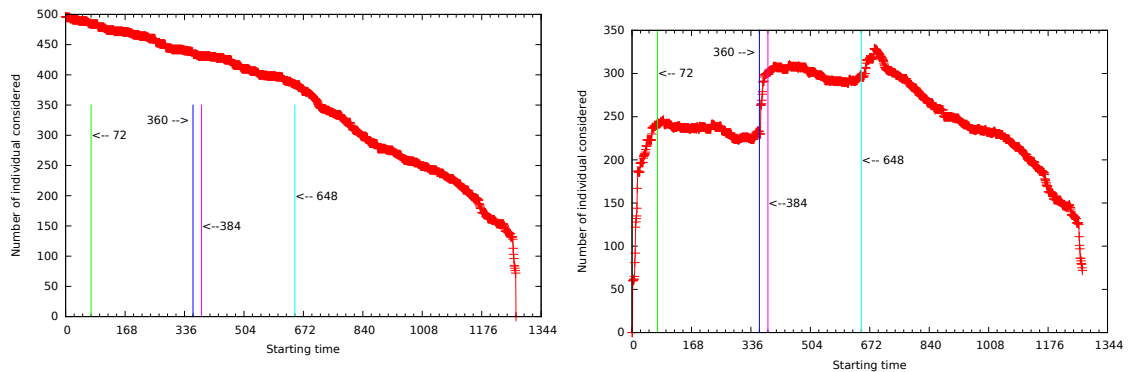
pendant l'expérience. Les graduations principales sont espacées de  $24 \times 7$  pas de temps, soit d'une semaine, et correspondent à minuit en terme d'heure de la journée. Les sous-graduations indiquent les jours ; ce sont donc des intervalles de 24 pas de temps.

- \* Tous les axes représentant le temps sur une période d'une semaine ont des graduations principales tous les 24 pas de temps (24 heures) et des sous-graduations tous les 4 pas de temps.

Dans la suite, nous nous appuyons sur deux types de populations, évoluant au cours du temps :

- la *population totale* à un temps  $t$  donné, qui comprend toutes les personnes présentes au temps  $t$  et celles qui seront présentes à un temps supérieur  $t + \tau$ . Cette population est celle qui participe à la diffusion.
- la *population partielle* à un temps  $t$  donné, qui comprend toutes les personnes dont l'intervalle de présence durant l'expérience contient  $t$ . On s'intéresse de près à la population partielle pour certaines analyses comme pour l'estimation de la durée des diffusions, afin de ne pas sous-estimer la vitesse de diffusion.

A titre indicatif, nous présentons l'évolution de la taille de la population totale (participant à la diffusion) au cours du temps (figure 4.1a) ainsi que l'évolution de la population partielle (figure 4.1b). L'axe des abscisses représente l'évolution temporelle exprimée en pas de temps. L'axe des ordonnées indique le nombre de personnes.



(a) Nombre de personnes participant prises en compte pour la diffusion en fonction du temps de départ. (b) Nombre de personnes ayant manifesté leur présence sans avoir encore quitté l'expérience.

FIGURE 4.1 – Évolution de la taille de la population prise en compte. La sous-figure 4.1a représente le nombre de personnes prises en compte dans la diffusion, c'est à dire le nombre de personnes présentes soit au temps  $t$ , soit à un temps  $t + \tau$ . La sous-figure 4.1b représente le nombre de personnes étant connues dans l'hôpital et participant toujours à l'expérience au temps  $t$  (des personnes qui peuvent être absentes au temps  $t$ , mais de retour à un temps  $t + \tau$ ).

On remarque sur la figure 4.1a, une diminution presque linéaire de la population pouvant intervenir dans la diffusion, du fait des arrivées et départs des personnes survenant au cours du temps. Cela montre que le nombre de personnes quittant l'hôpital définitivement est presque constant, mise à part les derniers jours. La taille de la population de l'hôpital ne

diminuant pas, on en conclut que le renouvellement de la population se fait régulièrement. Au contraire, sur la figure 4.1b, on remarque immédiatement deux périodes d'un peu plus d'une semaine dont la taille des populations est relativement stable. Lors d'analyses plus spécifiques et détaillées, nous nous intéresserons uniquement à ces périodes stables, afin de bénéficier d'une population comparable.

### 4.2.1 Détection des périodes de propagation

Commençons par repérer les temps les plus propices à la transmission, c'est à dire les temps durant lesquels il s'opère le plus d'échanges.

Intéressons nous d'abord à un premier critère qui peut nous permettre d'extraire les périodes importantes en terme de transmission : le taux d'utilisation des pas de temps par les différentes simulations. La figure 4.2 représente pour chaque heure, le nombre de simulations pour lesquelles il s'est produit au moins une transmission durant cette heure. On dira, par abus de langage, que la simulation utilise ce pas de temps ou cette heure.

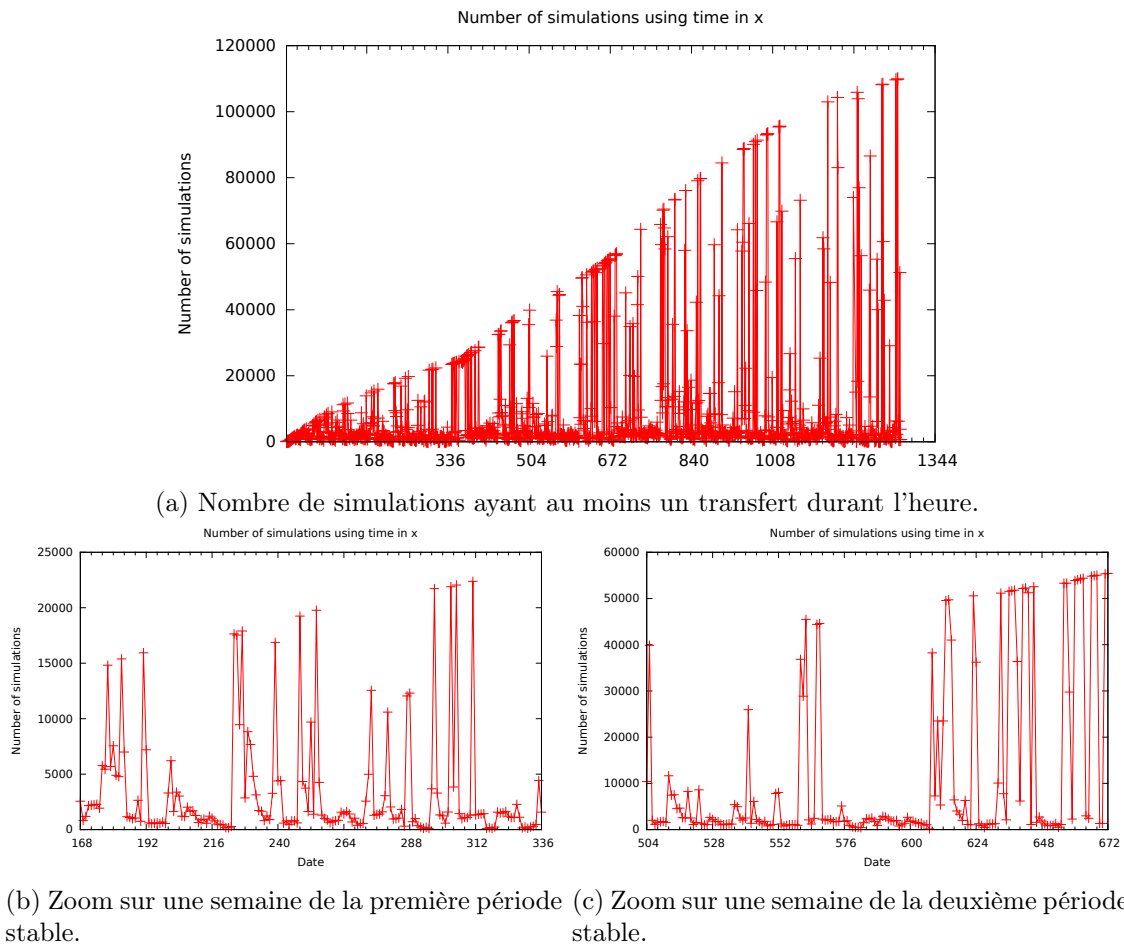


FIGURE 4.2 – Nombre de simulations ayant au moins un transfert durant l'heure.

Nous remarquons instantanément, sur la figure 4.2a, que certaines heures sont beaucoup plus propices à la transmission que d'autres, d'après le nombre considérable de simulations ayant au moins une transmission durant ces heures. En effet, un grand nombre de pics

ponctuels apparaissent. Sur la première moitié de la période, les pics prennent des valeurs de plus en plus importantes à mesure que le nombre de simulations en cours augmente. Sur la fin de la période, les valeurs maximales des pics ont tendance à diminuer, mises à part quelques exceptions. Les transmissions semblent donc se faire majoritairement à certains temps donnés. Le fait que la valeur des pics croît presque linéairement avec le nombre de simulations en cours laisse penser que la majorité des simulations se jouent sur la période entière et donc l'infection n'atteint pas 100% des personnes. Ceci reste à vérifier dans la suite.

Zoomons maintenant sur une semaine, pour mieux repérer les instants de la journée les plus actifs en terme de diffusion. La figure 4.2b comprend les valeurs de la première période stable identifiée sur la courbe 4.1. A cette période, moins de 25000 simulations sont en cours : nous avons donc réduit l'axe des ordonnées à cette échelle. Nous retrouvons les pics d'activité que nous avons identifiés lors de l'analyse temporelle des interactions. En effet, on peut repérer 3 ou 4 pics par jour, à savoir en début de matinée, aux alentours de midi, en milieu d'après-midi et vers minuit. La deuxième semaine dont la taille de la population est stable (cf. figure 4.2c) montre que ces tendances à 4 pics par jour sont moins marquées car les pics apparaissent moins régulièrement quand il y a plus de simulations en cours. Il existe donc des heures plus propices que d'autres pour la propagation. Cela n'exclut pas l'existence d'autres périodes de propagation plus ponctuelles.

Ainsi, les périodes durant lesquelles il y a le plus de simulations actives sont les heures de la journée et à minuit, ce qui correspond aux moments pendant lesquels il y a le plus d'interactions au sein de la population. Ces heures ne sont pas toujours les mêmes ; on ne peut donc pas, pour l'instant, définir précisément de temps cruciaux en terme de diffusion. Les premières heures de la période sont peu utilisées comparées à celles de fin de période. Cela s'explique en grande partie par le fait qu'au début de la période d'exécution, peu de simulations sont en cours.

Pour s'abstraire du nombre de simulations actives, considérons maintenant le pourcentage de simulations actives utilisant un temps  $T$  pour transmettre l'infection.

Nous pouvons observer les résultats sur la figure 4.3 qui, comme précédemment, présente d'abord l'évolution sur la période puis fait un zoom sur deux semaines appartenant aux deux périodes stables en terme de taille de population. On remarque que sur la globalité de la période de 8 semaines, très régulièrement, il existe des pas de temps utilisés par plus de 80% des simulations en cours d'exécution, quelque soit la semaine que l'on considère. Il semble donc exister des moments très importants en terme de diffusion et d'autres moins indispensables pour le processus de transmission.

En considérant les semaines avec plus d'attention (figure 4.3b et 4.3c) on remarque que la forme obtenue est la même que lorsque nous considérons le nombre de simulations actives. Comme à l'échelle plus détaillée, le nombre de simulations en cours d'exécution varie peu, il n'a que peu d'influence sur le pattern des pas de temps importants observés. Ces pas de temps importants pour la transmission sont répartis de manière relativement équilibrée pour la première semaine alors qu'ils se concentrent plutôt sur certains jours particuliers durant la deuxième semaine. Ainsi, nous remarquons de nouveaux des pas de temps qui ne sont pratiquement jamais utilisés alors que d'autres, particulièrement en journée ou à minuit, paraissent indispensables à la diffusion. Les patterns observés ne nous permettent malheureusement pas d'extraire des périodes systématiquement critiques en terme de diffusion, car, comme nous l'avons observé sur les deux semaines stables en taille

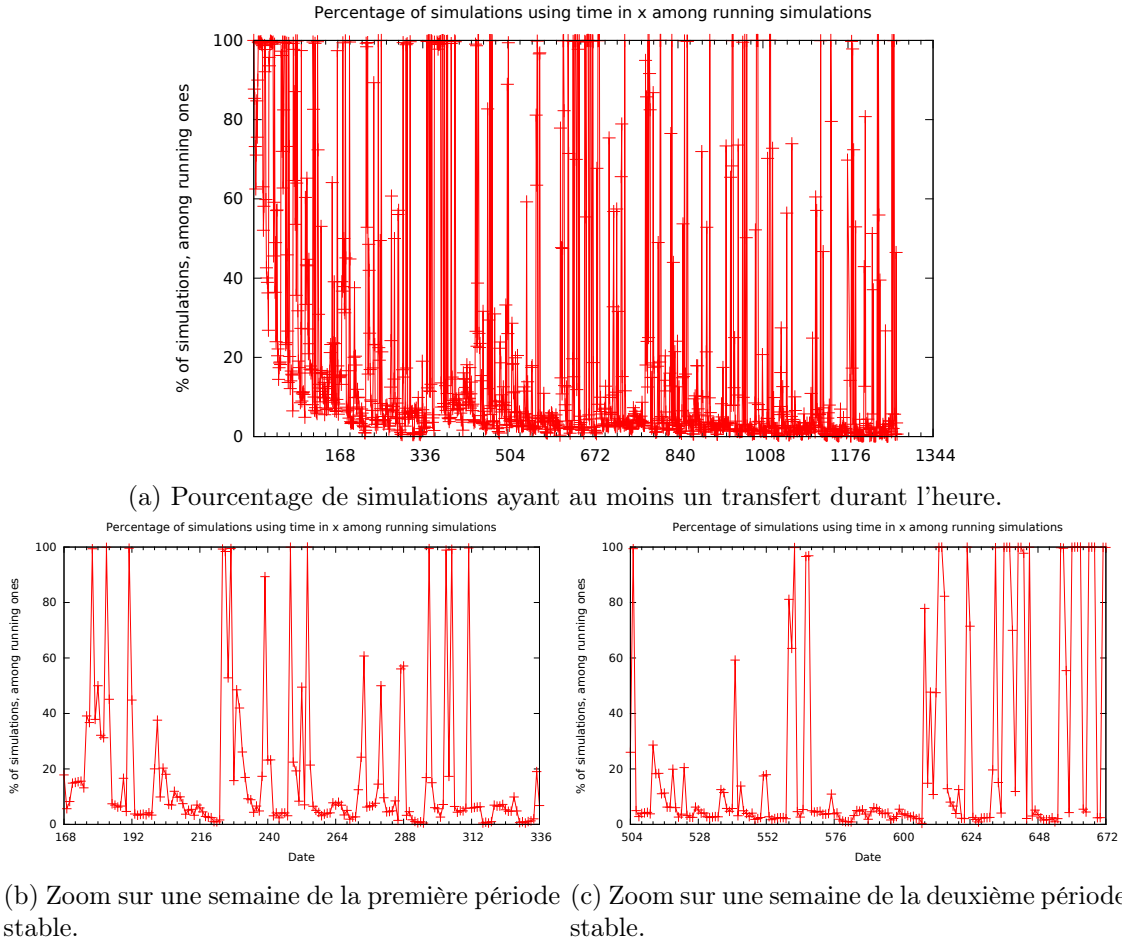


FIGURE 4.3 – Pourcentage de simulations ayant au moins un transfert durant l'heure, parmi l'ensemble des simulations ayant déjà commencé leur exécution.

de population, ces pas de temps importants n'ont pas toujours lieu aux mêmes moments dans la journée. Cette information ne nous suffit donc pas pour établir un modèle de motifs d'évolution de diffusion au cours du temps.

Un autre critère qui peut définir l'importance d'un pas de temps pour la transmission, est la répartition des moments durant lesquels, pour une simulation donnée, il y a eu le plus de transmissions. C'est pourquoi nous nous intéressons, pour chaque simulation  $S$ , au temps où la valeur des deux quantités suivantes atteint son maximum : le nombre de transmetteurs et le nombre de récepteurs. La définition de ces deux quantités est basée sur la notion de graphe acyclique orienté (DAG).

**Definition 4.2.1 (DAG)**

Un DAG est un graphe acyclique orienté (Directed Acyclic Graph en anglais) c'est à dire un graphe orienté ne contenant aucun cycle orienté.

**Définition 4.2.2** (Degré entrant)

Le *degré entrant* d'un nœud  $i$  est le nombre d'arêtes pour lesquelles il joue le rôle de destinataire et est noté  $\mathfrak{d}^-(i)$ . On parle d'arêtes entrant du nœud  $i$ .

**Définition 4.2.3** (Degré sortant)

Le *degré sortant* d'un nœud  $i$  est le nombre d'arêtes pour lesquels il joue le rôle de transmetteur et est noté  $\mathfrak{d}^+(i)$ . On parle d'arête sortante du nœud  $i$ .

**Définition 4.2.4** (Feuille)

Une feuille est un nœud n'ayant pas d'arête sortante.

Les simulations opérées sur le graphe des contacts peuvent être modélisées par un DAG. On notera  $G'$  un DAG obtenu par simulation de diffusion sur le réseau des contacts modélisé par un graphe  $G$ . Le DAG  $G'$  est construit par la méthode décrite ci-dessous. Pour chaque transmission de la diffusion de  $a$  vers  $b$  au temps  $t$  ( $b$  n'étant pas encore atteint par la diffusion au temps  $t - 1$ ) :

- nous construisons une arête  $e = (a, b)$  orientée de  $a$  vers  $b$ .
- nous étiquetons  $b$  par le temps de transmission  $t$

**Définition 4.2.5** (Transmetteurs et Récepteurs)

Soit  $G' = (V', E')$  un DAG. On dit qu'un nœud  $u' \in G'$  est un transmetteur au temps  $t$ , si et seulement si  $u'$  est infecté avant le temps  $t$  et donc étiqueté par  $d$ ,  $d < t$ , et qu'il existe un nœud  $v$  étiqueté par le temps  $t$ , voisin de  $u$ . En d'autres termes, soit  $G'' = (V'', E'')$  un sous-graphe orienté de  $G'$ , où  $V'' = V'$  et  $E''$  est l'ensemble des arêtes dont les destinations sont étiquetées par  $t$  :  $E'' = \{(u', v') \in E' \mid b' \text{ est étiqueté } t\}$ .  $u'$  est transmetteur dans  $G'$  au temps  $t$  si  $\mathfrak{d}^+(u'') > 0$ ,  $u'' \in G''$ .

On dit qu'un nœud  $v \in V'$  est un récepteur au temps  $t$ , si et seulement si  $v$  n'est pas infecté et qu'il existe un nœud  $i$  déjà infecté au temps  $t - 1$ , voisin de  $v$  au temps  $t$ . En d'autres termes,  $v$  est un récepteur si  $\mathfrak{d}'^-(v) > 0$

En effet, le nombre de transmetteurs nous indique quand les possibilités de transmission au même pas de temps sont les plus fortes et donc difficilement évitables pour les personnes touchées car elles peuvent être affectées par plusieurs sources. Le temps où le nombre de récepteurs est à son maximum nous indique les moments où le plus de personnes sont touchées par l'infection au même moment et donc les moments pour lesquels il faudrait réfléchir à une stratégie pour réduire les risques de transmission trop forts, en réduisant le nombre de contacts. Les temps forts de risque de transmission et les temps pour lesquels on peut difficilement éviter les infections sont-ils ou non les mêmes ? Nous allons nous en assurer dans la suite.

**Définition 4.2.6** (Largeur en transmetteurs et en récepteurs)

Nous appelons **largeur en transmetteurs** (resp. en récepteurs) au pas de temps  $T$  le nombre de transmetteurs (resp. en récepteurs) existant pour une simulation durant ce pas de temps  $T$ .

La **largeur maximale en transmetteurs** d'une simulation sera le nombre maximal de transmetteurs simultanés sur un pas de temps observé pour une simulation donnée.

Intéressons nous d'abord à la largeur maximale en transmetteurs des différentes simulations  $S$ .

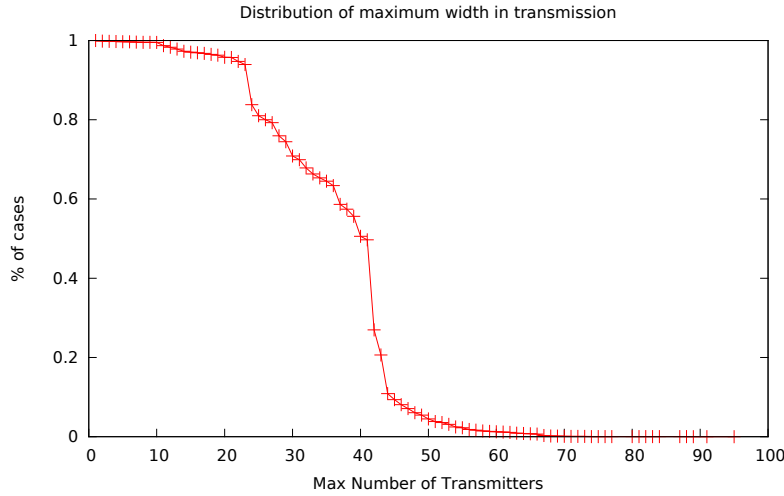


FIGURE 4.4 – Distribution du maximum de transmetteurs.

Afin de mieux interpréter la répartition de ces largeurs au cours du temps, commençons par considérer la figure 4.4, qui indique la distribution cumulative inversée des largeurs maximales en transmetteurs des simulations.

Nous sommes en présence d'une distribution pouvant être qualifiée d'uniforme centrée en 40. On observe que, dans plus de 80% des cas, le nombre maximum de transmetteurs est inférieur à 45. La moitié des simulations ont des largeurs maximales en transmetteurs inférieures à 40. Nous considérerons donc dans la suite qu'une largeur maximale en transmetteurs est importante quand elle dépasse 50 transmetteurs simultanés.

Considérons maintenant la répartition des dates pour lesquelles la largeur maximale de transmetteurs de chaque simulation est atteinte (cf. figure 4.5). Sur l'ensemble de la période (figure 4.5a), on remarque quelques pas de temps pour lesquels il y a beaucoup de valeurs différentes, dont certaines dépassent 45 transmetteurs différents. Toutes ces dates correspondent à des moments où un grand nombre de simulations (plus de 50% en cours) sont actives (figure 4.3a). Le contraire n'est par contre pas vrai. Donc ce deuxième critère, consistant à repérer les dates qui impliquent le maximum de transmetteurs des simulations permet d'affiner le premier critère qui détecte les dates les plus utilisées pour transmettre. Ainsi, certaines dates sont non seulement utilisées presque systématiquement pour transmettre, mais la transmission peut être très massive, quelle que soit l'heure de la journée. On remarque tout de même, sur les deux semaines dont la taille de la population est stable (cf figures 4.5b et 4.5c), qu'il existe peu de cas pour lesquels on atteint le maximum en transmetteur durant ces deux périodes. Les quelques dates pour lesquelles les largeurs maximales en transmetteurs sont atteintes se situent en milieu de journée, entre 9h00 et 16h00, heures de la journée durant lesquelles il y a le plus de monde au sein de l'hôpital car tous les corps de métiers sont représentés.

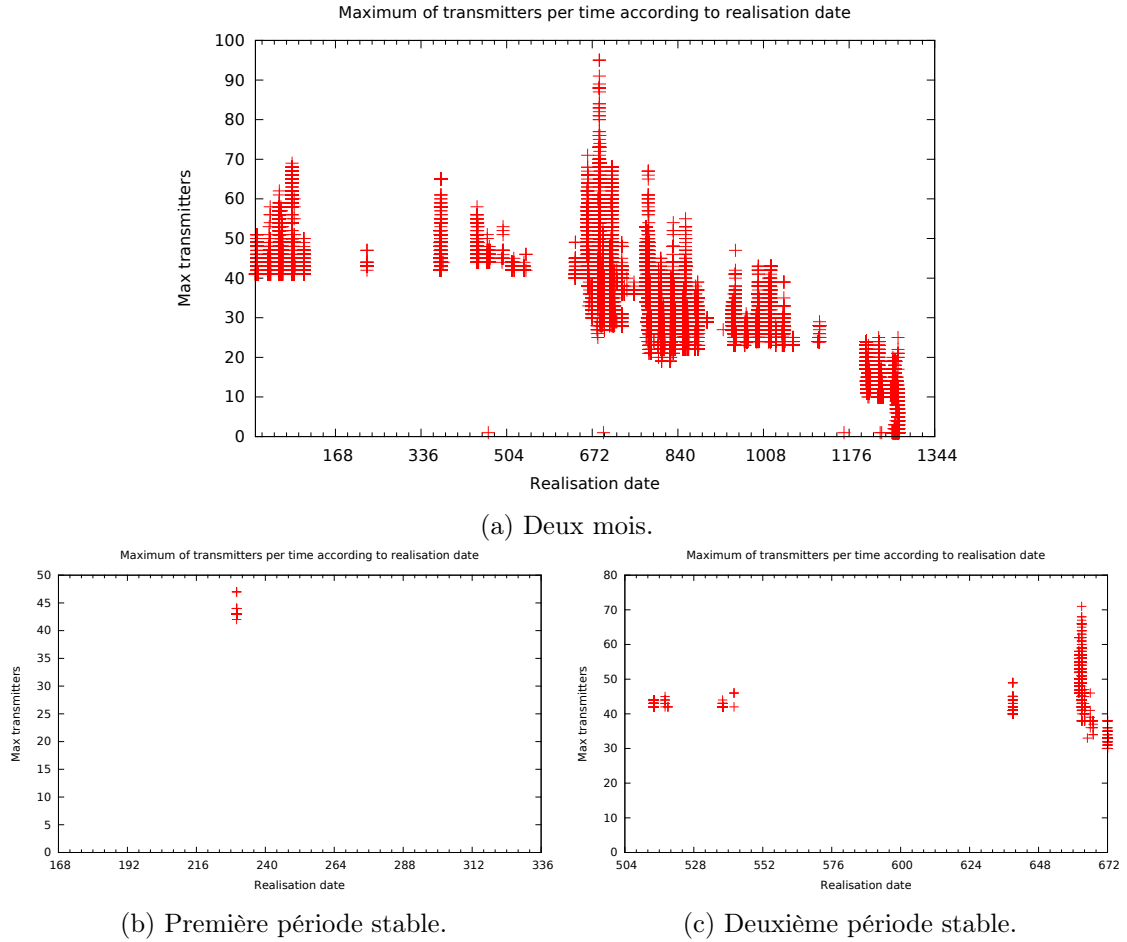


FIGURE 4.5 – Date d’atteinte de la valeur maximale du nombre de transmetteurs pour chaque simulation.

Intéressons nous maintenant, pour chacune des simulations  $S$ , à la date à laquelle leur largeur maximale en réception est atteinte, c’est à dire le moment durant lequel il y a le plus grand nombre de récepteurs simultanément, durant l’exécution de la simulation  $S$ .

Comme précédemment, commençons par regarder la distribution des largeurs maximales en réception (cf. figure 4.6). On constate que les valeurs obtenues ont un éventail un peu plus étendu puisque certaines d’entre elles dépassent la centaine de récepteurs simultanés. Les deux distributions de largeurs maximales se ressemblent, mais celle concernant les récepteurs est légèrement décalée pour donner des valeurs un peu plus élevées.

Concernant la répartition des largeurs maximales en récepteurs, nous observons des différences assez nettes par rapport à celle des transmetteurs. En effet, considérons la figure 4.7 qui présente les dates auxquelles les largeurs maximales en récepteurs de chaque simulation sont atteintes. La première sous-figure (4.7a) nous montre le résultat sur un mois et indique que certains temps sont plus propices que d’autres pour les transmissions en masse. Ces temps sont beaucoup plus nombreux que pour les transmetteurs et impliquent parfois beaucoup plus de personnes, même sur la fin de l’expérience.

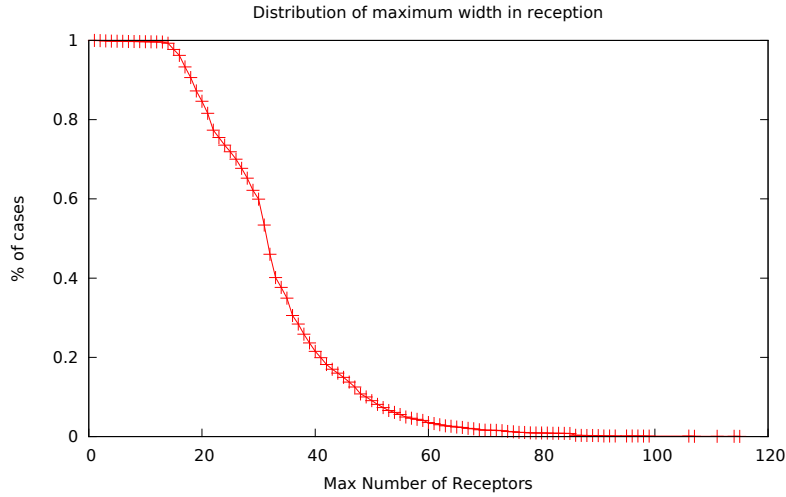
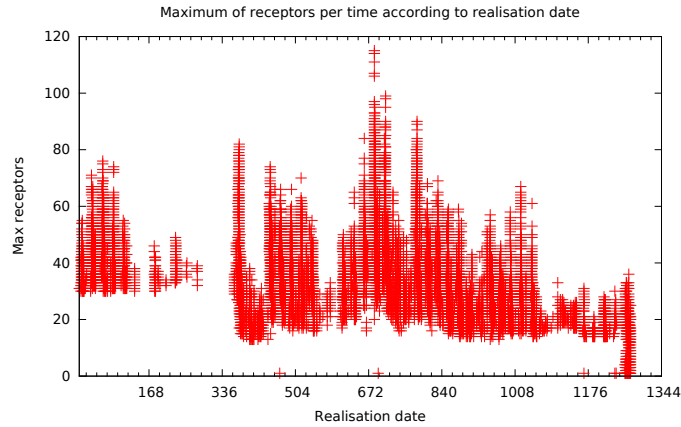
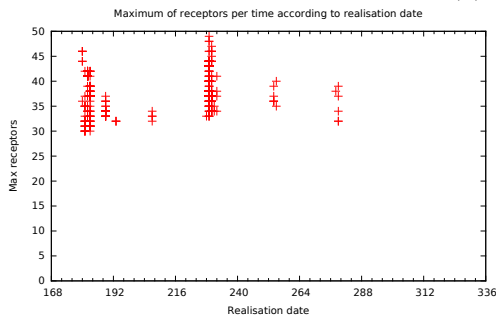


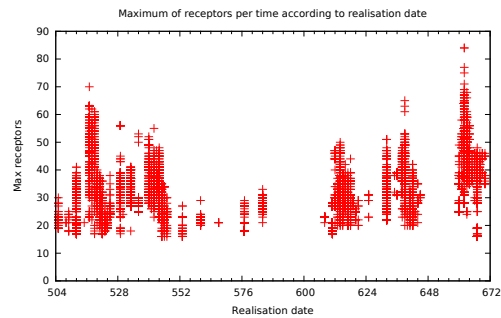
FIGURE 4.6 – Distribution du maximum de récepteurs.



(a) Deux mois.



(b) Première période stable.



(c) Deuxième période stable.

FIGURE 4.7 – Date d'atteinte de la valeur maximale du nombre de récepteurs pour chaque simulation.

Les deux sous-figures (figures 4.7b et 4.7c) présentent plus en détail sur une semaine de chacune des périodes stables la répartition des largeurs maximales en récepteurs au cours du temps. On peut remarquer qu'à certains pas de temps, les largeurs maximales en récepteurs sont importantes (supérieures à 30) mais ne concernent pas la moitié des simulations en cours d'après la figure 4.3a, comme par exemple sur la période entre les pas de temps 528 et 552.



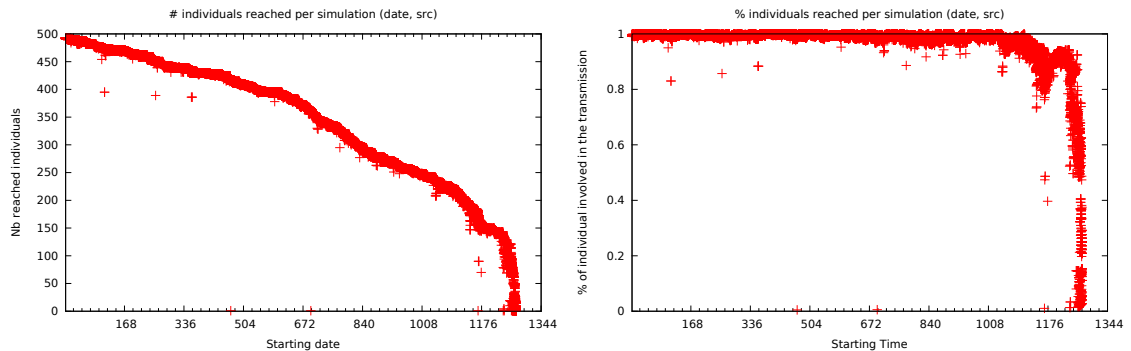
Donc les largeurs en récepteurs ne sont pas corrélées avec la fréquence d'utilisation des pas de temps par l'ensemble des simulations comme cela semblait être le cas. Il faut donc considérer séparément ces deux critères de sélection des moments importants en terme de diffusion.

#### 4.2.2 Efficacité de la transmission

Pour aller plus loin, nous travaillons sur l'évaluation de l'efficacité des simulations en terme de transmission. En d'autres termes, nous étudions la quantité de personnes touchées par la diffusion, la rapidité de transmission ainsi que la robustesse de la diffusion.

##### Quantité

Nous représentons le nombre de personnes infectées par simulation, présentées ici par heure de départ, sur la figure 4.8. Chaque point de la figure représente la valeur obtenu pour une simulation.



(a) Nombre de personnes infectées à la fin de chaque simulation. (b) Taux de personnes infectées à la fin de chaque simulation.

FIGURE 4.8 – Nombre et taux de personnes infectées à la fin de chaque simulation, ordonnées par heure de départ.

La sous-figure 4.8a présente le nombre de personnes infectées après le déroulement des simulations, ordonnées par heure de départ. On constate que la forme générale de la courbe obtenue ressemble très fortement à la forme de l'évolution de la taille de la population présente, en suivant les mêmes ordres de grandeur. Une grande majorité de la population est donc touchée par toutes les simulations. Nous remarquons tout de même quelques exceptions avec un nombre d'infections beaucoup plus faible, toujours corrélé au départ précoce de la personne source de la simulation, ce qui est difficilement explicable.

La distribution cumulative inversée du taux de personnes atteinte à la fin de chaque simulation sur la figure 4.9 confirme que plus de 90% des simulations atteignent plus de 60% de la population.

Le taux d'infection par simulation est très élevé, comme nous le laissait supposer la comparaison des courbes 4.8a et 4.1 car il est presque systématiquement supérieur à 95%. Un certain nombre de simulation n'atteint tout de même pas 100% de personnes infectées, ce

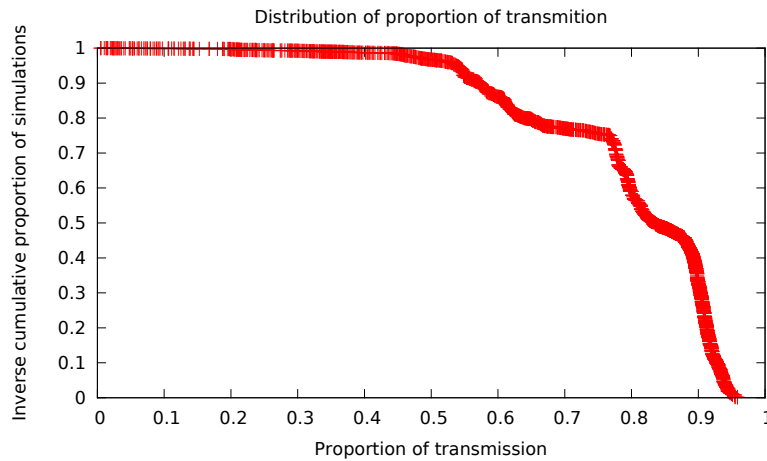


FIGURE 4.9 – Distribution cumulative inversée du taux de personnes atteintes par la diffusion, parmi les personnes présentes dès le début de la diffusion (les personnes arrivant dans l'expérience au cours d'une simulation ne sont pas prises en compte pour cette simulation.).

qui signifie qu'une diffusion importante peut s'opérer depuis n'importe quel nœud mais qu'il existe des paires de nœuds qu'aucun chemin temporel ne relie.

Ainsi, il semble que les temps de départ des simulations autant que la diversité des sources n'aient pas grande influence sur la propension de diffusion.

### Durée et nombre de sauts

La durée des simulations est un autre indice d'efficacité dans la transmission. Nous considérons qu'une simulation est terminée quand toutes les personnes considérées sont infectées ou quand nous avons atteint le dernier pas de temps avec des données d'interactions.

Bien que le taux d'infection soit très proche de 100% et atteint souvent toute la population, les simulations se terminent presque toutes à la fin de la période quelque soit leur date de départ. Cela signifie que certaines transmissions se font très tardivement, quelque soit la date de départ de la simulation. Ce phénomène peut en partie s'expliquer par l'absence, parfois longue de certains personnels.

La figure 4.10 nous montre les temps de terminaison de chacune des simulations en fonction de leur date de départ. La plupart des simulations terminent sur des pas de temps communs. En effet, la diversité des temps de terminaison est faible. On remarque aussi que les simulations commençant plus tôt (jusqu'au milieu de la première période de simulation) terminent légèrement avant la fin de la période de simulation.

La plupart des simulations commençant à partir de la moitié de la période de simulation se terminent à la fin de cette période, avec un pourcentage d'infection n'atteignant pas 100%. Il faut donc la quasi totalité de la durée de la période de simulation, soit 8 semaines, pour réussir à infecter toute la population, ce qui s'explique par la longue absence de certains personnels, présents en tout début et en toute fin de période de simulation.

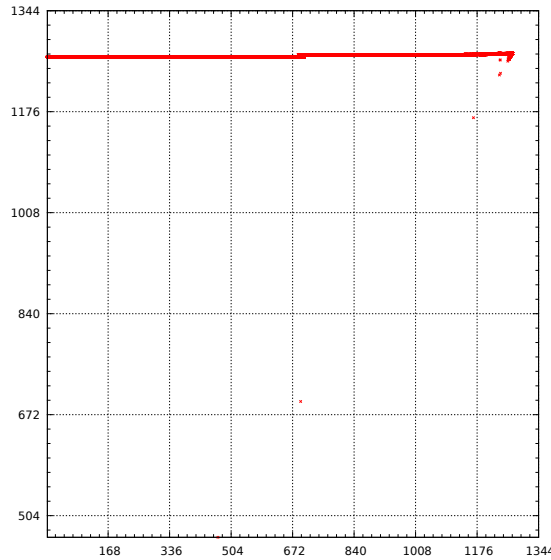


FIGURE 4.10 – Temps de terminaison de chacune des simulations en fonction de leur temps de départ.

La figure 4.11 représente la durée nécessaire aux simulations pour atteindre le temps auquel se réalise la largeur maximale en transmetteurs (sous-figure 4.11b), puis en récepteurs (sous-figure 4.11a), donc d’infections au même pas de temps.

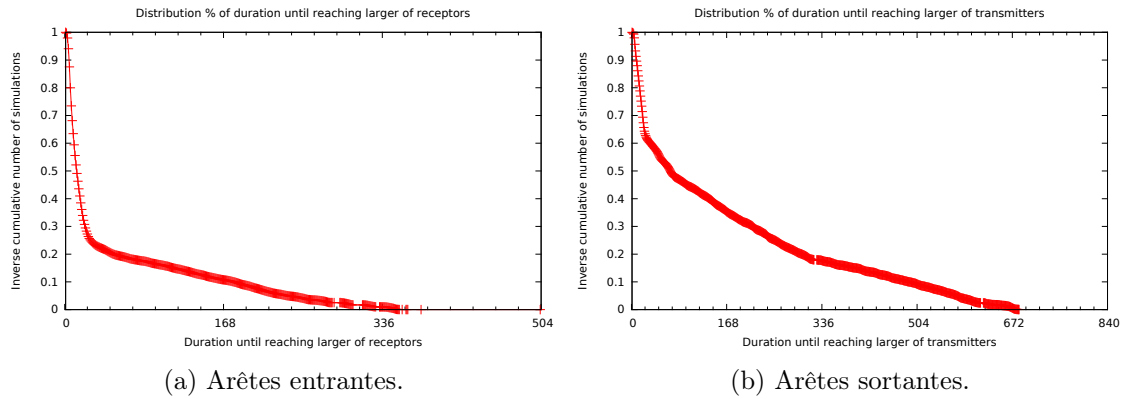


FIGURE 4.11 – Proportion de temps nécessaire pour atteindre la largeur maximale du DAG de diffusion.

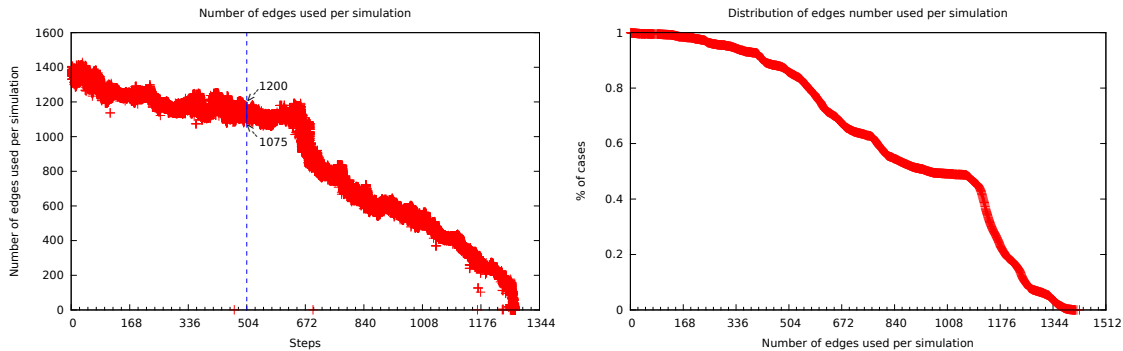
On remarque que moins de 30% des simulations dépassent une journée pour atteindre la date où le nombre de récepteurs au même pas de temps est le plus important (cf. figure 4.11a). On peut même souligner que la moitié des simulations atteint une largeur maximale en récepteurs au bout d’une demi-journée seulement. Les transmissions massives semblent donc se faire en début de simulation. Il existe tout de même un nombre non négligeable de cas (10%) pour lesquels la largeur maximale en récepteurs est atteinte plus d’une semaine après le début de la simulation. La largeur maximale en transmetteurs est, quant à elle, atteinte plus tardivement en moyenne. La moitié des simulations atteint ce maximum en moins de trois jours. Il est à noter que pour plus de 20% des simulations, cette largeur n’est pas atteinte avant une semaine et demie. Pour mettre en place une politique de vaccination,

il faudrait donc, dans le pire des cas, agir dès le deuxième jour de propagation, ce qui laisse peu de marge.

### Intensité des transmissions

Caractérisons maintenant la robustesse de la diffusion. Robustesse, dans ce contexte, signifie qu'il est difficile, voire impossible, d'enrayer la diffusion en modifiant légèrement le réseau, comme par exemple en enlevant une simple arête ou en retirant un nœud du réseau. Cela signifie donc que pour que l'infection passe d'un nœud à un autre, il existe plusieurs chemins dans le graphe, dont plusieurs liens sont différents. Plus le nombre de liens différents est important, plus la diffusion est robuste. En effet, pour empêcher la diffusion, il faut déconnecter le graphe et pour ce faire, il faut retirer un grand nombre d'éléments du graphe.

Pour étudier la robustesse, nous commençons ici par considérer le nombre d'*arêtes utilisées* du graphes  $G$  des interactions dans chacune des diffusions (figure 4.12), c'est à dire permettant une transmission entre deux nœuds, ce qui correspond au nombre d'arêtes du DAG représentant une simulation. La figure 4.12a nous permet de visualiser, pour chaque simulation ordonnée par date de départ, le nombre d'arêtes différentes utilisées durant son exécution. On constate que la forme générale de la courbe rappelle fortement celle décrivant la taille de la population prise en compte au cours du temps. Le nombre d'arêtes est donc fortement corrélé à la taille de la population et est entre 2 et 3 fois supérieure à cette taille. On observe tout de même de fortes différences, parfois plus de 100 arêtes de différence, parmi toutes les simulations ayant une même date de départ mais des sources différentes. C'est par exemple le cas au pas de temps 504 pour lequel certaines simulation n'ont que 1075 arêtes alors que d'autres en ont jusqu'à 1200. Le nombre d'arêtes utilisées durant la simulation dépend donc de la date de départ mais aussi de la source d'infection choisie. Les deux périodes stables en taille de population font partie des 80% de simulations utilisant plus de 400 arêtes comme l'atteste la distribution du nombre d'arêtes (figure 4.12b). Le nombre d'arêtes est donc assez important car il correspond environ à 4 arêtes par personne infectée.



(a) Nombre d'arêtes utilisées par simulations ordonnées par date de départ des simulations. (b) Distribution du nombre d'arêtes utilisées par simulation.

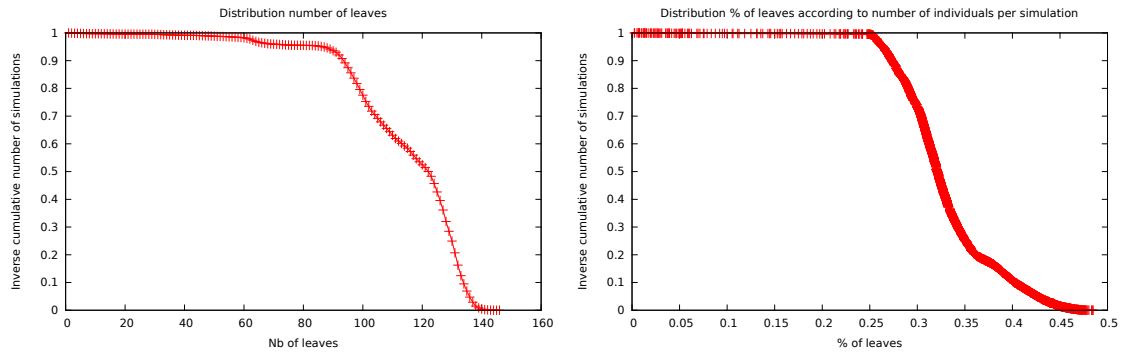
FIGURE 4.12 – Nombre d'arêtes utilisées par simulation.

Les arêtes utilisées sont orientées lors de la diffusion qui peut être représentée à l'aide d'un DAG (cf. définition 4.2.1). Nous pouvons distinguer deux types d'arêtes : celles qui se

trouvent au cœur de la diffusion et celles qui se trouvent en périphérie. Afin de les identifier, nous nous intéressons aux feuilles des DAG construits par les simulations.

La figure 4.13 nous renseigne sur la répartition des feuilles à la fin des simulations. Une feuille étant un nœud infecté par la simulation mais n'ayant pas eu l'occasion de retransmettre jusqu'à la fin de la simulation, par manque de temps ou de voisins non infectés, après le pas de temps l'ayant infecté lui-même. La sous-figure 4.13a représente la distribution du nombre de feuilles existant à la fin des simulations. On constate que la grande majorité (environ 80%) des simulations se terminent avec un nombre de feuilles compris entre 80 et 120. Cela signifie que 80 à 120 personnes ne retransmettent pas l'infection.

La sous-figure 4.13b représente la distribution de la proportion de feuilles par rapport à la population des simulations. Un fort pourcentage (80%) des simulations terminent avec un nombre important de feuilles : 32 et 44 % de la population est une feuille à la fin de l'exécution. Cela signifie que plus de 60% de la population joue le rôle de transmetteur, ce qui laisse une grande diversité de chemins de diffusion possible.



(a) Distribution du nombre de feuilles par simulation. (b) Distribution de la proportion de feuilles de la population d'une simulation.

FIGURE 4.13 – Répartition des feuilles selon les simulations.

### 4.2.3 Etude globale d'utilisation des chemins

Lors de la simulation, nous donnons un poids aux arêtes utilisées inversement proportionnel au nombre d'arêtes utilisées au temps  $t$  ayant la même destination. La figure 4.14 montre un exemple d'assignation du poids des arêtes.

La figure 4.15 représente la distribution des valeurs des liens de toutes les simulations. On remarque que moins d'un cinquième des arêtes ont une valeur à 1. En d'autres termes, moins d'un cinquième des arêtes sont d'unique vecteurs de transmission, et donc moins d'un cinquième des arêtes ont une destination qui n'a qu'un unique transmetteur. Nous savons déjà toutes les arêtes participant au départ d'une simulation (à  $t=1$ ) font parti de cet ensemble d'arêtes de poids  $w = 1$ , puisque nous utilisons une seule source à chaque fois.

Nous observons aussi que 15% des arêtes ont la valeur de 0.5, et qu'il existe donc exactement deux transmetteurs pour la destination de ces arêtes à un temps donné. Ainsi, plus de 50% des arêtes ont une valeur inférieure strictement à 0.5. Cela signifie que la plupart du temps,

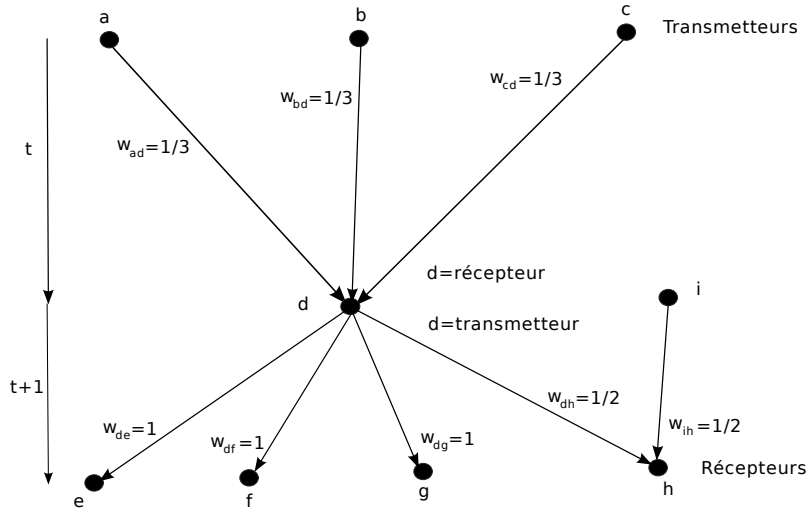


FIGURE 4.14 – Assignment du poids des arêtes.

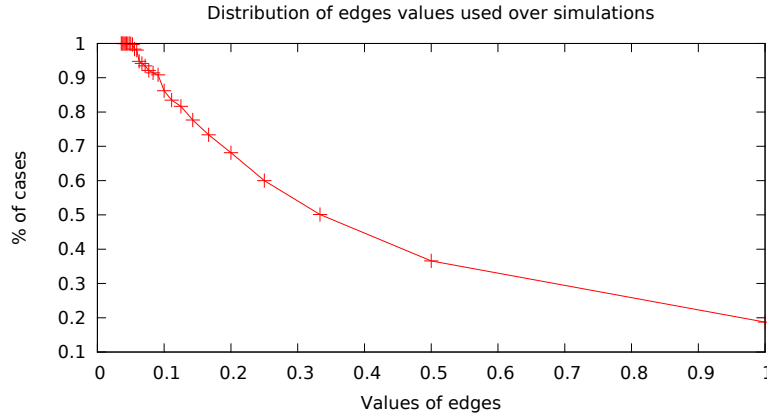


FIGURE 4.15 – Distribution de la valeur des arrêtes au cours des simulations.

un récepteur est infecté par plus de trois personnes à une heure donnée. Comme nous le redoutons, un grand nombre de possibilités permettant la transmission d'une diffusion semble exister.

Pour mieux cerner la diversité des sources de transmission, considérons la figure 4.16 qui représente la distribution du nombre de transmetteurs pour chaque récepteur. En d'autres termes, nous observons la distribution du nombre moyen par personnes d'arêtes entrantes pour chaque simulation.

On remarque que plus de 80% des destinations ont au moins 2 arêtes entrantes, ce qui signifie qu'au moins 2 sources différentes de transmission existent pour un même pas de temps. Les risques de propagation sont donc assez élevés car la plupart du temps, plusieurs sources sont à l'origine de la transmission.

La figure 4.17 donne la distribution du nombre moyen d'arêtes sortantes par personne et par date, c'est à dire le nombre de nœuds infectés par un nœud à un temps donné. L'immense majorité des simulations a une moyenne située entre 2 et 4 récepteurs par transmetteur : c'est à dire qu'ils transmettent l'infection à plusieurs nœuds durant la même heure.

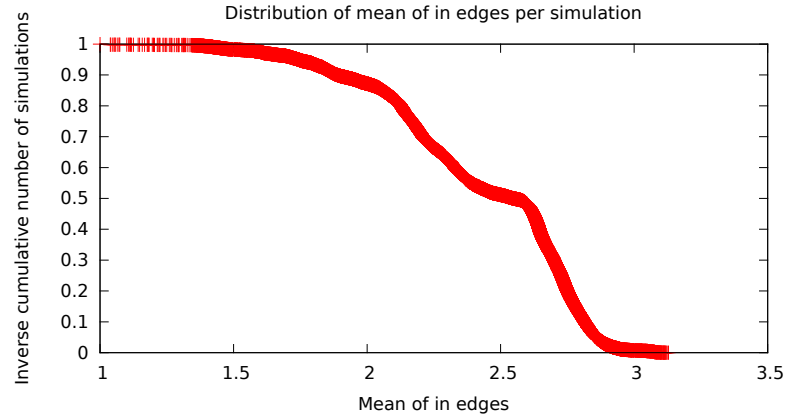


FIGURE 4.16 – Distribution du nombre moyen d’arêtes entrantes par personne pour chaque simulation, soit le nombre moyen de transmetteurs par récepteur.

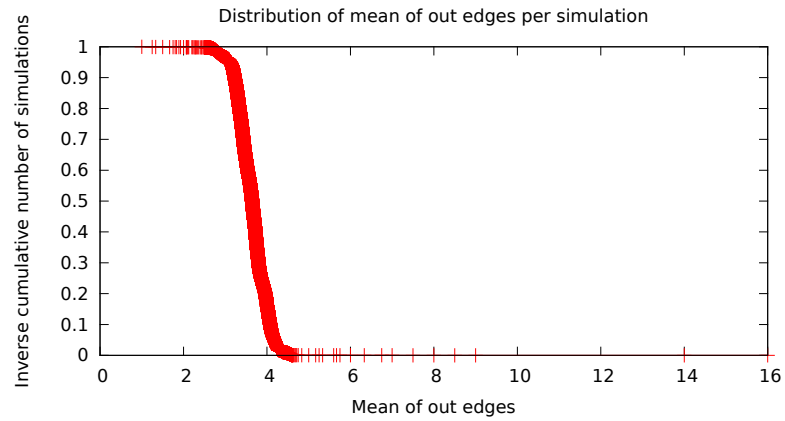


FIGURE 4.17 – Distribution du nombre d’arêtes sortantes, soit le nombre de récepteurs par transmetteur.

Ainsi, les sources de transmission sont multiples et variées ce qui dénote une certaine robustesse de la diffusion.

Les arêtes empruntées par la diffusion sont-elles toujours semblables ou bien varient-elles en fonction des sources d’infection et des temps de départ? Etudions la figure 4.18 qui représente la distribution, présentée sous la forme de courbe cumulative, de l’occurrence des arêtes au cours de toutes les simulations. Nous identifions une arête par sa date d’utilisation, sa source et sa destination. Considérons le nombre de cas où chacune des arêtes a servi dans la transmission. On constate que plus de 50% des arêtes ne sont pas utilisées par plus de dix simulations et donc sont négligeables dans la prise en compte du risque de contamination. Moins de 10% des arêtes sont utilisées presque systématiquement, mais elles existent tout de même et permettent la transmission. Ce sont des arêtes d’une importance capitale pour la diffusion car elles interviennent presque systématiquement. Dans le cas de mise en place de stratégies de vaccination afin d’enrayer une épidémie, il faudra identifier les sources de ces arêtes et les vacciner.

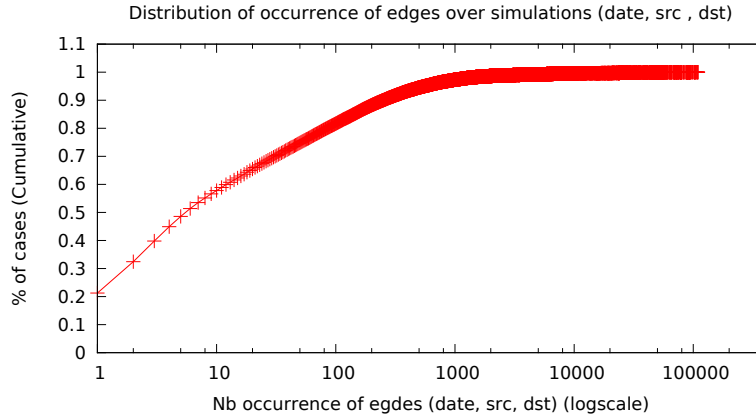


FIGURE 4.18 – Distribution (cumulative) du nombre d’occurrences des arêtes.

Nous avons constaté que quelques arêtes sont presque systématiquement utilisées pour la diffusion ce qui signifie qu’il y a peu d’alternatives dans les pas précédents pour atteindre la destination de ce type d’arêtes. Cela peut être le cas pour les personnes absentes pendant un long moment ou pour des arêtes faisant un pont entre deux groupes de nœuds plus connectés entre eux qu’avec l’extérieur de leur groupe (communauté). Nous savons aussi qu’il existe, la plupart du temps, plusieurs arêtes arrivant simultanément à une destination et plusieurs arêtes partant simultanément d’une même source. Nous cherchons à identifier les cas pour lesquels il est préférable d’isoler les contaminés, pour protéger les personnes encore saines et les cas où il faudrait vacciner les personnes non encore infectées, sachant qu’il est plus facile d’agir sur le plus petit groupe de personnes possible. Nous considérons les largeurs maximales en transmetteurs (beaucoup de sources d’infection sur un même pas de temps) et les largeurs maximales en récepteurs (beaucoup de nouveaux cas d’infection). La corrélation de ces deux valeurs, si l’une est plus importante que l’autre, nous permettrait d’émettre un premier avis sur la stratégie de vaccination à apporter.

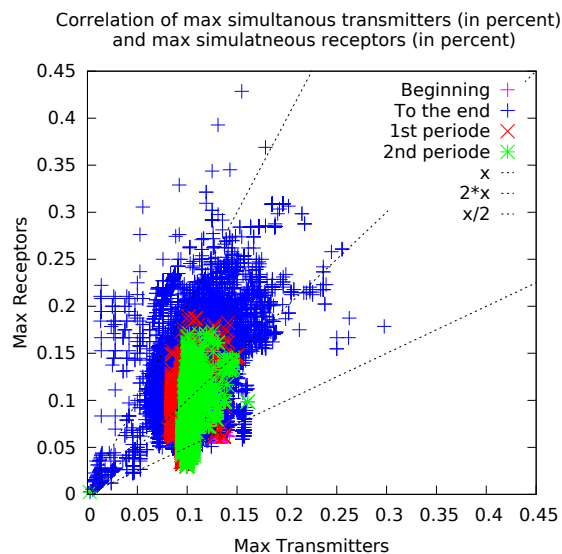
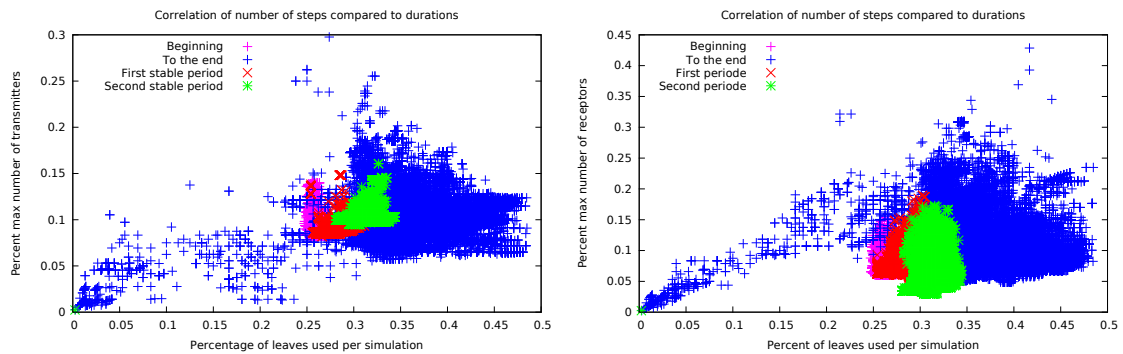


FIGURE 4.19 – Corrélation entre le maximum de transmetteurs simultanés et le maximum de récepteurs simultanés, en terme de pourcentage de population.



La figure 4.19 compare la largeur maximale en transmetteurs avec la largeur maximale en récepteurs, exprimée en proportion de la population. On constate que ces valeurs s'inscrivent dans un cône de corrélation. Considérons les périodes stables apparaissant en rouge et vert sur la figure. On constate qu'un grand nombre de simulations ont des largeurs maximales en récepteurs et transmetteurs assez équivalentes. Il existe aussi un grand nombre de simulations qui ont des largeurs maximales en transmetteurs plus de deux fois plus grande que les largeurs maximales en récepteurs ce qui signifie qu'il y a beaucoup de sources pour certaines transmissions. La diffusion est, dans ce cas, difficile à enrayer ou ralentir car les sources sont trop nombreuses. Au contraire, lorsque nous constatons que les largeurs maximales en réceptions sont plus de deux fois plus grandes qu'en transmissions, les mêmes sources de transmission sont sollicitées pour plusieurs récepteurs et une stratégie de ralentissement de la diffusion est possible : il suffit de vacciner ces quelques sources de transmission.

La figure 4.20 nous informe sur la corrélation qu'il peut exister entre le nombre de maximum de transmetteurs (sous-figure 4.20a) ou le nombre de récepteurs (sous-figure 4.20b) à un pas de temps pour une simulation et le nombre de feuilles à la fin de l'exécution. La figure



(a) Pourcentage du maximum de transmetteurs et de feuilles. (b) Pourcentage du maximum récepteurs et de feuilles.

FIGURE 4.20 – Pourcentage de population maximale étant transmetteurs ou récepteurs à un pas de temps donné comparé au pourcentage de la population étant une feuille à la fin de l'exécution.

4.20a compare la proportion de population de la largeur maximale en transmetteurs et la proportion de la population étant une feuille, ne devenant donc à aucun moment des transmetteurs, à la fin de la diffusion. La majorité des diffusions commençant durant une des deux périodes stables, représentées en rouge et en vert, ont une largeur maximale en transmetteurs comprise entre 5% de la population et 15% soit moins d'un quart de la population. En revanche, le pourcentage de la population étant des feuilles lors de la fin des simulations est compris entre 25 et 35 % de la population. La quantité de feuilles est, dans presque tous les cas, supérieure à la largeur maximale en transmetteurs.

Considérons la comparaison de la largeur maximale en récepteur des simulations avec la proportion de feuilles en fin de simulation (figure 4.20b). Les largeurs maximales en récepteurs concernant les deux périodes stables en terme de taille de population sont comprises entre 2 et 20 %. Certains de ces récepteurs sont susceptibles d'être aussi des feuilles. La proportion de feuilles est toujours comprise entre 25 et 35 % et est donc beaucoup plus importante que la largeur maximale en récepteurs. Cela signifie qu'un certain nombre de feuilles ont été infectées tout au long de la simulation et non pas sur les derniers pas de

temps. Les personnes infectées en milieu de simulation et restant des feuilles à la fin de la simulation sont des personnes qui ont un faible voisinage ou des voisins déjà touchés par la diffusion. Ces personnes sont sûrement peu connectées avec des personnes n'appartenant pas à leur communauté (service, par exemple) et ne sont donc pas les personnes recherchées pour les vacciner dans le but d'enrayer une épidémie.

La diffusion systématique appliquée aux données MOSAR nous donne ainsi une diffusion qui se joue sur presque toute la durée de la période testée, quelque soit la date et la source d'infection : la diffusion qui s'opère est donc lente à atteindre la totalité de la population du fait des absences de certains individus. Pourtant, la majorité des personnes sont touchées au début de l'exécution de cette diffusion qui par ailleurs est très robuste et donc semble difficile à enrayer si la transmission de l'infection est systématique. Nous allons étudier par la suite des diffusions avec des probabilités plus faibles de transmission.

### 4.3 Diffusions probabilistes et stratégies de vaccination

Dans cette partie, nous étudions plus en détail comment se déroule une simulation afin d'évaluer la vitesse de dissémination. Nous appuyons nos constatations, dans un premier temps, sur plusieurs simulations extraites des deux périodes stables identifiées précédemment. Sur ces deux périodes, nous avons en moyenne respectivement 35 et 37 % de la population étant des feuilles à la fin de l'exécution. Le degré entrant moyen pour les deux périodes se situe entre 3,32 et 3,34 alors que le degré sortant moyen se situe entre 2.0 et 2.1. Le fait que le degré entrant soit supérieur au degré sortant implique que l'infection est difficilement évitable car chaque personne transmet peu, dans un même temps, mais reçoit par contre l'infection de plusieurs sources différentes et de manière simultanée.

#### 4.3.1 Largeurs des DAG sur la diffusion systématique

Dans un premier temps, nous considérons quelques DAG obtenus par diffusion systématique et représentés sur la figure 4.21 qui résume l'évolution de la diffusion de quelques simulations de la première période stable en terme de population commençant à 9h00 pour certaines et 14h00 pour d'autres. Ces diffusions suivent les valeurs moyennes présentées en début de section. Le nombre de nœuds jouant le rôle de récepteurs à chaque pas de temps est représenté en rouge (valeurs négatives) et le nombre de transmetteurs en bleu (valeurs positives). Nous avons sélectionné différents temps de départ (9h et 14h) et différentes sources dont des patients et des personnels.

L'allure générale du DAG de diffusion est toujours la même, à savoir deux vagues durant lesquelles les largeurs croissent puis décroissent, l'une immédiatement après le départ de la diffusion et l'autre qui semble être proche du midi suivant le départ de la simulation. Ces constats n'étaient pas repérables lors de l'étude de l'ensemble de simulations, qui ne nous permettait pas de repérer les heures de repas. On observe en effet une première période d'augmentation des largeurs (pour les récepteurs comme pour les transmetteurs), puis une régression durant les premières heures suivant le début de la diffusion. Nous observons ensuite une phase très courte mais de très forte augmentation puis une décroissance des largeurs, et ce, toujours proche du milieu de journée. Ensuite nous observons un certain nombre de périodes durant lesquelles il n'y a pas d'évolution notable de la diffusion, à part quelques temps regroupés en îlots le matin ou proches de midi.

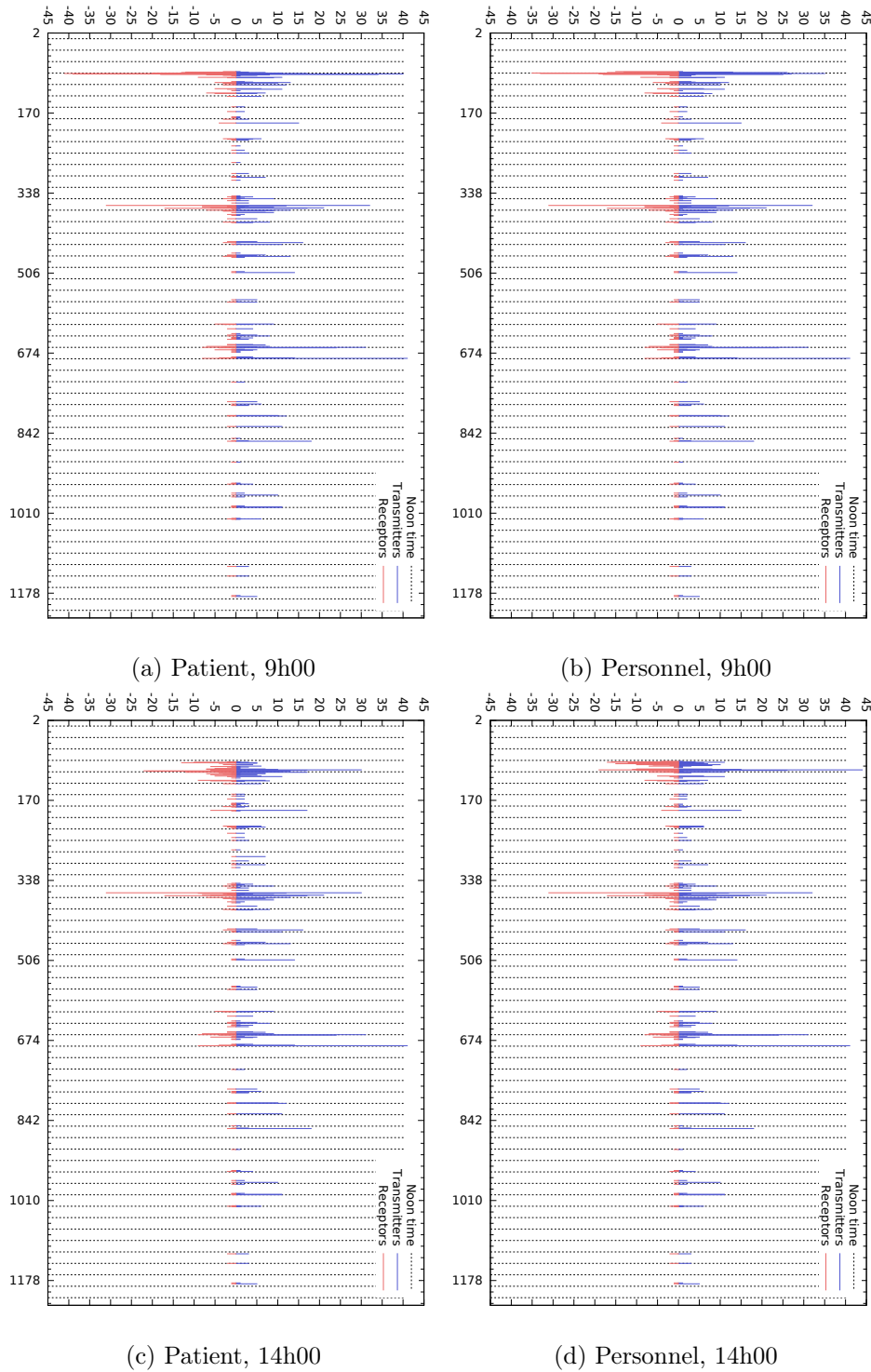


FIGURE 4.21 – Largeur des DAG au cours du temps, en terme de nombre de récepteurs (en haut) et nombre de transmetteurs (en bas). Toutes les simulations ont débuté durant la première période stable en terme de population. Les figures du haut ont commencé le matin contrairement à celle du bas qui ont commencé l'après-midi. Les diffusions apparaissant sur les figures de gauche sont initié à partir d'un patient, celle de droite, à partir d'un personnel.

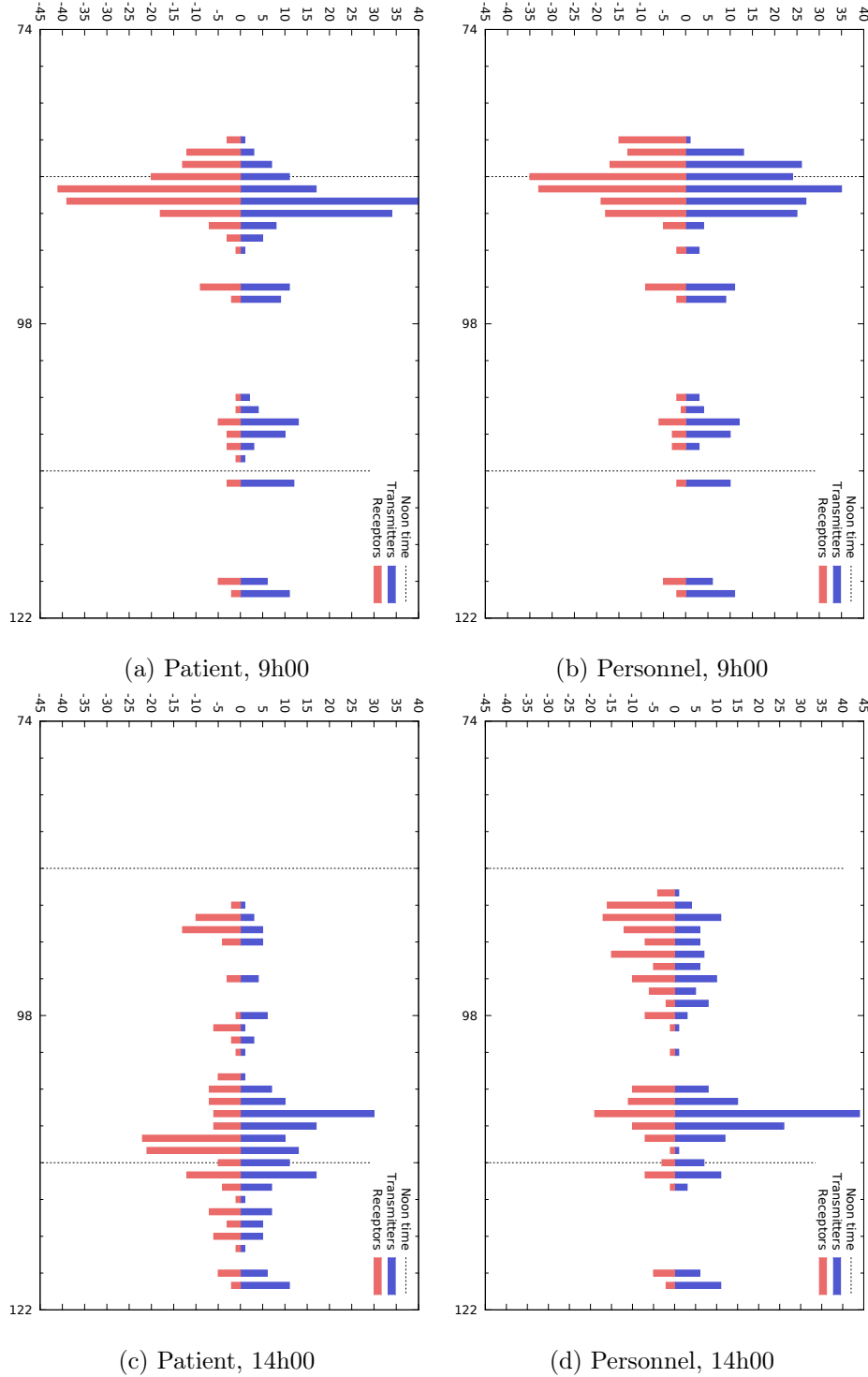


FIGURE 4.22 – Largeur des DAG au cours du temps, en terme de nombre de récepteurs (en haut) et nombre de transmetteurs (en bas). Toutes les simulations ont débuté durant la première période stable en terme de population.

Les évolutions du nombre de transmetteurs et du nombre de récepteurs semblent très corrélées sur toute la durée de la diffusion bien qu'elles ne suivent pas toujours les mêmes échelles. On observe aussi quelques décalages en temps des observations faites pour les transmetteurs par rapport à celles que l'on peut faire au sujet des récepteurs. On assiste

parfois à des temps de transmission très soutenue sans qu'il n'y ait énormément de récepteurs et inversement des temps pour lesquels il y a peu de transmetteurs, mais énormément de réceptions. Après le premier temps de midi suivant le départ de la diffusion, une grosse majorité des personnes sont touchées et il y a de plus en plus de temps durant lesquels la diffusion ne progresse pas.

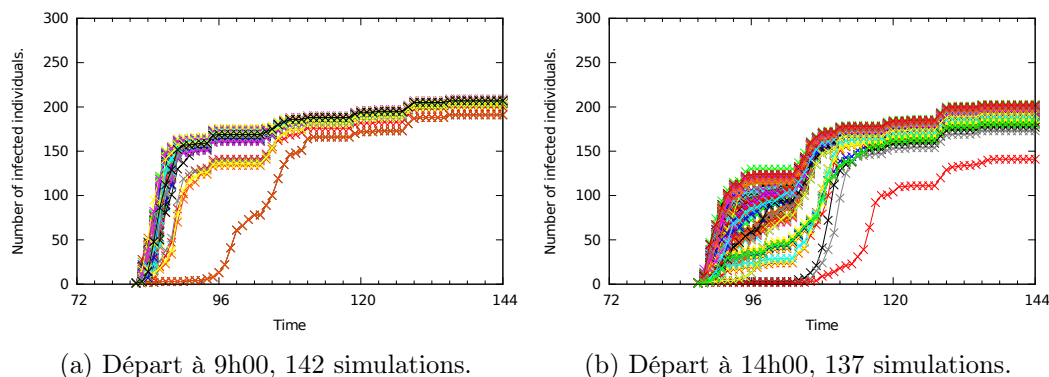


FIGURE 4.23 – Nombre de personnes touchées au cours du temps de toutes les simulations débutant aux pas de temps 81 et 86, durant la première période stable

Considérons maintenant le zoom (cf figure 4.22) sur les parties de forte diffusion de chacune des simulations que nous venons d'observer. Pour les simulations, commençant à 9h00, on remarque que les pics de transmission les plus importants, que ce soit en terme de récepteurs ou en terme de transmetteurs, ont lieu en début d'après-midi, entre 13h et 15h. Pour les simulations commençant en début d'après-midi, les plus gros pics s'observent plutôt entre 8h et 9h en terme de transmetteurs tandis qu'en terme de récepteurs, ces pics s'observent sur une plage horaire un peu plus étendue (8h et 11h00).

Ces deux plages de pics de transmissions correspondent aux pics d'activité que nous avons remarqués lors de l'analyse de l'évolution temporelle.

L'évolution de la diffusion, à une heure donnée, semble représentative pour les moments sélectionnés précédemment (des simulations commençant à 9h00 et des simulation débutant à 14h00). La plupart des diffusions sont assez rapides dans le sens où une bonne partie de la population est infectée en moins de deux jours, et souvent même moins d'une journée. En effet, nous présentons sur la figure 4.23 l'évolution du nombre d'individus touchés par la diffusion au cours du temps pour toutes les simulations commençant à 9h00 du matin, un jour durant la période stable (figure 4.23a).

Nous présentons ici uniquement la version zoomant sur les deux premiers jours, car cette période est suffisante pour atteindre un palier stable. Chaque graduation correspond à un jour à minuit ; les sous-graduations apparaissent toutes les trois heures. On remarque que pour toutes les simulations, une augmentation rapide dès le début, avec une pente très forte durant le régime transitoire, dénotant une très forte augmentation de la population atteinte par la diffusion, suivie d'un palier presque horizontal après que la majorité des individus aient été infectés. Le régime transitoire se fait presque immédiatement pour la quasi-totalité des simulations commençant à 9h00 (figure 4.23a). Seule deux simulation font exception et ont le même comportement tout au long de leur évolution : elle démarrent leur phase de transition le jour suivant leur départ et ce départ se fait à partir de deux personnels du service administratif. Le régime transitoire est un peu plus tardif et plus étalé

pour les simulations commençant à 14h (figure 4.23b). On observe en fait deux comportements différents : les simulations qui commencent presque immédiatement par leur phase de transitions et celles qui la commencent le jour suivant, vers 9h00. Il serait intéressant d'observer l'évolution de toutes les simulations commençant aux différentes heures de la journée, tous les jours confondus. Cela demanderait un travail poussé pour optimiser la mémoire et faire un affichage qui ne soit pas trop lourd, c'est pourquoi nous avons préféré présenter un exemple simple dans ce manuscrit.

Considérons maintenant l'évolution des simulations de la deuxième période stable, durant laquelle la population globale de l'hôpital est plus importante. Les résultats sont résumés par les figures 4.25 et 4.26. Les plages de fortes transmissions sont assez similaires à celles observées pour la première période stable. On constate tout de même que le temps de midi est aussi un temps très fort de transmission lorsque la simulation commence à 9h.

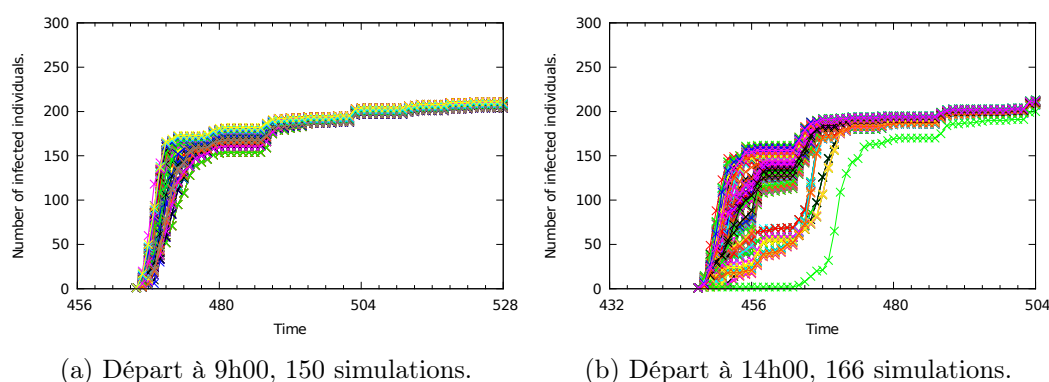


FIGURE 4.24 – Nombre de personnes touchées au cours du temps de toutes les simulations débutant aux pas de temps 81 et 86, durant la première période stable

La figure 4.25 représente l'évolution de la largeur des DAG de certaines simulations commençant lors de la deuxième période stable.

Comme pour la première période stable, nous présentons l'évolution de toutes les simulations commençant les jours de départ des simulations dont nous avons présenté les DAG. Ces simulations commencent toutes à 9h00 sur la figure 4.24a et toutes à 14h00 sur la figure 4.24b. On remarque là encore, que l'ensemble des simulations commençant en début de journées terminent leur régime transitoire à la fin de cette journée. Ce n'est pas toujours le cas des simulations commençant après le milieu de journée, qui ont deux types de comportements différents : celles qui se propagent dans la journée et celles qui commencent leur régime transitoire le lendemain. Les premières sont les plus nombreuses (138 simulation). Les deuxièmes sont en faible nombre et toutes ont un départ à partir d'une personne appartenant à un service de soins. La plupart de ces simulations démarrent à partir d'un patients (seules quatre simulations de ce type démarrent à partir d'un personnel qui est soit un infirmier soit un aides soignants). Dans tous les cas, la majorité de la population est infectée en moins de deux jours.

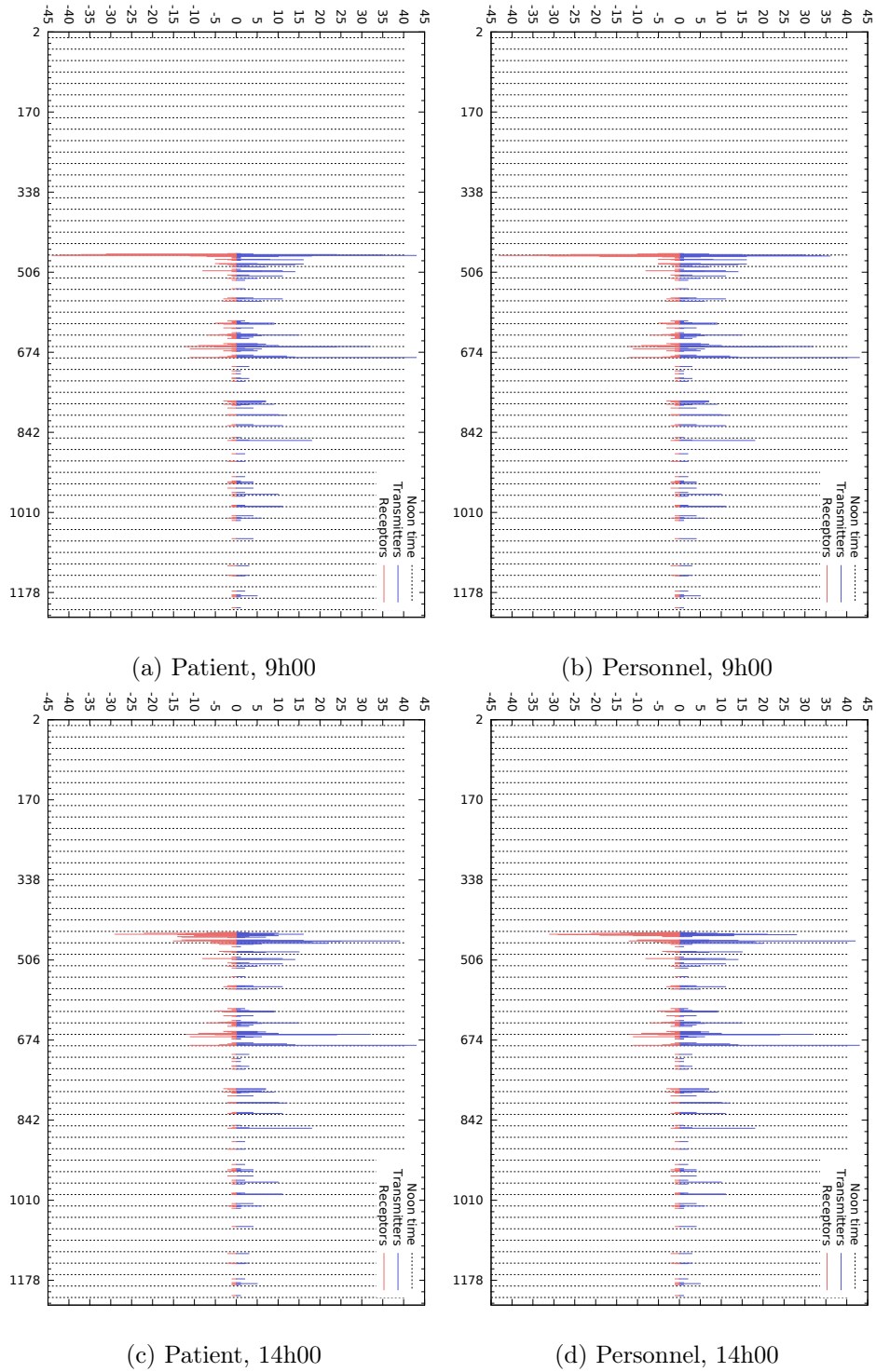


FIGURE 4.25 – Largeur des DAG au cours du temps, en terme de nombre de récepteurs (en haut) et nombre de transmetteurs (en bas). Toutes les simulations ont débuté durant la deuxième période stable en terme de population. Cette population est importante en regard de la moyenne habituelle de l'hôpital.

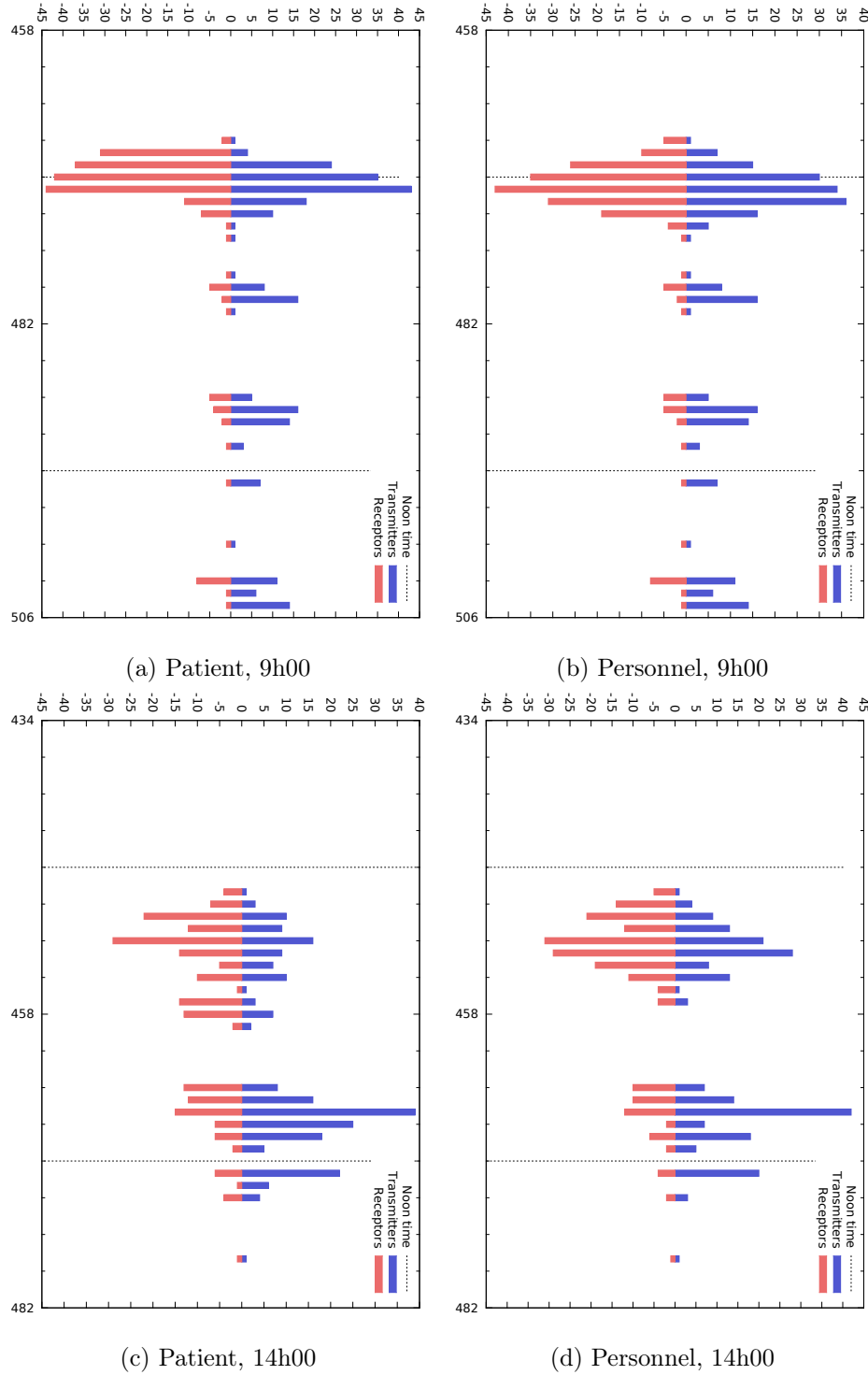


FIGURE 4.26 – Largeur des DAG au cours du temps, en terme de nombre de récepteurs (en haut) et nombre de transmetteurs (en bas). Toutes les simulations ont débuté durant la deuxième période stable en terme de population. Cette population est importante en regard de la moyenne habituelle de l’hôpital.



### 4.3.2 DAG obtenus par diffusion probabiliste.

Nous savons que la transmission de bactéries n'est jamais systématique et n'a donc pas nécessairement lieu au premier contact d'un individu avec un autre. Il est donc plus réaliste, lors de la simulation de diffusions à travers un réseau social, d'introduire des probabilités de transmission. Nous ne pouvons connaître quel est le taux réel exact des transmissions. Nous pouvons par contre comparer le comportement du réseau et de la diffusion en fonction de la probabilité  $p$  de transmission et ainsi repérer les temps et patterns importants du réseau pour qu'une transmission soit possible.

#### Principe de la diffusion probabiliste.

Les diffusions probabilistes que nous avons exécutées sur les données MOSAR fonctionnent comme suit.

##### Réseau

Nous avons agrégé les données à l'heure et travaillé sur le réseau agrégé non pondéré. Notre réseau temporel final est donc constitué d'arêtes (temps  $T$ , individu  $P1$ , individu  $P2$ ) existant entre paires de sommets si et seulement si il a existé une ou plusieurs interactions entre les deux individus dans l'intervalle  $[T, T+1h]$ . Chaque simulation est identifiable par sa source (première personne qu'on infecte arbitrairement au lancement de la simulation) et par son temps de départ.

##### Diffusion

La diffusion s'opère à partir du premier temps choisi comme départ de la simulation. Ensuite, tous les liens existant à ce temps sont parcourus. Si l'une des extrémité du lien est marquée comme infectée, et l'autre est saine, alors le nœud sain est marqué comme infecté si, lors du tirage au sort, le nombre obtenu est inférieur à la probabilité de transmission choisie pour la simulation.

Chaque exécution d'une simulation probabiliste étant différente de la précédente, nous nous intéressons au comportement de plusieurs simulations conduites à partir d'une même source et d'un même temps. Chaque simulation est ainsi répétée une centaine de fois afin d'avoir suffisamment d'échantillons pour que les moyennes soient significatives.

#### Résultats de la diffusion probabiliste.

Nous avons choisi d'observer le résultat de l'exécution d'une simulation, commençant à un instant  $T$ , avec une source  $S$  données, en faisant varier la probabilité  $p$  de transmission. Nous considérerons ensuite les résultats moyens de ces diffusions.

Nous considérons les probabilités 0.5 et 0.1 et nous comparons les largeurs de DAG obtenues pour ces 2 probabilités ainsi que le DAG obtenu avec transmission systématique.

Le résultat, sur la figure 4.27 fait penser à un signal atténué. C'est à dire que plus la probabilité de transmission est faible, plus faible est la largeur maximale, que ce soit en terme de transmetteur ou récepteur. Par contre les moments de forte transmission au cours du temps sont plus nombreux et ont toujours lieu dans la journée et la plupart du temps en milieu de journée. En comparant avec le DAG obtenu après une diffusion avec transmission

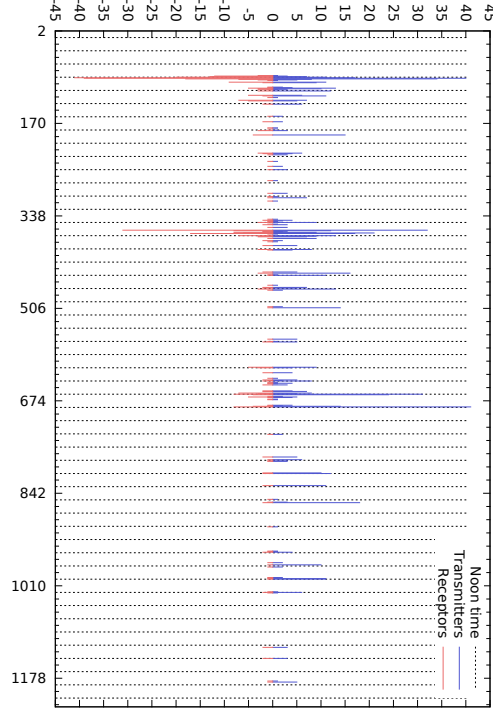
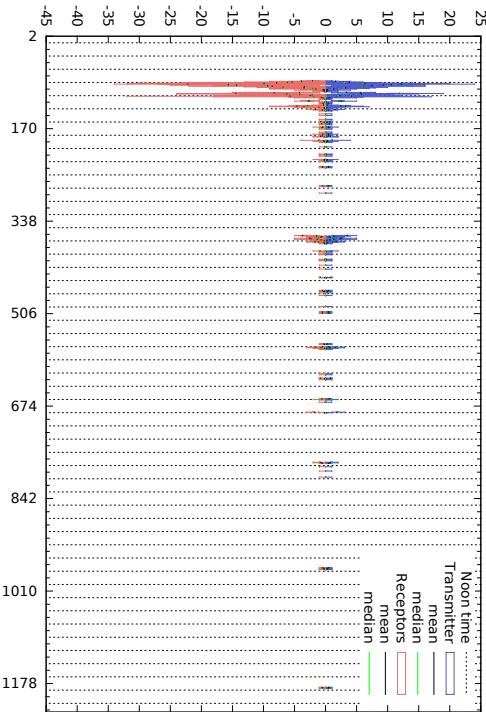
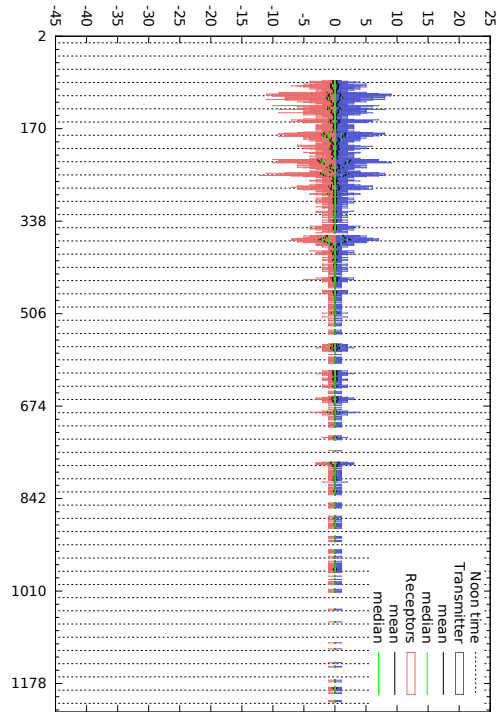

 (a) Probabilité  $p=1$ 

 (b) Probabilité  $p=0.5$ 

 (c) Probabilité  $p=0.1$ 

FIGURE 4.27 – Largeur de DAG pour les probabilités 1 (sous-figure 4.27a), 0,5 (sous-figure 4.27b) et 0,1 (sous-figure 4.27c) : le nombre de récepteurs en haut et le nombre de transmetteurs en bas. Les boîtes bleues et rouges donnent les valeurs minimales et maximales obtenues sur l'ensemble des simulations.

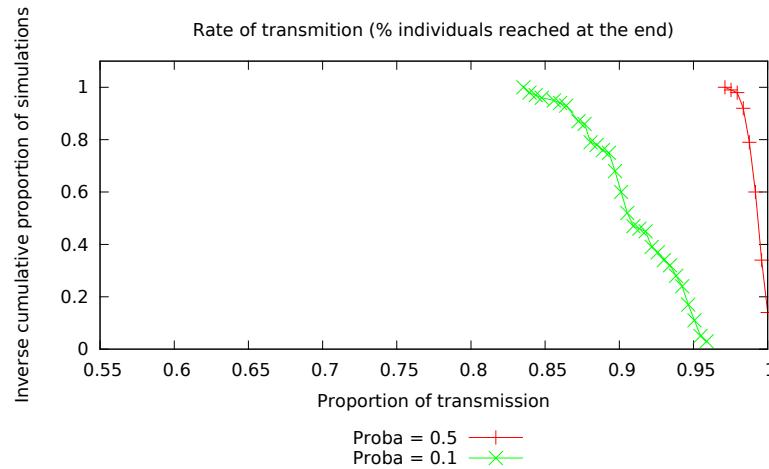


FIGURE 4.28 – Taux d’infection à la fin des simulations. Chaque courbe représente le résultat pour les simulations avec probabilité de transmission différente.

systématique, on constate que les temps pour lesquels l’activité était importante sont aussi ceux qui ressortent lors d’une diffusion probabiliste. Ainsi, pour ce réseau, afin de repérer le temps critique pour une diffusion, il serait presque suffisant d’étudier la forme du DAG d’une diffusion à transmission systématique ( $p = 1$ ), en considérant comme important tout temps durant lequel il existe au moins une transmission.

Le taux moyen d’infection à la fin des simulations reste très élevé, même pour les diffusions à probabilité de transmission 0.1 puisqu’il dépasse largement 80% pour toutes les simulations. Le taux d’infection est au-dessus de 95% lorsque le taux de transmission est fixé à 0.5, ce qui est très proche de ce que nous avons avec la transmission systématique. Les taux moyens observés pour une simulation sont résumés sur la figure 4.28. Réduire le taux d’infection de moitié ou le diviser par 10 ne semble pas suffisant pour enrayer une diffusion. Cela confirme qu’il existe un grand nombre de possibilités pour transmettre l’infection et donc que le réseau est relativement bien connecté, au moins sur les plages de temps stables en terme de taille de population auxquelles nous nous sommes intéressés.

### En résumé

Dans ce chapitre, nous avons proposé des critères permettant de caractériser en partie la propension du réseau de proximité humaine des données MOSAR à la diffusion. Nous avons d’abord considéré les caractéristiques d’une diffusion à transmission systématique pour identifier les temps forts, souvent situés en milieu de journée, la vitesse et le taux d’infection de la diffusion ainsi que sa robustesse. Nous avons ensuite constaté que la plupart des temps forts et importants de la diffusion identifiés dans le cas d’une transmission systématique le sont toujours dans le cas de diffusions probabilistes. Ces dernières s’étalent par contre sur des périodes d’activité plus longues et ont moins de temps de non transmission. Elles sont pourtant presque aussi efficaces en terme de taux d’infection final et font ressortir les mêmes pas de temps importants pour la transmission : principalement les heures de bureaux durant lesquelles toutes les catégories professionnelles employées par l’hôpital sont représentées et le soir, peu après le changement d’équipe de personnels. Il serait très intéressant de mettre en place un outil qui permette de comparer la structure des DAG obtenus par diffusion systématique pour décrire le phénomène de diffusion. Cela pourrait se faire à l’aide de la théorie du signal, en considérant le DAG comme un signal.



## Chapitre 5

# Conclusions et perspectives

Durant cette thèse, nous nous sommes intéressés aux propriétés des graphes dynamiques en terme de structure comme en terme de diffusion. Nous avons constaté que les modèles simples de populations n'ont pas les mêmes propriétés que les populations réelles qui sont souvent bien structurées, en fonction du contexte duquel elles sont extraites. Beaucoup de travaux sur l'étude des populations sont faits souvent dans le contexte d'étude épidémiologiques. Afin de mieux prévoir et comprendre les diffusions qui peuvent se jouer au sein de populations réelles, il est important d'identifier les propriétés particulières des populations dues au contexte particulier étudié et pour ce faire de continuer l'effort fait depuis quelques années pour collecter, analyser et classer différents types de populations et leurs risques spécifiques de diffusion. Nous avons ici travaillé sur l'extraction et la mise en valeur de la structure d'un graphe dynamique issu d'un contexte particulier : celui d'un hôpital où les séjours sont de longue durée.

### **Une structure spatiale et fonctionnelle**

L'étude de cette structure a pour but de mieux comprendre l'organisation interne d'un hôpital et d'identifier les éléments caractéristiques du réseau humain dans le but ensuite de pouvoir renseigner des modèles afin d'étudier les risques de diffusion de maladies nosocomiales transmissibles par voies orales. L'étude de plusieurs dimensions était nécessaire. Dans un premier chapitre, nous nous sommes intéressés à deux dimensions qui sont les dimensions spatiales, entre différents services d'un hôpital, puis la dimension fonctionnelle pour mieux comprendre les interactions entre les différents corps de métier ainsi qu'entre patients et personnels. Il en ressort que les services de soins ont une activité interne très importante comparée à l'activité qui peut exister entre différents services.

### **Rythmes d'activité**

Dans un deuxième chapitre, nous avons développé des analyses temporelles sur des données agrégées à l'heure qui mettent en avant un rythme circadien et hebdomadaire très clair. En effet, cela permet de retrouver, entre autre, les heures de changement d'équipe de personnel par de nombreux contacts courts entre personnels, ainsi que le milieu de journée par des pics d'activité chez les personnels comme les patients. Nous avons aussi considéré les données agrégées à la journée, qui aboutissent généralement à un réseau connexe avec un diamètre assez grand.

### **Des mesures de similarité**

Nous avons considéré plusieurs mesures pour évaluer le taux de similarité d'activité d'un

jour à l'autre. Les deux échelles d'agrégation sont toutes deux significatives car le rythme de l'hôpital est basé sur ces deux grandeurs. Il serait intéressant de travailler sur une méthode qui permette, si cela est possible, d'extraire d'autres échelles d'agrégation significatives. Une échelle significative, étant une échelle qui permette de résumer le réseau dynamique, sans perdre trop d'informations.

### **Des propriétés de diffusion**

Dans un dernier chapitre, nous avons commencé une étude de risques de diffusion sur les données agrégées à l'heure, afin de caractériser les propriétés de diffusion du réseau de contacts, en utilisant le modèle de diffusion le plus simple qui soit : le modèle SI. Ce modèle permet d'étudier dans le pire cas, le taux de transmission au sein d'une population. La vitesse de diffusion est aussi un point important lors d'études épidémiologiques, qui se font, dans notre cas en grande majorité un jour ou deux, ce qui laisse peu de temps pour réagir si le virus ou la bactérie a une capacité de transmission élevée, à partir d'un unique cas touché. Une notion abordée, qui demande à être investie plus en profondeur, est la robustesse de la transmission et l'extraction, si possible, de nœuds plus importants que d'autres dans la transmission. Nous avons proposé plusieurs mesures pour évaluer le taux d'infection, la robustesse de la diffusion ainsi que sa vitesse de diffusion dans le pire cas. La diffusion SI nous permet, pour chaque simulation, de construire un graphe orienté acyclique, dont les propriétés topologiques méritent d'être étudiées.

### **Perspectives**

L'analyse du comportement de la dissémination sur les données MOSAR mérite un travail plus approfondi qui permettrait de mieux appréhender la vitesse de propagation des diffusions. Le nombre de diffusions jouées étant considérable (plus de 100000), il est nécessaire d'utiliser des outils optimisés pour étudier les comportements moyens des simulations. Beaucoup de travaux peuvent encore aussi être conduits sur les propriétés topologiques des graphes acycliques issus de simulations de diffusion. Nous aimerions en particulier établir une méthode qui permette de comparer les graphes acycliques obtenus et étudier la robustesse de leur connexité. Une autre piste de réflexion, est d'établir une méthode pour les réseaux dynamiques permettant de repérer des nœuds plus actifs dans la diffusion que d'autres, ce qui n'est à priori pas trivial après les analyses déjà poursuivies.

Un grand nombre de travaux visant à définir la centralité ou l'importance des nœuds dans un réseau ont été conduits sur des réseaux statiques [14, 58]. Établir des mesures pour les réseaux dynamiques, inspirées des méthodes classiques sur les graphes statiques est un premier pas. Une deuxième étape serait de vérifier que ces mesures mènent à des résultats cohérents et plus précis que ceux acquis avec les méthodes statiques, et confirme encore une fois l'importance de considérer l'aspect temporel des réseaux, encore trop souvent laissé de côté aujourd'hui. Il serait intéressant d'établir une mesure de centralité permettant de caractériser la centralité des nœuds en temps et en espace.

La validation de politiques de vaccination est aussi un enjeu important de l'étude des diffusions épidémiologiques. Il serait donc judicieux, après avoir identifié les individus centraux pour la diffusion, de construire des modèles de diffusion en introduisant une politique de vaccination. L'analyse du comportement des systèmes ainsi construits permet, d'une part, de vérifier la mesure de centralité spatiaux-temporelle et, d'autre part, de qualifier l'efficacité de certaines stratégies de vaccination.

# Bibliographie

- [1] R. ALBERT et A.-L. BARABÁSI : Statistical mechanics of complex networks. *Rev. Modern Phys.*, 74(1):47–97, 2002.
- [2] R. ALBERT, H. JEONG et A.-L. BARABÁSI : Internet : Diameter of the world-wide web. *Nature*, 401(6749):130–131, sept. 1999.
- [3] G. AN : In silico experiments of existing and hypothetical cytokine-directed clinical trials using agent-based modeling. *Crit Care Med*, 32(10):2050–2060, oct. 2004.
- [4] L. BACKSTROM, P. BOLDI, M. ROSA, J. UGANDER et S. VIGNA : Four degrees of separation. *CoRR*, abs/1111.4570, 2011.
- [5] S. C. BANKES : Agent-based modeling : A revolution ? *Knowledge Creation Diffusion Utilization*, 99:7199–7200, 2002.
- [6] S. BANSAL, J. READ, B. POURBOHLOUL et L. A. MEYERS : The dynamic nature of contact networks in infectious disease epidemiology. *J Biol Dyn*, 4(5):478–489, Sep 2010.
- [7] F. BAQUERO, M.-C. NEGRI, M.-I. MOROSINI et J. BLÁZQUEZ : Antibiotic-selective environments. *Clinical Infectious Diseases*, 27:pp. S5–S11, 1998.
- [8] A.-L. BARABÁSI : *Linked : The New Science of Networks*. Perseus Publishing, Cambridge, MA, 2002.
- [9] A. BARRAT, M. BARTHÉLEMY et A. VESPIGNANI : *Dynamical processes on complex networks*. Cambridge University Press, Cambridge, 2008.
- [10] E. A. BENDER et E. CANFIELD : The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296 – 307, 1978.
- [11] B. J. L. BERRY, L. D. KIEL et E. ELLIOTT : Adaptive agents, intelligence, and emergent human organization : Capturing complexity through agent-based modeling. *Proceedings of the National Academy of Sciences*, 99:7187–7188, 2002.
- [12] E. BERTIN, G. BESLON, O. GANDRILLON, S. GRAUWIN, P. JENSEN et N. SCHABANEL : Les complexités : point de vue d’un institut des systèmes complexes. *Hermès*, (60):145–150, juin 2011.
- [13] V. D. BLONDEL, J.-L. GUILLAUME, R. LAMBIOTTE et E. LEFEBVRE : Fast unfolding of communities in large networks. *Journal of Statistical Mechanics : Theory and Experiment*, 2008(10):P10008 (12pp), 2008.
- [14] P. BONACICH et P. LLOYD : Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23:191–201, 2001.
- [15] M. BONTEN, S. SLAUGHTER, A. A. ans M.K. HAYDEN, J. van VOORHIS et AL. : The role of "colonization pressure" in the spread of vancomycin-resistant enterococci : an important infection control variable. *Arch. Intern. Med.*, 158:1127–32, 1998.

- 
- [16] M. J. BONTEN, D. J. AUSTIN et M. LIPSITCH : Understanding the spread of antibiotic resistant pathogens in hospitals : mathematical models as tools for control. *Clin Infect Dis*, 33(10):1739–1746, Nov 2001.
  - [17] M. BOOTSMA, M. BONTEN, S. NIJSSEN, A. FLUIT et O. DIEKMANN : An algorithm to estimate the importance of bacterial acquisition routes in hospital settings. *Am. J. Epidemiol.*, 166:841–51, 2007.
  - [18] E. BROOKS-POLLOCK, G. O. ROBERTS et M. J. KEELING : A dynamic model of bovine tuberculosis spread and control in great britain. *Nature*, 511(7508):228–231, Jul 2014.
  - [19] G. CALDARELLI, D. GARLASCHELLI et L. PIETRONERO : Food web structure and the evolution of complex networks. In R. PASTOR-SATORRAS, M. RUBI et A. DÍAZ-GUILERA, édés : *Statistical Mechanics of Complex Networks*, num. 625 de Lecture Notes in Physics, p. 148–166. Springer-Verlag, 2003.
  - [20] C. CATTUTO, W. V. den BROECK, A. BARRAT, V. COLIZZA, J.-F. PINTON et A. VESPIGNANI : Using sensor networks to study the effect of peripatetic health-care workers on the spread of hospital-associated infections. *PLoS ONE*, 5:e11596, 2010.
  - [21] S.-H. CHA : Comprehensive survey on distance/similarity measures between probability density functions, 2007.
  - [22] F. R. CHUNG et L. LU : *Complex graphs and networks*, vol. 107. American mathematical society Providence, 2006.
  - [23] A. CLAUSET : Finding local community structure in networks. *Phys. Rev. E*, 72:026132, Aug 2005.
  - [24] V. COLIZZA, A. BARRAT, M. BARTHÉLEMY et A. VESPIGNANI : The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci U S A*, 103(7):2015–2020, Feb 2006.
  - [25] A. J. K. CONLAN, K. T. D. EAMES, J. A. GAGE, J. C. von KIRCHBACH, J. V. ROSS, R. A. SAENZ et J. R. GOG : Measuring social networks in British primary schools through scientific engagement. *Proceedings of the Royal Society B : Biological Sciences*, 278(1711):1467–1475, mai 2011.
  - [26] T. H. CORMEN, C. E. LEISERSON, R. L. RIVEST et C. STEIN : *Introduction to algorithms*. MIT Press, Cambridge, MA, third éd., 2009.
  - [27] E. DAVID et K. JON : *Networks, Crowds, and Markets : Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA, 2010.
  - [28] P. de MAN, E. van DER VEEKE, M. LEEMREIJZE, W. van LEEUWEN, G. VOS et AL. : Enterobacter species in a pediatric hospital : horizontal transfer or selection in individual patients ? *J. Infect. Dis.*, 184:211–4, 2001.
  - [29] K. EAMES, S. BANSAL, S. FROST et S. RILEY : Six challenges in measuring contact networks for use in modelling. *Epidemics*, 10:72–77, Mar 2015.
  - [30] N. ERIKSEN, F. ESPERSEN, V. ROSDAHL et K. JENSEN : Carriage of staphylococcus aureus among 104 healthy persons during a 19-month period. *Epidemiology and Infection.*, 115:51–60, 1995.
  - [31] S. EUBANK, H. GUCLU, V. S. ANIL KUMAR, M. V. MARATHE, A. SRINIVASAN, Z. TOROCZKAI et N. WANG : Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, mai 2004.
-



- [32] N. FERGUSON : Capturing human behaviour. *Nature*, 446(7137):733, Apr 2007.
- [33] N. M. FERGUSON, D. A. T. CUMMINGS, C. FRASER, J. C. CAJKA, P. C. COOLEY et D. S. BURKE : Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452, avr. 2006.
- [34] S. GANDON et P. A. VAN ZANDT : Local adaptation and host-parasite interactions. *Trends in Ecology and Evolution*, 13:214–216, 1998.
- [35] D. M. GORMAN, J. MEZIC, I. MEZIC et P. J. GRUENEWALD : Agent-based modeling of drinking behavior : a preliminary model and potential applications to theory and practice. *American journal of public health*, 96(11):2055–2060, nov. 2006.
- [36] J. F. GRACIA, J.-P. ONNELA, M. L. BARNETT, V. M. EGUÁLUZ et N. A. CHRISTAKIS : Spread of pathogens in the patient transfer network of US hospitals, avr. 2015.
- [37] S. GRAUWIN, G. BESLON, E. FLEURY, S. FRANCESCHELLI, C. ROBARDET, J.-B. ROUQUIER et P. JENSEN : Complex Systems Science : Dreams of Universality, Reality of Interdisciplinarity. *CoRR*, abs/1206.2216, 2012.
- [38] A. HARRIS, M. KOTETISHVILI, S. SHURLAND, J. JOHNSON, J. MORRIS et AL. : How important is patient-to-patient transmission in extended-spectrum beta-lactamase escherichia coli acquisition. *Am. J. Infect. Control.*, 35:97–101, 2007.
- [39] A. HARRIS, E. PERENCEVICH, J. JOHNSON, D. PATERSON, J. MORRIS et AL. : Patient-to-patient transmission is important in extended-spectrum beta-lactamase-producing klebsiella pneumoniae acquisition. *Clin. Infect. Dis.*, 45:1347–50, 2007.
- [40] M. HASHEMIAN, K. STANLEY et N. OSGOOD : Leveraging h1n1 infection transmission modeling with proximity sensor microdata. *BMC Medical Informatics and Decision Making*, 12(35), 05 2012.
- [41] P. HOLME : Network reachability of real-world contact sequences. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 71(4), 2005.
- [42] T. HORNBECK, D. NAYLOR, A. SEGRE, G. THOMAS, T. HERMAN et P. POLGREEN. : Using sensor networks to study the effect of peripatetic healthcare workers on the spread of hospital-associated infections. *The Journal of Infectious Diseases*, 206:1549–1557, 10 2012.
- [43] N. HUPERT, W. XIONG et A. MUSHLIN : The virtue of virtuality : the promise of agent-based epidemic modeling. *Transl Res*, 151(6):273–274, Jun 2008.
- [44] L. ISELLA, M. ROMANO, A. BARRAT, C. CATTUTO, V. COLIZZA, W. V. den BROECK, F. GESUALDO, E. PANDOLFI, L. RAVÀ, C. RIZZO et A. E. TOZZI : Close encounters in a pediatric ward : Measuring face-to-face proximity and mixing patterns with wearable sensors. *PLoS ONE*, 6(2):e17144, 02 2011.
- [45] L. KELLER et M. G. SURETTE : Communication in bacteria : an ecological and evolutionary perspective. *Nat Rev Microbiol*, 4(4):249–258, Apr 2006.
- [46] J. KIM, D. H. LEE, Y. S. SONG, S. W. KANG et S. W. KIM : The within-host population dynamics of normal flora in the presence of an invading pathogen and antibiotic treatments. *J Microbiol Biotechnol*, 17(1):146–153, Jan 2007.
- [47] J. KLEINBERG : Complex networks and decentralized search algorithms. In *In Proceedings of the International Congress of Mathematicians (ICM)*, 2006.
- [48] J. KLUYTMANS, A. van BELKUM et H. VERBRUGH : Nasal carriage of staphylococcus aureus : epidemiology, underlying mechanisms, and associated risks. *Clin. Microbiol. Rev.*, 10:505–520, 07 1997.

- 
- [49] J. LESKOVEC, K. J. LANG et M. MAHONEY : Empirical comparison of algorithms for network community detection. *In Proceedings of the 19th International Conference on World Wide Web*, WWW '10, p. 631–640, New York, NY, USA, 2010. ACM.
  - [50] J. LESKOVEC, M. MCGLOHON, C. FALOUTSOS, N. GLANCE et M. HURST : Cascading behavior in large blog graphs : Patterns and a model. *Rap. tech.*, 2006.
  - [51] B. R. LEVIN et R. ANTIA : Why we don't get sick : the within-host population dynamics of bacterial infections. *Science*, 292(5519):1112–1115, May 2001.
  - [52] J.-C. LUCET, C. LAOUEANAN, G. CHELIUS, N. VEZIRIS, D. LEPELLETIER, A. FRIGGERI, D. ABITEBOUL, E. BOUVET, F. MENTRÉ et E. FLEURY : Electronic sensors for assessing interactions between healthcare workers and patients under airborne precautions. *PLoS ONE*, 7(5):e37893, 2012.
  - [53] L. MARTINET, C. CRESPELLE et E. FLEURY : Dynamic contact network analysis in hospital wards. *In P. CONTUCCI, R. MENEZES, A. OMICINI et J. PONCELA-CASASNOVAS, édés : Complex Networks V*, vol. 549 de *Studies in Computational Intelligence*, p. 241–249. Springer International Publishing, 2014.
  - [54] J. MARTINEZ et F. BAQUERO : Interactions among strategies associated with bacterial infection : pathogenicity, epidemicity, and antibiotic resistance. *Clin Microbiol*, 15:647–79, 2002.
  - [55] J. MERRER, F. SANTOLI, C. A. de VECCHI, B. TRAN, B. D. JONGHE et AL. : "Colonization pressure" and risk of acquisition of methicillin-resistant staphylococcus aureus in a medical intensive care unit. *Infect. Control. Hosp. Epidemiol.*, 21:718–23, 2000.
  - [56] R. T. MIKOLAJCZYK, M. K. AKMATOV, S. RASTIN et M. KRETZSCHMAR : Social contacts of school children and the transmission of respiratory-spread pathogens. *Epidemiology and Infection*, 136(06):813–822, 2008.
  - [57] A. NARDY, J.-P. CHEVROT et S. BARBU : Sociolinguistic convergence and social interactions within a group of preschoolers : A longitudinal study. *Language Variation and Change*, 26:273–301, 10 2014.
  - [58] M. NEWMAN : mathematics of networks. *In S. N. DURLAUF et L. E. BLUME, édés : The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke, 2008.
  - [59] M. E. NEWMAN, S. H. STROGATZ et D. J. WATTS : Random graphs with arbitrary degree distributions and their applications. *Physical review E*, 64(2):026118, 2001.
  - [60] M. E. J. NEWMAN : The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2):404–409, January 2001.
  - [61] M. E. J. NEWMAN : Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6):066133+, juin 2004.
  - [62] M. E. J. NEWMAN : *Networks : An Introduction*. Oxford University Press, New York, 2010.
  - [63] T. OBADIA, R. SILHOL, L. OPATOWSKI, L. TEMIME, J. LEGRAND, A. C. M. THIÉBAUT, J.-L. HERRMANN, E. FLEURY, D. GUILLEMOT et P.-Y. BOËLLE : Detailed Contact Data and the Dissemination of Staphylococcus aureus in Hospitals. *PLoS Computational Biology*, 11(3), 2015.
  - [64] J. P. ONNELA, J. SARAMÄKI, J. HYVÖNEN, G. SZABÓ, D. LAZER, K. KASKI, J. KERTÉSZ et A. L. BARABÁSI : Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, mai 2007.
-

- [65] A. PAAUW, A. FLUIT, J. VERHOEF et M. L. van HALL : Enterobacter cloacae outbreak and emergence of quinolone resistance gene in dutch hospital. *Emerg. Infect. Dis.*, 10:807–12, 2006.
- [66] R. PASTOR-SATORRAS et A. VESPIGNANI : Epidemic dynamics and endemic states in complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 63(6 Pt 2):066117, Jun 2001.
- [67] I. PELUPESSY, M. J. M. BONTEN et O. DIEKMANN : How to assess the relative importance of different colonization routes of pathogens within hospital settings. *Proc Natl Acad Sci U S A*, 99(8):5601–5605, Apr 2002.
- [68] L. PUZNIAK, T. LEET, J. MAYFIELD, M. KOLLEF et L. MUNDY : To gown or not to gown : the effect on acquisition of vancomycin-resistant enterococci. *Clin. Infect. Dis.*, 35:18–25, 2002.
- [69] X. RODIER, M. LE COUÉDIC, F. HAUTEFEUILLE, S. LETURCQ, B. JOUVE et E. FIEUX : From space to graphs to understand spatial changes using medieval and modern fiscal sources. *Archaeology in the Digital Era*, p. 420, 2014.
- [70] J. V. SCHINDLER et T. MRAZ : Agent-based modeling for real-time decision-support for point-of-distribution managers during influenza mass vaccination. *AMIA Annu Symp Proc*, p. 1124, 2008.
- [71] J. SCHLESINGER, S. NAVON-VENEZIA, I. CHMELNITSKY, O. HAMMER-MUNZ, A. LEAVITT et AL. : Extended-spectrum beta-lactamases among enterobacter isolates obtained in tel aviv, israel. *Antimicrob. Agents Chemother*, 49:1150–6, 2005.
- [72] P. SERMPEZIS et T. SPYROPOULOS : Information diffusion in heterogeneous networks : The configuration model approach. In *INFOCOM*, p. 3261–3266. IEEE, 2013.
- [73] T. SMIESZEK, E. U. BURRI, R. SCHERZINGER et R. W. SCHOLZ : Collecting close-contact social mixing data with contact diaries : reporting errors and biases. *Epidemiol Infect*, 140(4):744–752, Apr 2012.
- [74] J. STEHLE, N. VOIRIN, A. BARRAT, C. CATTUTO, V. COLIZZA, L. ISELLA, C. REGIS, J.-F. PINTON, N. KHANAFAER, W. VAN DEN BROECK et P. VANHEMS : Simulation of an seir infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine*, 9(1):87, 2011.
- [75] J. STEHLE, N. VOIRIN, A. BARRAT, C. CATTUTO, L. ISELLA, J.-F. PINTON, M. QUAGGIOTTO, W. V. den BROECK, C. RÉGIS, B. LINA et P. VANHEMS : High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE*, 6(8):e23176, 08 2011.
- [76] J. STEHLÉ, N. VOIRIN, A. BARRAT, C. CATTUTO, L. ISELLA, J.-F. PINTON, M. QUAGGIOTTO, W. VAN DEN BROECK, C. RÉGIS, B. LINA et P. VANHEMS : High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS One*, 6(8):e23176, 2011.
- [77] L. TEMIME, L. KARDAS-SLOMA, L. OPATOWSKI, C. BRUN-BUISSON, P.-Y. BOËLLE et D. GUILLEMOT : Nososim : an agent-based model of nosocomial pathogens circulation in hospitals. *Procedia Computer Science*, 1(1):2245 – 2252, 2010. {ICCS} 2010.
- [78] L. TEMIME, L. OPATOWSKI, Y. PANNET, C. BRUN-BUISSON, P. Y. BOËLLE et D. GUILLEMOT : Peripatetic health-care workers as potential superspreaders. *Proc Natl Acad Sci U S A*, 106(43):18420–18425, Oct 2009.
- [79] J. TRAVERS et S. MILGRAM : An experimental study of the small world problem. *Sociometry*, 32:425–443, 1969.

- [80] J. UGANDER, B. KARRER, L. BACKSTROM et C. MARLOW : The anatomy of the facebook social graph. *CoRR*, abs/1111.4503, 2011.
- [81] E. VALDANO, L. FERRERI, C. POLETO et V. COLIZZA : Analytical computation of the epidemic threshold on temporal networks, juin 2014.
- [82] P. VANHEMS, A. BARRAT, C. CATTUTO, J.-F. PINTON, N. KHANAFER, C. RÃ©GIS, B. a KIM, B. COMTE et N. VOIRIN : Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. *PLoS ONE* 8(9) : e73970 (2013), sept. 2013.
- [83] J. WALLINGA, P. TEUNIS et M. KRETZSCHMAR : Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am J Epidemiol*, 164(10):936–944, Nov 2006.
- [84] S. WASSERMAN et K. FAUST : *Social network analysis : Methods and applications*, vol. 8. Cambridge university press, 1994.
- [85] D. J. WATTS et S. H. STROGATZ : Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, juin 1998.
- [86] H. WERTHEIM, D. MELLES, M. VOS, W. van LEEUWEN, A. van BELKUM *et al.* : The role of nasal carriage in staphylococcus aureus infections. *The Lancet Infectious Diseases*, 5:751–762, 2005.
- [87] D. B. WEST : *Introduction to graph theory*. Prentice Hall, Inc., Upper Saddle River, NJ, 1996.

# Table des figures

1.1	Agencement de l'hôpital de Berk-sur-Mer. . . . .	17
1.2	Écouvillon nasal. . . . .	18
1.3	Règle de modélisation des contacts d'un réseau de proximité. . . . .	19
1.4	Présence quotidienne des individus. . . . .	20
1.5	Exemple d'une liste de contacts à trois nœuds. . . . .	23
1.6	Construction d'un réseau dynamique. . . . .	23
1.7	Exemple d'agrégation par groupe d'un réseau simple. . . . .	27
1.8	Distribution des durées de contacts au cours de la période juillet/août . . .	28
1.9	Distribution des quantités quotidiennes par individu. . . . .	29
1.10	Corrélations entre les différentes quantités. . . . .	30
2.1	Nombre quotidien de personnes présentes en moyenne dans chaque service. .	32
2.2	Moyenne des quantités d'activité de contacts entre les services. . . . .	34
2.3	Moyenne par personne des quantités d'activité de contacts entre les services.	35
2.4	Corrélation entre les quantités d'activité. . . . .	36
2.5	Exemple simple de graphe dynamique. . . . .	39
2.6	Distribution cumulative des quantités quotidiennes par personne. . . . .	40
2.7	Moyenne par personne des quantités d'activité de contacts entre les services.	42
2.8	Introversion des services. . . . .	44
2.9	Exemple simple de l'utilisation des modèles de configuration. . . . .	46
2.10	Déviaton d'introversion selon le modèle de configuration simple. . . . .	48
2.11	Déviaton d'introversion selon le modèle de configuration fin. . . . .	50
2.12	Nombre de personnes interagissant quotidiennement entre les services. . . .	53
2.13	Exemple de répartition des contacts entre les services. . . . .	54
2.14	Matrice d'activité quotidienne entre les services. . . . .	55
2.15	Matrice d'activité quotidienne par personne entre les services. . . . .	56
2.16	Pourcentage d'occupation des services. . . . .	58
2.17	Matrices de déviaton d'activité des services. . . . .	60
2.18	Matrice de déviaton des contacts externes aux services. . . . .	61
2.19	Nombre de patients et de personnels interagissant entre les services. . . . .	63
2.20	Pourcentage de patients et de personnels interagissant entre les services. . .	64
2.21	Matrices d'activité entre les services selon les degrés quotidiens par personnes.	65
2.22	Matrices d'activité entre les services selon le nombre de contacts quotidiens par personnes. . . . .	67
2.23	Matrices d'activité entre les services selon les durées de contact quotidiennes par personnes. . . . .	68
2.24	Nombre moyen quotidien de personnes présentes par catégories. . . . .	69

2.25	Valeurs moyennes des catégories socioprofessionnelles par jour des différentes quantités. . . . .	71
2.26	Valeurs moyennes par personnes organisées en catégories socioprofessionnelles par jour des différentes quantités. . . . .	72
2.27	Corrélation des différents types d'activité par catégories socio-professionnelles, n'impliquant que les personnels de l'hôpital. . . . .	73
2.28	Introversion des différentes catégories selon les 3 quantités habituelles. . . .	75
2.29	Déviaton d'introversion des différentes catégories selon les 3 quantités habituelles.	76
2.30	Nombre de personnes interagissant quotidiennement entre les différentes catégories socio-professionnelles. . . . .	78
2.31	Matrices d'activité entre les catégories par personne jour, selon les degrés quotidiens, le nombre de contacts et la durée des contacts. . . . .	79
2.32	Exemple de sélection d'activité. . . . .	80
2.33	Taux d'occupation des différentes catégories selon le degré des individus. . .	80
2.34	Taux d'occupation des différentes catégories selon le nombre de contacts des individus. . . . .	81
2.35	Taux d'occupation des différentes catégories selon la durée de contact des individus. . . . .	82
2.36	Matrices de déviation des degrés des catégories. . . . .	83
2.37	Matrice de déviation du nombre de contacts des catégories. . . . .	84
2.38	Matrices de déviation de l'activité entre les catégories, basé sur la durée cumulée des contacts. . . . .	86
3.1	Évolution de la population au cours du temps. . . . .	90
3.2	Composantes connexes et diamètres journaliers. . . . .	91
3.3	Corrélation entre le nombre de composantes connexes et la taille de la population. . . . .	92
3.4	Corrélation entre le diamètre et journalier et la taille chacune des composantes.	93
3.5	Nombre de personnes actives au cours du temps. . . . .	94
3.6	Evolution des quantités d'activité moyenne par individu. . . . .	95
3.7	Evolution et corrélation entre le degré des personnes et la durée moyenne de contacts. . . . .	96
3.8	Taux de conservation des paires entre jours consécutifs (équation 3.1) . . . .	103
3.9	Différence symétrique sur les paires. . . . .	103
3.10	Indice de Jaccard sur les paires. . . . .	104
3.11	Similarité cosinus sur les paires. . . . .	105
3.12	Différence d'activité des paires moyennées sur l'ensemble des personnes présentes le jour 1 (eq. 3.9). . . . .	106
3.13	Similarité cosinus renormé sur les paires. . . . .	106
3.14	Différence asymétrique appliqué aux nombres et aux durées de contact. . . .	109
3.15	Indice de Jaccard appliqué aux nombres et aux durées de contact. . . . .	109
3.16	Similarité cosinus appliquée aux nombres et aux durées de contact. . . . .	109
3.17	Différence d'activité appliquée aux nombres et aux durées de contact. . . .	110
3.18	Similarité cosinus renormé sur les paires. . . . .	110
3.19	Agrégation des données à l'heure. . . . .	111
3.20	Évolution à l'heure de la quantité d'individus actifs. . . . .	111
3.21	Évolution à l'heure du nombre de contacts et de leur durée au cours d'une semaine. . . . .	113

3.22	Comparaison de l'évolution à l'heure du degré des personnes et de la durée de contacts des personnes. . . . .	114
4.1	Evolution de la taille de la population prise en compte pour les simulations.	119
4.2	Nombre de simulations ayant au moins un transfert durant l'heure. . . . .	120
4.3	Taux de simulations actives par pas de temps. . . . .	122
4.4	Distribution du maximum de transmetteurs. . . . .	124
4.5	Date d'atteinte de la valeur maximale du nombre de transmetteurs pour chaque simulation. . . . .	125
4.6	Distribution du maximum de récepteurs. . . . .	126
4.7	Date d'atteinte de la valeur maximale du nombre de récepteurs pour chaque simulation. . . . .	126
4.8	Nombre et taux de personnes infectées à la fin de chaque simulation, ordonnées par heure de départ. . . . .	127
4.9	Distribution du taux d'infection. . . . .	128
4.10	Temps de terminaison de chacune des simulations en fonction de leur temps de départ. . . . .	129
4.11	Proportion de temps nécessaire pour atteindre la largeur maximale du DAG de diffusion. . . . .	129
4.12	Nombre d'arêtes utilisées par simulation. . . . .	130
4.13	Répartition des feuilles selon les simulations. . . . .	131
4.14	Assignation du poids des arêtes. . . . .	132
4.15	Distribution de la valeur des arrêtes au cours des simulations. . . . .	132
4.16	Distribution du nombre moyen d'arêtes entrantes par personne pour chaque simulation, soit le nombre moyen de transmetteurs par récepteur. . . . .	133
4.17	Distribution du nombre d'arêtes sortantes, soit le nombre de récepteurs par transmetteur. . . . .	133
4.18	Distribution (cumulative) du nombre d'occurrences des arêtes. . . . .	134
4.19	Corrélation entre le maximum de transmetteurs simultanés et le maximum de récepteurs simultanés, en terme de pourcentage de population. . . . .	134
4.20	Comparaison du taux maximal de transmetteurs et récepteurs avec le taux de feuilles. . . . .	135
4.21	Largeur de DAG issus de la première période stable au cours du temps. . .	137
4.22	Largeur de DAG issus de la première période stable au cours du temps : zoom.	138
4.23	Nombre de personnes touchées au cours du temps de toutes les simulations débutant aux pas de temps 81 et 86. . . . .	139
4.24	Nombre de personnes touchées au cours du temps de toutes les simulations débutant aux pas de temps 81 et 86. . . . .	140
4.25	Largeur de DAG issus de la deuxième période stable au cours du temps. . .	141
4.26	Largeur de DAG issus de la deuxième période stable au cours du temps : zoom. . . . .	142
4.27	Largeur de DAG probabilistes au cours du temps. . . . .	144
4.28	Taux d'infection. . . . .	145