École doctorale : Sciences et Technologies

Doctorat en Informatique

THÈSE

pour obtenir le grade de docteur délivré par

Université Gaston Berger de Saint Louis

Spécialité : Informatique

présentée et soutenue publiquement par

Djibril Mboup

le 12 Juin 2019

Mobilité et phénomènes de propagation dans les réseaux complexes temporels

---

Directeur de thèse : **Pr. Moussa LO**
Co-encadrant de thèse : **Dr. Cherif DIALLO**

Jury

Pr. Hocine CHERIFI,	Université de Bourgogne	Rapporteur
Pr. Mamadou BOUSSO,	Université de Thiès	Rapporteur
Pr. Bi Tra GOORE,	Institut National Polytechnique Félix Houphouët Boigny	Rapporteur
Pr. Ibrahima NIANG,	Université Cheikh Anta Diop	Examinateur
Pr. Ousmane THIARE,	Université Gaston Berger	Examinateur

REMERCIEMENTS

Au Nom d'ALLAH, le Tout Miséricordieux, le Très Miséricordieux !
Je formule mes sincères prières sur Son Prophète Mouhamad (PSL).
Je rends grâce à notre vénéré guide et référence Cheikh A. Bamba Khadimou Rassoul.

Je tiens avant tout à exprimer ma profonde gratitude à Dr. Cherif Diallo d'avoir accepté à superviser les travaux de ma thèse. Je lui remercie beaucoup pour son accompagnement, sa disponibilité et sa patience. Je remercie aussi le Professeur Moussa Lo, mon directeur de thèse, pour le soutient et la confiance qu'il a donné à ma modeste personne.

Je tiens à remercier les Professeurs Hocine Cherifi, Bi Tra Goore et Mamadou Bousso d'avoir accepté à rapporter mon travail. Les jugements et les remarques qu'ils ont fournis à l'égard de mon travail ont été pertinents et très enrichissants.

Je remercie aussi les membres du jury particulièrement le Professeur Ousmane Thiaré et le Professeur Ibrahima Niang pour leur disponibilité et pour l'honneur qu'ils me font en acceptant de participer à la validation de mes travaux.

Je remercie l'ensemble des Professeurs de l'UFR SAT qui ont participé activement et sans faille à ma formation.

Merci au Centre d'Excellence Africain en Mathématiques, Informatique et TIC (CEA-MITIC) pour leur accompagnement durant cette thèse.

Mes remerciements vont aussi à l'endroit de tous mes camarades doctorants de l'UFR SAT particulièrement à ceux de la section Informatique.

Je remercie Babacar Diop d'avoir pris le temps de relire mon document.

Mention spéciale à mes camarades et frères Sanarois : Moussa Thiaw, Aly Ngoné Ngom, Gorgui Gning, Lamine Diop, Abdou Aziz Ndiaye, Samba Diop, Modou Ndiaye, Yahya Traoré, Kwamé APPOH, Ousmane Dieng, Ablaye Sy, Seynabou Touré, Papa Birane Seye. La liste est loin d'être exhaustive.

Je remercie particulièrement tous les membres du Daara Mafaatihoul Bichri.

Mention spéciale à ma mère et tous les membres de ma famille pour leur compréhension et leur assistance durant toutes ces années d'étude.

DEDICACES

A Mon pére Elhadj Ngagne Mboup.
A Mon ami d'enfance Ngagne Sylla.
Que leurs demeures soient le Paradis.

Abstract

A network is said to be complex when it brings into interaction a large number of nodes whose global behavior is not deductible from individual behaviors. This collective behavior is called "*emergent*" behavior. This type of network concerns the Internet, social networks, metabolic networks, telephony networks, transport networks, networks of human contacts, information networks like the Web. In addition, networks are the locus of many dynamic phenomena such as the birth of community structures, opinion formation, information dissemination, epidemic spreading, synchronization, and so on. In the contexts of temporal networks of contacts or proximity, it happens that the external dynamics becomes non negligible and influences the internal dynamics of the network. In this point of view, understanding the mobility of individuals is crucial when we want to know how the epidemic in a person-to-person contact network spreads. Recently, with the emergence of new data sources (*Mobile Phone traces, Online Social Media, GPS, IoT, Bluetooth traces* and so on) accompanied by new modeling approaches (*Temporal Networks, Multiplex Networks, Adaptive Networks, Preferential Attachment and variants*), new research perspectives are opening up in the study of dynamics in complex networks.

Our aim is essentially to study the spreading process, SIS (*Suspected-Infected-Suspected*) model, in a temporal network (*Time-Varying Graph*) by focusing on the impact of mobility in dynamic evolution of the structure of the contact graph and the contamination process. Using traces generated from synthetic mobility models (Random Waypoint, Gauss-Markov, Truncated Levy Walk and Spatio-TEmporal Parametric Stepping), we study the topology of temporal contact networks in comparison with those obtained at from the real traces. In addition, we study an epidemic model that takes into account the relationship which exists between mobility and spreading process by redefining a new coefficient of *spreading rate* depending on the distance. We also propose another SIS model that adds the impact of distance proximity and the neighborhood with infected nodes in the probability of infection.

Keywords : Complex networks, Dynamical networks, Evolving networks, Temporal networks, Dynamic process, Epidemic spreading, Mobility model

Résumé

Un réseau est dit complexe lorsqu'il met en interaction un grand nombre de nœud dont le comportement global n'est pas déductible des comportements individuels : on parle ainsi de phénomène d'émergence. Ce type de réseau concerne Internet, les réseaux sociaux, les réseaux métaboliques, les réseaux de téléphonie, les réseaux de transport, les réseaux de contacts humains, les réseaux d'informations comme le Web. En plus, les réseaux sont le siège de plusieurs phénomènes dynamiques tels que la naissance des structures de communauté, la formation d'opinion, la diffusion d'information, la propagation d'épidémie, la synchronisation, etc. Dans le contexte des réseaux temporels de contacts ou de proximité, il arrive que la dynamique externe devienne non négligeable et influe sur la dynamique interne du réseau. La mobilité des individus est un élément crucial pour savoir par exemple comment un agent pathogène se propage dans un réseau de contact personne à personne. De nouvelles perspectives de recherche s'ouvrent dans ce domaine avec l'apparition de nouvelles sources de données provenant des réseaux de téléphonie mobile, des réseaux sociaux en ligne, des GPS, des capteurs et des traces de Bluetooth mais aussi avec les nouvelles approches de modélisation (*Temporal Networks*, *Multiplex Networks*, *Réseaux Adaptatifs*, *Attachement Préférentiel* et variantes) permettant d'étudier la dynamique dans ces réseaux.

L'objectif de cette thèse s'adresse essentiellement à l'étude de processus dynamiques (modèle SIS, *Suspected-Infected-Suspected*) dans un réseau complexe temporel (*Time-Varying Graph*) en se focalisant sur l'impact de la mobilité dans l'évolution du graphe de contacts temporels ainsi que sur le processus de contamination et de propagation. En s'appuyant sur des traces générées à partir des modèles de mobilité synthétiques (*Random Waypoint*, *Gauss-Markov*, *Truncated Levy Walk*, *Spatio-TEmporal Parametric Stepping*), nous avons étudié la topologie des réseaux de contacts temporels en comparaison avec ceux obtenus à partir des traces réelles. Nous avons, d'une part, développé un modèle qui prend en compte la relation qui existe entre la mobilité et le processus de propagation en redéfinissant un nouveau taux de transmissibilité (*spreading rate*), noté ici par β, dépendant du facteur de distance-proximité. D'autre part, nous avons proposé un autre modèle SIS qui ajoute l'impact du voisinage avec les nœuds infectés, en plus de la proximité, sur la probabilité d'infection.

Mots-clés : Réseaux complexes, Réseaux dynamiques, Réseaux temporels, Processus épidémique, Propagation épidémique, Phénomène dynamique, Modèle de mobilité

Liste de publications

Les travaux de cette thèse ont abouti à trois (03) publications dans des revues internationales avec comité de lecture et indexées par les abstracts (*ACM, IEEE, Atlantis Press*). Il s'agit :

[1] Djibril Mboup, Cherif Diallo et Moussa Lo. "Mobility and Epidemic Process in Temporal Networks". In : *Proceedings of the Second International Conference on Advanced Wireless Information, Data, and Communication Technologies, AWICT 2017, Paris, France, November 13-14, 2017*. 2017, 5:1-5:7

[2] Djibril Mboup, Cherif Diallo et Moussa Lo. "Impact of distance-proximity in spreading process in mobile networks". In : *2017 Global Information Infrastructure and Networking Symposium, GIIS 2017, Saint Pierre, France, October 25-27, 2017*. 2017, p. 71-74

[3] Djibril Mboup, Cherif Diallo et Moussa Lo. "Structural and dynamical analysis of complex networks". In : *The 2nd Information Technology and Mechatronics Engineering Conference (ITOEC 2016)*. Atlantis Press, mai 2016

Table des Matières

Page

Liste des tableaux **xiii**

Table des figures **xv**

1 Introduction Générale **1**

2 Réseaux Complexes **5**

2.1 Introduction . . . 5
2.2 Historique et Définition . . . 6
2.3 Les Réseaux Réels . . . 8
2.3.1 Les réseaux sociaux . . . 8
2.3.2 Les réseaux d'information . . . 9
2.3.3 Les réseaux technologiques . . . 9
2.3.4 Les réseaux biologiques . . . 10
2.4 Concepts Fondamentaux . . . 11
2.4.1 Graphe . . . 12
2.4.2 Degré d'un nœud . . . 12
2.4.3 Connexité . . . 13
2.4.4 Plus court chemin (*shortest path*), Moyenne des plus courts chemins (*Average path length*) . . . 14
2.4.5 Effet petit-monde . . . 14
2.4.6 Coefficient de clustering . . . 15
2.4.7 Distribution des degrés . . . 16
2.4.8 Assortativité . . . 17
2.4.9 Centralité . . . 18
2.4.10 Motifs . . . 19
2.4.11 Structure de communauté . . . 20
2.5 Modélisation des Réseaux Complexes . . . 20
2.5.1 Erdös Rényi Random graphs . . . 21
2.5.2 Modèle de Watts et Strogatz . . . 23
2.5.3 Scale-free de Barabási & Albert . . . 25
2.5.4 Réseaux Adaptatifs . . . 27
2.5.5 Réseaux Multiplexes . . . 28
2.5.6 Réseaux Temporels . . . 31
2.6 Conclusion . . . 40

3 Phénomène de propagation et mobilité dans les réseaux complexes 41
3.1 Introduction . . . 41
3.2 La propagation d'épidémie . . . 42
3.2.1 Modèles classiques de propagation d'épidémie . . . 42
3.2.2 Épidémiologie et réseaux complexes . . . 47
3.3 La Mobilité . . . 52
3.3.1 Caractéristiques de la mobilité humaine . . . 53
3.3.2 Modèles de mobilité . . . 56
3.4 Conclusion . . . 63

4 Impact de la mobilité dans la propagation d'épidémie 65
4.1 Introduction . . . 65
4.2 Travaux connexes (*Related Works*) . . . 66
4.3 Méthodologie . . . 67
4.4 *Déplacement Quadratique Moyen* (MSD) . . . 70
4.5 Analyse topologique des propriétés des réseaux de contacts temporels obtenus . 72
4.5.1 Densité et composantes connexes . . . 72
4.5.2 Chemin temporel et accessibilité . . . 75
4.5.3 Coefficient de corrélation temporelle et l'efficience temporelle . . . 78
4.5.4 Les mesures de centralité dans les réseaux temporels . . . 80
4.5.5 Occurrence des paires de contacts . . . 82
4.5.6 Durée de contact (*Contact Time*) et Durée d'inter-contact (*Inter-Contact-Time*) . . . 84
4.5.7 Fluctuabilité et Volatilité . . . 87
4.6 Modèles épidémiques . . . 87
4.6.1 Modèle 1 : Description . . . 87
4.6.2 Résultats . . . 88
4.6.3 Modèle 2 : Description . . . 90
4.6.4 Résultats . . . 92
4.7 Discussion . . . 93

5 Conclusion & Perspectives 95

Bibliographie 97

Appendix A 119

Liste des tableaux

Table **Page**

2.1 Mesures du coefficient de clustering et de la moyenne des longueurs de plus courts chemins des réseaux biologiques, sociaux et technologiques : C, coefficient de clustering ; C_{rand}, coefficient de clustering dans un réseau aléatoire équivalent ; L, la moyenne des longueurs de plus courts chemins ; L_{rand}, la moyenne des longueurs de plus courts chemins dans un réseau aléatoire équivalent. D'après [Tel+11] 24

4.1 Paramètres de simulation 70
4.2 Mesures de corrélation entre D^t et N^t_{Comp} 74
4.3 Mesures de coefficient de corrélation temporelle $\mathscr{C}$ et l'efficience temporelle $\mathscr{E}$ . . . 80
4.4 Mesures de corrélation entre la centralité temporelle de degré (TCD) et la centralité temporelle de proximité (TCP) 82
4.5 Quelques distributions à queue lourde (*heavy tailed*) 84
4.6 Mesures de Fluctuabilité et Volatilité 87
4.7 Paramètres de Simulation 89

Table des figures

Figure **Page**

2.1 Ponts de Königsberg 6
2.2 La structure du réseau World Wide Web [Bro+00] 10
2.3 Représentation du réseau trophique des animaux de l'Arctique [Bul12] 11
2.4 Matrice et Liste d'adjacence 13
2.5 Illustration du coefficient de *clustering* : $C^1 = 3 \times \frac{1}{8} = \frac{3}{8}$ car nous avons un triangle et huit triplets alors que l'ensemble des coefficients de clustering locaux est de $C_{1,2,3,4,5} = \{0, 0, 1, 1, \frac{1}{6}\}$, ce qui donne la moyenne $C^2 = \frac{13}{30}$ 16
2.6 Exemple de 13 motifs d'un sous-graphe composé de 3 nœuds. D'après [Boc+06] . . 19
2.7 Exemple de réseau montrant trois structures de communauté. D'après [GN02] . . . 21
2.8 Exemple du processus de création du modèle de ER Random Graph $G^{ER}_{(N,p)}$, extrait dans Réf [AB02] 22
2.9 Procédure de recâblage de Watts et Strogatz. D'après [WS98] 23
2.10 Caractéristiques de la moyenne de longueurs l et du coefficient de clustering C dans le modèle de WS 24
2.11 Distribution des degrés généré par le modèle de Barabási-Albert sur un réseau de N nœuds avec $m = 3$ ($P(k) \sim k^{-3}$) 25
2.12 Processus de régénération de GNA en fonction de l'algorithme $\langle E, R, I \rangle$: (a) Évolution du modèle en commençant par la configuration initiale I. (b) Détection des événements de réécriture à chaque pas de temps ts. (c) Mécanisme d'extraction E . (d) Mécanisme de remplacement R. (e, f) Construction de modèles de E et R sur la base des données d'apprentissage. (g) Modèle GNA final. D'après [Say+13] . 29
2.13 Une représentation d'un réseau multi-couche composé de deux graphes : G1 et G2. Les connexions inter-couches sont en rouge tandis que les connexions intra-couches sont en vert pour G1 et en bleu pour G2. La matrice d'adjacence du réseau de projection associé $proj(M)$ est affichée dans le coin inférieur droit [Cas+16] 31
2.14 Deux exemples de réseaux multicouches : (a) Exemple de réseau multi-couche de différentes compagnies aériennes (Lufthansa, Air-France, British-Airways), (b) Exemple de réseau social d'amitié multi-couche (Facebook) [Kiv+14] 32
2.15 Comparaison entre un réseau temporel et un réseau statique [Tan+10a] 33
2.16 Représentation d'un réseau temporel (TVG) selon [HS12] 33
2.17 (a) Phénomène de *burstiness*, (b) Phénomène non *bursty* (processus de poisson) [ML16b] 37

2.18 Fluctuabilité et Volatilité : Les réseaux temporels illustrés dans les panneaux A, B et C ont tous le même nombre de nœuds et d'arêtes, mais ils diffèrent par leur fluctuabilité et leur volatilité. (A) Ce réseau a une faible fluctuabilité (F = 0,125) et une volatilité (V = 0,73). (B) Ce réseau présente la volatilité la plus élevée (V = 2,55) des trois réseaux et une fluctuabilité (F = 0,25) égale à celle du réseau correspondant à C. (C) Ce réseau présente une volatilité inférieure à B (V = 1,27) avec une fluctuabilité égale (F = 0,25). [TBF17] 40

3.1 Modèle compartimental SI 43
3.2 Modèle SI 44
3.3 Modèle compartimental SIS 44
3.4 Modèle SIS 45
3.5 Modèle compartimental SIR 46
3.6 Modèle SIR 47
3.7 Phases de transition caractérisées par la valeur critique du seuil épidémique λ_c d'un modèle SIS [Pas+15] 49
3.8 Random Brownian Motion vs Lévy Flight Motion. D'après [Bov15] 53
3.9 Trajectoire d'un individu suivant le modèle *Random Waypoint Mobility* (RWP) . . 58
3.10 Dispersion spatiale des noeuds suivant le modèle RWP [BW02] 59
3.11 Trajectoire d'un agent suivant le modèle de mobilité *Gauss-Markov* (GM) 60
3.12 Mouvement d'un individu suivant la mobilité *Truncated Lévy Walk* (TLW) 62
3.13 Mouvement d'un agent suivant la mobilité *Spatio-TEmporal Parametric Stepping* (STEPS) 63

4.1 Exemple de Réseau Temporel (TVG) $\mathcal{G} = \{G_1, G_2, G_3, G_4\}$ 68
4.2 *Déplacement Quadratique Moyen* (MSD) des differentes modèles de mobilités utilisés 71
4.3 Évolution temporelle de la densité et du nombre de composantes connexes 73
4.4 Corrélation entre la densité et le nombre de composantes connexes 74
4.5 Évolution de D^t dans le graphe $\mathcal{G}_{RWP}$ 75
4.6 Évolution de D^t dans le graphe $\mathcal{G}_{GM}$ 75
4.7 Évolution de D^t dans le graphe $\mathcal{G}_{TLW}$ 76
4.8 Évolution de D^t dans le graphe $\mathcal{G}_{STEPS}$ 76
4.9 Moyenne et Écart-type de la densité D^t des graphes temporels en fonction de la taille N 77
4.10 Densité temporelle des graphes de contacts réels 78
4.11 Reachability Latency 79
4.12 L'effet *petit monde*'dans les graphes temporels 80
4.13 Centralité de temporelle de degré 81
4.14 centralité temporelle de proximité 82
4.15 Corrélation entre la centralité de degré et la centralité de proximité 83
4.16 Distribution des occurrences de paires de contacts sur une échelle logarithmique . 84
4.17 Distribution de la durée des contact sur une échelle log-log 85
4.18 Distribution de la durée d'inter-contact sur une échelle log-log 86
4.19 Mesures du coefficient de Burstiness B en fonction de la variation locale LV 86

4.20 Le nouveau taux de transmissibilité $\bar{\beta}_{ij}$ dépendant du taux β et de la force du lien s_{ij}. $\bar{\beta}_{ij} = \beta * s_{ij}$ 88
4.21 Taux de prévalence obtenus sur les différents TVG $\mathcal{G}_{RWP}, \mathcal{G}_{GM}, \mathcal{G}_{TLW}$ 90
4.22 Comparaison entre les moyenne totales de nœuds infectés 90
4.23 Taux de prévalence obtenus sur $\mathcal{G}_{RWP}, \mathcal{G}_{GM}, \mathcal{G}_{TLW}$ lorsque β est remplacé par le nouveau coefficient $\bar{\beta}$ 91
4.24 Comparaison des moyennes totales des nœuds infectés lorsque β devient $\bar{\beta}$ sur les différents TVG utilisés 91
4.25 Taux de prévalence dans $\mathcal{G}_{RWP}$ en fonction du paramètre α 92
4.26 Taux de prévalence dans $\mathcal{G}_{GM}$ en fonction du paramètre α 93
4.27 Taux de prévalence dans $\mathcal{G}_{TLW}$ en fonction du paramètre α 93

2 La distribution de la centralité de proximité des noeuds suivant le modèle RWP . . 121

Chapitre 1

Introduction Générale

Notre monde réel a toujours été le lieu de plusieurs phénomènes dynamiques naturels dont leurs complexités ne nous permettent pas de comprendre facilement leurs émergences. Ces phénomènes sont souvent liés par l'effet des interactions entre les différentes composantes d'un système donné. Quelques exemples de ces phénomènes émergents sont par exemple les crises financières, le changement climatique, les goulots d'étranglement, les nuées d'oiseaux, les bancs de poissons, les colonies d'insectes, les structures de fractales, les structures de communautés dans les réseaux sociaux, la diffusion d'informations, la propagation des épidémies etc. L'étude de ces systèmes dits complexes a pour but de comprendre, de prédire et de contrôler ces phénomènes dynamiques.

Par ailleurs, les modèles sont utilisés pour expliquer une partie de la réalité mais pas la réalité en tant que telle. Cette idée rejoint la proposition de George Box qui dit que *"tous les modèles sont faux mais certains sont utiles"*[M15]. Partant de ce point de vue, les chercheurs ont créé différents types de modèles permettant d'étudier les phénomènes dynamiques dans les systèmes complexes. Parmi ces modèles, certains sont basés sur des systèmes dynamiques [Str15 ; HSD13 ; Lu+16], sur des processus stochastiques [HB15 ; Lu+16], sur des systèmes multi-agents [Sie14 ; Lu+16], sur des algorithmes génétiques [SBS98 ; Lu+16] ou sur des automates cellulaires [KSH10 ; Chi+10]. Récemment, avec le développement de ce que l'on appelle la *Science des Réseaux* [BP16] ou réseaux complexes, l'étude des systèmes complexes a connu une nouvelle ère.

Dans les réseaux complexes, les phénomènes dynamiques peuvent être résumés en trois catégories différentes : l'évolution dynamique de la topologie du réseau (niveau macroscopique), la dynamique sur le réseau (niveau microscopique) et l'interaction entre l'évolution de la topologie et les processus qui s'y déroulent. L'évolution dynamique est caractérisée par l'apparition et la disparition des nœuds et/ou des liens de manière temporelle. C'est le cas des réseaux de contacts humains, du Web et de l'Internet. Dans le cas de la dynamique sur le réseau, il s'agit de phénomènes de propagation, de diffusion, de routage, de synchronisation, des processus de recherche, de la percolation, entre autres. Nous savons que dans les réseaux complexes réels la dynamique interne (*processus dynamique*) et celle externe (*évolution*

topologique) sont intrinsèquement liées. Il peut arriver que la dynamique interne soit largement plus rapide que l'évolution topologique. Par conséquent, la dynamique externe n'influe pas beaucoup sur le processus qui a lieu dans le réseau, dans ce cas l'approche des réseaux complexes statiques nous permet d'analyser et de répondre aux questions qui seront posées. Dans le cas contraire, c'est-à-dire si l'évolution temporelle de la structure est non négligeable devant le processus dynamique alors la modélisation avec les réseaux complexes statiques ne suffisent plus pour comprendre la dynamique du phénomène étudié. Ainsi, naissent d'autres approches de modélisation des réseaux complexes : les réseaux multicouches, les réseaux multiplexes, les réseaux adaptatifs et les réseaux temporels (*Time-Varying Networks*).

Ces travaux de thèse s'adressent essentiellement à l'étude de processus dynamique (modèle **SIS** : *Suspected-Infected-Suspected*) dans un réseau complexe temporel en se focalisant sur l'impact de la mobilité dans l'évolution du graphe de contacts temporels ainsi que sur le processus de contamination et de propagation. En utilisant des contacts temporels générés par des modèles de mobilité synthétiques (***Random Waypoint***, ***Gauss-Markov***, ***Truncated Lévy Walk***, ***Spatio-TEmporal Parametric Stepping***) nous tenterons d'analyser et de répondre aux questions suivantes : Comment les contacts peuvent-ils être créés à partir de la mobilité ? Quel sera l'impact de la mobilité sur l'évolution de la structure du réseau de contact temporel ainsi que sur la propagation d'une épidémie ? La distance de proximité entre les individus, peut-elle être considérée comme un élément d'importance sur le facteur de transmission de certains cas de maladie ? La probabilité d'infection, dépend-t-il du nombre de voisins rencontrés ou bien d'autres facteurs comme la durée des contacts et d'inter-contacts ?

Outre l'introduction générale (**Chapitre 1**) et la conclusion (**Chapitre 5**), cette thèse s'articule en trois grands chapitres :

- Dans le **Chapitre 2**, nous présentons une étude approfondie sur l'état de l'art des réseaux complexes en passant par leurs descriptions, les différentes propriétés permettant de les caractériser enfin les différentes approches de modélisation surtout celles qui permettent de prendre en compte la dynamique temporelle.
- Le **Chapitre 3**, qui synthétise nos premières contributions, expose en détails la modélisation d'un processus épidémique dans un réseau complexe en se focalisant sur les modèles de références utilisés par la communauté scientifique (SI, SIS, SIR). En plus, nous y avons évoqué les propriétés de la mobilité locale et globale avec une description détaillée des modèles de mobilité utilisés comme support dans cette thèse pour générer des réseaux temporels de contacts synthétiques.
- Quant au **Chapitre 4**, il est consacré à d'autres contributions qui s'articulent essentiellement autour trois points :
  - Dans un premier temps, nous avons fait une analyse statistique des caractéristiques topologiques des réseaux de contact temporel qui ont été générés à partir des modèles de mobilité utilisé.
  - Deuxièmement, nous avons développé un modèle qui prend en compte la relation qui existe entre la mobilité et le processus de propagation en redéfinissant un nouveau taux de transmissibilité (**spreading rate**) souvent noté par β dépendant du facteur de distance-proximité.
  - En dernier lieu, nous avons proposé un autre modèle SIS qui ajoute l'impact

du voisinage avec les nœuds infectés, en plus de la proximité, sur la probabilité d'infection.

Afin de conclure, une synthèse globale a été faite avec une discussion sur les résultats obtenus en répertoriant certaines faiblesses qui ont été rencontrées durant notre étude. Enfin, nous avons dégagé quelques perspectives pouvant ouvrir de nouvelles questions de recherche.

Chapitre 2

Réseaux Complexes

2.1 Introduction

Depuis très longtemps, les réseaux ont été étudiés avec un grand intérêt dans plusieurs domaines de la science. Leur expansion a été surtout connue dans les domaines comme les sciences sociales, la physique, la biologie moléculaire et l'informatique. Un réseau peut être défini de manière abstraite comme un système dont les entités ou éléments peuvent être représentés par des nœuds (sommets) et les relations ou interactions entre ces éléments comme étant des liens. Mathématiquement connu sous le nom de graphe, un réseau permet de représenter et d'expliquer de divers phénomènes des systèmes complexes par exemple la diffusion d'une information ou celle d'un rumeur dans un réseau social, la congestion dans un trafic routier, la propagation d'une maladie au sein d'une population, le routage dans un réseau comme Internet.

La fin des années 1990 a été marquée par l'explosion d'un vaste mouvement de recherche autour des réseaux dits "complexes", appelés aussi grands graphes de terrain (réseau défini comme étant une structure irrégulière, complexe et évoluant dynamiquement dans le temps [Boc+06]). En effet, ceci est le résultat d'une évolution décentralisée et non planifiée. Récemment, l'étude des réseaux complexes a été encouragée du fait de l'apparition de grandes bases de données permettant d'étudier systématiquement la topologie des grands réseaux complexes mais aussi l'augmentation de la vitesse de calcul des ordinateurs facilitant l'exploration des réseaux contenant des millions de nœuds.

Vu les limites observées dans la modélisation de ces systèmes qui s'est toujours appuyée sur la théorie des graphes, les chercheurs ont introduit différents concepts permettant de mieux comprendre leurs structures et leurs dynamiques. C'est dans cette optique que nous allons, dans cette partie, étudier, analyser et comprendre ces deux aspects que sont la dynamique et la structure des réseaux complexes.

Dans ce chapitre nous allons faire une présentation des réseaux complexes afin d'expliquer leurs origines et où est-ce que nous pouvons les trouver. Nous allons aussi présenter les concepts fondamentaux permettant de décrire les caractéristiques de ces réseaux. Enfin, nous allons

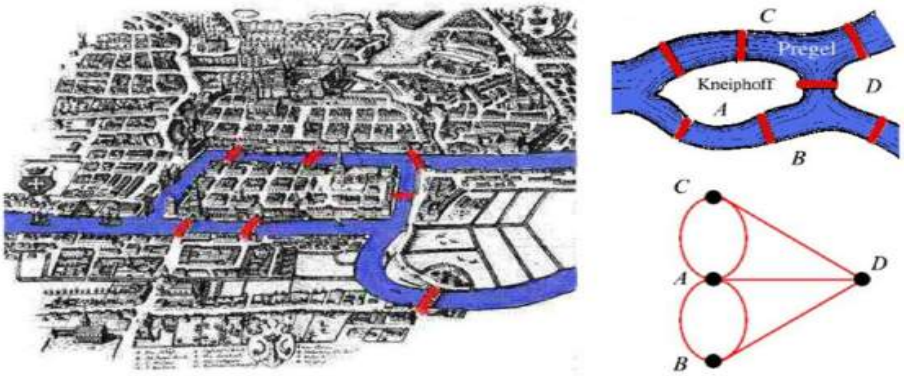


FIGURE 2.1: Ponts de Königsberg

passer en revue les différents types de modélisation des réseaux complexes.

2.2 Historique et Définition

Historiquement, la théorie des graphes est née en 1936 lorsque Léonard Euler, devenu vieux, voulait traverser chacun des sept ponts de la ville de Königsberg (aujourd'hui Kaliningrad) une seule fois et de revenir au point de départ. Euler modélisa le problème sous forme de graphe dont les sommets représentaient les terres et les arêtes les ponts comme le montre la Figure 2.1. Les ponts enjambent les bras du fleuve Pregel qui coulent de part et d'autre de l'île de Kneiphoff. Il prouva que, pour que la traversée soit possible il faut un nombre pair de ponts ce qui n'est pas le cas pour Pregel.

La théorie des graphes constitue l'un des piliers fondamentaux des mathématiques discrètes. Elle s'est développée au fil des années entre les disciplines telles que la physique (physique statistique), la chimie (la modélisation de la matière), la biologie (le génome), les neurosciences (la modélisation du réseau de connexion des neurones), les sciences sociales (la modélisation des relations sociales), et tant d'autres domaines. Elle est aussi un outil puissant dans la modélisation des problèmes en Informatique.

Durant cette dernière décennie, l'étude des réseaux complexes, basée en grande partie sur la théorie des graphes, a connu une grande réputation dans la modélisation de ce que l'on appelle les systèmes "*complexes*". Cette notion de système complexe a intéressé un grand nombre de chercheurs de diverses disciplines. Biologistes, physiciens, linguistes, géographes ou informaticiens, tous étudient, à leur manière et avec leurs outils de tels systèmes. Depuis les années 1980, de nombreux efforts ont été fournis, utilisant les théories basées sur la physique et la simulation numérique. Selon Newman dans [New11], il n'existe pas encore une définition unanime d'un système complexe, mais beaucoup de chercheurs sont d'accords que c'est un système composé de plusieurs parties en interaction appelées "*agents*", tel que le comportement collectif de ces parties est plus que la somme de leurs différents comportements. Le comportement collectif est, parfois, appelé comportement "*émergent*" [New11]. Pour quantifier les détails d'un tel système, il faut premièrement spécifier sa topologie - qui interagit avec qui ? - et ensuite sa dynamique - comment les agents se comportent individuellement et comment ils interagissent ?

La théorie des systèmes complexes peut être divisée entre deux approches [New11]. La

première concerne l'étude et la création de modèles mathématiques qui ne représentent pas le comportement exact de ces systèmes, mais donnent une abstraction des éléments les plus importants qualitativement sur le plan scientifique. La seconde approche est guidée par la création de modèles compréhensibles et réalisables, généralement sous la forme de simulations informatiques, qui représentent les parties en interaction d'un système complexe. Cette approche permet de surveiller et de mesurer les comportements qui émergent du système [New11]. C'est dans ce contexte que les réseaux sont devenus comme un élément incontournable dans la modélisation des systèmes complexes.

La recherche sur les réseaux dits complexes a franchi, vers la fin des années 1990, une étape importante avec les travaux de Watts et Strogatz qui ont abouti à la découverte des graphes petit-monde (*small-world*) [WS98], en 1998, et de Barabási et Albert aux graphes sans échelles (*scale-free*) [BA99], en 1999. Ces travaux constituent, aujourd'hui, l'un des éléments fondamentaux dans l'étude et la compréhension des réseaux complexes. Après ces pionniers, de nombreuses revues d'articles et de livres sont apparus dans la littérature avec des thématiques parlant aussi bien de la structure que de la dynamique de ces réseaux [PV04] [BP16] [BBV08] [CV07] [LNR17] [New03] [AB02] [Boc+06] [Pas+15].

Le terme réseau complexe dérive des systèmes complexes. Il est récemment appelé Science des Réseaux (*Network Science*) par Barabási [BP16]. Barabási le définit comme la nouvelle science qui permet d'encoder les interactions entre les composantes d'un système complexe : interactions entre les gènes, connexion entre les neurones, relation entre des entités sociales etc. [BP16]. Selon Newman, un réseau complexe peut être défini comme étant un réseau d'interactions entre entités dont le comportement global n'est pas déductible des comportements individuels des dites entités, d'où l'émergence de nouvelles propriétés [New03]. En effet, ces deux définitions se réfèrent à la notion de système complexe. Cependant, dans un exposé intitulé *Applications of Complex Network* [Bul12], Buldù définit un réseau complexe comme un réseau avec des caractéristiques topologiques non triviales, avec un mode de connexion entre les éléments qui n'est ni régulier ni purement aléatoire. Enfin, Boccaletti le définit dans [Boc+06] comme un réseau dont la structure est irrégulière, "complexe " et qui évolue dynamiquement dans le temps.

Nous pouvons apprendre de ces définitions qu'un réseau complexe peut-être vu, d'une part, comme un *mapping* d'un système complexe, d'autre part, comme un graphe qui n'est pas statique dont sa structure évolue dynamiquement dans le temps. L'intérêt d'étudier ces types de réseaux se présente d'une part dans le but de comprendre à partir de simples interactions l'émergence de phénomènes dynamiques qui s'y déroulent et d'autre part la complexité que peut engendrer l'évolution de sa structure. Nous pouvons prendre comme exemple d'Internet dont chaque jour des milliards de flux de données transitent entre les millions de nœuds que constituent ce réseau. Internet fait partie des plus grands réseaux complexes dont la compréhension des caractéristiques qui gouvernent ce réseau intéressent toujours la communauté scientifique. De même, des données de contact humain récoltées dans des traces peuvent constituer un réseau complexe qui permet d'étudier la propagation dynamique des maladies infectieuses dans une population. En effet, plusieurs autres réseaux réels assimilés à des réseaux complexes sont aussi étudiés dans la littérature. Nous pouvons citer parmi ces réseaux : les réseaux sociaux[SSL16] [Nef13], les réseaux de transport électriques

[WS98], les réseaux de marchés financiers [CV07 ; KS17], les composants chimiques [AL00 ; Jeo+01], les réseaux de neurones du cerveau [AB02 ; TBF17], les systèmes de transports [Lon+15], le réseaux de contacts humain [Gui+03 ; Van+13], les réseaux métaboliques [AL00 ; Jeo+01], les réseaux téléphonie mobiles, les réseaux trophiques (*food web*)[MS02 ; WM00], le *World Wide Web* (WWW) [AJB99] etc.

2.3 Les Réseaux Réels

Les réseaux réels sont les réseaux construits à partir des données du monde réel. Le traitement et l'analyse de ces données qui peuvent provenir de diverses sources sont encouragées par l'augmentation de la performance des CPUs et la capacité de stockage des machines. Pourtant, depuis très longtemps, l'étude de ces réseaux a toujours intéressé la communauté scientifique. Dans cette partie, nous allons passer en revue les différentes catégories de réseaux réels. Ceci nous amène à adopter la catégorisation faite par Newman dans [New03]. Ainsi, nous pouvons subdiviser les réseaux réels en quatre grandes catégories : les réseaux sociaux, les réseaux d'informations, les réseaux technologiques et les réseaux biologiques.

2.3.1 Les réseaux sociaux

Un réseau social est formellement défini comme un ensemble d'individus ou d'entités sociales qui sont liés entre eux à travers des relations sociales [Boc+06]. Ces relations peuvent être de nature diverse : liens amicaux, liens de parenté, collaborations, contacts sexuels, relations commerciales etc. L'analyse des réseaux sociaux a commencé au début des années 1920 avec les travaux de Simmel, Wellman, Bouglé et d'autres. En effet, elle est due en grande partie par l'apport de la sociométrie qui a très vite connu un essor grâce aux travaux de Moreno, qui est perçu comme l'un des précurseurs de l'analyse des réseaux sociaux et de la psychologie sociale. Elle s'est, ensuite développée, très vite grâce à l'expérience du psychologue social américain en 1967, Milgram, qui a introduit le fameux concept de "*six degrés de séparation*" (voir détails dans la section 2.4.5) [Mil67].

Les données de réseaux sociaux sont souvent obtenues à partir des fichiers d'enquêtes qui peuvent être limités parfois par la taille de l'échantillon d'étude. Cependant, aujourd'hui, avec l'expansion des réseaux sociaux en ligne comme Facebook, Twitter, LinkedIn et l'accessibilité des données provenant des capteurs, des CDRs (*Call Data Records*) etc., l'analyse des réseaux sociaux devient un domaine très convoité par la communauté des réseaux complexes. Le *Open Data* ou l'ouverture de données sur Internet a considérablement facilité l'étude de la structure et des propriétés dynamiques de ces réseaux.

Parmi les propriétés dynamiques qui sont étudiées dans les réseaux sociaux nous pouvons citer la formation et l'évolution des structures de communauté [SSL16] [Nef13] [Alv+16] [Bas+16], la popularité [KPY16] [Gao+18], la diffusion d'information ou de rumeur [HXS15] [Nek+07], la formation d'opinion [Gra09] [PT17], la propagation d'épidémie [Kup13] [ZHA+15] [Pas+15].

Un exemple classique de ce type de réseau est le réseau de collaboration des acteurs de film dont la source des données est disponible dans les IMDB (*Internet Movie Database*). Dans

ce type de réseau, deux acteurs sont considérés connectés s'ils apparaissent ensemble dans un même film. Les propriétés statistiques de ce réseau sont analysées par un certain nombre d'auteurs [AH00] [Ama+00] [New01]. D'autres exemples de réseaux sociaux ont également été explorés par les chercheurs : les réseaux de co-auteurs dans une académie [CFK17 ; Che+15], les réseaux d'appel téléphonique [GHB08 ; RSM18], les réseaux d'email [BP16], les réseaux de contact sexuel [RLH10 ; AB02], etc.

2.3.2 Les réseaux d'information

Notre seconde catégorie constitue les *réseaux d'information*. Selon Newman, elle se rapporte à l'exemple classique des réseaux de citations entre papiers scientifiques. On constate que beaucoup d'articles citent d'autres articles précédents par rapport à un sujet qui les relie. Ces citations forment un réseau dont les nœuds sont les papiers et un lien orienté d'un nœud A vers un nœud B indiquant que B est cité dans A. La structure des informations est stockée dans les nœuds d'où le terme de réseau d'informations. Le réseau de citations est un réseau acyclique car les papiers qu'on peut citer ne peuvent être que ceux qui existent déjà. Redner, en étudiant la distribution des citations de 783339 papiers catalogués par *the Institute of Scientifc Information* et de 24296 publiés dans *Physical Review D* entre 1975 et 1994, a trouvé que la probabilité pour qu'un papier soit cité k fois suivait une loi puissance ($p_k \simeq k^{-\lambda}$, avec $\lambda \sim 3$) [Red98]. Récemment, Goldberg et *al.* utilisent les données de *High Energy Physics - Theory ArXiv* (hep-th ArXiv) récoltées entre 1992 et 2002 pour montrer que le modèle statistique du nombre de papiers ayant c citations suit une distribution log-normale [GAE14].

Un autre réseau important dans cette catégorie est le *World Wide Web* (WWW). Ce réseau est formé par les pages web, contenant des milliers d'information, qui constituent les nœuds et les hyperliens qui les relient entre eux. En effet, le WWW contient aujourd'hui des milliards de nœuds et de liens. Contrairement au réseau de citations, le World Wide Web (WWW) est cyclique car il n'existe pas un ordre naturel des sites et il n'y a pas aussi une contrainte qui prévient l'apparition des boucles [New03]. Pourtant, les études sur le réseau WWW ont commencé depuis la fin des années 1990 avec Albert, Jeong, Barabási [AJB99] et Kumar et al. [Kum+99]. Une étude faite par Broder et *al.* sur 2 millions de pages et 1.5 milliards de liens capturés par un *'web crawler'* confirme que la distribution des degrés sortant et des degrés entrant suit une loi puissance et montre aussi que la structure du web peut être assimilée comme un nœud de papillon (voir Figure 2.2) [Bro+00]. En effet, la complexité de sa structure à une forte influence sur l'étude d'un quelconque phénomène dynamique dans ce réseau. D'autres recherches continuent d'explorer le réseau WWW qui ne cesse de s'agrandir.

2.3.3 Les réseaux technologiques

Cette catégorie regroupe les réseaux qui sont artificiellement créés par l'homme pour la distribution d'un service ou de l'énergie. Ces réseaux concernent le réseau de distribution d'électricité, le transport ferroviaire, le trafic routier, le réseau de transport aérien, les réseaux téléphoniques, les réseaux mobiles sans fils (MANETS : *Mobile Ad hoc Networks*), Internet etc. Le premier réseau étudié dans cette catégorie est la topologie du réseau électrique du *Western States Power Grid* aux États Unis [WS98]. Ce réseau de haute tension permettait la distribution électrique entre les États de l'Ouest. L'étude de ce type de réseau a permis de

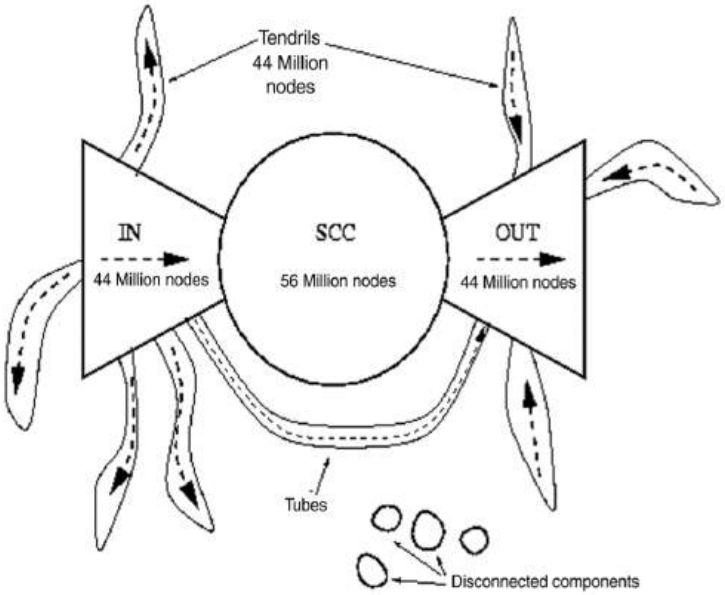


FIGURE 2.2: La structure du réseau World Wide Web [Bro+00]

comprendre ses défaillances et sa robustesse mais aussi d'anticiper sur certains événements stratégiques tels que les éventuelles attaques ciblées. Les propriétés statistiques de ce réseau sont étudiées par Watts et Strogatz [WS98], Pagani et Aiello [PA13]. Les réseaux de transport sont aussi d'une grande importance dans la mesure où ils permettent d'étudier plusieurs phénomènes des systèmes complexes dépendant de la mobilité humaine.

Il y a un autre réseau technologique qui attire plus notre attention : Internet. Différent du WWW, Internet est le réseau physique représenté par des machines et d'autres appareils de télécommunication qui sont interconnectés entre eux par des câbles ou l'air. La structure d'Internet a évolué rapidement durant ces deux dernières décennies. Cette évolution à grande échelle augmente la complexité d'étude de ce réseau. La topologie d'Internet est étudiée sur deux niveaux différents : au niveau des routeurs qui constituent les nœuds et les arêtes les supports de connexions, au niveau inter-domaine ou système autonome (AS : *Autonomous System*) dans lequel chaque domaine composé par des centaines de routeurs et de machines est représenté par un seul nœud et il existe une arête si deux domaines sont reliés par un routeur [AB02]. Faloutsos et *al.* ont étudié ces deux niveaux et ont conclu que le degré de distribution de chacun de ces réseaux suivait une loi puissance [FFF99].

2.3.4 Les réseaux biologiques

Enfin la dernière catégorie est constituée des réseaux biologiques. Ils concernent tous les réseaux liés aux êtres vivants. Ce type de réseau varie des réseaux de régulateur génétique aux réseaux d'interactions protéiques en passant par les réseaux écologiques. L'exemple le plus classique dans cette classe est le réseau des voies métaboliques (*metabolic pathways*). Dans ce réseau les substrats (tel que ATP, ADP, H2O) sont considérés comme des nœuds et il y a un lien orienté s'il existe une réaction chimique dont le résultat agit sur les substrats. Jeong et *al.* [Jeo+00 ; Pod+01], Guimerà et Amaral [GN05] ont très tôt étudié certaines propriétés structurelles de ces réseaux.

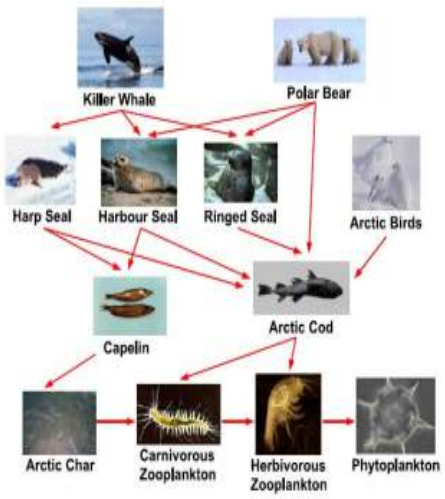


FIGURE 2.3: Représentation du réseau trophique des animaux de l'Arctique [Bul12]

Par ailleurs, le réseau trophique permettant de représenter la chaines alimentaire (*food web*) entre les êtres vivants constitue un intérêt majeur dans l'étude de notre écosystème. Ainsi, nous pouvons le représenter comme un graphe orienté dont les liens expriment la relation proie-prédateur comme le montre l'exemple de la Figure 2.3. Les propriétés statistiques de ce réseaux sont étudiées par Solé et Montoya [MS02], Williams et Martinez [WM00], Elhesha et *al.* [EKB17] et d'autres. En outre, il existe d'autres classes importantes de réseaux biologiques qui ont toujours intéressé la communauté : les réseaux d'interaction entre protéines [AL00 ; Jeo+01], les réseaux de régulation génétique [DP15 ; Pod+01], les réseaux cellulaires [AB02 ; TBF17] etc.

Tous ces réseaux réels évoqués au-dessus semblent être trop différents du point de vue de leurs origines. Par contre, les chercheurs ont découvert qu'ils présentent en effet plusieurs caractéristiques communes. Par exemple, l'effet petit-monde (*small-world*) qui considère que la moyenne de distance des plus courts chemins (*average path length*) entre deux nœuds dans un réseau complexe est très faible. En outre, on a aussi constaté que ces réseaux sont en général sans échelle (*scale-free*) c'est à dire marqué par la présence d'un petit nombre de nœuds avec un fort degré, appelé *hub*, par rapport à une majorité qui a un degré faible. Dans ces réseaux, on remarque aussi que les nœuds ont tendance à se regrouper en *cluster* ou structure de communauté. Il y a aussi la notion de centralité qui joue un rôle très important dans la diffusion ou la propagation d'un phénomène dynamique dans les réseaux. D'autres propriétés ont été révélées dans ces réseaux et celles-ci impliquent plus de difficulté à mettre en place un modèle qui prendra en compte l'essentiel des éléments de complexité.

2.4 Concepts Fondamentaux

La structure d'un réseau complexe est un peu particulière à cause parfois de sa grande dimension et des nombreuses interactions qui influent sur ses caractéristiques. En plus, c'est un réseau qui évolue dynamiquement dans le temps et s'y déroulent plusieurs phénomènes dynamiques. Ainsi, tout modèle de réseau complexe doit permettre au moins de comprendre l'une des caractéristiques suivantes : sa topologie et sa dynamique. En effet, la modélisation

des réseaux complexes était essentiellement basée sur les concepts de théorie des graphes. Cependant, ces concepts ne suffisent pas pour faire une étude exhaustive de tels systèmes. Il a fallu aux chercheurs de ce domaine de mettre en place d'autres moyens plus adaptés pour compléter l'étude fondée sur cette théorie. Dans cette section, nous allons aborder les principaux éléments permettant de comprendre les réseaux complexes.

2.4.1 Graphe

Definition 2.1. Mathématiquement, un graphe G peut être défini comme un couple (V,E) où :

- V est un ensemble (fini) d'objets. Les éléments de V sont appelés sommets ou nœuds du graphe.
- E est un sous ensemble de $V \times V$. Les éléments de E sont les arêtes du graphe.

Un graphe peut être orienté (*directed graph*) ou non orienté (*undirected graph*). Parfois, il peut être important de quantifier les liens par une valeur ou un poids. Dans ce cas, on parle de graphe pondéré (*weighted graph*). Un *sous-graphe* $G' = (V', E')$ est un graphe tel que $V' \subset V$, $E' \subset E$. Si G' contient tous les liens de G qui relient deux nœuds dans E', on dit que G' est un *sous-graphe induit* par E'. Il est noté par $G' = G\left[E'\right]$. En effet, plusieurs types de graphes ont été définis en fonction de leur structure. Par exemple, un *multigraphe* est un graphe pour lequel il peut exister plusieurs arêtes entre deux sommets. Un graphe *biparti* est un graphe non-orienté pour lequel on peut identifier une partition des nœuds en deux sous-ensembles tels qu'il n'existe pas de lien entre des nœuds du même sous-ensemble. Un graphe *simple* n'est rien d'autre qu'un graphe qui n'est pas multi-graphe et ne possède aucune boucle.

Un graphe $G = (V,E)$ est exactement équivalent à sa matrice adjacence. Soit $A = (a_{ij})$, la matrice d'adjacence de G de dimension $n \times n$ telle que

$$a_{ij} = \begin{cases} 1 & \text{si} \quad (i,j) \in E \\ 0 & \text{sinon} \end{cases} \tag{2.1}$$

On peut aussi représenter un graphe par une *liste d'adjacence* qui est une liste de $n = |V|$ éléments où chaque élément est constitué d'un nœud $v \in V$ et de ses voisins dans une liste chaînée. La Figure 2.4 montre un exemple de graphe orienté avec sa matrice et sa liste d'adjacence.

2.4.2 Degré d'un nœud

Dans un graphe, on dit que le nœud i est voisin du nœud (ou adjacent à) j (et vice versa) s'il existe un lien entre i et j. Le degré d'un nœud i est le nombre de liens adjacents à ce nœud. Le degré d'un nœud i est noté en général par k_i et donné par l'expression suivante :

$$k_i = \sum_{j \in V} a_{ij} \text{ ; avec } a_{ij} \in A \tag{2.2}$$

Dans le cas des graphes orientés, nous avons la notion de degré sortant k^{out} et de degré entrant k^{in} d'un nœud i. Le degré sortant (ou extérieur) k_i^{out} est le nombre d'arcs ayant i comme extrémité initiale tandis que le degré entrant (ou intérieur) k_i^{in} est le nombre d'arcs ayant i

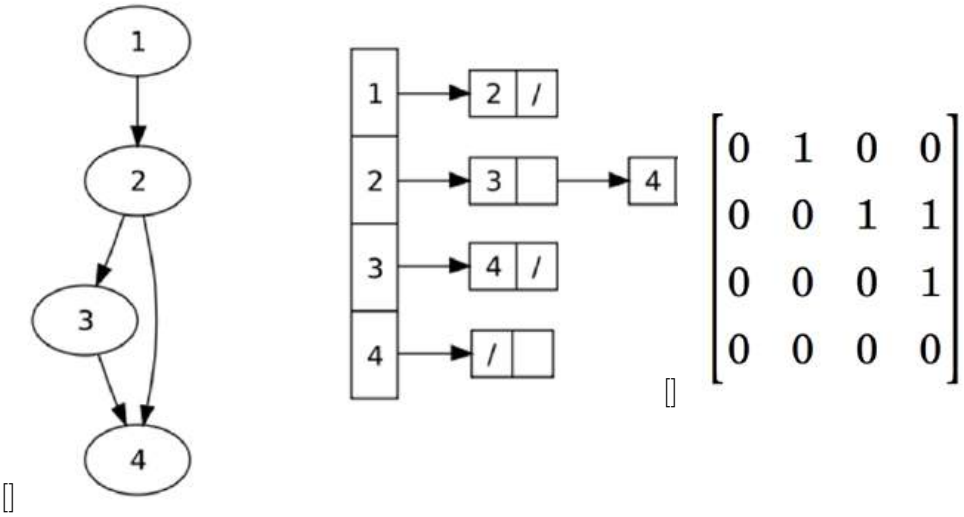


FIGURE 2.4: Représentation d'un graphe orienté avec sa matrice et sa liste d'adjacence.

comme extrémité finale.

$$k_i^{in} = \sum_{j \in V} a_{ij} \tag{2.3}$$

$$k_i^{out} = \sum_{j \in V} a_{ji} \tag{2.4}$$

Ainsi, le degré total du nœud i est défini par $k_i = k_i^{in} + k_i^{out}$.

2.4.3 Connexité

Le fait qu'un nœud quelconque soit joignable dans un réseau à partir de n'importe quel autre nœud constitue un aspect important dans beaucoup de réseaux réels. Dans un graphe, une *chaîne* est une séquence finie et alternée de sommets et d'arêtes, débutant et finissant par des sommets, telle que chaque arête est incidente avec les sommets qui l'encadrent dans la séquence. Un *cycle* est une chaine dont les extrémités coïncident. Une chaîne est *élémentaire* si tous les sommets qui la composent sont tous distincts. Une chaîne est *simple* si toutes les arêtes qui la composent sont toutes distinctes. Un *chemin* est une chaine (dans le cas d'un graphe orienté) dont les arcs sont orientés dans le même sens. Un *circuit* n'est rien d'autre qu'un chemin fermé.

Un graphe $G = (V, E)$ est *connexe* si pour tout $i, j \in V$ il existe une chaîne entre i et j. Une *composante connexe* est un sous-ensemble de sommets qui est connexe. Autrement dit, un graphe est connexe s'il comporte une composante connexe maximale et une seule. Une composante géante est une composante dont la taille est d'ordre N.

Les graphes *complets* sont les graphes dont tous les sommets sont "reliés " deux à deux. Ce sont des graphes qui sont toujours connexes. Si G est un graphe complet avec un nombre de nœuds égal à N alors le nombre de liens $K = N(N-1)/2$. Dans ce cas, G est noté K_N. Un *clique* est un sous graphe complet.

La densité D d'un graphe peut être définie comme étant le coefficient entre le nombre d'arêtes existantes et le nombre d'arêtes possibles. Un graphe G est dit *dense* si $K = O(N^2)$, par contre il est éparse si $D \ll 1$ ou $K \ll N^2$. Nous ne pouvons pas nier, l'effet de la densité sur la connexité du graphe. En effet la plupart des réseaux réels étudiés sont, par nature, éparses.

2.4.4 Plus court chemin (*shortest path*), Moyenne des plus courts chemins (*Average path length*)

Le plus court chemin ou distance géodésique est un concept fondamental de la théorie des graphes. Il joue un rôle important dans le transport et la communication dans un réseau [Boc+06], par exemple dans le routage de paquets dans un réseau comme Internet. Le plus court chemin d_{ij} entre deux nœuds i et j n'est rien d'autre que la distance minimale entre tous les chemins qui connectent i et j. Ainsi, le maximum entre tous les plus courts chemins est appelé diamètre du graphe, $Diam(G) = max(d_{ij})$.

Nous pouvons définir la moyenne des longueurs (*average path length*) entre deux nœuds i et j comme étant la longueur moyenne des plus courts chemins. On l'appelle aussi *la moyenne géodésique*. Considérons l la moyenne des longueurs, l peut être défini par la relation mathématique suivante :

$$l = \frac{1}{N(N-1)} \sum_{\substack{i,j \\ i \neq j}} d_{ij} \tag{2.5}$$

où d_{ij} est la distance géodésique entre i et j, N est le nombre total de nœuds du graphe G. En réalité, cette définition n'est valable que si G est connexe c'est à dire s'il existe un chemin entre deux nœuds quelconques du graphe G. Par contre, si le graphe est déconnecté alors l diverge. Pour éviter ce problème, il est préférable d'utiliser dans certains cas la *moyenne harmonique*, encore appelée l'*efficience globale* E_{glob} [LM01].

$$E_{glob}(G) = \frac{1}{N(N-1)} \sum_{\substack{i,j \\ i \neq j}} \frac{1}{d_{ij}} \tag{2.6}$$

Ainsi, tout couple de nœuds appartenant à des composantes déconnectées du graphe contribue à la sommation dans $E_{glob}(G)$ à une valeur égale à zéro. Nous avons, par définition, $0 \leq E_{glob}(G) \leq 1$.

2.4.5 Effet petit-monde

L'effet petit-monde est né de l'expérience du psychologue social américain Stanley Milgram [Mil67]. Dans les années 1960, Milgram reprend une idée développée par Frigyes Karinthy, écrivain hongrois, dans sa nouvelle "*Chains* " de 1929. Cette théorie dit que chacun d'entre nous sur la planète peut être connecté à une autre personne en suivant une chaine de connaissances ne contenant pas plus de six intermédiaires.

Milgram donne 300 lettres à des participants américains habitants à Omaha au Nebraska, à qui il fut demandé de faire parvenir les lettres à un agent de change vivant à Boston. Les participants ne pouvaient faire transiter les lettres que de main à main à des personnes qu'ils jugeaient le plus susceptibles de connaître la personne cible (directement ou par le biais de connaissances). Milgram trouve que la longueur moyenne de la chaine qui permet de relier une personne à une autre était de six. Ainsi, l'expression de "*six degrés de séparation* " fut donné à cette expérience par John Gaure quelques années plus tard. Cette découverte était tellement révolutionnaire que Milgram a conclu que la société humaine n'est constituée que d'un "petit-monde ", type de réseau caractérisé par des courts chemins [ZAI10]. Dans la littérature, ce

concept fait souvent référence à une faible valeur de la moyenne des longueurs (*Average path length*), notée l. Cette valeur est une donnée très utile dans l'analyse des réseaux complexes. On a découvert que l présente une valeur faible aussi bien dans les réseaux aléatoires que dans les réseaux réels. Pour les graphes non connexes, il a été montré par Latora et Marchiori que la grande valeur de $E_{glob}(G)$ permet de caractériser l'effet petit-monde [LM01].

2.4.6 Coefficient de clustering

Le coefficient de *"clustering "*, encore appelé *Transitivité* [Boc+06], est une propriété qui explique le phénomène suivant constaté dans les réseaux réels : si un nœud A est connecté à un nœud B et le nœud B connecté à un nœud C alors il y a une forte probabilité pour que A soit connecté avec C. Dans le jargon de la sociologie, il se traduit par le fameux concept "*les amis de mes amis sont mes amis* ". En effet, la racine de cette idée vient des travaux de George Simmel qui a introduit le concept de *triad* comme une structure fondamentale des réseaux sociaux [ZAI10]. Un *triad* est un groupe social composé de trois membres tandis qu'un *dyad* est un groupe social de deux membres. En terme de topologie de réseaux, la transitivité signifie la forte présence du nombre de triangle dans le réseau [New03]. Dans la littérature, deux définitions du coefficient de clustering sont utilisées : le clustering **global** et le clustering **local**.

Clustering global il est défini comme étant trois fois le nombre de triangle observés sur le nombre de triplets de nœuds connectés, mesurés sur tout le graphe.
Soit le graphe G, le coefficient de clustering global C^1 est donnée par l'expression :

$$C^1 = \frac{3 \times \text{nombre de triangles dans } G}{\text{nombre de triplets connectés dans } G} \tag{2.7}$$

Le facteur 3 du numérateur compense le fait que chaque triangle composé par trois nœuds contribue à la formation de trois triples connectés et assure que $0 \leq C^1 \leq 1$. Pour $C^1 = 1$, nous avons $G = K_N$. Cette formule peut être encore écrite sous la forme suivante [New03] :

$$C^1 = \frac{6 \times \text{nombre de triangles dans } G}{\text{nombre de chemins de longueur deux}} \tag{2.8}$$

Clustering local il se rapporte à un nœud spécifique i du graphe. C'est le nombre de triangles connectés à i sur le nombre de triplets centrés sur i.

$$C_i = \frac{\text{nombre de triangles connectés à } i}{\text{nombre de triplets centrés sur } i} \tag{2.9}$$

Cette définition est donnée par Watts et Strogatz dans [WS98]. En considérant e_i comme le nombre de liens entre les voisins de i et k_i le degré du nœud i. Nous obtenons la formule 2.9 autrement.

$$C_i = \frac{2 \times e_i}{k_i(k_i - 1)} \tag{2.10}$$

En définitif, le **clustering local** nous permet de calculer le nouveau coefficient de clustering C^2 du graphe comme étant la moyenne des C_i.

$$C^2 = \frac{1}{N} \sum_i C_i \tag{2.11}$$

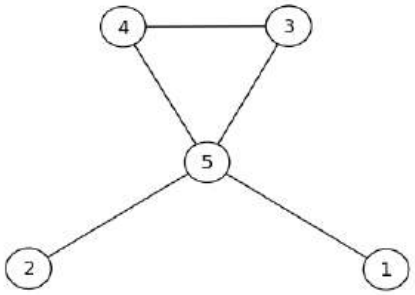


FIGURE 2.5: Illustration du coefficient de *clustering* : $C^1 = 3 \times \frac{1}{8} = \frac{3}{8}$ car nous avons un triangle et huit triplets alors que l'ensemble des coefficients de clustering locaux est de $C_{1,2,3,4,5} = \{0, 0, 1, 1, \frac{1}{6}\}$, ce qui donne la moyenne $C^2 = \frac{13}{30}$

Par définition, nous avons $0 \leq C_i \leq 1$ et $0 \leq C^2 \leq 1$. Les deux définitions du coefficient de clustering fournissent des résultats différents comme le montre l'exemple de la figure 2.5. Les détails de ces différences sont expliqués dans [New03]. D'autres extensions du coefficient de clustering sont faites dans les graphes pondérés. Les travaux de Tore Opsahl et Pietro Panzarasa [T O09] en sont un exemple. Une autre mesure alternative des propriétés de clustering d'un graphe G est *l'efficience locale* définie par :

$$E_{loc} = \frac{1}{N} \sum_{i \in V} E(G_i) \tag{2.12}$$

où $E(G_i)$ est l'efficience de G_i évaluée par la formule 2.6.
Nous avons remarqué que la présence des triples ou triads est très faible dans les graphes aléatoires par apport aux réseaux du monde réel où un fort coefficient de clustering est observé [ZAI10].

2.4.7 Distribution des degrés

Rappelons d'abord que le degré d'un nœud dans un réseau est le nombre de liens qui lui sont incidents. Nous définissons p_k ou $P(k)$ comme étant la fraction de nœuds de degré k sur le nombre total des nœuds dans un graphe. Le terme $P(k)$ signifie également la probabilité qu'un nœud choisi de manière aléatoire soit de degré k. La courbe $P(k)$ d'un réseau peut être obtenu en traçant l'histogramme correspondant en fonction des degrés des nœuds du graphe. Cet histogramme est appelé le *degré de distribution*. Le degré de distribution $P(k)$ satisfait la condition de normalisation

$$\sum_{k=0}^{\infty} P(k) = 1 \tag{2.13}$$

Dans le cas des graphes orientés, il y aura deux types de distribution qui seront pris en compte, $P(k^{in})$ et $P(k^{out})$. Le *moment de distribution* des nœuds d'un graphe est défini comme étant :

$$\langle k^n \rangle = \sum_k k^n P(k) \tag{2.14}$$

Le moment d'ordre 1, $\langle k \rangle$, définit le degré moyen du graphe. Le second moment $\langle k^2 \rangle$ permet de mesurer les fluctuations des connexions observées au niveau de la distribution des nœuds. Par conséquent, la divergence de $\langle k^2 \rangle$ change radicalement le comportement du processus dynamique qui a lieu sur le graphe [Boc+06].

Depuis longtemps, on avait cru que le degré de distribution de la plupart des réseaux suivait une distribution de Poisson ($p_k \sim \frac{z^k e^{-z}}{k!}$ où $z = \langle k \rangle$), mais en réalité, des expériences ont montré que la majeur partie des réseaux réels ont une distribution qui suit une loi puissance ($p_k \sim k^{-\lambda}$) [New03]. Cette caractéristique implique que dans les réseaux réels, certains nœuds ayant des degrés plus élevés jouent des rôles particuliers par rapport à d'autres nœuds possédant des degrés plus faibles. Ces types de réseaux sont appelés *scale-free*.

2.4.8 Assortativité

Le degré de distribution est toujours utilisé pour déterminer les propriétés statistiques des graphes non corrélés. Néanmoins, on a constaté qu'un grand nombre de réseaux présente des corrélations entre ses différents nœuds dans la mesure où la probabilité pour qu'un nœud de degré k se connecte avec un autre nœud de degré k' dépend de k et de k'. C'est pour cela qu'il est nécessaire d'introduire la notion de probabilité conditionnelle $P(k'|k)$, définie comme étant la probabilité qu'un lien venant d'un nœud de degré k pointe sur un nœud de degré k'. $P(k'|k)$ satisfait la normalisation $\sum_{k'} P(k'|k) = 1$ et la condition de balance [Boc+06] :

$$kP(k'|k) = k'P(k|k') \tag{2.15}$$

Si un graphe est non corrélé, $P(k'|k)$ ne dépend plus de k. D'après la normalisation et l'équation 2.15, on peut obtenir facilement la relation suivante :

$$P(k'|k) = \frac{k'P(k')}{\langle k \rangle} \tag{2.16}$$

En général, l'évaluation directe des mesures de probabilités conditionnelles donne parfois des résultats bruités (difficiles à analyser) pour la plupart des réseaux. Ce problème peut être surmonté en définissant le degré moyen des voisins d'un sommet i (*Average nearest neighbors degree*).
Considérons un graphe G avec sa matrice d'adjacence $\{a_{ij}\}$, le degré moyen des voisins d'un sommet i, $k_{nn,i}$, est donné par :

$$k_{nn,i} = \frac{1}{k_i} \sum_{j=1}^{N} a_{ij} k_j \tag{2.17}$$

Pour interpréter la distribution de corrélations des nœuds, nous avons besoins de définir le degré moyen des plus proches voisins d'un nœud (*Average degree of the nearest neighbors of a vertex of degree*), $k_{nn}(k)$, encore appelé *assortativité* du graphe. En utilisant l'équation 2.17, nous pouvons définir $k_{nn}(k)$ ainsi

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i/k_i=k} k_{nn,i} \tag{2.18}$$

où N_k est le nombre de nœuds de degré k. Une autre définition de $k_{nn}(k)$ peut être aussi obtenue en utilisant les probabilités conditionnelles :

$$k_{nn}(k) = \sum_{k} k'P(k'|k) \tag{2.19}$$

Si le graphe n'est pas corrélé alors la formule 2.19 donne $k_{nn}(k) = \langle k^2 \rangle / \langle k \rangle$. Dans le cas des graphes corrélés, un graphe est dit *"assortative "* si $k_{nn}(k)$ augmente en fonction de k.

En d'autres termes, les nœuds fortement connectés tendent à se relier avec d'autres nœuds fortement connectés. Cette caractéristique s'observe, par exemple, dans les réseaux sociaux (popularité). En revanche, un graphe est *"disassortative "* si $k_{nn}(k)$ décroît en fonction de k, ce qui signifie que les nœuds fortement connectés tendent à se relier avec d'autres nœuds faiblement connectés. Ce phénomène est observé dans certains réseaux comme Internet.

2.4.9 Centralité

La notion de *centralité* est une propriété qui est typiquement des réseaux sociaux. Les chercheurs avaient remarqué que certains acteurs jouaient un rôle plus "important " que d'autres. Par exemple, certaines personnes avaient plus de contact au sein du réseau par rapport à d'autres [CHI10]. Cette importance que manifestent certains nœuds est quantifiée dans un graphe par *la mesure de centralité*. En effet, la destruction des nœuds centraux aura, par exemple, des conséquences sur la diffusion d'une information dans un réseau social.
Différentes mesures de centralité ont été proposées que ça soit dans les graphes orientés ou dans les graphes non orientés. En 1979, Freeman, dans son article intitulé *"Centrality in social networks : Conceptual clarification"*, avait proposé trois définitions formelles du concept de centralité : la centralité de degré (*degree centrality*), la centralité de proximité (*closeness centrality*) et la centralité intermédiaire (*betweenness centrality*) [Fre78].

Centralité de degré (*degree centrality*) La centralité de degré est basée sur l'idée que les acteurs centraux sont les acteurs les plus actifs et qui ont le plus de liens avec les autres acteurs. Cette mesure considère le degré du nœud dans le graphe d'où le nom de *centralité de degré*. Considérons un graphe $G=(V,E)$ d'ordre N représenté par sa matrice d'adjacence $\{a_{ij}\}$, nous pouvons définir la centralité de degré d'un nœud i, $C_D(i)$, par

$$C_D(i) = \frac{1}{N}\sum_j a_{ij} \tag{2.20}$$

Dans le cas des graphes orientés, on pourra calculer la centralité de degré en tenant compte des liens sortants ou entrants. Toutefois, cette mesure de centralité est locale, elle ne prend pas en compte la structure du graphe mais elle est malgré tout beaucoup utilisée dans certains contextes, notamment en sociologie [SEI12].

Centralité de proximité (*closeness centrality*) Cette deuxième mesure proposée par Freeman est basée sur l'idée qu'un nœud i est central s'il peut rapidement interagir avec l'ensemble des autres nœuds du réseau, c'est-à-dire s'il est situé à proximité de l'ensemble des autres nœuds du graphe. Dans un graphe $G=(V,E)$ d'ordre N, cette mesure peut être calculée par :

$$C_C(i) = \frac{N-1}{\sum_{j\in V\setminus\{i\}} dist(i,j)} \tag{2.21}$$

où $dist(i,j)$ est la distance entre i et j. Pour le calcul des distances, plusieurs métriques peuvent être utilisées. Freeman propose par exemple d'utiliser la distance géodésique [CHI10]. Par conséquent, plus la centralité de proximité est élevée, plus le nœud est proche de l'ensemble des autres nœuds et plus il est central. De la même manière, on peut étendre cette formule dans les graphes orientés.

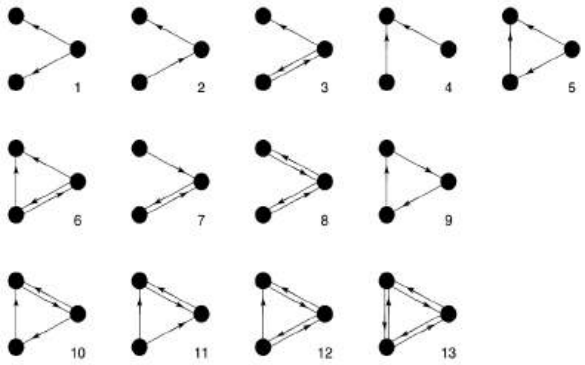


FIGURE 2.6: Exemple de 13 motifs d'un sous-graphe composé de 3 nœuds. D'après [Boc+06]

Centralité intermédiaire (*betweenness centrality*) La centralité intermédiaire mesure l'utilité d'un nœud dans la transmission de l'information au sein du réseau. Plus précisément, un nœud ayant une forte centralité intermédiaire est un nœud par lequel passe un grand nombre de chemins géodésiques (i.e. chemins les plus courts) dans le graphe. Par exemple, dans un réseau d'informations, la centralité intermédiaire peut être considérée comme la probabilité qu'une information transmise entre deux nœuds passe par ce nœud intermédiaire [CHI10]. Elle est définie par la relation suivante :

$$C_B(i) = \sum_{j,k \in V, j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \tag{2.22}$$

où $\sigma_{jk}(i)$ est le nombre de plus courts chemins allant de j à k passant par i et σ_{jk} le nombre total de plus courts chemins allant de j à k.

2.4.10 Motifs

Souvent, on retrouve dans de nombreux réseaux complexes des structures fréquentes de sous-graphes appelées *motifs*. Ce concept a été introduit par Milo et *al.* pour caractériser le nombre fréquent d'apparition de ces "*patterns*" dans les réseaux réels comparé à celui des réseaux aléatoires [Mil+02] de même taille. Ces motifs étaient étudiés par les auteurs dans les réseaux de régulation génétiques, les réseaux de neurones, les réseaux trophiques, les réseaux électriques, Internet et le Web. Ce qui est intéressant dans cette découverte, nous remarquons qu'il y a une répétition fréquente de certains motifs dans ces différents types de réseaux. Selon les auteurs, un motif $\boldsymbol{M}$ est défini comme un sous graphe (orienté ou non) de $\boldsymbol{n}$ nœuds connectés [Mil+02]. La figure 2.6 montre un exemple de la liste de tous les motifs de 3 nœuds connectés selon Milo et *al.*.

Les motifs peuvent être identifiés en comparant les occurrences de ces sous-graphes apparaissant fréquemment dans le réseau complexe réel considéré par rapport à un réseau aléatoire de même taille qui a fondamentalement une topologie similaire. La signification statistique de $\boldsymbol{M}$ est alors décrite par la définition de *Z-score* :

$$Z_M = \frac{n_M - \langle n_M^{rand} \rangle}{\langle \sigma_M^{rand} \rangle} \tag{2.23}$$

où n_M représente le nombre de fois le motif M apparaît dans G, $\langle n_M^{rand} \rangle$ et $\langle \sigma_M^{rand} \rangle$ sont respectivement la moyenne et l'écart-type du nombre d'apparition dans un graphe aléatoire qui lui est similaire.

2.4.11 Structure de communauté

La recherche de communauté dans les réseaux a longtemps intéressé les sociologues. Intuitivement, dans un réseau social, les individus se regroupent naturellement en communautés qui sont liés par l'amitié, le travail, le loisir, la parenté etc. Aujourd'hui, la recherche sur la *structure de communauté* dans les réseaux attire beaucoup l'attention des chercheurs de divers domaines : physiciens, biologistes, informaticiens etc. Proposer une définition formelle de ce qu'est une communauté ne sera pas facile car il existe une panoplie de définitions qui sont différentes selon leur contexte [For10]. Toutefois, une définition largement acceptée, liée à la topologie du réseau, considère une communauté comme un sous-graphe dans lequel il y a une forte concentration de liens par rapport au reste du réseau (Girvan-Newman dans [GN02 ; NG03]). Cette définition met l'accent sur un élément essentiel qui est la forte densité de connexion des nœuds internes comparée aux nœuds externes comme l'indique la figure 2.7. Dans les perspectives d'identifier des structures de communauté sur les graphes, plusieurs algorithmes ont été proposés par les chercheurs et de nouvelles méthodes ne cessent d'apparaître régulièrement. Toutes les techniques proposées utilisent en général la méthode de partitionnement du graphe. Il faut aussi souligner qu'ils fonctionnent pour des graphes qui sont connexes [CHI10].

Parmi les méthodes de détection de communauté proposées, nous avons celles qui s'appuient sur les techniques traditionnelles de divisions successives du graphe en des ensembles (de tailles prédéterminées) selon le critère suivant : le partage est choisi de manière à conserver le plus de liens possibles dans chacun des groupes créés, et ceux qui maximisent la modularité Q (qualité d'une partition) introduite par Girvan et Newman [NG03]. Q est donnée par la formule suivante :

$$Q = \sum_J (e_{II} - a_I^2) \tag{2.24}$$

où e_{ij} représente la fraction des liens associant un nœud de la communauté I un nœud de la communauté J et $a_I = \sum_J e_{IJ}$ est la fraction des liens ayant une extrémité au moins dans I.

Par ailleurs, Palla et *al.* proposent une autre approche de structure de communauté avec la notion de communauté recouvrante (*overlapping community*) qui considère qu'un nœud peut appartenir à plusieurs communautés [GV05]. Cette approche de communauté recouvrante devient de plus en plus utilisée dans la détection de communauté.

2.5 Modélisation des Réseaux Complexes

La modélisation des réseaux complexes réels est un peu délicate du fait de leur évolution macro- et micro-échelle. Au niveau macroscopique, la structure du réseau peut être altérée, d'une part, par l'absence ou la présence de nouvelles interactions et d'autre part, par l'apparition et la disparition de nœuds. De même on note la formation de structure de communauté qui peuvent évoluer ou bien subir un déclin en fonction du temps. Au niveau microscopique, l'état

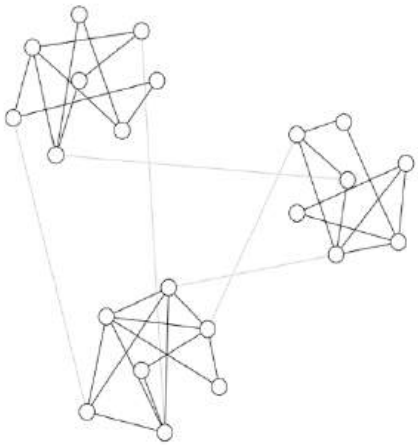


FIGURE 2.7: Exemple de réseau montrant trois structures de communauté. D'après [GN02]

du système peut être modifié par une dynamique intrinsèque qui est un changement d'états des liens ou des nœuds. Cependant, les premiers modèles tentant de modéliser ces systèmes sont les graphes aléatoires statiques d'Erdös et de Rényi, les modèles de Watts et Strogatz et les graphes évoluants de Barabási & Albert ensuite viennent d'autres approches de modélisation permettant de prendre mieux en compte l'évolution topologique et la dynamique interne du réseau.

2.5.1 Erdös Rényi Random graphs

L'étude des graphes aléatoires ou *random graphs* a été initiée par Erdös et Rényi en 1959 dans leur premier article intitulé *On random graphs* [ER59]. Dans ces travaux, ils proposent un premier modèle de graphe aléatoire avec N nœuds et E arêtes appelé *Erdös Rényi (ER) Random Graph* que nous noterons $G^{ER}_{(N,E)}$. Ce graphe est défini en partant d'un graphe sans arêtes, donc N nœuds totalement déconnectés, puis on sélectionne un couple de nœud tiré de manière *aléatoire uniforme* pour former les arêtes jusqu'à avoir E liens. Un autre modèle alternatif équivalant à la première, dénommé $G^{ER}_{(N,p)}$, semble le plus étudié. Dans ce dernier, on part de N nœuds et chaque arête est indépendamment présente avec une probabilité p et absente avec une probabilité $1-p$. $G^{ER}_{(N,p)}$ est ainsi construit dans son ensemble avec une probabilité $p^E(1-p)^{N(N-1)/2-E}$. Un exemple de ce modèle est illustré par la figure 2.8. Cependant, il est montré que la complexité pour générer un graphe $G^{ER}_{(N,p)}$ est en $O(N^2)$ car il faut décider pour chaque paire de sommets si le lien associé existe ou pas, alors que la complexité vaut $O(E)$ pour générer un graphe du type $G^{ER}_{(N,E)}$ [New03 ; AB02].

Souvent, les caractéristiques du graphe $G^{ER}_{(N,p)}$ ne sont pas étudiées en fonction de p mais plutôt en fonction du degré moyen $\langle k \rangle$. À la fin de la procédure, le nombre d'arêtes moyen est de $E = pN(N-1)/2$ et le nombre moyen de connexion correspond à deux fois ce résultat puisque que chaque arête a deux extrémités. Nous pouvons en déduire le degré moyen :

$$\langle k \rangle = \frac{N(N-1)p}{N} = (N-1)p \simeq Np \tag{2.25}$$

Erdös et Rényi étaient les premiers à étudier la distribution du maximum et du minimum de degrés des graphes aléatoires alors que la distribution complète des degrés sera trouvée plus tard par Bollobás en 1981 [Bol81]. Dans un *Random Graph*, la probabilité qu'un nœud i ait

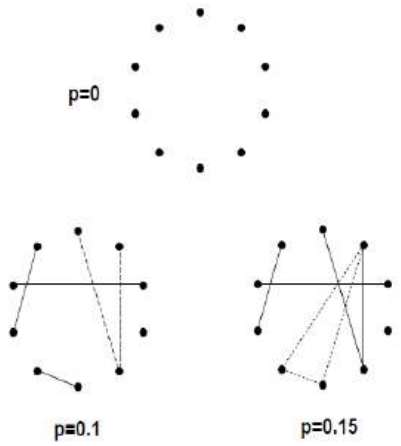


FIGURE 2.8: Exemple du processus de création du modèle de ER Random Graph $G^{ER}_{(N,p)}$, extrait dans Réf [AB02]

un degré $k = k_i$ suit une loi binomiale de paramètres $N-1$ et p.

$$P(k = k_i) = C^k_{N-1} p^k (1-p)^{N-1-k} \tag{2.26}$$

où p^k est la probabilité d'avoir k arêtes, $(1-p)^{N-1-k}$ est la probabilité d'absence des $N-1-k$ arêtes restantes et C^k_{N-1} est le nombre de fois que ces arêtes peuvent être sélectionnées. Lorsque $N \to \infty$, et $p \to 0$, pour une valeur de $\langle k \rangle$ fixée, nous avons la distribution binomiale qui tend vers une distribution de Poisson

$$P(k = k_i) \simeq \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle} \tag{2.27}$$

C'est la raison pour laquelle cette distribution homogène est souvent appelée *Poisson Random Graph*. Pourtant, cette distribution est tout à fait différente de celle des réseaux réels dont la distribution suit, en général, la loi puissance.
L'étude des *ER Random Graph* révèle une propriété intéressante qui est le phénomène des valeurs critiques, la percolation. Ce phénomène révèle, à une certaine valeur seuil de probabilité, $p_c = \frac{1}{N}$ (correspondant à $\langle k \rangle_c = 1$), l'apparition d'une composante géante.

Les graphes aléatoires reproduisent bien l'une des principales caractéristiques des réseaux réels discutée dans la section 2.4.5, l'effet petit-monde. Le diamètre d'un graphe aléatoire tourne autour de $Diam = \ln N / ln(pN) = \ln N / ln\langle k \rangle$. Par conséquent, la longueur moyenne des plus courts chemins l est aussi une fonction de N et a le même comportement que le diamètre $l \sim \ln N / ln\langle k \rangle$. Ainsi, la longueur l présente une valeur faible qui correspond à l'effet petit-monde. Par contre, les autres propriétés des graphes aléatoires ne riment pas avec celles observées dans les réseaux réels. C'est le cas du coefficient de clustering qui varie autour de $C = p = \langle k \rangle / N$, cette valeur devient de plus en plus faible lorsque N est grand. Ce modèle présente aussi un degré de distribution de type *"Poissonnière"* qui est contraire à celui des réseaux réels. En plus, il n y a pas de corrélation entre le degré des nœuds adjacents, ni la présence de structure de communauté [New03].

Malgré ces manquements, Bender et Canfield (1978) proposent une généralisation de ce modèle, appelée modèle configurationnel [BC78]. C'est un modèle basé sur la génération

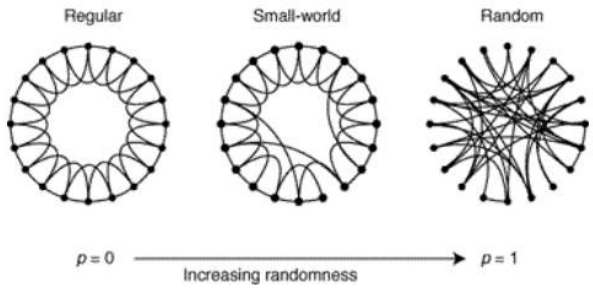


FIGURE 2.9: Procédure de recâblage de Watts et Strogatz. D'après [WS98]

de graphe avec une distribution des degrés contrôlée (pas forcément *"poissonnième"*). Par conséquent, ce modèle permettra de construire un graphe aléatoire avec une distribution hétérogène des degrés (c'est-à-dire on peut faire de sorte que la distribution des degrés suive une loi puissance). Malgré cette distribution hétérogène, ce modèle présente toujours un coefficient de clustering faible mais aussi ne possède pas de structure de communauté.

2.5.2 Modèle de Watts et Strogatz

C'est l'un des premiers modèles utilisés pour modéliser les graphes dits "petit-monde " (*small-world*). Selon Watts et Strogatz, un small-world est un réseau caractérisé par un fort coefficient de clustering et une faible valeur de la moyenne des longueurs (*Average path length*) [WS98]. Ce modèle tente d'expliquer ce phénomène de *small-world* qui à été découvert dans plusieurs réseaux réels (voir tableau 2.1). Le principe de ce modèle est de passer d'un graphe régulier à un graphe aléatoire afin d'obtenir un graphe petit-monde, en utilisant une procédure de recâblage des liens selon une probabilité p (voir figure 2.9). Initialement, les auteurs démarrent le processus avec un graphe régulier sous forme d'anneau de N nœuds fixés dont chaque nœud est symétriquement lié avec ces $2m$ plus proches voisins. Les nœuds sont visités l'un après l'autre : chaque lien connectant un nœud à son voisin dans le sens des aiguilles d'une montre est replacé vers un autre nœud choisi aléatoirement avec une probabilité p, ou bien préservé avec une probabilité $1-p$. Il faut souligner que dans le cas où $p=0$, nous avons un graphe régulier tandis que pour $p=1$, le modèle produit un graphe complètement aléatoire qui a au moins un degré minimum égal à m. Entre les deux valeurs extrêmes, nous avons un graphe qui présente un effet petit-monde avec un coefficient de clustering non trivial. Pour comprendre l'existence de l'effet petit-monde et de clustering dans le modèle de Watts et Strogatz (WS), il faut étudier le comportement du coefficient de clustering $C(p)$ et celui de la moyenne des longueurs $l(p)$ en fonction de la probabilité de recâblage p. Dans le cas où l'anneau est régulier, le calcul du coefficient de clustering donne une grande valeur, $C(0) = (3m-3)/(4m-2)$, qui tend vers $3/4$ lorsque m devient large. En plus, la moyenne des longueurs tourne autour de $l(0) \simeq N/4m$, ce qui fait que les graphes réguliers ne présentent pas un effet petit-monde. Si $p \to 1$, le modèle de WS converge vers un graphe aléatoire de moyenne géodésique $l(1) \simeq ln(N)/ln(2m)$ et de coefficient de clustering $C(1) \simeq 2m/N$. Ainsi entre les deux étapes, on note une descente logarithmique de $l(p)$ et une décroissance de $C(p)$. Mais ce qui attire plus l'attention de Watts et Strogatz est la rapide descente de la longueur des plus courts chemins $l(p)$, pendant que $C(p)$ reste encore élevé entre un intervalle de valeurs de

TABLE 2.1: Mesures du coefficient de clustering et de la moyenne des longueurs de plus courts chemins des réseaux biologiques, sociaux et technologiques : C, coefficient de clustering ; C_{rand}, coefficient de clustering dans un réseau aléatoire équivalent ; L, la moyenne des longueurs de plus courts chemins ; L_{rand}, la moyenne des longueurs de plus courts chemins dans un réseau aléatoire équivalent. D'après [Tel+11]

Réseaux	Nombre de noeuds	C	C_{rand}	L	L_{rand}
Karate club[Zac77]	35	0.55	0.31	2.41	2.24
Word adjacency [NG03]	112	0.17	0.19	2.54	2.49
Caenorhabditis elegans [DA05]	453	0.65	0.28	2.66	2.50
E-mail [Gui+03]	1133	0.22	0.09	3.60	3.27
Protein interactions [Jeo+01]	1539	0.07	0.04	6.81	5.69
Internet	22 963	0.23	0.09	3.84	3.58

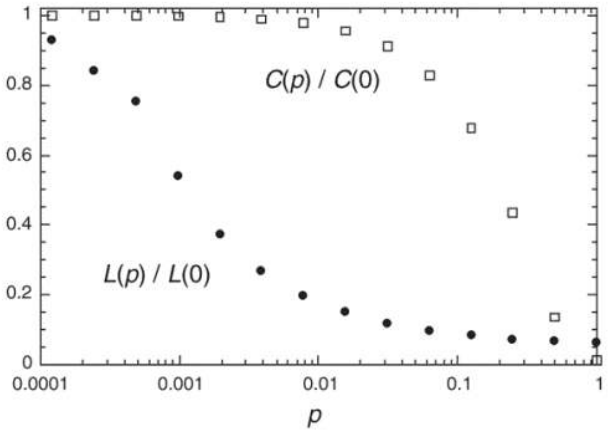


FIGURE 2.10: Caractéristiques de la moyenne de longueurs l et du coefficient de clustering C dans le modèle de WS

p donné (voir figure 2.10) [WS98]. En effet, selon Watts, l'origine de cette rapide descente de l caractérise l'apparition des "raccourcis " (*shortcuts*) entre les nœuds. Chaque "raccourci ", créé aléatoirement, permet la jonction de deux parties éloignées du graphe. Ainsi, il aura un impact significatif sur la caractéristique de la moyenne des longueurs l du graphe entier. Comme les arêtes originales du graphe sont remplacées de manière aléatoire, sur un intervalle $0 \ll p \ll 1$, alors le coefficient de clustering reste encore proche de $C(0)$. Par conséquent, dans la région où $1/N \ll p \ll 1$, les auteurs ont remarqué la coexistence d'une faible valeur de l et d'un fort coefficient de clustering C qui constitue une des caractéristiques présentes dans la plupart des réseaux réels [AB02].Ce qui explique, dans ces réseaux la disposition des nœud n'est pas complètement aléatoire mais aussi on peut atteindre n'importe qu'elle point avec un nombre faible de lien.

Malgré les nombreuses avancées qu'il apporte dans l'étude des réseaux complexes, le modèle original proposé par Watts et Strogatz présente quelques difficultés. La distribution des degrés du modèle de Watts et Strogatz ne correspond pas à la plupart de celle des réseaux du monde réel, mais ce n'est pas surprenant du fait que ce n'était pas l'objectif initial de ce modèle [New03]. Par ailleurs, le fait qu'une seule des extrémités d'une arête soit redirigée, qu'aucun

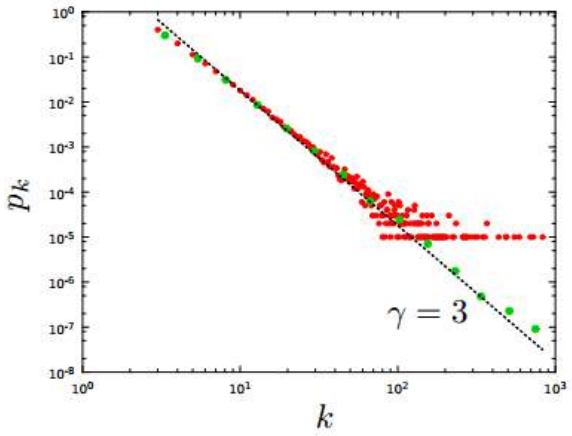


FIGURE 2.11: Distribution des degrés généré par le modèle de Barabási-Albert sur un réseau de N nœuds avec $m = 3$ ($P(k) \sim k^{-3}$)

nœud ne puisse être connecté à lui-même, et qu'une arête ne soit jamais ajoutée entre des paires de nœuds où il en existe déjà une, rend difficile l'énumération ou le calcul de la moyenne des longueurs sur l'ensemble du graphe. C'est pourquoi plusieurs variantes ont été proposées parmi lesquelles le modèle de Newman, Moore et Watts [MW00]. Ce modèle reproduit le modèle de WS à la seule différence qu'ici les arêtes ne sont plus remplacées mais de nouvelles arêtes servant de "raccourcis " sont ajoutées entre deux nœuds choisis suivant une probabilité p.

2.5.3 Scale-free de Barabási & Albert

Le modèle de Barabási & Albert est l'un des premiers modèles qui utilisent l'*attachement préférentiel* pour générer un réseau qui évolue dans le temps [Col14]. Pourtant, les réseaux "invariants d'échelle " (*scale-free*) sont étudiés, en premier, par le physicien anglais Derek de Solla Price en 1965. Dans son article intitulé *Networks of scientific papers*, Price montre que le réseau de citations des papiers scientifiques suivait une distribution de loi puissance [Sol65]. Il découvrit étonnamment qu'au fur et à mesure que le réseau s'agrandit, le degré des nœuds les plus connectés devient de plus en plus grand par rapport à ceux moins connectés, ce qu'il appela *l'avantage cumulatif*. Ce terme est connu, aujourd'hui, sous le nom d'*attachement préférentiel*, introduit par Barabási et Albert [BA99]. L'idée de Price était que le nombre de citations reçues par un papier était proportionnel au nombre de citations qui le désignaient auparavant. Ce même comportement est remarqué dans certains réseaux comme le World Wide Web [New03].

Barabási et Albert créent en 1999 un modèle de réseau qui ressemble beaucoup à celui découvert par Price. Contrairement au modèle de Price, le modèle de Barabási et Albert (BA) présente des arêtes non orientées, ce qui fait perdre une information importante dans le réseau. Pourtant, en ignorant la nature des orientations du réseau, le modèle de BA permet de contourner le problème que rencontrait Price c'est-à-dire comment un papier reçoit sa première citation (ou un site web son premier lien). Barabási et Albert étudient le réseau World Wide Web en s'appuyant sur deux règles fondamentales : *la croissance du réseau* et *l'attachement préférentiel* [BA99]. Ces deux règles sont utilisées pour générer, à partir d'un ensemble m_0

de nœuds connectés, un graphe non orienté $G^{BA}_{(N,E)}$. À chaque étape $t = 1,2,...,N - m_0$ de l'algorithme, un nouveau nœud j avec $m \leq m_0$ liens est ajouté dans le réseau. La probabilité qu'un lien connecte un nœud j à un nœud existant i est linéairement proportionnelle au degré de i.

$$\prod_{j \to i} = \frac{k_i}{\sum_l k_l} \tag{2.28}$$

Le réseau aura à chaque instant t un nombre de nœuds $N = t + m_o$ et un nombre d'arêtes $E = mt$, avec un degré moyen correspondant à $\langle k \rangle = 2m$. Des simulations numériques indiquent que le graphe généré par le modèle de Barabási et Albert évolue spontanément sous la forme d'une distribution de loi puissance $P(k) \sim k^{-\lambda}$, avec $\lambda = 3$ (figure 2.11). Selon les auteurs, l'origine de la loi puissance est essentiellement due à l'attachement préférentiel qui entraîne la présence d'un faible groupe de nœuds de fortes connexions appelés "*hubs* ". Cette faible présence de "*hubs* " a pour conséquence la résistance aux pannes mais laisse le réseau vulnérable aux attaques ciblées.

Des expériences ont montré que la moyenne des longueurs l (*Average path length*) est plus faible dans les *scale-free* que dans les *random graphs*, $l \sim \frac{logN}{loglogN}$ [AB02].

Le modèle de Barabási-Albert (BA) [BA99] est l'un des modèles de réseau évolutif et dynamique qui présentent une caractéristique longtemps observée dans les réseaux réels : la distribution des degrés en loi puissance (*power law*). Cette caractéristique découle, selon les auteurs, principalement de l'attachement préférentiel. Néanmoins, ce modèle produit un réseau de structure arborescente alors que les réseaux réels ne présentent jamais leurs structures sous cette forme. Plusieurs modifications ou extensions ont été apporté pour améliorer ce modèle. Parmi les variantes du modèle de Barabási-Albert, Krapivsky et Redner considèrent la version orientée de ce modèle [KR01]. Ils ont étudié l'âge a d'un nœud i, défini comme le nombre de nœuds ajouté après i, son degré et la corrélation avec le degré des nœuds adjacents. Ils ont montré que le nombre de nœuds qui ont un âge a avec un nombre de liens égal à k est de :

$$N_a(k) = \sqrt{1 - \frac{a}{N}} \left(1 - \sqrt{1 - \frac{a}{N}}\right)^k$$

Une autre alternative du modèle Barabási-Albert fut proposée plus tard par les mêmes auteurs [R A99]. La méthode est basée sur l'algorithme itérative suivante : a) m nouveau nœuds sont ajoutés avec une probabilité p ; b) m arêtes sont reconnectées avec une probabilité q ; c) un nouveau nœud est ajouté avec m arêtes qui se connectent préférentiellement avec m nœuds existants selon une probabilité $1 - p - q$.

Il y a aussi le modèle de Dorogovtsev-Mendes-Samukhin(DMS) [DMS00] qui ajoute les nœuds selon l'attachement préférentiel suivant :

$$\prod_{j \to i} = \frac{k_i + k_0}{\sum_l (k_l + k_0)} \text{ avec } -m < k_0 < \infty$$

Krapivsky et *al.* prouvent qu'en utilisant l'attachement préférentiel généralisé $\prod_{j \to i} = \frac{k_i^\beta}{\sum_l k_l^\beta}$ le degré de distribution du réseau émerge en loi puissance seulement si $\beta = 1$. Lorsque $\beta > 1$ le degré de distribution est exponentiel tandis que si $\beta < 1$ un seul site se connecte presque à tous les autres [KR01].

2.5.4 Réseaux Adaptatifs

Comme nous l'avons mentionné au début, les réseaux complexes sont marqués par une coévolution de la dynamique du niveau macroscopique qu'au niveau microscopique. Il n'est pas toujours facile de prendre en compte en même temps, dans certains modèles, la transformation de la topologie du réseau (niveau macroscopique) et le changement d'état des nœuds du réseaux (niveau microscopique). Dans les réseaux adaptatifs, la transition de chaque composante et la transformation topologique y sont couplées pour comprendre et modéliser l'évolution dynamique dans les systèmes complexes. Les premiers modèles de Réseaux Adaptatifs sont initiés par Bornholdt et Rohlf pour étudier l'auto-organisation du cerveau dans ses états critiques [BR00]. Bornholdt et Rohlf ont défini un concept de réseau qui évolue avec une activation/désactivation des liens pour expliquer comment le niveau critique du cerveau peut être atteint [BR00]. Ensuite, les Réseaux adaptatifs sont utilisés pour modéliser plusieurs phénomènes dynamiques tels que l'auto-organisation de systèmes biologiques [Fri+09], la synchronisation [ZK06], la formation d'opinion [HN06] [DG10], la propagation épidémique [GDB06] [Fun+09] etc. Dans [Say+13], H. Sayama et al. utilisent le concept de réseau d'automate généré ou *Generated Network Automata* (GNA) pour représenter les réseaux adaptatifs. Selon [Say+13], un GNA est défini comme un graph formé de nœuds qui sont dynamiques et des liens orientés. Les relations non orientées sont représentées par une paire de liens symétriquement orientés entre deux nœuds. Chaque nœud du graphe a un état s qui prend sa valeur dans l'ensemble des états des nœuds, S. Ensuite, les liens décrivent les relations référentielles entre les nœuds en spécifiant comment les nœuds s'influencent mutuellement dans la transition d'état et la transformation topologique. Chaque lien peut aussi prendre une valeur possible dans l'ensemble des états des liens, $S^{'}$. Ainsi, la configuration de GNA pour un temps t donné, constituée par l'ensemble des états et des topologies du réseau, est formellement définie par les concepts suivants :

- V_t : l'ensemble des nœuds à l'instant t. On suppose ici que cet ensemble peut varier en fonction du temps à cause des ajouts et suppression de nœuds.
- $C_t : V_t \rightarrow S$: Une application définie de l'ensemble V_t vers l'ensemble des états des nœuds, S, qui décrit l'état global du réseau à l'instant t. Si les états locaux sont des scalaires alors C_t peut être représentée par un vecteur dont la taille peut varier en fonction du temps.
- $L_t : V_t \rightarrow \{V_t \times S^{'}\}$: Une application définie d'un ensemble des nœuds de départ vers un ensemble de nœuds d'arrivés et l'ensemble des états des liens. Elle représente la topologie globale du réseau à l'instant t, qui peut varier en fonction du temps.

Dans cette représentation, la modélisation du Réseaux Adaptatifs correspond la mise à la jour des états et de la topologie à travers des séquences répétitives d'événement décrites comme suit :

- Extraction dans le graphe GNA du sous-graphe (subGNA) qui est supposé avoir des modifications.
- Production d'un nouveau subGNA qui remplacera le sous-graphe subGNA qui était déjà sélectionné
- Recollage du nouveau subGNA dans le graphe GNA.

Par conséquent, la dynamique dans GNA peut être définie, selon les auteurs, par le triplet

$\langle \boldsymbol{E}, \boldsymbol{R}, \boldsymbol{I} \rangle$:

- $\boldsymbol{E}$: Un mécanisme d'extraction déterminant la partie du GNA qui doit être sélectionnée pour la mise à jour. Il est défini comme une fonction qui prend toute la configuration de GNA et renvoie un sous-graphe subGNA spécifique qui doit être remplacé. Il peut être déterministe ou stochastique.
- $\boldsymbol{R}$: Un mécanisme d'extraction qui produit un nouveau subGNA à partir du subGNA extraite par E et qui spécifie la correspondance entre les anciens et les nouveaux nœuds dans subGNA. Il est défini comme une fonction qui prend une configuration subGNA et renvoie une paire d'une nouvelle configuration subGNA et une correspondance entre les nœuds de l'ancien subGNA et les nœuds du nouveau subGNA. Il peut être déterministe ou stochastique.
- $\boldsymbol{I}$: Configuration initiale de GNA.

L'exemple de la figure 2.12 illustre en détails le processus de régénération du graphe GNA en fonction du triplet $\langle \boldsymbol{E}, \boldsymbol{R}, \boldsymbol{I} \rangle$. Les graphes GNA sont flexibles et adaptés pour reproduire d'autres modèles dynamiques de réseaux complexes. Par exemple, il est très facile d'utiliser GNA pour représenter le phénomène d'attachement préférentiel dans les réseaux *scale-free*. H. Sayama et *al.* ont utilisé ce modèle de représentation afin de mettre en place un système de recherche et de sauvetage dans l'Arctique canadien appelé SARnet (*Arctic Search and Rescue Network model*). Cet application implémentée dans le simulateur *OpNetSim* [PSW12] a permis aux chercheurs de faire la simulation d'un incident SAR réel dans l'Arctique survenu en décembre 2008 [Say+13] en créant un réseau opérationnel basé sur cet approche. De même, un autre exemple d'application utilisant le graphe GNA a été proposé par Yamanoi et Sayama pour montrer le processus d'intégration culturelle en utilisant le concept de réseau social adaptatif [YS13]. Zhou et Xia étudient la propagation épidémique d'un modèle SIS (*Suspected-Infected-Suspected*) [ZX14] sur un réseau adaptatif pondéré par la force de contacts entre deux individus. Ils ont développé un processus de transmission basé sur la méthode d'adaptation du poids des contacts tel que, d'une part les individus sains ont tendance à réduire leurs poids des liens avec les individus infectés et, d'autre part, à augmenter ce poids de connexion avec d'autres individus sains. Par conséquent, ils ont montré que cet approche aurait un impact considérable sur le taux de prévalence et peut être utile dans la stratégie de contrôle de la propagation [ZX14].

Par ailleurs, d'autres travaux ont été effectués dans les Réseaux Adaptatifs pour modéliser les systèmes complexes. Le livre de Gross et Sayama a introduit une compilation d'articles, d'état de l'art, regroupant les principaux thématiques concernant les Réseaux Adaptatifs [GS09]. Malgré les efforts qui ont été fournis dans ce domaine, les Réseaux adaptatifs restent encore peu utilisés dans la modélisation des réseaux complexes dynamiques.

2.5.5 Réseaux Multiplexes

Les nombreuses contraintes constatées dans l'approche traditionnelle de modélisation des réseaux complexes qui considère les réseaux réels comme des systèmes statiques et isolés, rendent difficile la compréhension de certains phénomènes émergents. Cependant, il faut comprendre que les réseaux réels sont composés en général de différents entités dynamiques, liées entre elles par diverses relations ou interactions. Ceci avait encouragé très tôt l'étude

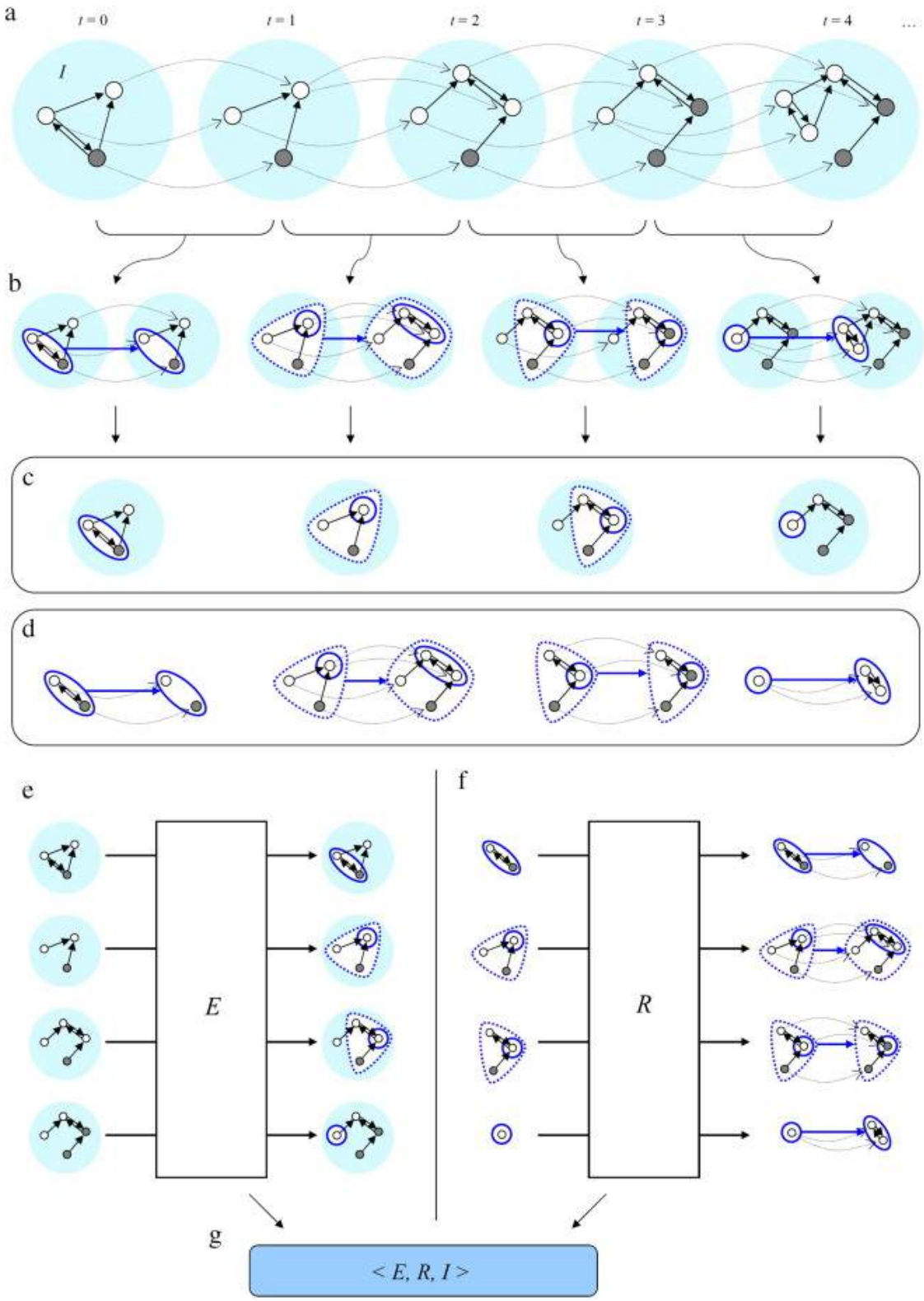


FIGURE 2.12: Processus de régénération de GNA en fonction de l'algorithme $\langle \boldsymbol{E}, \boldsymbol{R}, \boldsymbol{I} \rangle$: (a) Évolution du modèle en commençant par la configuration initiale I. (b) Détection des événements de réécriture à chaque pas de temps $\boldsymbol{ts}$. (c) Mécanisme d'extraction E . (d) Mécanisme de remplacement R. (e, f) Construction de modèles de E et R sur la base des données d'apprentissage. (g) Modèle GNA final. D'après [Say+13]

des réseaux comportant plusieurs types de nœuds et différents types de connexions, appelés "réseaux de réseaux", dans le domaine de la sociologie et de l'ingénierie. Cependant, cette approche traditionnelle ne rendait pas facile leurs interprétations. C'est dans cet optique qu'a émergé une nouvelle approche de représentation des réseaux complexes appelée réseaux *"multi-couches"* (*Multilayer Networks*). Les réseaux multi-couches explicitent plusieurs canaux de connexité et constituent un système idéal pour décrire les réseaux interconnectés à travers différentes catégories de connexions dont chaque canal (relation, activité, catégorie) est représenté par une couche et un même nœud ou entité peut avoir différents types d'interactions c'est-à-dire différents types de voisins dans chaque couche[Boc+14]. Selon Boccaletti, ces réseaux multi-couches constituent en effet, une généralité de tous les réseaux qui peuvent se représenter sous forme de couches comme par exemple les réseaux *multiplexes*, les réseaux *multidimensionnelles*, les réseaux *multi-niveaux*, réseaux *inter-connectés*, les *hypergraphes*, les réseaux *hétérogènes* etc.[Boc+14]. En effet, pour être plus général, nous allons annoncer la définition mathématique des réseaux multi-couches en passant par la définition des réseaux dit multiplexes qui deviennent de plus en plus utilisés dans la modélisation de l'évolution dynamique des nœuds de même nature.

Definition 2.2. Un réseau *multi-couche* est un pair $\mathscr{M} = (\mathscr{G}, \mathscr{C})$ tel que

$$\mathscr{G} = \{G_\alpha ; \alpha \in \{1, \dots, M\}\} \tag{2.29}$$

est une famille de graphes (orientée, non-orienté ou pondéré) $G_\alpha = (X_\alpha, E_\alpha)$ et

$$\mathscr{C} = \{E_{\alpha\beta} \subseteq X_\alpha \times X_\beta ; \alpha \in \{1, \dots, M\}, \alpha \neq \beta\} \tag{2.30}$$

est l'ensemble des interconnexions reliant les couches G_α et G_β avec $\alpha \neq \beta$. Les éléments de $\mathscr{C}$ sont appelés couches croisées (*crossed layers*) et les éléments de E_α sont les connexions intra-couches de $\mathscr{M}$ contrairement aux éléments de $E_{\alpha\beta}$ $(\alpha \neq \beta)$ qui représentent les connexions inter-couches [Boc+14].

Les réseaux multi-couches peuvent être représentés sous forme matricielle ou bien sous forme tensorielle. Ici, pour plus de simplicité nous allons nous limiter dans la représentation matricielle (pour plus de détails dans la représentation tensorielle voir [Kiv+14] [De +13]). Si nous considérons, les nœuds de la couche G_α par $X_\alpha = \{x_1^\alpha, \dots, x_{N_\alpha}^\alpha\}$ alors la matrice d'adjacence de la couche G_α est notée par $A^{[\alpha]} = (a_{ij}^\alpha) \in \mathbb{R}^{N_\alpha \times N_\alpha}$ telle que :

$$a_{ij}^\alpha = \begin{cases} 1 & \text{si} \quad (x_i^\alpha, x_j^\alpha) \in E_\alpha, \\ 0 & \text{sinon} \end{cases} \tag{2.31}$$

avec $1 \leq i, j \leq N_\alpha$ et $1 \leq \alpha \leq M$. De plus, la matrice adjacence correspondant à l'inter-couche $E_{\alpha\beta}$ est donnée par la matrice $A^{[\alpha\beta]} = (a_{ij}^{\alpha\beta}) \in \mathbb{R}^{N_\alpha \times N_\beta}$ définie par :

$$a_{ij}^\alpha = \begin{cases} 1 & \text{si} \quad (x_i^\alpha, x_j^\beta) \in E_{\alpha\beta}, \\ 0 & \text{sinon} \end{cases} \tag{2.32}$$

La projection du réseau multi-couche $\mathscr{M}$ donne un graphe (*mono-layer*) $\mathtt{proj}(\mathscr{M})$ équivalent à $G_{\mathscr{M}} = (X_{\mathscr{M}}, E_{\mathscr{M}})$ tel que :

$$X_{\mathscr{M}} = \bigcup_{\alpha=1}^{M} X_\alpha, \quad E_{\mathscr{M}} = \left(\bigcup_{\alpha=1}^{M} E_\alpha\right) \left(\bigcup_{\substack{\alpha,\beta=1 \\ \alpha \neq \beta}}^{M} E_{\alpha\beta}\right) \tag{2.33}$$

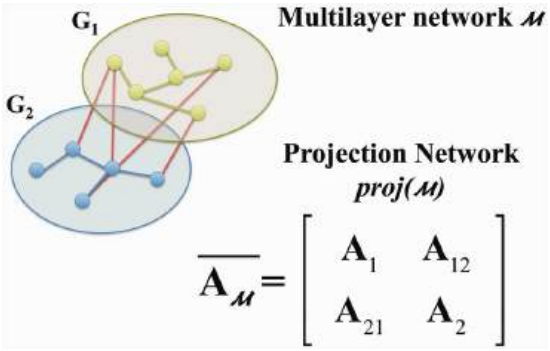


FIGURE 2.13: Une représentation d'un réseau multi-couche composé de deux graphes : G1 et G2. Les connexions inter-couches sont en rouge tandis que les connexions intra-couches sont en vert pour G1 et en bleu pour G2. La matrice d'adjacence du réseau de projection associé $proj(M)$ est affichée dans le coin inférieur droit [Cas+16]

.

La Figure 2.13 illustre la projection matricielle $\bar{A}_{\mathcal{M}}$ correspondant à un réseau multi-couche $\mathcal{M}$ composé de deux 2 couches (G1 et G2) [Cas+16].

Dans la littérature, il existe plusieurs réseaux de réseaux inter-connectés pouvant être assimilés comme étant des réseaux multi-couches : les réseaux de métabolisme génétiques[KS18 ; Kiv+14], réseaux de contacts temporels [Sah+19], réseaux de transports [GB15], les réseaux sociaux [BK12 ; Kiv+14], les réseaux d'informations [Che+18] etc. La Figure 2.14a montre un exemple de réseau multi-couche qui est le réseau de transport aérien constitué de 3 couches, dont chacune représente une compagnie aérienne (Lufthansa, Air-France, British-Airways). Un autre exemple de réseau multi-couche correspondant à l'interconnexion de différentes catégories de réseaux sociaux est montré dans la Figure 2.14b.

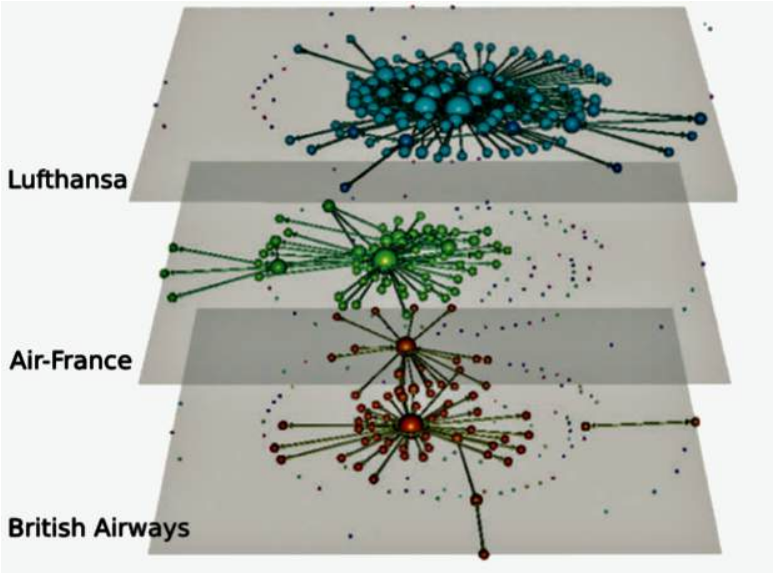
Definition 2.3. Un Réseau Multiplexe est typiquement un réseau multi-couche dont le nombre de nœuds par couche est le même ($X_1 = X_2 = \ldots = X_M = X$) et les seuls connexions inter-couches qui existent sont ceux qui lient un nœud avec leur semblable dans les couches homologues [Boc+14]. Autrement dit,

$$E_{\alpha\beta} = \{(x,x); x \in X\} \text{ for every } \alpha,\beta \in \{1,\ldots,M\}, \alpha \neq \beta \tag{2.34}$$

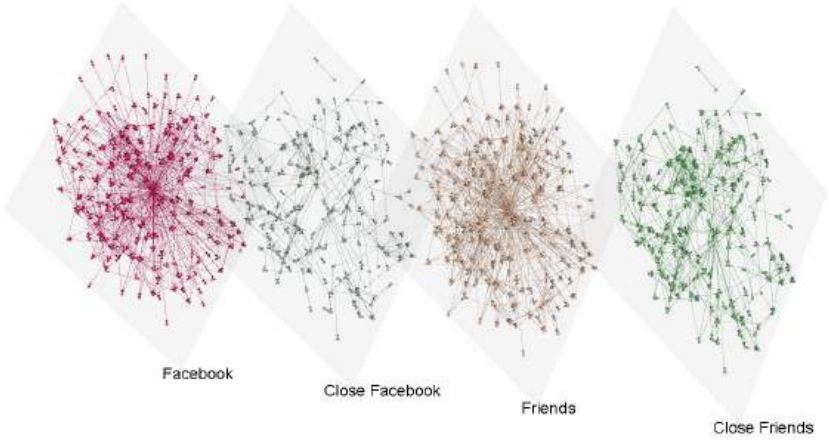
De même pour les réseaux multi-couches, tous les concepts définis dans les réseaux complexes ont été étendus dans les réseaux multiplexes : la modularité, le coefficient de clustering, la moyenne des plus courts chemin, l'efficience locale et globale, l'effet petit-monde, l'assortativité, l'attachement préférentiel, la corrélation des degrés et d'autres mesures permettant d'analyser ces graphes [Kiv+14 ; Boc+14 ; De +13 ; AM18].

2.5.6 Réseaux Temporels

Les réseaux temporels ou *Time Varying Graphs* (TVG) sont définis comme des séquences instantanés ou *snapshots* de graphes statiques évoluant sur une séquence temporelle $\{t_1, t_2, \ldots, t_i, t_{i+1}, \ldots\}$ tel que $t_{i+1} = t_i + \Delta t$ [HS13]. En dépit de la difficulté de prendre en compte la dynamique dans un système complexe, ce type de graphe peut être un outil très adapté dans l'étude de la dynamique évoluant dans les réseaux complexes. Beaucoup de systèmes complexes peuvent être modélisés à partir des réseaux temporels : réseaux de contact et de proximité,



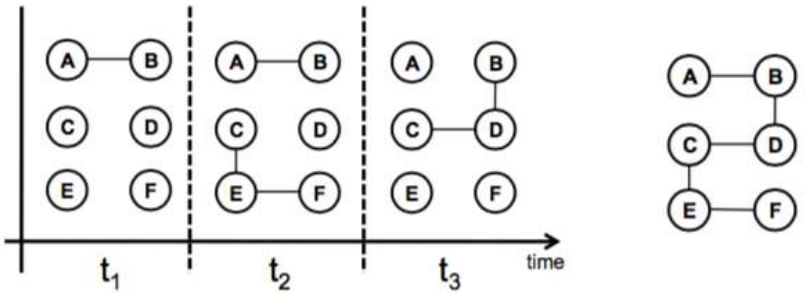
(a)



(b)

FIGURE 2.14: Deux exemples de réseaux multicouches : (a) Exemple de réseau multi-couche de différentes compagnies aériennes (Lufthansa, Air-France, British-Airways), (b) Exemple de réseau social d'amitié multi-couche (Facebook) [Kiv+14]

réseaux de communication, réseaux de collaboration, réseaux de citations, réseaux de mobilité humaine, réseaux écologiques, réseau de cellules nerveuses, réseaux sociaux etc.[Hol15 ; HS13]. Pour statuer sur la motivation d'utiliser un réseau temporel, nous considérons un exemple de processus de dissémination d'une information temporelle dans un réseau complexe. D'abords, en considérant le graphe statique dans la Fig. 2.15b, il existe alors un chemin entre A et F via (A, B, C, D, E, F). Cependant, si l'on tient compte de l'ordre chronologique du graphe temporel (voir Figure 2.15a) A ne peut pas atteindre F. Ce constat est très important si nous voulons suivre par exemple la propagation ou la dissémination d'un pathogène dans le réseau. Si on considère le TVG (Figure 2.15a), l'agent pathogène peut se propager de A vers B, de B vers C, de C vers D, mais pas de D à E ou F. Parce que, la transitivité n'existe pas dans les réseaux temporels.



(a) Graphe Temporel (b) Graphe Statique

FIGURE 2.15: Comparaison entre un réseau temporel et un réseau statique [Tan+10a]

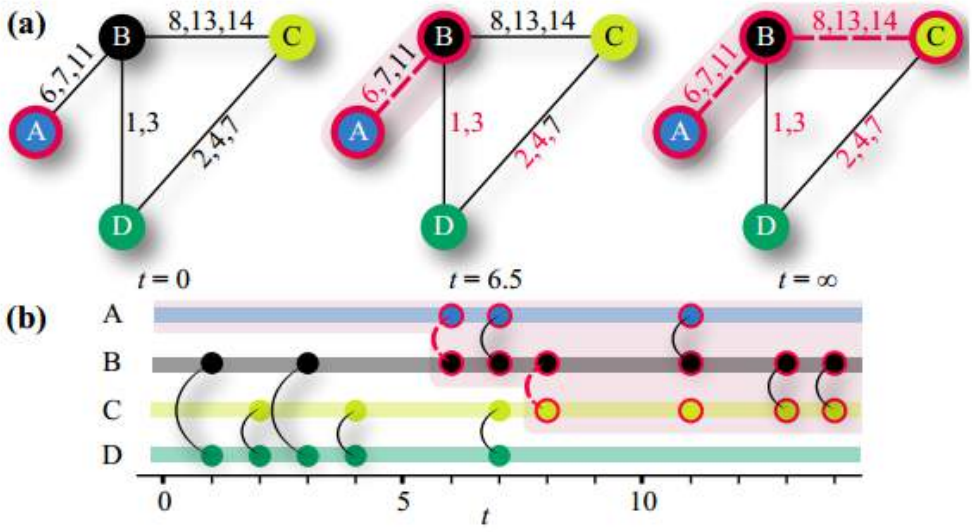


FIGURE 2.16: Représentation d'un réseau temporel (TVG) selon [HS12]

Nous avons trouvé de nombreuses descriptions acceptables de TVGs dans la littérature. Boccaletti suggère dans [Boc+14] qu'un réseau temporel peut être vu comme une extension d'un réseau multi-couche $\mathcal{G} = G(t)_{t=1}^{T}$ dont les couches sont représentés par l'ensemble $\{G_1,\ldots,G_T\}$ où $G_t = G(t)$, $E_{\alpha\beta} = \emptyset$ si $\beta \neq \alpha + 1$, avec

$$E_{\alpha\alpha+1} = \{(x,x); x \in X_\alpha \cap X_{\alpha+1}\}$$

Cette définition est un peu semblable à celle donnée par Tang qui définit un réseau temporel comme une séquence ordonnée de graphes statiques $(G_1,G_2\ldots,G_T)$ telle que $G_t = (V_t,E_t)$ où V_t et E_t sont respectivement l'ensemble des nœuds et l'ensemble de liens dans une fenêtre temporelle t (voir Figure 2.16) [Tan+10a]. Il existe un lien appartenant à E_t entre i et j lorsque i et j sont en relation dans le graphe observé sur une intervalle de fenêtre temporelle. Dans cette définition, l'auteur suppose que l'ensemble des nœuds est maintenu fixe, $V_t = V, \forall t \in [\tau,T[$ et $|V| = N$. Cette définition correspond à la représentation montrée dans la figure 2.15a.

Par ailleurs, Si nous considérons le triplet $(V,E,\mathbb{T})$ tel que V soit l'ensemble des nœuds, E l'ensemble des liens et $\mathbb{T}$ l'échelle de temps, alors un réseau temporel peut être défini comme une *séquence de contacts* (c'est-à-dire un ensemble de contacts C), dont chaque contact est représenté par le triplet (i,j,t) avec $i,j \in V$ et $t \in \mathbb{T} = \{t_1,t_2,\ldots,t_n\}$ (voir l'exemple de cette

définition dans la Figure 2.16) [HS12 ; HS13 ; Hol15]. Selon Holme et Saramäki [HS12], cette représentation, est, d'une part, appropriée au contexte où la durée d'un contact n'est pas très importante par exemple dans les données de communication (SMS, les e-mails, les appels téléphoniques etc.) les réseaux de contacts sexuels etc. [HS12]. D'autre part, si la durée de contact est prise en compte dans la modélisation alors le TVG est représenté par le quadruplet $(i,j,t,\delta t)$ et est appelé *interval graph*. Par exemple, δt peut être considéré comme la durée de communication de deux personnes, durée d'un vol aérien, durée de vie d'une ligne dans un réseau de transport, délai de stimulus d'un flux nerveux, durée de contact lorsque deux personnes sont en proximité etc.

Dans la littérature, il existe d'autres définitions de réseaux temporels. Par exemple, Harary et Gupta définissent un TVG comme un graphe dépendant d'une fonction d'activation des nœuds ($f : V \rightarrow N$ avec $N = |V|$) et une fonction d'activation des liens ($g : E \rightarrow M$ avec $M = |E|$) [HG97]. Casteights et *al.* le définissent comme un quintuplet (V,E,τ,ρ,ζ) tel que V représente l'ensemble des nœuds, E l'ensemble des liens, L un ensemble de label (c.-à-d. un label l pouvant permettre de représenter le poids d'un lien, l'intensité d'une relation sociale, type du transport dans un réseau de transport etc.) tel que $E \subset V \times V \times L$, une durée de vie $\tau \subseteq \mathbb{T}$ (avec $\mathbb{T}$ représentant l'échelle temporelle qui peut être discrète ou continue), une fonction de présence $\rho : E \times \tau :\rightarrow \{0,1\}$ et une fonction de latence $\zeta : E \times \tau : \mathbb{T}$ indiquant la durée d'activation d'un lien [AS12]. En effet, la variation de la définition de réseaux temporels s'explique par le contexte et le domaine dans lequel est utilisé [D B10 ; PE08 ; Col14 ; LM05 ; M K14 ; SM06]. Cependant, toutes ces définitions ont en commun l'aspect temporalité des liens et/ou les nœuds. Par conséquent, dans notre exposé nous allons surtout nous focaliser sur les définitions données par Holme et *al.* [HS12 ; HS13 ; Hol15], présentées ci-dessus, qui donnent une présentation claire des réseaux temporels et plutôt facile à implémenter dans le contexte de l'étude de la propagation dans un réseau complexe. Ainsi, toutes les métriques définies dans les réseaux statiques ont été redéfinies dans le cadre des réseaux temporels.

2.5.6.1 Chemin et connexité dans les TVG

Dans les réseaux statiques, la condition d'accessibilité est basée sur la notion de *chemin* (*path*). On dit qu'un nœud j est joignable à partir d'un nœud i, s'il existe un chemin commençant par i et se terminant par j. Ici, le concept de chemin ne dépend que de la *connexité* entre les nœuds. Contrairement au réseaux statiques, la notion de chemin dans les réseaux temporels dépend de l'ordre chronologique des liens d'où le nom chemin temporel (*'time-respecting path'*) [HS12]. Ainsi, nous pouvons définir le chemin temporel (*'time-respecting path'*) entre un nœud source v et un nœud cible w comme toute séquence de liens marqués par le temps (ou séquence de contacts) telle que :

$$(v_0,v_1,t_1),(v_1,v_2,t_2)\ldots(v_{l-1},v_l,t_l) \tag{2.35}$$

avec $v_0 = v$ et $v_l = w$ et une séquence croissante de valeurs discrètes du temps c.-à-d. $t_1 < t_2 < \ldots < t_l$. Par exemple, dans la figure 2.16, il existe un chemin temporel entre A et B, $(A,B,7),(B,C,8)$, mais pas entre A et D. Dans la littérature, le terme *"time-respecting path"* est aussi appelé *"journey"* [AS12] ou *'temporal path'* [Nic+13 ; Tan+10a]. Comme les réseaux statiques, les notions de *connexité* et de *composants connexes* sont très importantes dans les

TVG. En outre, on peut définir qu'un nœud i est temporellement connecté à j s'il existe un chemin temporel quittant i vers j. Comme la relation de connexité n'est pas symétrique dans les réseaux temporels, la *forte connexité* est utilisée pour dire que deux nœuds i et j sont connectés temporellement dans les deux directions. Dans [Nic+12], Nicosia et *al.* définissent en détails les concepts de connexité et composante dans les réseaux temporels. Ceci nous donne l'opportunité de définir les termes suivantes : longueur du chemin temporel (*temporal path length*), plus court chemin (*shortest path*) , plus rapide chemin (*fastest path*), distance temporelle (*temporal distance*).

Différent du chemin (topologique) temporel (*'time-respecting path'*), la longueur ou la durée du chemin temporel peut être mesurée comme l'intervalle de temps entre le premier et le dernier contact dans le chemin, c.-à-d. $t_l - t_1 + 1$. Nous pouvons également définir le *plus court chemin temporel* entre deux nœuds v et w comme le chemin temporel avec la plus petite longueur (topologique) alors que *le plus rapide chemin temporel* est le chemin temporel avec la plus courte durée (temporelle). De plus, si l'on considère deux nœuds i et j, la *distance temporelle* τ_{ij}, appelée aussi *latence*, est définie comme étant le temps le plus court pour atteindre j à partir de i à l'instant t en considérant les chemins temporels. C'est aussi la longueur temporelle du plus court chemin temporel de i à j. A partir de la distance temporelle, Tang [Tan+10a] définit la moyenne des longueurs du chemin temporel (*characteristic* or *average temporal path length*) $\mathscr{L}$, similaire à celui de Watts & Strogatz :

$$\mathscr{L} = \frac{1}{N(N-1)} \sum_{ij} \tau_{ij} \tag{2.36}$$

Les auteurs définissent aussi, dans le contexte des réseaux temporels non connexes, la notion de l'efficience globale temporelle (*temporal global efficiency*) [Tan+10a].

$$\mathscr{E} = \frac{1}{N(N-1)} \sum_{ij} \frac{1}{\tau_{ij}} \tag{2.37}$$

2.5.6.2 Coefficient de Corrélation Temporelle

Les réseaux réels dynamiques sont souvent marqués par l'effet *"burstiness"* caractérisé par une fréquence des contacts entre les individus suivant une loi puissance [HS12]. Ce phénomène est dû à une faible fréquence du nombre de contacts durant une longue période suivie d'une augmentation sur une courte période. En plus, si deux nœuds sont liés à un instant t alors il existe une probabilité non négligeable qu'ils soient en relation à un instant $t + \Delta t$. Cette caractéristique appelée *corrélation temporelle* permet de réadapter le concept de clustering déjà rencontré dans les graphes statiques [Nic+13]. Par conséquent, cette mesure a permis d'étendre l'étude de l'effet *petit-monde* dans les réseaux temporels [Tan+10a]. Le *coefficient de corrélation temporelle* ou *clustering temporel* $\mathscr{C}$ est défini ainsi

$$\mathscr{C}_i = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{\sum a_{ij}(t) a_{ij}(t+1)}{\sqrt{\left(\sum a_{ij}(t)\right)\left(\sum a_{ij}(t+1)\right)}} \tag{2.38}$$

$$\mathscr{C} = \frac{1}{N} \sum \mathscr{C}_i \tag{2.39}$$

avec $a_{ij}(t) \in \mathscr{A}^t$, la matrice d'adjacence du graphe G_t. Dans [BSK16], Büttner et *al.* ont défini, récemment, le *coefficient de corrélation temporelle* dans les réseaux temporels orientés. Cui et

al. proposent une définition du coefficient de clustering temporel qui est plus proche de celle déjà rencontrée dans la section 2.4.6 [CZL13]. Cependant, nous conservons au courant de cette thèse le concept de coefficient de corrélation temporelle dans l'analyse des résultats obtenus.

2.5.6.3 Centralité dans les TVGs

La centralité fait partie de l'une des plus importantes mesures étudiées dans les réseaux complexes. Elle permet de comprendre à quel point un nœud est central dans un réseau. Cette mesure déjà étudiée dans les réseaux complexes statiques a été étendue dans les TVGs. Dans cette partie, nous allons définir la centralité temporelle de degré, la centralité temporelle de proximité et la centralité temporelle intermédiaire.

— **Centralité temporelle de degré *temporal degree centrality***

Dans les graphes statiques, l'influence d'un nœud peut-être mesurée par le nombre de voisins par lesquels il est connecté (c'est-à-dire le degré du nœud). Intuitivement, les nœuds ayant un nombre de voisins supérieur devrait mieux contribuer à la propagation que les nœuds avec peu de connexions. C'est la raison pour laquelle, la notion de centralité de degré connue dans les réseaux statiques est redéfinie dans les réseaux temporels pour capturer comment un nœud est influent ou central dans un TVG. Formellement, la centralité temporelle de degré, $C_D^t(i)$ d'un nœud i est calculée par la formule suivante [TBF17] :

$$C_D^t(i) = \sum_{j=1}^{N} \sum_{t=t_1}^{T} a_{ij}(t) \tag{2.40}$$

où N représente le nombre de nœuds, $T \in \mathbb{T}$ la valeur maximale d'une séquence temporelle, et a_{ij}^t un élément de matrice d'adjacence $\mathscr{A}^t$ du graphe $G_t = (E_t, V_t)$ obtenu à l'instant t tel que $t_1 \leq t \leq T$. L'équation 2.40 peut-être normalisée en divisant l'expression par $2(|V|-1)\Delta t$, avec $\Delta t = T - t_1$.

— **Centralité temporelle de proximité et intermédiaire (*temporal closeness and betweenness centrality*)**

Le degré d'un nœud constitue un élément important pour caractériser la centralité d'un nœud. Cependant, si nous approfondissons notre analyse, nous observons que le nombre important de voisins n'est pas seulement suffisant pour dire si un nœud est un bon influenceur (*'spreaders'*), car l'organisation structurelle et temporelle des contacts a également un impact sur la vitesse du processus de propagation. En effet, les nœuds par où passent un nombre important de plus courts chemins sont ceux qui contribuent le plus à la transmission d'une épidémie où de l'information dans le réseau. Par conséquent, la centralité de proximité et la centralité intermédiaire sont reprises dans les réseaux temporels pour capturer ce phénomène. La centralité temporelle de proximité est définie par la relation suivante [PS11].

$$C_C^t(i) = \frac{1}{N-1} \sum_{j=1}^{N} \frac{1}{\langle \tau_{ij} \rangle} \tag{2.41}$$

où $\langle \tau_{ij} \rangle$ représente la moyenne des plus courts chemins temporels entre deux nœuds i et j. De manière similaire aux graphes statiques, plus la distance temporelle τ_{ij} est faible plus la valeur $C_C^t(i)$ est élevée. Par conséquent, un processus dynamique peut

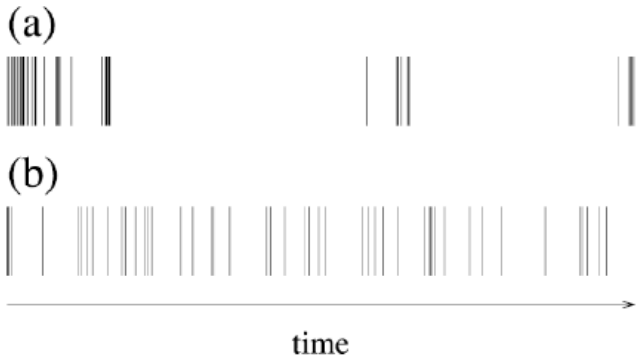


FIGURE 2.17: (a) Phénomène de *burstiness*, (b) Phénomène non *bursty* (processus de poisson) [ML16b]

atteindre rapidement les autres nœuds du TVG s'il a une proximité temporelle avec le nœud i.

La mesure de centralité intermédiaire temporelle ou *'temporal betweenness'* nous renseigne plus sur la capacité que certains nœuds ont pour propager l'information rapidement dans le réseau. De ce fait, un nœud qui a une forte centralité intermédiaire est un nœud par lequel passent plusieurs chemins temporels géodésiques. Selon Tang et ses collègues [Tan+10b], la centralité intermédiaire *betweenness* d'un nœud i dans un réseau temporel est définie par :

$$C_B^t(i) = \frac{1}{(N-1)(N-2)} \sum_{j=1, j\neq i}^{N} \sum_{k=1, k\neq i}^{N} \frac{\sigma_{jk}^i(t)}{\sigma_{jk}} \tag{2.42}$$

où σ_{jk} est le nombre de plus courts chemins temporels entre j et k tandis que $\sigma_{jk}^i(t)$ représente le nombre de plus courts chemins temporels entre j et k en passant par i à une période t, avec $t_1 \leq t \leq T$.

2.5.6.4 La durée d'Inter-Contacts (*Inter-contact time*) et le phénomène de *'Burstiness'*

La durée d'inter-contacts (*inter-contact time*, *inter-event time* ou *waiting time*) est définie comme étant la durée entre deux événements consécutifs sur une paire de nœuds. Elle permet de caractériser la période d'inactivité d'un lien. Elle a été très tôt utilisée dans les réseaux opportunistes pour analyser la performance de certains protocoles de routage [KBV10 ; CLF07 ; AS09 ; Hon+08]. Dans le monde réel, nous constatons souvent que plusieurs événements peuvent avoir lieu sur une courte période suivie d'une longue durée de pause avant de réapparaître (voir Figure 2.17). Ce phénomène appelé *'bursty'* est présent dans plusieurs réseaux complexes réels. Longtemps supposé comme un processus de poisson, Barabasi et Goh montrent que ce phénomène sporadique est caractérisé par une durée des inter-contacts ou des évènements intermédiaires suivant une distribution de loi puissance [GB06].

Si nous considérons une séquence de contacts ou d'événement entre deux nœuds $\{1,\ldots,i,i+1,\ldots\}$ alors τ_i la durée d'inter-contact entre l'interval t_i et t_{i+1} est définie formellement par :

$$\tau_i = t_{i+1} - t_i \tag{2.43}$$

La figure 2.17 b) montre une séquence d'événements suivant un processus de poisson dont la durée d'inter-contact est donnée par la distribution suivante [ML16b] :

$$\psi(\tau) = \lambda e^{-\lambda t}, \tag{2.44}$$

avec λ, un paramètre représentant le taux de contacts sur un intervalle de temps. Dans la Figure 2.17 a), nous observons un phénomène de bursty caractérisé par une distribution de puissance [ML16b] :

$$\psi(\tau) \propto \tau^{\alpha} \tag{2.45}$$

avec $\alpha > 0$. Plusieurs résultats empiriques montrent que la valeur α se trouve dans l'intervalle $1 \leq \alpha \leq 2.7$ [Váz+06 ; Kar+12]. Pour quantifier le degré de *'burstiness'*, le moyen le plus simple c'est de calculer le *Coefficient de Variation* (CV), défini comme le rapport de l'écart-type σ_τ à la moyenne $\langle\tau\rangle$ de la distribution $\psi(\tau)$ [ML16b].

$$CV = \frac{\sigma_\tau}{\langle\tau\rangle} = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(\tau_i - \langle\tau\rangle)^2}}{\langle\tau\rangle} \tag{2.46}$$

avec,

$$\langle\tau\rangle = \frac{1}{n-1}\sum_{i=1}^{n-1}\tau_i \tag{2.47}$$

D'une part, nous pouvons déduire de l'équation 2.46 que $CV = 1$ si la durée d'inter-contact suit un processus de Poisson car sa moyenne et son écart type sont égaux $(1/\lambda)$. D'autre part, si les séquences d'événements ou de contacts sont périodiques c'est à dire tous les τ_i sont les mêmes alors $CV = 0$. Par ailleurs, dans certaine distribution de puissance où $\langle\tau\rangle$ existe et l'écart-type $\sigma_\tau \longrightarrow \infty$ également, alors $CV \longrightarrow \infty$. La formule de l'équation 2.46 peut-être normalisée pour donner ce qu'on appelle le coefficient de *burstiness* [GB06] :

$$B = \frac{CV - 1}{CV + 1} = \frac{\sigma_\tau - \langle\tau\rangle}{\sigma_\tau + \langle\tau\rangle} \tag{2.48}$$

avec $-1 \leq B < 1$. Nous avons $B = 0$ $(CV = 1)$ si le processus est de type Poisson, $B = -1$ $(CV = 0)$ si la séquence est périodique. Un phénomène de bursty est caractérisé par une valeur de $B \longrightarrow 1$ $(CV \longrightarrow \infty)$.

Une autre alternative de mesure du phénomène de bursty est la *Variation Locale* (LV), introduite par Shinomoto et *al.* en 2003 dans l'analyse des intervalles entre les pics d'influx nerveux (*inter-spike intervals*) [SMK05]. La LV est définie par formule ci-dessous :

$$LV = \frac{3}{n-1}\sum_{i=1}^{n-1}\left(\frac{\tau_i - \tau_{i+1}}{\tau_i + \tau_{i+1}}\right)^2 \tag{2.49}$$

Avec le facteur de normalisation $3/(n-1)$. Ainsi, un processus de Poisson et une séquence d'événements périodiques donnent respectivement les valeurs approximatives $LV = 1$ et $LV = 0$. Pour une séquence d'évènements montrant un phénomène de bursty, nous obtenons une valeur de LV comprise entre 1 et 3 [ML16b].

2.5.6.5 Fluctuabilité (*Fluctuablity*)

Dans les réseaux où la topologie est hautement dynamique c'est-à-dire les réseaux dont les liens de connexion peuvent apparaître et disparaître à tout moment (par exemple, dans le cas des réseaux mobiles sans fils), nous avons besoin d'avoir une information globale sur la variabilité des connexions au cours du temps. Dans un contexte similaire, Thompson et *al.* [TBF17] proposent la mesure de *Fluctuabilité* pour quantifier la variabilité de la connexité temporelle des neurones au niveau macroscopique. Elle est définie comme étant le ratio entre la somme des liens présents dans la matrice d'adjacence agrégée $\mathscr{A}$ sur la somme globale des matrices $\mathscr{A}_t$ correspondant à G_t.

$$F = \frac{\sum_i \sum_j U(a_{ij})}{\sum_{i=1} \sum_j \sum_t a_{ij}^t} \tag{2.50}$$

avec,

$$U(a_{ij}) = \begin{cases} 1 & \text{si} \quad \sum_t^T a_{i,j}^t > 0 \\ 0 & \text{si} \quad \sum_t^T a_{i,j}^t = 0 \end{cases} \tag{2.51}$$

Nous pouvons déduire à partir de la formule 2.50 que la valeur maximale de F ($F = 1$) est atteinte lorsque chaque arête est unique et n'apparaît qu'une seule fois dans le temps. Par conséquent, la *fluctuabilité* permet de mesurer la diversité des liens temporels dans un TVG. Si nous considérons les TVGs montrés dans les Figures 2.18.a, 2.18.b, nous remarquons qu'il n y a que deux liens présents dans chaque pas de temps (*timestep*). Dans la Figure 2.18a, il y a trois liens uniques donc la somme de U dans la formule 2.50 est égale à 3. Cependant, nous avons 6 possibilités de connexions en considérant la Figure 2.18.b (ce qui fait que nous avons une somme des U égale à 6). Étant données que les deux TVGs ont le même nombre de liens temporels (au total 24) au cours du temps, il est facile de voir que le réseau illustré dans la Figure 2.18.b a une valeur de fluctuabilité F deux fois plus importante que le réseau illustré dans la Figure 2.18.a.

En dépit de cette caractéristique importante, la mesure de fluctuabilité est, au contraire, insensible à l'ordre temporel de la connexité. En guise d'illustration, les réseaux représentés sur les Figures 2.18.b et 2.18.c ont la même mesure de fluctuabilité, malgré les ordres temporels de connexité des liens considérablement différents. Pour pallier ce problème, les auteurs ont étendus la définition au niveau nodal (plus de détails voir l'article [TBF17]).

2.5.6.6 Volatilité (*Volatility*)

La *volatilité* est suggérée par les mêmes auteurs de l'article [TBF17] pour quantifier l'ordre temporel de la connexité non prise en compte par la définition de *fluctuabilité*. Cette mesure évalue, en moyenne, le taux de changement des liens entre deux snapshots de graphe consécutif. La *volatilité* peut être définie par la relation mathématique suivante :

$$V = \frac{1}{T-1} \sum_{t=1}^{T-1} D(G_t, G_{t+1}) \tag{2.52}$$

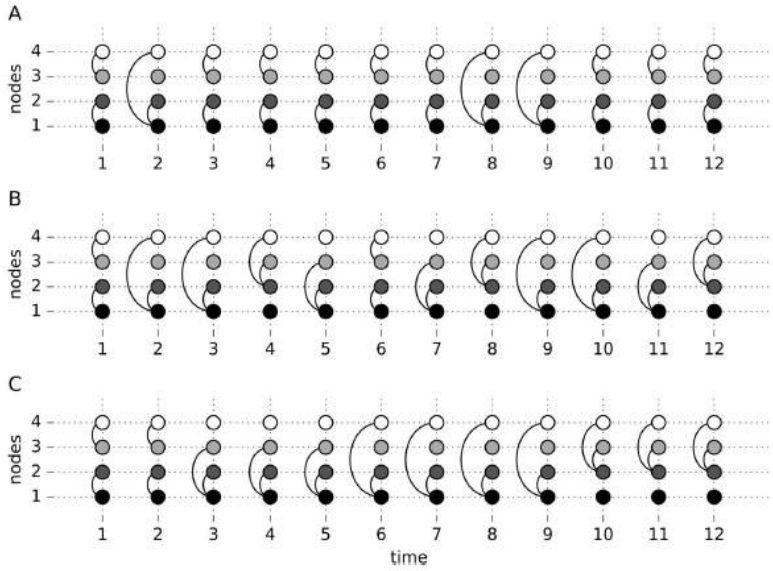


FIGURE 2.18: Fluctuabilité et Volatilité : Les réseaux temporels illustrés dans les panneaux A, B et C ont tous le même nombre de nœuds et d'arêtes, mais ils diffèrent par leur fluctuabilité et leur volatilité. (A) Ce réseau a une faible fluctuabilité (F = 0,125) et une volatilité (V = 0,73). (B) Ce réseau présente la volatilité la plus élevée (V = 2,55) des trois réseaux et une fluctuabilité (F = 0,25) égale à celle du réseau correspondant à C. (C) Ce réseau présente une volatilité inférieure à B (V = 1,27) avec une fluctuabilité égale (F = 0,25). [TBF17]

où $D(G_t, G_{t+1})$ est une fonction de distance qui mesure la différence entre un snapshot de graphe G_t et un autre G_{t+1}. Dans [TBF17], les auteurs utilisent la distance de Hamming [1] du fait qu'ils manipulent des données binaires. La conséquence de cette mesure est illustrée dans les Figures 2.18.b et 2.18.c où la *fluctuabilité* ne change pas tandis que la mesure de la *volatilité* nous montre une grande différence sur l'ordre temporel de connexité des deux TVGs.

2.6 Conclusion

Dans ce chapitre, nous avons passé en revue les concepts fondamentaux permettant de comprendre et d'étudier la topologie et la dynamique des réseaux complexes. Étant donné que les réseaux complexes réels sont des structures qui évoluent dans le temps caractérisées par différents types d'interactions pouvant donner lieu divers phénomènes dynamiques. C'est ce qui motive la recherche croissante de nouveaux concepts et de nouvelles approches de modélisation, permettant de mieux prendre en compte les caractéristiques des systèmes étudiées. Dans le reste du document, nous nous focalisons dans l'étude de la dynamique interne particulièrement à la propagation épidémique dans un réseau temporel de contacts.

1. La distance de Hamming est la somme des différence entre les matrices d'adjacence des deux graphes G_t et G_{t+1}

Chapitre 3

Phénomène de propagation et mobilité dans les réseaux complexes

3.1 Introduction

Depuis très longtemps, les chercheurs ont été animés par la curiosité de savoir, de comprendre, et surtout de prédire les phénomènes qui émergent des systèmes complexes. En effet, comprendre les lois qui gouvernent ces différents phénomènes peut impacter sur l'amélioration des conditions de vie sur le plan social, économique, sanitaire et environnemental. Par exemple, la diffusion d'une technique d'innovation ou la promotion d'un produit peut avoir un impact socio-économique, la prédiction des goulots d'étranglement ou des grandes affluences dans le réseau de transport peut permettre d'y apporter des solutions stratégiques, enfin la compréhension de la propagation d'une maladie peut permettre de freiner son expansion dans la population. Les phénomènes de diffusion et de propagation dans les systèmes complexes ont très tôt intéressé les chercheurs surtout ceux de la physique statistique. De vastes études ont été menées, commençant par de simples modèles mathématiques jusqu'à des modèles plus complexes. Bernoulli, depuis 1760, avait établi le premier modèle mathématique permettant d'expliquer la propagation de l'épidémie *Smallpox* [Ber60]. Ensuite, viennent d'autres modèles qui sont parfois spécifiques à des cas de maladies ou génériques (adaptable à différents types d'épidémie). Aujourd'hui, avec le développement des nouvelles technologies s'accompagnant de la disponibilité de larges datasets, nous assistons à de nouvelles approches de modélisation des phénomènes épidémiologiques qui sont adaptées à différents contextes [MH17 ; AM91 ; KR07]. Par ailleurs, le processus de contamination dépend généralement de trois grand facteurs que sont le contact direct, indirect ou par voie aérienne. Ces facteurs sont surtout favorisés par la mobilité des individus. Ainsi, la mobilité joue un rôle crucial dans le déclenchement et la persistance d'une épidémie dans une population. Nous pouvons donner l'exemple du virus H1N1 dont les scientifiques ont découvert que la propagation rapide de l'épidémie est due en grande partie par une mobilité non contrôlée des personnes atteintes du virus [CVS11 ; Cha+17]. De même, l'épidémie de SARS (*Severe Acute Respiratory Syndrome*) qui a fait des ravages en Asie de l'Est en 2003 [WNW]. Récemment, le virus Ébola est l'exemple

le plus célèbre de phénomène de propagation épidémique dont le mode de transmission est basé sur le contact direct (avec le sang ou les fluides corporels tels que la salive, le sperme et le mucus) entre les individus [Bar16]. En effet, la mobilité y a joué un rôle clé dans l'expansion du virus [Kra+16]. Aujourd'hui, avec la prolifération des smartphones et des objets connectés s'accompagnant des larges avancées obtenues sur la technologie de communication sans files (Wifi, Bluetooth, RFID etc.), les chercheurs ont montré que les données récoltées à partir de ces derniers peuvent servir de proxy pour étudier la vulnérabilité de systèmes complexes face à la propagation de virus ou d'épidémie [Pas+15 ; RSM18 ; Sap+18 ; Van+13].

Dans ce chapitre, nous allons, d'une part, faire une introduction sur la modélisation d'épidémie dans les réseaux complexes en passant par les modèles utilisés comme références dans cette thèse. D'autre part, nous allons revenir un peu en détail sur la mobilité locale et globale avec quelques exemples de modèles de mobilité servant de support pour étudier la topologie de la structure temporelle de notre réseau de contact. Enfin, nous discuterons sur l'influence de la mobilité sur les phénomènes de propagation.

3.2 La propagation d'épidémie

3.2.1 Modèles classiques de propagation d'épidémie

L'expansion rapide et parfois même à grande échelle des maladies épidémiques a été une problématique qui est toujours d'actualité. Le premier modèle classique mathématique d'épidémiologie a été mise en place par deux chercheurs anglais Kermack et McKendrick en 1927 [KM27]. Ce modèle subdivise la population en des groupes distincts appelés '*compartiments*'. Chaque compartiment permet de distinguer et de suivre l'*état* d'un individu donné de la population dans le temps. Ainsi, Kermack et McKendrick initient leur modèle SIR qui suppose que la population est divisée en trois classes : les *Susceptibles*, noté par S, représentent les individus en bonne santé et qui peuvent être infectés par la maladie ; les *Infectieux* (I) sont ceux qui sont infectés et peuvent contaminer d'autres ; enfin les *Retirés* (R) concernent les personnes qui sont guéries, immunisées ou mortes. Le modèle SIR a toujours servi de référence pour étudier les phénomènes de propagation dans une population. C'est pour cela que ce modèle est très flexible et peut être étendue pour étudier différents cas de maladie. En plus, le choix des compartiments et des paramètres à inclure dans le modèle peut dépendre des caractéristiques particulières liées à la maladie ou au phénomène de propagation étudié. Dans la littérature, nous rencontrons divers modèles compartimentaux qui varient en fonction de la propagation d'épidémie étudiée, par exemple SI, SIS, SIRS, SEIR, SEIS, MSIR, MSEIR etc. [**Chowell2009** ; Pas+15 ; BBV08 ; KR07 ; AM91]. Dans cette partie, nous nous focalisons sur les modèles de références qui sont largement utilisés dans la communauté scientifique : SI, SIS, SIR.

Le phénomène de contamination ou de transmission est un processus très complexe qui demeure toujours un challenge à chaque type d'épidémie qui s'installe dans une population. Pour répondre à cette question, les praticiens de la santé donnent un certain nombre de facteurs correspondant à chaque cas de maladie, par exemple le contact interhumain, le contact entre une personne et un animal, la diffusion par gouttelettes, la transmission aérienne, les objets contaminés, les piqûres d'insectes (maladies vectorielles), etc. Vu la complexité de la

représentation de la réalité, les scientifiques émettent des hypothèses simples qui permettront de construire un modèle capable de reproduire certains aspects importants du phénomène étudié. Dans le cadre de la modélisation de la méthode compartimentale, l'hypothèse de mélange homogène (*homogeneous mixing hypothesis*) suppose que chaque individu a la même chance d'entrer en contact avec un autre individu qui est infecté [BBV08]. Cette hypothèse élimine la nécessité de savoir la topologie précise du réseau de contact sur lequel l'épidémie se propage, en le remplaçant avec l'hypothèse qui dit que "*n'importe qui peut être infecté par n'importe qui*" [Pas+15 ; BBV08]. En vertu de cette hypothèse, nous pouvons voir que la dissémination d'un pathogène est lié au nombre de contacts par unité de temps c que fait un individu infecté et la probabilité p pour que la maladie soit transmise au contact [KM27 ; BBV08]. Par conséquent, la probabilité pour qu'un individu sain soit contaminé par un individu malade durant un instant dt est égal à la valeur βdt (avec $\beta = cp$). A savoir, β est appelé paramètre de transmissibilité (*spreading rate*). Ensuite, les individus infectés se rétablissent en devenant susceptibles (modèle SIS) ou retirés (modèle SIR) avec la probabilité μdt. Le paramètre μ représente le taux de guérison (*recovery rate*). Il correspond à l'inverse de la période infectieuse.

En considérant $S(t)$ le nombre total d'individus susceptibles à un instant t, $I(t)$ le nombre d'individus infectés au temps t et $R(t)$ le nombre de personnes retiré, nous supposons dans ce qui suit que la taille de la population ne change pas au cours du temps c'est à dire qu'il n'y a pas de naissance, ni de décès, ni de phénomène de migrations. On obtient alors $N = N(0) = N(t)$.

3.2.1.1 Modèle SI

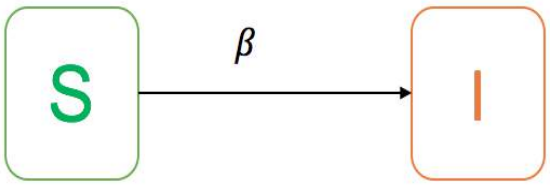


FIGURE 3.1: Modèle compartimental SI

Considérons le plus simple modèle épidémiologique SI (*Susceptible-Infectieux*) tel que les individus ne peuvent être que dans deux états discrets, susceptible ou infecté. L'évolution du modèle est étudié, en général, en fonction du nombre d'individus infectés $I(t)$ ou bien de sa densité correspondante $i = I(t)/N$ qui est aussi appelée *taux de prévalence*. Nous pouvons aussi définir $s(t) = S(t)/N$ le taux d'individu sain dans la population en fonction du temps t. Comme nous supposons qu'il n y a pas l'avènement d'un individu externe, nous avons à l'étape initiale $N = S(0) + I(0)$. L'évolution du modèle SI montré dans la figure 3.1 peut s'écrire sous le formalisme d'une réaction-diffusion [Pas+15].

$$S + I \xrightarrow{\beta} 2I \tag{3.1}$$

où β est le taux de transition infectieuse. La *force de l'infection donnée* par $\alpha = \beta i(t)$ représente le taux moyen d'individus pouvant infecter leurs voisins sains à un temps t [Pas+15]. Ainsi,

le processus de propagation du modèle SI est gouverné par l'équation différentielle ordinaire (EDO) suivante [BBV08] :

$$\frac{di(t)}{dt} = \beta i(t)(1 - i(t)) \tag{3.2}$$

lorsque $i(t)$ exprime la densité de la population infectée alors la fraction du nombre d'individus sains est donnée par la relation $s(t) = 1 - i(t)$. La solution de ce EDO est donnée par

$$i(t) = \frac{1}{1 + \left(i_0^{-1} - 1\right) e^{-\beta t}} \tag{3.3}$$

avec les conditions initiales, $i_0 = i(t = 0)$. Le résultat 3.3 prédit une augmentation exponentielle du nombre d'individu infecté au début de la propagation [1]. En plus, $i(t) \longrightarrow 1$ lorsque $t \longrightarrow \infty$, c'est à dire que nous pouvons arriver à un moment où toute la population sera infectée (voir figure 3.2).

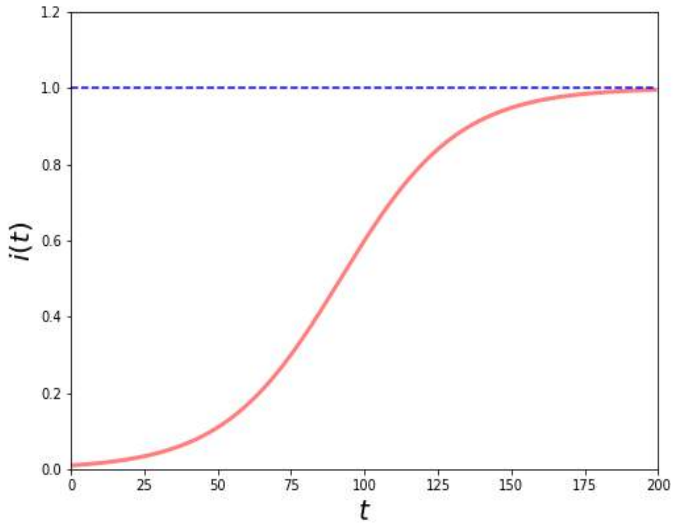


FIGURE 3.2: Modèle SI

3.2.1.2 Modèle SIS

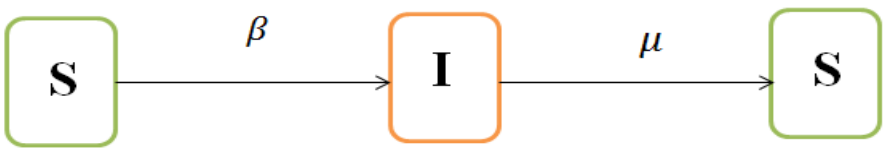


FIGURE 3.3: Modèle compartimental SIS

Le modèle SIS (*Susceptible-Infectieux-Susceptible*) est largement utilisé dans la modélisation de certaines maladies comme la *grippe saisonnière*, *blennorrhagia* etc. [Wan+17]. Il constitue

1. On suppose qu'au début de la simulation $i \ll 1$.

une extension du modèle SI revue précédemment à la seule différence qu'ici les individus infectés peuvent redevenir susceptibles avec une probabilité μdt. Ce qui fait que ce modèle est cyclique comme le montre la figure 3.3. Les paramètres β et μ sont respectivement le taux de transmissibilité et le taux de guérison. En supposant les mêmes hypothèses que précédemment, l'équation de réaction-diffusion de SIS est donnée par [Pas+15] :

$$S + I \xrightarrow{\beta} 2I \tag{3.4}$$

$$I \xrightarrow{\mu} S \tag{3.5}$$

Ainsi, l'évolution dynamique du taux d'individus infectés est donnée par l'équation suivante[BBV08] :

$$\frac{di(t)}{dt} = \beta i(t)(1 - i(t)) - \mu i(t) \tag{3.6}$$

La condition de normalisation $s(t) = 1 - i(t)$ permet d'obtenir la fraction de la population susceptible au cours du temps. La solution analytique du ODE 3.6 est donnée par [SKS10].

$$i(t) = \frac{\beta - \mu}{1 + (\beta - \mu)Ce^{-(\beta-\mu)t}} \tag{3.7}$$

La constante C est obtenue à partir des conditions initiales $i(t = 0) = i_0$,

$$C = \frac{\beta - \mu - i_0}{(\beta - \mu)i_0} \tag{3.8}$$

La figure 3.4 permet de simuler l'évolution du taux de cas d'infection d'un modèle SIS 3.3.

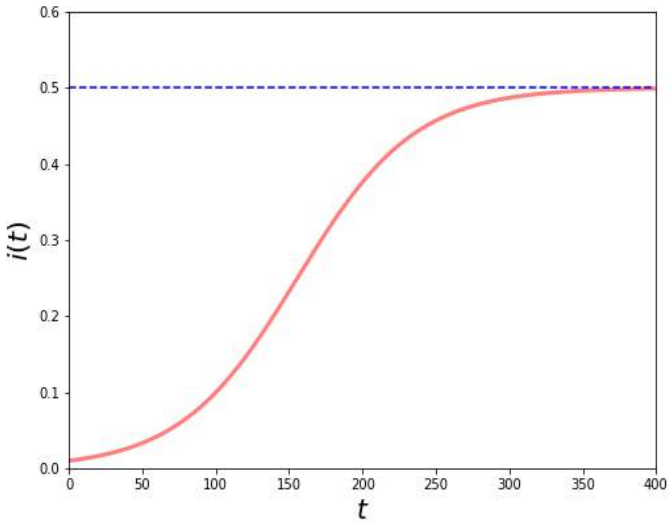


FIGURE 3.4: Modèle SIS

Cette courbe évoluant très rapidement atteint une limite appelée *seuil endémique* : état où les deux proportions de la population ne peuvent plus varier. En effet, nous pouvons approcher l'équation de propagation 3.6 au début de l'épidémie, en supposant la faible proportion du taux de malades ($i \ll 1$) et en négligeant les termes i^2. Ainsi, nous obtenons l'EDO linéaire suivante [BBV08].

$$\frac{di(t)}{dt} \simeq \beta i(t) - \mu i(t) \tag{3.9}$$

dont la solution est $i(t) \simeq i_0 e^{(\beta-\mu)t}$, avec $i_0 = i(t=0)$. Ce résultat nous permet de remarquer que l'infection est maintenue dans la population pour une valeur β suffisamment grande ou une valeur de μ très petite. Autrement dit,

$$\beta - \mu > 0 \Rightarrow R_0 = \frac{\beta}{\mu} > 1 \tag{3.10}$$

avec R_0 représentant le taux de reproduction de base [**Chowell2009** ; Die93 ; KR07 ; AM91]. La valeur de R_0 définit le nombre moyen d'individus qui pourra être infecté par une personne infectieuse durant sa période d'infection, dans une population entièrement susceptible. Ceci nous donne l'occasion de définir la notion de seuil épidémique (*epidemic threshold*) [Cha+08 ; Pas+15 ; HA13 ; KR07] permettant d'étudier le déclenchement et l'extinction de l'épidémie dans une population donnée. Si $R_0 < 1$, l'épidémie ne va pas survivre longtemps dans la population tandis que si $R_0 > 1$ alors l'épidémie persiste jusqu'à atteindre l'état endémique. Cette valeur seuil de R_0 constitue le premier paramètre permettant aux épidémiologistes de vérifier la sévérité de l'entrée d'un nouveau pathogène. De manière générale, le concept de seuil épidémique (*epidemic threshold*) est un élément central en épidémiologie car il fournit un cadre de référence aidant les acteurs du domaine de comprendre l'évolution des processus de contagion et de mettre en place des politiques visant à une réduction efficace du taux de reproduction de base afin d'arrêter les épidémies.

3.2.1.3 Modèle SIR

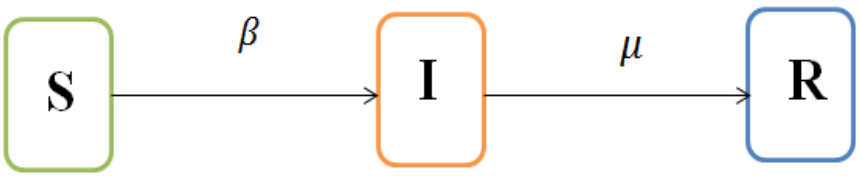


FIGURE 3.5: Modèle compartimental SIR

Le modèle de Kermack-McKendrick [KM27], représenté par le schéma 3.5, suppose que les individus infectieux migrent avec une probabilité μ dans le compartiment R : individus qui sont guéris, immunisés ou morts. Il faut noter que la durée moyenne qu'un individu passe dans le compartiment I sera μ^{-1} d'unités de temps. Le modèle SIR est idéal pour étudier l'évolution dynamique de certaines maladies comme la grippe, la rage, la varicelle et la rubéole [KR07] car lorsque les individus deviennent retirés (R) ils n'entrent plus dans le processus de contamination. En considérant qu'il n'y a pas de phénomène de migration, ni de naissance, ni de décès, le modèle SIR peut être représenté, de manière similaire à *SIS*, comme une réaction-diffusion

$$S + I \xrightarrow{\beta} 2I \tag{3.11}$$

$$I \xrightarrow{\mu} R \tag{3.12}$$

L'évolution du modèle SIR est gouverné par les EDOs suivantes.

$$\frac{ds(t)}{dt} = -\beta i(t)(1 - r(t) - i(t)) \tag{3.13}$$

$$\frac{di(t)}{dt} = \beta i(t)(1 - r(t) - i(t)) - \mu i(t) \tag{3.14}$$

$$\frac{dr(t)}{dt} = \mu i(t) \tag{3.15}$$

La fraction des individus susceptibles est donnée par la relation $s(t) = 1 - r(t) - i(t)$. En considérant une population N composée d'une majorité qui est susceptible et une faible proportion d'individus infectieux ($i(0) \ll 1$) nous pouvons suivre l'évolution du modèle SIR en fixant les paramètres β et μ (voir Figure 3.6). Dans l'exemple dans la figure 3.6, la densité de la population infectée ($i(t)$) augmente jusqu'à atteindre un seuil de **70%** avant de tendre vers l'extinction de l'épidémie. Comme le modèle SIS, la quantité $R_0 = \beta/\mu$ représente le taux de reproduction de base de ce modèle et détermine essentiellement le comportement du modèle.

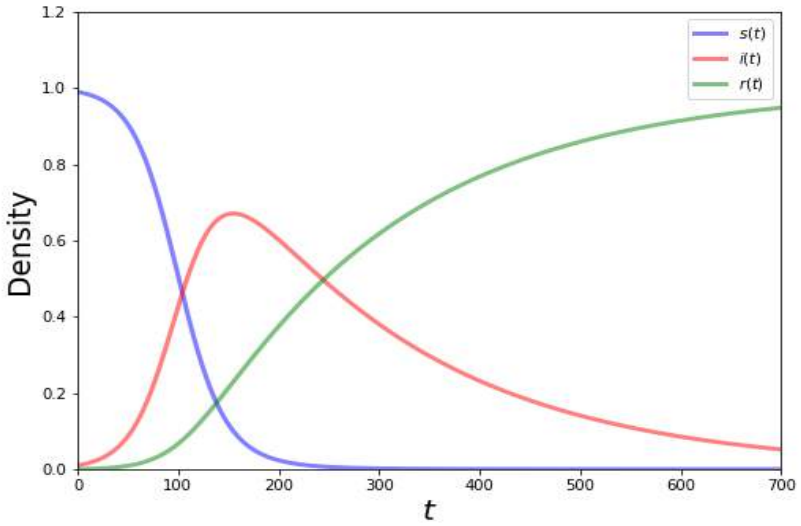


FIGURE 3.6: Modèle SIR

3.2.2 Épidémiologie et réseaux complexes

Malgré l'assomption de l'hypothèse de *mélange homogène* appliquée dans les modèles classiques d'épidémiologie, le processus de contamination tel qu'on l'avait défini en haut n'est pas toujours adéquate dans la plupart des cas réels. Par conséquent, l'épidémiologie va connaitre une nouvelle ère avec le développement rapide de la science des réseaux. Ici, la notion de contact est liée intrinsèquement au voisinage qui peut être homogène ou hétérogène. La propagation des épidémies a été étudiée dans les réseaux complexes présentant des structures homogènes (*random graphes*, *small-world*) [MN00 ; LW07], hétérogènes (*scale-free*) [PV02 ; PV01b] et récemment dans les réseaux dynamiques [Pas+15 ; ML16b ; MH17]. Dans cette partie nous allons parler du processus de propagation dans les réseaux complexes en nous focalisant sur les modèles SIS et SIR.

3.2.2.1 Processus de propagation dans les réseaux réguliers et homogènes

Rappelons d'abord que la topologie des réseaux réguliers ou k-réguliers comme les *lattices* conservent le même nombre de voisins pour les tous nœuds tandis que les réseaux homogènes ont une distribution de degré homogène c'est à dire centrée autour d'une moyenne $\langle k \rangle = pN$, avec p la probabilité que deux nœuds se connectent [AB02 ; Boc+06]. Dans ce type de réseaux nous pouvons supposer que le nombre de contacts qu'un individu infectieux peut contaminer est égal à k. Comme chaque nœud infectieux a une probabilité βdt d'infecter son voisin susceptible durant un instant dt alors un nœud susceptible ayant n voisins infectieux peut devenir infectieux lui même à une probabilité $1-(1-\beta dt)^n$ [BBV08]. En plus, si on considère que chaque nœud a k connexions alors le nombre de voisins infectieux $n = ki$ [BBV08]. En supposant $\beta dt \ll 1$, la probabilité d'être infectée peut être estimée à $1-(1-\beta dt)^{ki} \simeq \beta k i(t) dt$. Dans le cas des réseaux homogènes, le nombre de contacts par nœud peut approximativement être égal à la moyenne des connexions, $k \simeq \langle k \rangle$. L'évolution du taux de prévalence des modèles SIS et SIR peut être redéfinie par l'équation 3.16.

$$\frac{di(t)}{dt} = \beta \langle k \rangle i(t) s(t) - \mu i(t) \tag{3.16}$$

où $s(t)$ peut être déduit facilement avec la condition de normalisation $s(t) = 1 - i(t)$ et $s(t) = 1 - r(t) - i(t)$ respectivement dans SIS et SIR. Dans l'équation 3.16, le premier terme représente le taux moyen d'individus qui ont la probabilité d'être infectée par un individu infectieux tandis que le deuxième terme est la proportion de personnes retirées par unité de temps.

En considérant le taux de transmission effective (*effective spreading rate*) λ comme étant le rapport β/μ, le déclenchement de l'épidémie est toujours déterminé par une valeur seuil λ_c appelée *epidemic threshold* [HA13 ; BBV08 ; Pas+15]. La valeur critique λ_c permet de distinguer deux phases de transitions dans le cycle de vie de l'épidémie : la phase d'absorption et la phase active (voir Figure 3.7). Lorsque $\lambda \leq \lambda_c$, il y a une phase d'absorption qui signifie une extinction rapide de l'épidémie. Si $\lambda \geq \lambda_c$ alors nous avons les conditions nécessaires pour que la maladie soit maintenue dans la population et de déclencher une épidémie globale. D'après [Wan+17], la détermination de la valeur seuil peut parfois dépendre caractéristiques plus ou moins complexes du réseau de contacts. C'est pour cela que nous distinguons dans la littérature différents approches pour étudier les fluctuations de la valeur seuil, λ_c : l'approche *Mean-field (MF) (avec l'hypothèse du mélange homogène)*, l'approche *Heterogeneous Mean-field (HMF)*, l'approche *Quench mean-field (QMF)*, l'approche *Dynamical message passing (DMP)* enfin avec l'approche *Link percolation (LP)* [Pas+15 ; Wan+17].

Dans le cas des graphes homogènes (l'approche MF), la valeur de λ_c peut être calculée en considérant l'état stationnaire de l'équation 3.16 [BBV08]

$$\frac{di(t)}{dt} = 0 \tag{3.17}$$

En l'appliquant au modèle SIS, nous obtenons.

$$i(\infty) - \lambda \langle k \rangle i(\infty)\big(1 - i(\infty)\big) = 0 \tag{3.18}$$

Ainsi, nous avons

$$i(\infty)[1 - \lambda \langle k \rangle (1 - i(\infty))] = 0 \tag{3.19}$$

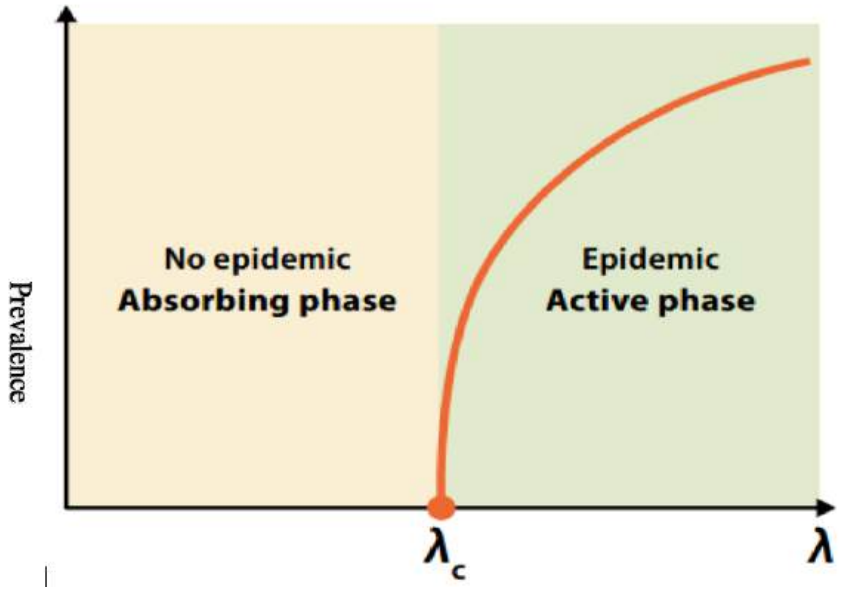


FIGURE 3.7: Phases de transition caractérisées par la valeur critique du seuil épidémique λ_c d'un modèle SIS [Pas+15]

avec $\lambda = \beta/\mu$, $i(\infty) = i(t \to \infty)$ représente la fraction de nœuds à l'état stationnaire. L'équation 3.19 admet deux solutions : une racine triviale $i(\infty) = 0$ et une racine non triviale $i(\infty) > 0$ qui n'existe que lorsque le taux de transmission effective est supérieur à une valeur seuil :

$$\lambda > \frac{1}{\langle k \rangle} = \lambda_c \tag{3.20}$$

— Si $\lambda < \lambda_c$, le nombre de nœuds infecté va décroitre exponentiellement. Par conséquent, l'épidémie ne va plus survivre longtemps dans la population et finira par s'éteindre.
— Si $\lambda > \lambda_c$, l'infection va s'étendre sur tout le réseau jusqu'à atteindre la phase endémique avec un taux de prévalence approximativement égale à $\lambda - \lambda_c$ [PV01a]. Dans, ce cas on dit que l'épidémie est en phase active comme le montre la figure 3.7.

L'approche *MF* apporte une contribution permettant de suivre de manière déterministe l'évolution de l'épidémie au sein de la population par contre elle ne prend pas en compte l'hétérogénéité de la structure de certains réseaux réels. L'hétérogénéité des degrés, la tendance des nœuds se regrouper en communauté, l'assortativité, et l'évolution dynamique du réseau ont un impact considérable sur le processus de propagation. Ces propriétés doivent être prises en compte dans la modélisation et la mise en place des politiques d'extinction de l'épidémie.

3.2.2.2 Processus de propagation dans les réseaux hétérogènes

Au regard de ces contraintes, deux chercheurs italiens, Pastor-Satorras et Vespignani, ont été les premiers à intégrer l'hétérogénéité des degrés constatée dans plusieurs réseaux réels comme Internet dans la modélisation de la propagation d'épidémie [PV01b ; PV01a ; PV02]. Cette hétérogénéité est caractérisée par la loi puissance de la distribution de degrés ($P(k) \sim k^{-\lambda}, 2 < \lambda \leq 3$) [BA99 ; AB02 ; FFF99 ; BP16]. Les auteurs ont montré que la vitesse de propagation devient plus rapide dans les réseaux hétérogènes sans échelles (scale-free) que dans les réseaux homogènes [PV01b ; PV01a]. Ce phénomène s'explique par la présence de *"hubs"* qui jouent le rôle de *"super-spreaders"* [ML16a ; Zha+16]. En plus, le concept de seuil

épidémique dépend entièrement de la topologie du réseau. Dans ce contexte, l'hypothèse du mélange homogène sera remise en cause et le processus de contamination entre les nœuds est redéfini par Pastor-Satorras et Vespignani [PV01b]. Cette découverte constitue une nouvelle étape dans l'épidémiologie moderne.

Le modèle de Pastor-Satorras et Vespignani [PV01a] était basé sur l'hypothèse que tous les nœuds de même degré sont statistiquement équivalents (*Degree-based Mean-Field*). Cette approche permettra d'étudier l'évolution de l'épidémie en fonction des groupes de nœuds pouvant être subdivisés sous forme de classe de degré k.

$$i_k = \frac{I_k}{N_k}; \qquad s_k = \frac{S_k}{N_k}$$

où N_k représente le nombre de nœuds de degré k, I_k et S_k sont respectivement les individus infectés et susceptibles dans cette classe. Ainsi, la densité des nœuds infectés et susceptibles peut être obtenu par :

$$i = \sum_k P(k) i_k; \qquad s = \sum_k P(k) s_k$$

L'équation déterminant le modèle SIS est donnée par la relation suivante [Pas+15]

$$\frac{di(t)}{dt} = -i_k(t) + \lambda k (1 - i_k(t)) \sum_{k'} P(k'|k) i_{k'}(t) \tag{3.21}$$

avec $\lambda = \beta/\mu$ représentant le taux de propagation effective et $P(k'|k)$ la probabilité pour qu'un nœud de degré k soit connecté avec un nœud de degré k'. Dans l'équation 3.21, le premier terme de droite prend en compte la probabilité de rétablissement des nœuds de degré k tandis que le second terme exprime la portion de nœuds qui pourra être infectée par le reste des nœuds infectés. Si le réseau est non corrélé alors $P(k'|k) = k' P(k')/\langle k \rangle$ et l'équation 3.21 devient

$$\frac{di(t)}{dt} = -i_k(t) + \lambda k (1 - i_k(t)) \Theta(i_k(t)), \tag{3.22}$$

avec

$$\Theta(i_k(t)) = \sum_{k'} \frac{k' P(k')}{\langle k \rangle} i_{k'}(t) \tag{3.23}$$

Cette dernière expression signifie la probabilité qu'un lien choisi aléatoirement se connecte à un nœud infecté. A l'état endémique ($t \longrightarrow \infty$), la densité $i_k(t)$ sera une fonction i_k dépendant de λ [PV01a]. Avec la condition de stationnarité du modèle SIS, nous avons :

$$\frac{di(t)}{dt} = 0 \tag{3.24}$$

On obtient :

$$i_k = \frac{\lambda k \Theta}{1 + \lambda k \Theta} \tag{3.25}$$

où la probabilité Θ devient une constante qui dépend du taux de transmission effective, λ. Dans cette expression, nous pouvons remarquer que plus le degré d'un nœud est grand plus sa probabilité d'infection est grande. Ceci montre la conséquence de l'hétérogénéité des degrés dans la dissémination de l'épidémie. Ainsi, en combinant les équations 3.25 et 3.23 nous pouvons en déduire la constante Θ [PV01a] :

$$\Theta = \frac{1}{\langle k \rangle} \sum_k k P(k) \frac{\lambda k \Theta}{1 + \lambda k \Theta} \tag{3.26}$$

L'auto-consistance de cette équation entraîne un seuil épidémique λ_c qui n'est obtenu que lorsque Θ admet une solution non nulle [PV02]. Enfin, λ_c est déterminé si la condition suivante est satisfaite :

$$\frac{d}{d\Theta}\left(\frac{1}{\langle k\rangle}\sum_k kP(k)\frac{\lambda k\Theta}{1+\lambda k\Theta}\right)\Bigg|_{\Theta=0} \geq 1, \tag{3.27}$$

nous avons,

$$\frac{1}{\langle k\rangle}\sum_k kP(k)\lambda k = \frac{\langle k^2\rangle}{\langle k\rangle}\lambda \geq 1, \tag{3.28}$$

Ce qui donne le résultat suivant :

$$\lambda_c = \frac{\langle k\rangle}{\langle k^2\rangle} \tag{3.29}$$

Pastoras-satorras et Vespignani ont montré, dans le cas des réseaux sans échelle avec une distribution de puissance γ telle que $2 < \gamma \leq 3$, que le seuil épidémique, λ_c, tend asymptotiquement vers une valeur nulle (car $\langle k^2\rangle \longrightarrow \infty$) [PV01a]. Ce résultat a été illustré dans le cas d'un modèle de Barabasi-Albert présentant une distribution de degré $P(k) = 2mk^{-3}$ [PV01a]. Par conséquent, ils prouvent que la densité de la population infectée peut être estimée par la relation suivante :

$$i \sim 2e^{-1/(m\lambda)} \tag{3.30}$$

Les travaux de Satorras et Vespignani ont rapidement eu une conséquence sur les réseaux présentant une structure sans-échelle (*scale-free*) comme Internet, et montrant une rapide propagation des virus due essentiellement à la présence de hubs. Cependant, ces travaux n'ont pas pris en compte la corrélation entre le degré des nœuds qui sera plus tard étudiée par Bogũná et *al.* [BPV03]. Basés sur les travaux de [PV01a], Bogũná et *al.* [BPV03] ont remplacé l'expression de l'équation 3.22 par :

$$\Theta(i_k(t)) = \sum_{k'} P(k'|k) i_{k'}(t) \tag{3.31}$$

c'est à dire la probabilité pour qu'un nouveau nœud de degré k soit infecté par un quelconque nœud k' dépend de sa corrélation avec celui-ci et de sa probabilité d'infection, $i_{k'}(t)$. Les résultats des travaux prouvent qu'un seuil épidémique est obtenu lorsque :

$$\lambda > \lambda_c = \frac{1}{\Lambda_M} \tag{3.32}$$

avec Λ_M la plus grande valeur propre de la matrice de connectivité définie par $C_{kk'} = [kP(k'|k)]$ [BPV03]. Par conséquent, les auteurs ont conclu que les corrélations de degré ne sont pas en mesure d'arrêter une épidémie dans les réseaux *scale-free*. Ces résultats sont importants dans la mesure où plusieurs réseaux réels montrent une corrélation non triviale entre le degré des nœuds [BPV03].

Cependant, l'approche HMF se base essentiellement sur l'hypothèse *Degree-based Mean-Field* qui utilise l'équivalence des nœuds de même degré pour décrire la topologie du réseau. En effet, d'autres approches ont été utilisées pour étudier la propagation d'épidémie dans les réseaux complexes telle que l'approche *Quench Mean-Field* qui stipule que les les états des nœuds voisins sont statistiquement indépendants [Pas+15]. L'approche *Quench Mean-Field* (QMF) prend en compte la structure globale du réseau en utilisant sa matrice de

d'adjacence. Ainsi, l'évolution dynamique du modèle SIS basée sur QMF est décrite par l'équation différentielle ordinaire (EDO) suivante [Wan+17]

$$\frac{d\rho_i(t)}{dt} = -\rho_i(t) + \lambda[1-\rho_i(t)]\sum_i a_{ij}\rho_j(t) \tag{3.33}$$

dont $\rho_i(t)$ caractérise la probabilité du nœud i d'être infecté par ses voisins, λ représente le taux de transmission effective donné par le rapport β/μ et a_{ij} les coefficients de la matrice d'adjacence A du réseau correspondant. Dans ce modèle, le seuil épidémique est obtenu par la relation suivante.

$$\lambda_c = \frac{1}{\Lambda_A} \tag{3.34}$$

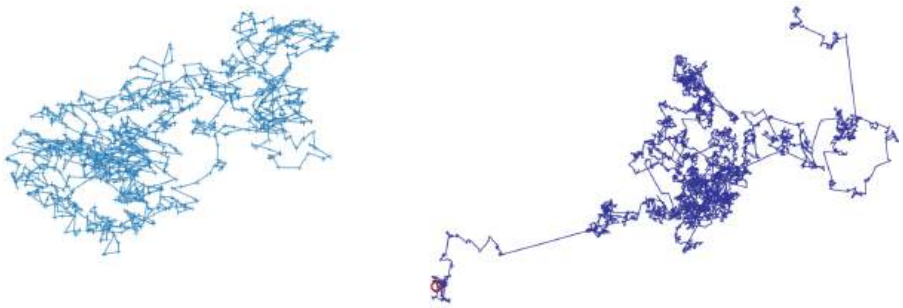
où Λ_A est la plus grande valeur propre de la matrice d'adjacence A. Cette valeur est estimée dans les *scale-free* ($P(k) \sim k^\gamma$) à $\lambda_c \propto \langle k\rangle/\langle k^2\rangle$ si $2 < \gamma < 5/2$, équivalent à la valeur seuil de l'équation 3.29, et de $\lambda_c \propto 1/\sqrt{k_{max}}$ lorsque $\gamma > 5/2$ [CP10].

En plus des modèles théoriques, les données de contacts constituent aujourd'hui un support important dans l'étude de phénomènes épidémiques. La nature de ces données est basée directement ou indirectement sur la mobilité des individus qui sont à l'origine de ces données. C'est dans cette optique que nous allons essayer de comprendre les caractéristiques de la mobilité et comment nous pouvons la modéliser.

3.3 La Mobilité

Aujourd'hui avec la prolifération des technologies sans fils, la mobilité devient un élément crucial surtout dans la conception de protocoles de routages destinés aux réseaux mobiles cellulaires et Ad hoc (MANETs). L'homme est de nature mobile, et cette mobilité est étroitement liée à son environnement et à ses activités socio-économiques. C'est ce qui fait que comprendre la mobilité devient d'une part une nécessité dans l'amélioration des conditions de transports (routiers, ferroviaires, aériennes) [Jia+17], dans la prédiction les flux de trafic intra et inter urbains [Tan+15 ; Kan+15] et dans la mise en place de politique d'urbanisation [Nou+12]. D'autre part, elle est un élément crucial dans la prédiction, le suivi et le contrôle des phénomènes de propagation et de diffusion dans les systèmes complexes (maladies, virus, rumeurs, opinions, mêmes etc.). Entre autre, elle a été la cause de la prolifération rapide de certaines maladies épidémiques (SARS) [WNW] et des grandes pandémies comme la grippe [Cha+17]. Par ailleurs, dans le cadre du projet GLEAMviz (*Global Epidemic and Mobility Model*), les auteurs ont combiné des données réelles sur la population et la mobilité humaine dans le but de concevoir de nouveaux modèles qui permettent de prédire la transmission des maladies à grande échelle et d'aider à minimiser l'impact des épidémies potentiellement dévastatrices [Gle18]. Le cas Ébola est l'exemple le plus récent de maladie épidémique dont le mode de transmission est basé sur le contact direct et/ou indirect entre les individus, et la mobilité y a joué un rôle important dans son expansion [Kra+16].

Étudier la mobilité humaine s'avère un peu délicat du fait que l'homme ne se déplace pas de manière aléatoire dans l'environnement. Des recherches ont montré que la fréquence de visites des personnes dépend fortement de leur relation sociale [GHB08 ; Son+10 ; Bar+18]. L'homme visite en général un nombre limité de lieux où il passe la majeure partie de son temps (maison, lieu de travail, salle de sport etc.) [Bar+18]. Il se déplace souvent en courte



(a) Random Brownian Motion (b) Lévy Flight Motion

FIGURE 3.8: Random Brownian Motion vs Lévy Flight Motion. D'après [Bov15]

distance et effectue rarement de longs voyages [Son+10]. C'est ainsi que beaucoup d'efforts ont été fournis pour mettre en place des modèles qui tentent de reproduire les caractéristiques de la mobilité humaine et pouvant être utilisés dans plusieurs applications dans la vie réel. Dans cette partie, nous allons passer en revue quelques modèles de mobilité individuelle et globale en passant par les différentes propriétés caractérisant la mobilité humaine.

3.3.1 Caractéristiques de la mobilité humaine

Différentes propriétés spatio-temporelles et sociales permettent de caractériser le mouvement des individus : distance de voyage (*trip distance)*, motifs de mobilité (*mobility pattern*), fréquence de visites (*visitation frequency*) etc. Très tôt, Einstein a utilisé un modèle de marche aléatoire (*Brownian Motion*) pour décrire la mobilité des particules suspendues dans un liquide [Ein05]. Et, il a démontré que la probabilité d'observer une particule a une distance d à partir de l'origine après une durée t suivait une distribution gaussienne [Bar+18] (voir 3.8a). Contrairement au mouvement Brownien, le modèle de Lévy (*Lévy Flight or Lévy Walk*) est une marche aléatoire non gaussienne décrivant un ensemble de petits déplacements, entrecoupés occasionnellement par des déplacements de longues distances (voir Figure 3.8b). Bien vrai que les modèles aléatoires ne peuvent pas mimer le comportement humain, mais ils peuvent néanmoins servir de support pour capturer certaines caractéristiques primitives de la mobilité humaine. Du point vue de la physique, le mouvement d'un agent mobile est quantifié par les paramètres suivants : la position, la direction, la vélocité et l'accélération. Cependant, les caractéristiques du mouvement humain peuvent être révélées par la distribution de la distance de déplacement (*Jump Lenght*), de la moyenne quadratique (*Mean Square Displacement*, MSD), du rayon de giration (*Giration Radius*), de la durée des arrêts (*Waiting Time*), de la fréquence de visites (*Visitation Frequency*) de certains lieux.

3.3.1.1 Distance des déplacements ou *Jump Length*

La distance qu'un individu parcourt entre deux points distincts durant une période bien déterminée fait partie des facteurs clés qui permettent de caractériser sa mobilité. Dans la littérature, nous avons rencontré plusieurs termes qui décrivent ce concept : *jump length*, *jump size*, déplacement, *trip distance*. En plus, cette description dépend du contexte et des sources de données qui sont utilisées (GPS, Call Data Records ou CDR, traces Wifi etc.). Cette

distance, notée par Δr, est définie comme étant la distance euclidienne entre $r(t)$ et $r(t+dt)$ correspondant aux endroits où l'individu a été observé durant l'intervalle t et $t+dt$[Bar+18]. En étudiant la dispersion des billets de banque aux États-Unies, Brockmann et *al.* ont étudié la trajectoire du billet vert et ont démontré que la distribution de la distance de parcours des billets observée suivait une loi puissance, $P(\Delta r) \sim \Delta r^{-(1+\beta)}$ avec $\beta = 0.59 \pm 0.02$ [BHG06]. Song et *al.* ont montré dans [Son+10] que la distribution des traces CDR (*Call Data Records*) suit aussi la loi puissance avec $0 \leq \beta \leq 2$. De même, González et *al.* ont trouvé que la distribution empiriques des déplacements obtenus à partir de données CDR révèlent approximativement une loi puissance tronquée [GHB08] :

$$P(\Delta r) \simeq (\Delta r + \Delta r_0)^{-\beta} \exp(\Delta r/\kappa) \tag{3.35}$$

où $\beta = 1.75 \pm 0.15$ (moyenne ± écart type), $\Delta r_0 = 1.15km$ et κ représentant le paramètre de coupure (*cutoff value*). En plus, cette fameuse loi de puissance est aussi confirmée dans d'autres recherches [Nou+12][Tan+15]. Récemment avec les données GPS de GeoLife et de Nokia MDC, Zhao et *al.* ont étudié la distribution $P(\Delta r)$ de plusieurs modes de transport (marche, vélo, voiture, train) et ont trouvé que la trajectoire de chaque type de transport suivait une loi log-normale tandis que la distribution globale des déplacements obtenus sur les deux datasets (GeoLife et de Nokia MDC) serait une loi de puissance tronquée avec $\beta = 0.5$ et $\beta = 0.39$ [Zha+15]. En conséquence, le constat fait sur les résultats de ces expériences montre que les individus se déplacent souvent en courte distance et très rarement en longue distance. En effet, elle est due à la loi de puissance ou la loi de puissance tronquée observée dans certains déplacements des individus. C'est pour cela, González et *al.* [GHB08] et Song et *al.* [Son+10] ont suggéré que la mobilité humaine peut être décrite comme une trajectoire de Lévy (*Lévy Flight*) caractérisée par plusieurs sauts de courtes distances et quelques rares trajets de longues distances.

Également, nous pouvons aussi noter que le temps de pause (*waiting time or pause time*), Δt, dans le déplacement des individus est une autre propriété étudiée en parallèle dans ces différents types de datasets évoqués précédemment. Il a été aussi vérifié de même que leur distribution est de loi puissance ou de loi puissance tronquée. Par exemple dans des données de mobilités collectées sur quelques opérateurs téléphoniques, Song et *al.* [Son+10] ont pu découvrir que la distribution du temps de pause est approximativement $P(\Delta t) \sim \Delta t^{-(1+\alpha)}$ avec $0 \leq \alpha \leq 1$. Brockmann et *al.* ont montré aussi que la distribution de la durée de pause des billets verts est égale à $P(\Delta t) \sim \Delta t^{-(1+\alpha)}$ avec $\alpha = 0,6 \pm 0,03$ [BHG06].

3.3.1.2 Déplacement Quadratique Moyen

Le déplacement quadratique moyen ou *Mean Squared Displacement* (MSD) est un concept physique permettant de mesurer la déviation de la position d'une particule par rapport à une position de référence dans le temps. Il est utilisé dans la mobilité humaine pour étudier le comportement de la trajectoire des individus par rapport à une origine quelconque. Formellement, le *déplacement quadratique moyen* peut être défini par la formule suivante :

$$MSD(t) = \langle (r(t) - r_0)^2 \rangle = \langle \Delta r(t)^2 \rangle \tag{3.36}$$

Autrement dit, $\langle \Delta r(t)^2 \rangle$ correspond à la moyenne des carrés des distances par rapport à l'origine de la trajectoire r_0 et la position du mobile mesurée à l'instant t, $r(t)$. Dans plusieurs

traces de mobilité, nous avons tendance à observer un phénomène de *super-diffusion* du mouvement humain caractérisé par une distribution de puissance du MSD, $MSD(t) \sim t^{\gamma}$, tel que $\gamma > 1$ [KLE10a ; Son+10]. D'autres études montrent un comportement de diffusion ultra-faible (*ultra-slow diffusif* ou *sub-diffusion*) dans des traces de mobilité individuelle caractérisé par $MSD(t) \sim t^{2\alpha/\beta}$ ($\beta \sim 0{,}59$ et $\alpha \sim 0.6$ sont approximativement égales aux coefficients des distributions respectives $P(\Delta r)$ et $P(\Delta t)$ des billets de banques) [BHG06] [Bar+18]. De même, Rhee et *al.* découvrent ce comportement de diffusion ultra-faible dans certaines données réelles [Rhe+11]. Ils avancent que la tendance de l'homme à revenir à son point de départ est à l'origine de ce phénomène [Rhe+11]. En effet, les mouvements humains ne sont pas aléatoires et l'homme retourne souvent à la maison en fin de journée ou bien revient à un lieu donné. Cette tendance à "retourner au foyer" ralentit excessivement le phénomène de diffusion, entraînant une diffusion ultra-faible.

3.3.1.3 *Rayon de giration*

Le Rayon de giration est une autre grandeur physique qui résume bien l'étendue spatiale de la mobilité totale d'un utilisateur observé sur une période donnée. Selon [Bar+18], le rayon de giration peut être défini par la relation suivante :

$$r_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |r_i - r_{cm}|^2} \tag{3.37}$$

où r_i sont les positions des N lieux visités, r_{cm} représentent le centre de masse de l'ensemble de la trajectoire, $r_{cm} = \sum_{i=1}^{N} r_i/N$. Cette mesure, appelée *distance caractéristique* par González et *al.* [GHB08], se révèle très importante dans le cas où elle permet de délimiter la trajectoire d'un individu sur un intervalle de temps. Nous pouvons remarquer que la valeur de r_g peut être très petite lorsqu'on se déplace dans plusieurs endroits proches, très grande lorsqu'on parcourt par exemple deux endroits éloignés. En calculant le rayon de giration de chaque utilisateur mobile $r_g^{(a)}$, González et *al.* ont trouvé que sa distribution suivait approximativement la loi de puissance tronquée [GHB08].

$$P(r_g) = (r_g + r_g^0)^{\beta_r} \exp(r_g/\kappa) \tag{3.38}$$

avec $r_g^0 \sim 5.8$ km, $\beta_r = 1.65 \pm 0.15$ and $\kappa \simeq 350$ km. Ce résultat montre encore une fois que la plupart des individus se déplacent très souvent à des endroits proches de leur lieu d'habitation et rarement à de longues distances. Dans [GHB08], les auteurs vont plus loin en étudiant l'évolution temporelle du rayon de giration. Ils ont montré que la moyenne du rayon de giration $rg(t)$ des données d'utilisateurs de téléphonie mobile évolue de manière logarithmique ($\langle rg(t) \rangle \sim A + B \ln t$) contrairement à ceux observés dans les marches aléatoires ($rg(t)|_{RW} \sim t^{1/2}$), Lévy Flights ($rg(t)|_{FL/TLF} \sim t^{3/1+\beta_{LF}}$ avec $\beta_{LF} = 1{,}75 \pm 0.75$)[GHB08]. Par conséquent, nous pouvons en déduire que la mobilité humaine est caractérisée par une périodicité, une tendance de retour (*Preferential Return* [PS16 ; Son+10]) dans les lieux déjà visités, une probabilité très faible de se déplacer vers de nouvelles destinations et une limitation due aux contraintes spatiales.

3.3.1.4 Fréquence de visite

Il est évident de constater que l'homme se déplace naturellement sur un ensemble limité de lieux de manière fréquente. Il a tendance à retourner périodiquement vers les lieux qui ont été déjà visités, et à explorer parfois de nouvelles localités. Parmi ses lieux de visite, la périodicité peut être quotidienne, hebdomadaire ou mensuelle. Pour étudier cette fréquence de visite, Gonzalez et *al.* ont classé les localités en fonction d'un rang L représentant le L^{ieme} site le plus visité par l'individu, par exemple la maison peut prendre le rang $L = 1$, le bureau 2 et le supermarché 3 [GHB08]. Ces auteurs montrent, en utilisant des données de téléphonie mobile que la probabilité de trouver un utilisateur à un endroit de rang L est approximativement égale à $P(L) \sim 1/L$. En plus, ce résultat est indépendant du nombre d'emplacements visités par l'utilisateur. La fréquence f du lieu L^{ieme} le plus visité suit une loi de Zipf selon [Son+10] [Jur+15].

$$f_L \sim L^{\zeta} \tag{3.39}$$

Avec une valeur de $\zeta = 1.2 \pm 0.1$. De même, la distribution de la fréquence de visite estimée par $P(f) \sim f^{1+1/\zeta}$ [Son+10]. Song et *al.*, en analysant les CDRs, montre que le nombre distinct de sites S qu'un individu peut visiter durant un temps t est donné par $S(t) \sim t^{\mu}$, avec $\mu = 0.6 \pm 0.02$. Ces résultats permettent de confirmer, effectivement, les constats relevés au début sur la mobilité humaine.

3.3.2 Modèles de mobilité

Comme évoqué précédemment, la mobilité humaine est si complexe qu'on ne peut pas capturer toutes les caractéristiques inhérentes. Et pourtant, nous avons besoin des modèles qui pourront nous servir dans certaines applications ou bien de comprendre leurs impacts sur certains phénomènes de la vie réel. C'est dans cette perspective que les chercheurs n'ont jamais cessé d'étudier et de proposer des modèles aidant à comprendre cette mobilité. Parmi ces modèles, nous pouvons distinguer deux types de mobilité : la mobilité individuelle et la mobilité globale. La mobilité globale est orientée à une grande échelle et permet d'étudier par exemple les trafics urbains et les flux migratoires entre les populations tandis que la mobilité individuelle permet de décrire le comportement des individus sous formes d'agents suivant un certain nombre de règles dans un environnement spécifique. Et ces règles dépendent de certaines propriétés physiques que sont la position, la distance, la direction, la durée de pause etc. Par conséquent, la mobilité individuelle peut être séparée en deux classes : les modèles synthétiques basés sur des équations mathématiques qui tentent de reproduire le mouvement du mobile dans une zone de simulation et les modèles empiriques basés sur des traces (logs, traceroute, GPS, Bluetooth, Wifi etc.) obtenus à partir de systèmes déployés dans l'environnement.

3.3.2.1 Modèle de Gravité

Proposé par George K. Zipf en 1946 [Zip46], le modèle de gravité s'inspire de la loi de Newton pour calculer approximativement les flux migratoires entre deux populations données. Il fût parti des premiers modèle de mobilité et s'est rapidement répandu dans la communauté scientifique. Le modèle de gravité a été ensuite appliqué dans différents contextes tels que

le transport, le trafic des biens et services, les transactions financières et économiques, les communications téléphoniques etc. [Roy11a]. Dans sa première version, Zipf assume que le flux T_{ij} est proportionnel au produit de la population P_i P_j vivant respectivement dans les sites i et j multiplié par l'inverse de la distance r_{ij} qui les sépare.

$$T_{ij} \propto \frac{P_i P_j}{r_{ij}} \tag{3.40}$$

L'avantage de ce modèle est qu'il permet facilement de capturer l'attraction des grandes agglomérations. L'inconsistance et les contraintes spatiales liées à la notion de distance font que ce modèle présente des limites. Par ailleurs, une généralisation du modèle de gravité est donnée par la formule suivante [Bar+18] :

$$T_{ij} = K P_i^{\alpha} P_j^{\beta} f(r_{ij}) \tag{3.41}$$

tel que K est une constante, α et β sont des paramètres de calibration et $f(r_{ij})$ est appelée facteur de "friction" qui diminue en fonction de la distance. Cette fonction $f(r_{ij})$ peut être adaptée selon notre contexte d'étude et peut-être estimée comme une fonction puissance, exponentielle ou bien la combinaison des deux par exemple [Bar+18].

$$f(r_{ij}) = r_{ij}^{-\gamma}; \;\; f(r_{ij}) = e^{-\beta r_{ij}}; \;\; f(r_{ij}) = \alpha r_{ij}^{-n} e^{-\beta r_{ij}} \tag{3.42}$$

Le véritable défi du modèle c'est de trouver une bonne calibration de ces paramètres α, β et γ. Plus de détails ont été développés dans les articles [Bar+18 ; Roy11a ; Bar11] pour faire le *'fitting'* de ces paramètres. En dehors des applications précitées, le modèle de gravité a été intégré dans certains modèles épidémiques pour prendre en compte le mouvement des populations. Dans [Bal+09], Vespignani et *al.* intègrent le modèle de gravité dans leur modèle épidémiologique étudié sur une méta-population avec des données de mobilité collectées dans presque 29 pays par IATA (*International Air Transport Association*) afin d'évaluer la force d'infection due aux processus de mobilité multi-échelles dans la dynamique de la maladie.

3.3.2.2 Modèle de Radiation

A cause des nombreux manquements qui sont relevés dans [Sim+12], Simini et *al.* suggèrent une alternative au modèle de gravité. Le modèle de Simini et *al.* ou modèle de *Radiation* s'appuie sur le concept du "choix des opportunités" (*'opportunity selection'*)[Sim+12]. Ce concept se résume en deux hypothèses : premièrement, le nombre de possibilités d'emploi dans chaque région est proportionnel à la population résidente. Deuxièmement, l'individu choisit l'opportunité la plus proche de son domicile, dont les avantages sont supérieurs à la meilleure offre disponible dans sa région d'origine. A partir de ces hypothèses, les auteurs proposent le modèle de Radiation défini comme suit [Sim+12] :

$$\langle T_{ij} \rangle = T_i \frac{m_i n_i}{(m_i + s_{ij})(m_i + n_i + s_{ij})} \tag{3.43}$$

où T_i représente le nombre total de flux quittant le site i, m_i et n_i sont respectivement les populations des régions i et j, enfin s_{ij} représente la taille de la population étendue dans un cercle de rayon r_{ij} centré sur i. Contrairement au modèle de gravité, ce modèle est sans paramètres et permet facilement de prédire différents types de mobilité. Cependant, la

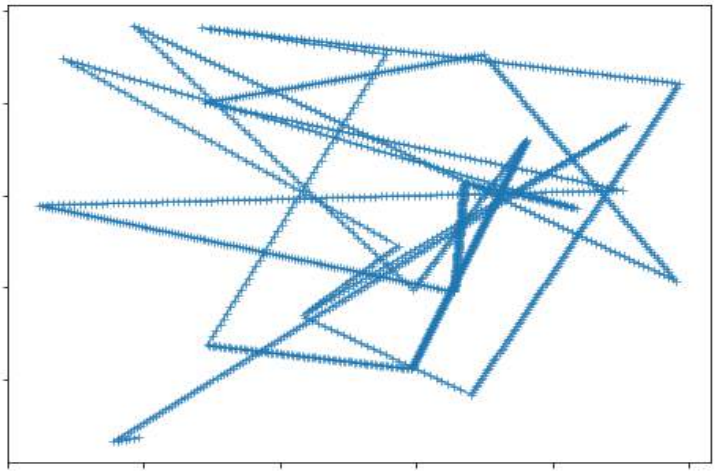


FIGURE 3.9: Trajectoire d'un individu suivant le modèle *Random Waypoint Mobility* (RWP)

nature sans paramètre du modèle limite malheureusement ses capacités à mesurer les diverses motivations qui initient la mobilité humaine individuelle [Kan+15]. C'est ainsi que Kang et *al.* [Kan+15] ont proposé récemment une série de modèles paramétrables et généralisables dans différents contextes. Comme le modèle de gravité, le modèle de radiation est utilisé sur plusieurs modèles épidémiques à grande échelle [Sal+17 ; Cho+16 ; Tiz+14].

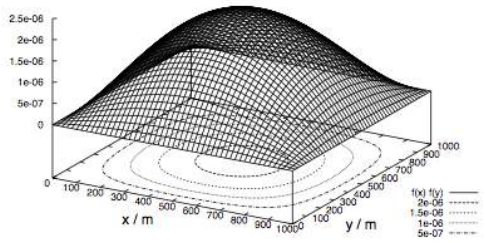
3.3.2.3 Modèles synthétiques de mobilité

Dans cette partie, nous nous intéressons essentiellement aux modèles synthétiques de mobilité tels que *Random Waypoint* (RW), *Gauss Markov* (GM), *Truncated Levy Walk* (TLW), *Spatio-TEmporal Parametric Stepping* (STEPS) que nous allons utiliser pour générer les données de contacts temporelles servant de support dans notre étude.

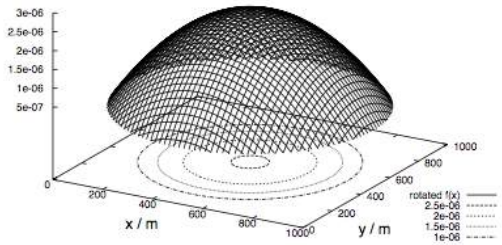
— **Random Waypoint**

Le Random Waypoint (RWP) a été proposé pour la première fois par Johnson et Maltz [JM96] et est devenu le principal modèle de mobilité utilisé pour les simulations dans les MANETs, surtout pour tester certains algorithmes de routage. Étant une marche aléatoire, le RWP détermine la position du mobile évoluant dans un espace de simulation en fonction de sa vitesse, de sa direction et du temps de pause consommé à chaque étape. Itérativement, le nœud choisit de manière aléatoire une nouvelle destination (*'waypoint'*) avec une vitesse qui est uniformément distribuée dans l'intervalle $[v_{min}, v_{max}]$. Arrivée à destination, chaque mobile observe un temps de pause choisie aléatoirement entre 0 et w_n (avec w_n la durée de pause maximale) avant de se déplacer vers un autre lieu. La Figure 3.9 montre la trajectoire d'un mobile suivant une mobilité RWP.

Du fait de sa popularité, le modèle RWP a fait l'objet de plusieurs investigations dans la littérature [BRS03 ; BW02 ; HLV06]. Bettstetter et Wagner ont montré que la densité de probabilité de la distribution des nœuds spatiaux du modèle RWP dans un espace carré est donnée par la formule 3.45 [BW02] :



(a) Zone de simulation carrée



(b) Zone de simulation circulaire

FIGURE 3.10: Dispersion spatiale des noeuds suivant le modèle RWP [BW02]

$$f(x,y) \approx f(x)f(y) = \frac{9}{16x_m^3 y_m^3}(x^2 - x_m^2)(y^2 - y_m^2) \tag{3.44}$$

tels que $0 \leq x \leq x_m$ et $0 \leq y \leq y_m$. Tandis que sa distribution dans un espace circulaire de rayon r d'angle ϕ est donnée par la relation ci-dessous [BW02] :

$$f(r,\phi) \approx f(r) = -\frac{2}{r_m^4 \pi} r^2 + \frac{2}{r_m^4 \pi} \tag{3.45}$$

où $0 \leq r \leq r_m$. La figure 3.10 fait une comparaison de la dispersion des nœuds suivant RWP entre un espace carré et un espace circulaire. Cependant, ce modèle souffre de quelques problèmes dans son fonctionnement et ne reproduit pas les caractéristiques de la mobilité humaine. Par exemple, les nœuds ont tendances à se concentrer au centre de l'espace de simulation au fil du temps [BRS03]. En outre, RWP souffre du fait que le régime transitoire (c'est-à-dire non stationnaire) peut durer très longtemps [M M09]. Il est également prouvé que le RWP présente une décroissance de la vitesse du mobile au fil du temps [BW02 ; JN03]. Malgré les problèmes évoqués ci-dessus, ce modèle reste encore très utilisé dans le domaine des réseaux opportunistes.

— **Mobilité de Gauss-Markov**

Ce modèle, développé par Liang et Haas [LH03], tente d'imiter le fameux processus aléatoire de Gauss-Markov. En effet, l'implémentation de ce modèle dite Gauss-Markov est motivée par le principe qui est un peu plus proche de la réalité, dans le sens où un nœud peut s'accélérer, se ralentir ou tourner progressivement dans n'importe quelle direction. Il est montré qu'en temps continu, un processus Gaussien stationnaire $v(t)$ est un processus de Gauss-Markov s'il possède la fonction d'auto-corrélation suivante [LH03]

$$R_v(\tau) = E[v(t)v(t+\tau)] = \sigma^2 e^{-\gamma|\tau|} \tag{3.46}$$

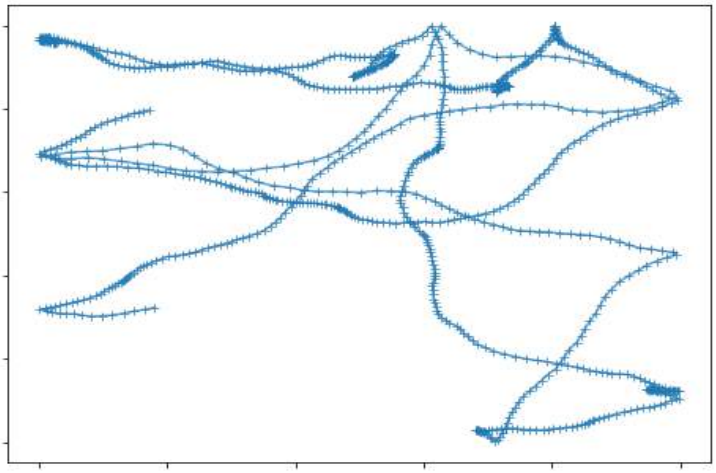


FIGURE 3.11: Trajectoire d'un agent suivant le modèle de mobilité *Gauss-Markov* (GM)

où σ^2 est la variance de $v(t)$ et γ détermine le degré de mémoire dans le modèle de mobilité. Les auteurs ont montré que la vitesse et la direction à la nième instance du processus discret de Gauss-Markov sont calculées en fonction de la valeur de la vitesse et de la direction à la (n-1) -*ème* instance. Par conséquent, ces deux variables aléatoires sont obtenues en utilisant les équations suivantes [Roy11b].

$$v_n = \alpha v_{n-1} + (1-\alpha)\mu_v + \sqrt{1-\alpha^2}\, w_{v_{n-1}} \tag{3.47}$$

$$\theta_n = \alpha \theta_{n-1} + (1-\alpha)\mu_\theta + \sqrt{1-\alpha^2}\, w_{\theta_{n-1}} \tag{3.48}$$

où v_n et θ_n sont respectivement la nouvelle vitesse et la nouvelle direction du mobile à l'instant t_n, α est un paramètre qu'on pourrait modifier pour garder le caractère aléatoire du modèle $(0 \leq \alpha \leq 1)$, μ_v et μ_θ sont deux constantes représentant la valeur moyenne de la vitesse et de la direction lorsque n tend vers l'infini, et enfin $w_{v_{n-1}}$ et $w_{\theta_{n-1}}$ sont des variables aléatoires suivant une distribution gaussienne. De ce fait, la position du mobile à chaque intervalle de temps est calculée en fonction de son emplacement actuel, de la vitesse et de la direction du mouvement. Elle est donnée par les équations ci-dessous [Roy11b].

$$x_n = x_{n-1} + v_{n-1}\cos\theta_{n-1} \tag{3.49}$$

$$y_n = y_{n-1} + v_{n-1}\cos\theta_{n-1} \tag{3.50}$$

où (x_n, y_n) et (x_{n-1}, y_{n-1}) sont respectivement les coordonnées x et y de la position du mobile aux n-ème et $(n-1)$ ème intervalles de temps et v_{n-1} et θ_{n-1} sont respectivement la vitesse et la direction du mobile à l'intervalle de temps $(n-1)$. La figure 3.11 montre la trajectoire d'un individu suivant une mobilité Gauss-Markov.

Nous savons que la mobilité Gauss-Markov est un modèle à mémoire contrôlé par le paramètre α. Par conséquent, pour différentes valeurs de α, nous pouvons constater différentes étapes du modèle.

— Si $\alpha = 0$, alors le modèle devient sans mémoire ; la vitesse v_n et la direction θ_n dépendent entièrement de vitesse moyenne et de la direction moyenne et des variables aléatoires gaussiennes.

$$v_n = \mu_v + w_{v_{n-1}} \tag{3.51}$$

$$\theta_n = \mu_\theta + w_{\theta_{n-1}} \tag{3.52}$$

— Si $\alpha = 1$, alors le mouvement devient prévisible, perdant tout caractère aléatoire. Les nouvelles valeurs de direction et de vitesse sont identiques aux valeurs précédentes de direction et de vitesse. En bref, le nœud continue en ligne droite.

$$v_n = v_{n-1} \tag{3.53}$$

$$\theta_n = \theta_{n-1} \tag{3.54}$$

— Si $0 < \alpha < 1$, alors le calibrage du paramètre α permet de garder le caractère aléatoire et de varier le degré de mémoire du modèle. Outre le paramètre α, la dynamique du modèle de mobilité de Gauss-Markov est fortement influencée par d'autres variables telles que le pas de temps (*time step*), la sélection de la vitesse moyenne et de la direction, ainsi que la moyenne et l'écart type choisis pour les variables aléatoires de Gauss.

— **Truncated Lévy Walk (TLW)**

Truncated Lévy Walk est un modèle développé par Rhee et *al.* [Rhe+11] dans le but de reproduire certains caractéristiques de mobilité humaine dans leur modèle. Les caractéristiques de ce modèle sont marquées par la distribution de Lévy sur la trajectoire des individus (*heavy-taild or Levy flight*), la loi de puissance sur la durée de pause et la nature de super-diffusion dans le MSD [Son+10]. Les auteurs ont défini TLW comme une marche aléatoire utilisant la loi de Pareto tronquée pour la distribution des *flights* (déplacements) et des temps de pause afin de simuler la mobilité dans une zone confinée. Dans cette définition, Rhee et *al.* [Rhe+11] utilise le plus souvent le terme *'flight'* qui peut être défini comme étant le plus long trajet en ligne droite d'un endroit à un autre qu'un mobile effectue, sans changement de direction ni de pause. Ainsi, nous pouvons définir mathématiquement un TLW comme une marche aléatoire constituée par une séquence d'étapes telle que chaque étape est représentée par un pas $S = (l, \theta, \Delta t_f, \Delta t_p)$ tel que $l > 0$ est la longueur du *flight* (*flight lenght*), θ représente la direction du *flight*, enfin $\Delta t_f > 0$ et Δt_p sont respectivement la durée du *flight* et le temps de pause. La Figure 3.12 montre l'exemple de la trajectoire d'un mobile suivant TLW.

À chaque étape, le mobile choisit une direction aléatoire d'angle θ à partir d'une distribution uniforme sur l'intervalle $[0, 360]$, une durée Δt_f finie tiré au hasard sur une certaine distribution avec une longueur de *flight* l et un temps de pause Δt_p choisis de manière aléatoire à partir des densité de probabilités $p(l)$ et $\psi(\Delta t_p)$ qui suivent la loi de Lévy avec respectivement les coefficients α et β. La distribution de Lévy de paramètres c et α est donnée par la transformation de Fourier suivante [Rhe+11] :

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx - |ct|^\alpha} \tag{3.55}$$

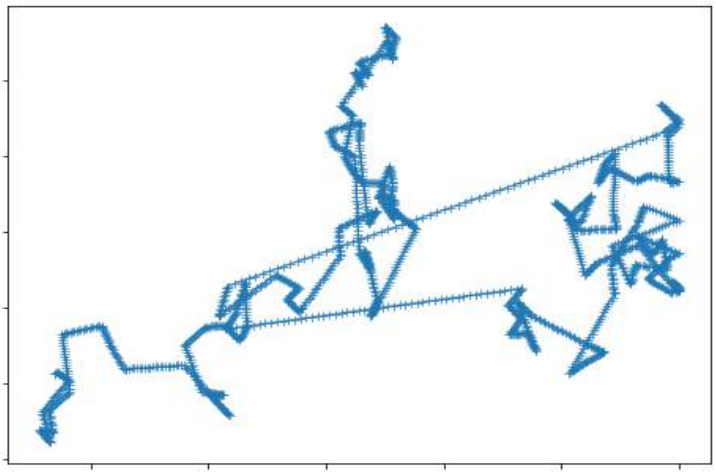


FIGURE 3.12: Mouvement d'un individu suivant la mobilité *Truncated Lévy Walk* (TLW)

La constante $c > 0$ est un facteur d'échelle. Si $\alpha = 2$, nous avons une distribution gaussienne avec un écart-type $\sigma = \sqrt{2}c$. Pour une valeur de $\alpha = 1$, on obtient une distribution de Cauchy.

A partir des paramètres α, β et c, le modèle Truncated Levy-Walk génère des traces de mobilité avec des facteurs de troncatures τ_l et τ_p pour la longueur du *flight* et le temps de pause, respectivement. La durée du flight Δt_f est donnée par la relation $\Delta t_f = k l^{1-\rho}$, avec k et ρ des constantes. A chaque pas S, si l et Δt_p sont négatifs, $l > \tau_l$ ou $\Delta t_p > \tau_p$, alors S est ignoré et un nouveau pas est généré à nouveau. Et ce processus est répété jusqu'à la fin de la simulation. Il est important de noter dans ce modèle que lorsqu'un *'flight'* franchit une limite de la zone prédéfinie, alors le mobile est réfléchi par le mur.

— ***Spatio-TEmporal Parametric Stepping (STEPS)***

De la même façons que TLW, le modèle STEPS [Ngu+11] a été élaboré dans le but d'y intégrer certaines caractéristiques rencontrées dans le mobilité humaine. En plus de la propriété de la loi puissance observée dans le déplacement des individus (*heavy-taild or Levy flight*), sur le temps de pause [Son+10] et sur la durée de contacts et inter-contacts [KBV10], Nguyen et *al.* ajoutent l'attachement préférentiel dans le choix des lieux de visite dans leur modèle [Ngu+11].

Dans le modèle STEPS, l'espace de simulation est constitué par d'un ensemble d'emplacements, chacun étant considéré comme une zone dans laquelle un nœud peut se déplacer librement selon un modèle de mobilité aléatoire tel que *Random Waypoint* (RWP). Un emplacement peut représenter la maison, le lieu de travail, la salle de sport etc. Ensuite, les déplacements entre les zones et la durée des temps de pauses dans une zone sont tous deux tirés à partir de distributions de loi de puissance.

Initialement, le nœud choisit une zone Z_0. Répétitivement, il sélectionne aléatoirement

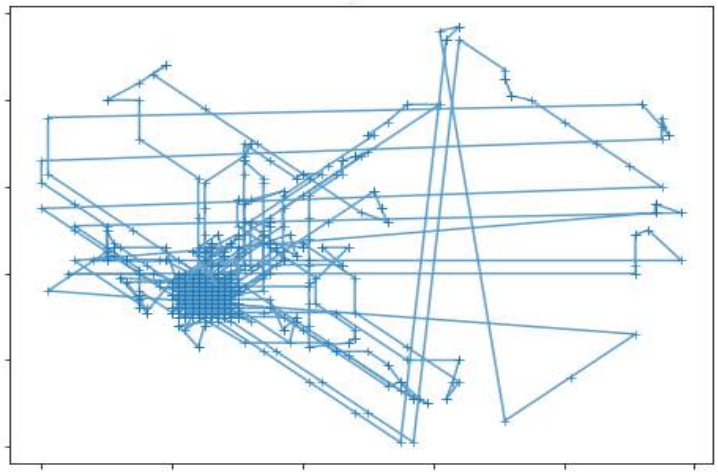


FIGURE 3.13: Mouvement d'un agent suivant la mobilité *Spatio-TEmporal Parametric Stepping* (STEPS)

une distance d (distance de Chebyshev [2]) selon la distribution de probabilité suivante :

$$P(D = d) = \frac{\beta}{(1+d)^{\alpha}} \tag{3.56}$$

avec d une distance à partir de Z_0, α représentant l'exposant de la loi de puissance et β une constante de normalisation. Ensuite, il sélectionne au hasard une zone Z_i parmi les zones dont leurs distances entre Z_0 est de d et un point dans cette zone où il va y accéder à une vitesse qui varie entre $[v_{min}, v_{max}]$. À l'intérieur de Z_i, le nœud va se déplacer aléatoirement en suivant un mouvement RWP. Nous, notons que le mobile va rester dans cette zone jusqu'à une durée t, tirée de la distribution de loi puissance suivante.

$$P(T = t) = \frac{\omega}{t^{\tau}} \tag{3.57}$$

avec τ le degré de préférence temporelle du nœud et ω une constante de normalisation. La Figure 3.13 illustre la trajectoire d'un mobile suivant le modèle de mobilité STEPS.

3.4 Conclusion

La compréhension et la prédiction des comportements des phénomènes dynamiques émergents dans les systèmes complexes, représentent toujours un enjeu crucial dans la communauté scientifique. Avec le développement des réseaux complexes, l'épidémiologie moderne prend en compte plusieurs paramètres dans sa modélisation : la topologie du réseau (réseaux aléatoires, *scale-free*, *small-world*), les différents types d'inter-actions, la dynamique interne et externe (évolution temporelle de la structure, modification des états des nœuds et des connexions), la '*clusterisation*' des nœuds en communauté, la proximité des agents, la durée des contacts, le

2. Entre deux points A et B, de coordonnées respectives $(A_0, \ldots, A_n)$ et $(B_0, \ldots, B_n)$, la distance de Tchebychev est définie par : $d(A,B) = \max_{i \in [[0,n]]} (|A_i - B_i|)$.

phénomène de bursty etc. Dans ce chapitre, notre objectif était d'une part de comprendre la modélisation de la propagation épidémiques particulièrement les modèles SI, SIS, SIS, d'autre part d'étudier les principales caractéristiques de la mobilité humaine qui ont permis de construire des modèles qui pourront être adaptés dans différentes situations du monde réel. En bref, nous constatons que beaucoup d'efforts ont été fournis dans ces deux domaines cependant la question sur l'impact de la mobilité dans la diffusion des maladies épidémiques reste encore ouverte.

Chapitre 4

Impact de la mobilité dans la propagation d'épidémie

4.1 Introduction

Nous savons que les réseaux complexes ont toujours servi de support pour modéliser les phénomènes dynamiques dans les systèmes complexes. Dans plusieurs travaux, nous notons un apport considérable des réseaux complexes statiques dans la compréhension de la dynamique interne telle que les phénomènes de diffusion et de propagation [BBV08 ; Pas+15]. Dans certains contextes, il faut considérer l'évolution de la topologie du réseau en fonction de la dynamique externe qui devient non négligeable surtout dans les réseaux de contacts humains. Par exemple, le nombre de contacts personne-à-personne dans un aéroport ou dans un hôpital peut être très important pour chaque individu entre des intervalles de temps très proches. Et ce fait est essentiellement dû à la mobilité des individus. Dans ce type de scénario, nous pouvons constater que la dispersion d'un virus dans un environnement pareil pourrait être très rapide et difficile à stopper. Ceci nous interpelle à nous demander comment appréhender ce processus de propagation en tenant compte de la mobilité humaine. Récemment, avec l'opportunité qu'offrent les données d'utilisateurs collectées dans les supports sans fil (Wifi, RFID, GPS, Bluetooth ou Beacon), nous assistons à une nouvelle vague de recherches associant données temporelles et processus dynamiques dans les systèmes complexes. Ce vaste mouvement est encouragé par les nouvelles approches de modélisations des réseaux complexes telles que les réseaux multiplexes, les réseaux adaptatifs et les réseaux temporels (voir la Section 2). Cependant, la plupart de ces études se focalisent sur des données de contacts physiques ou des liens sociaux ce qui fait perdre l'information sur la mobilité des individus. Le cas Sociopattern [1] qui est un projet permettant de récupérer des données de contacts face-à-face avec de la radio-identification (RFID) en est une parfaite illustration [MFB15 ; GBC14]. Considérant les limites émanant de la protection de données personnelles, il y a eu peu d'études orientés dans la collecte des traces de positions exactes des individus. Le projet Geolife de *Microsoft Research Asia* avait permis de collecter les données GPS de 178 utilisateurs sur une période de plus de quatre ans (d'Avril 2007 à Octobre 2011) [Zhe+09].

1. `http://www.sociopatterns.org`

Toutefois, ces données sont très éparses puisqu'elles sont à l'échelle de 30 villes situées en Chine et quelques régions localisés aux États-Unis et en Europe. En plus, le nombre de participants est très limité ce qui diminue la possibilité d'avoir des contacts directs en temps réels entre les individus. Sous l'initiative de Orange D4D (*Data for Développement*), les données des utilisateurs mobiles de Côte d'ivoire (2012) et Sénégal (2014) étaient aussi ouvertes dans le but de permettre aux chercheurs du monde entier d'apporter des réponses aux défis liés à l'environnement, à la santé, aux transports et d'autres thématiques socio-économiques [Blo+12 ; Mon+14]. Cependant, ces datasets de mobilité ne donnent pas la position exacte des utilisateurs, au contraire leurs postions à l'échelle de la couverture des antennes, leurs fréquences d'appels etc. A défaut de la disponibilité de ces données de mobilité, nous avons saisi, pour une première expérience, l'opportunité qu'offre les modèles synthétiques de mobilité afin de comprendre et de répondre aux questions adressées ci-dessous :

Comment les contacts sont-ils créés en fonction de la mobilité ? Quel sera l'impact de la mobilité sur la structure du réseau de contact temporel ainsi que sur la propagation d'épidémie ? Quelle est l'importance de la distance de proximité dans le facteur de transmission de certains cas de maladies (par exemple les maladies qui se transmettent par voie aérienne comme la tuberculose, la grippe, l'Anthrax, SARS) ? Comment la probabilité d'infection peut-elle dépendre du nombre de voisins infectés rencontrés ou bien de la durée de leurs contacts ? Les réponses à toutes ces questions pourraient constituer des hypothèses à prendre en compte pour bien comprendre la relation qui existe entre mobilité et propagation d'épidémie.

Dans cette partie, en nous appuyant sur le modèle épidémiologique SIS, nous avons utilisé différentes traces générées par les modèles synthétiques de mobilité tels que *Random Waypoint*(RWP), *Gauss-Markov*(GM), *Truncated Lévy Walk*(TLW), *Spatio-TEmporal Parametric Stepping* (STEPS) pour observer l'impact de la mobilité sur les phénomènes dynamiques. Après une brève revue des travaux précédents autour de ce sujet, nous allons effectuer une analyse statistique de la topologie des graphes dynamiques obtenus après une description l'environnement de travail. Ensuite, nous allons proposer un modèle de propagation permettant d'intégrer la notion de proximité et de voisinage dans le processus de transmission. Enfin, nous terminerons par une discussion des résultats obtenus.

4.2 Travaux connexes (*Related Works*)

Plusieurs travaux ont été menés dans le domaine des réseaux complexes pour comprendre l'impact des interactions sociales et physiques dans le processus de contamination et de la propagation des épidémies. Dans [SPL15], Stopczynski et *al.* ont montré l'impact critique de la distance de proximité des contacts sur la transmission des maladies. Ils ont étudié les réseaux de contacts d'une population entre deux niveaux : un réseau d'interactions '*intimes*' défini sur une distance d qui est inférieure ou égale à un mètre et un réseau ambiant constitué par l'ensemble des interactions entre les individus situés sur une distance $d < 10$ mètres. En utilisant des données *Bluetooth* collectées sur un mois, ils ont pu découvrir que les processus de propagation sur les réseaux de proximité étaient très sensibles à la proximité des distances. Dans ce papier, les auteurs ne se sont pas intéressés sur l'influence de la distance d'interaction entre les individus dans la probabilité d'infection [SPL15]. Dans [Fen+16], les auteurs ont

tenté d'expliquer le fait que les personnes en bonne santé réduisent leurs contacts avec leurs voisins infectés et ont tendance à les rétablir une fois qu'ils sont guéris. En utilisant un modèle SIS sur une graphe dynamique pondéré, ils ont montré que les interactions de poids fort ont tendance à augmenter plus le seuil épidémique, à rendre la moyenne du taux de prévalence faible et à diminuer la durée de vie de l'épidémie. En plus, les individus qui réagissent selon les informations locales de l'infection diminuent plus la probabilité de déclenchement d'épidémie par rapport à ceux qui prennent en compte les informations globales de l'infection. De manière similaire, Sun et *al.* [Sun+14] ont développé un modèle épidémique basé sur les réseaux dynamiques pondérés pour intégrer le rôle des relations multiples que peuvent entretenir les individus dans le phénomène de propagation. Cependant, leurs travaux ignorent l'influence de la dispersion spatiale et de la proximité des individus. Valler et *al.* ont étudié la dynamique de propagation des virus dans le contexte d'un réseau temporel de proximité basé sur des modèles de mobilité *Random Walk*, *Random Waypoint* et *Lévy Flight* [Val+11]. A partir du modèle SIS, ils ont démontré que la vitesse des nœuds dans le réseau ad hoc dépendant de la mobilité n'affectait pas le seuil épidémique obtenu [Val+11]. Cependant, cette étude n'intègre pas dans le processus de contamination l'influence de la distance, du voisinage et de la durée des contacts [Val+11]. Récemment, Vestergaard et Génois ont utilisé la méthode stochastique de Gillespie pour simuler le processus de contagion SIS et SIR sur un TVG avec une probabilité de guérison qui n'est pas constant [VG15]. Ils ont aussi montré que leur modèle est extensible à d'autres processus dynamiques et s'avère plusieurs fois plus rapide que les modèles déjà connus. Néanmoins, nous remarquons qu'ils ne prennent pas en compte les contraintes spatiales. Dans [Hon+16], les auteurs utilisent le modèle de mobilité RWP pour montrer l'influence de plusieurs paramètres d'un réseau ad hoc tels que le rayon de couverture, la vitesse de déplacement et la densité de nœuds sur le taux de prévalence des nœuds dans le modèle SEIR.

4.3 Méthodologie

En se fixant comme objectif de comprendre et de répondre aux formulations précitées, nous avons supposé, dès le début, de travailler avec des données provenant de traces synthétiques pour mieux mettre en évidence la notion de proximité dans le processus de contamination et de propagation. Ce qui nous a poussé, hormis les traces réelles existantes, à générer des données à partir de différents modèles de mobilité que nous allons utiliser, pour enfin construire des réseaux ou graphes temporels aléatoires de proximité ou plutôt de contacts. Nous considérons que deux individus sont supposés être en contacts s'ils sont en proximité et que la distance qui les sépare est inférieure à une distance seuil, noté d_{th}. A partir de différents modèles de mobilité utilisés (*RWP*, *GM*, *TLW* et *STEPS*) nous allons étudier la nature des réseaux de contacts temporels qui ont été générés et l'évolution du phénomène épidémique qui a lieu dans ces différents réseaux.

Nous considérons un réseau temporel (TVG) définit par $\mathcal{G} = \{G_1, G_2, \ldots, G_T\}$ est constitué d'un ensemble de *snapshots* de graphe $G_t = (V_t, E_t)$ généré à chaque intervalle de temps $t \in \mathbb{T}$ (avec $\mathbb{T}$ une mesure de temps discret $\mathbb{T} \subset \mathbb{N}$) à partir de l'un des modèles de mobilité suivants : *RWP*, *GM*, *TLW* et *STEPS*. Dans la suite de l'exposé, chaque TVG obtenu sera indexé par

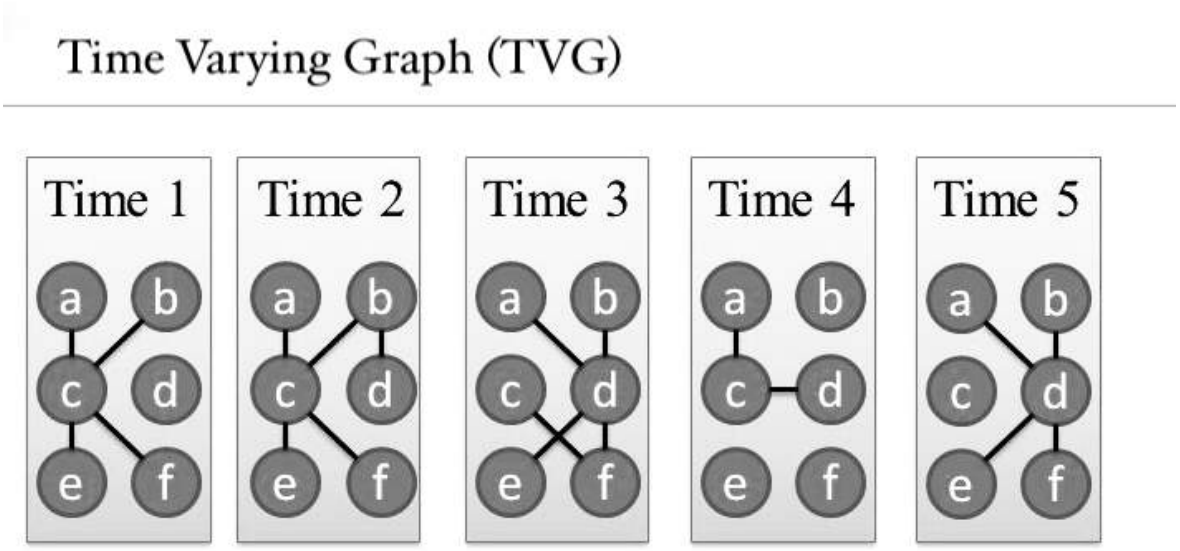


FIGURE 4.1: Exemple de Réseau Temporel (TVG) $\mathcal{G} = \{G_1, G_2, G_3, G_4\}$

le modèle de mobilité utilisé ; par exemple $\mathcal{G}_{TLW}$ correspond à celui dont les données sont générées par le modèle de mobilité *Truncated Lévy Walk*. Nous notons que la dynamique topologique de $\mathcal{G}$ ne concerne que l'activité des contacts entre les individus, ce qui implique que le nombre de nœuds reste invariant $V_t = V$ durant toute la période de simulation. La Figure 4.1 montre un exemple de réseau temporel de 6 nœuds (a, b, c, d, e, f) constitué par une séquence de *snapshots* de graphe.

Dans la phase de génération des traces concernant les modèles de mobilité *RWP*, *GM* et *TLW*, nous précisons que nous avons utilisé le simulateur open source *pymobility* développé en Python par André Panisson [Pan14], tandis que les traces du modèle *STEPS* ont été générées à partir de la librairie du même nom développée par les auteurs de l'article [Ngu+11]. Le code source de ce modèle est écrit en MATLAB. Les algorithmes 1 et 2 décrivent en détails les différentes étapes de génération des graphes temporels que nous avons obtenues. Dans l'algorithme 1, nous avons développé une fonction qui permet de créer un graphe instantané à partir des positions des mobiles. L'algorithme 2 utilise cette fonction pour générer une liste de snapshots de graphe représentant le réseau temporel $\mathcal{G}$ décrit ci-dessus.

Nous précisons que le tableau 4.7 permet de récapituler les différentes paramètres de simulation que nous avons utilisés dans cette phase de génération de graphes temporels.

En effet, nous savons que la probabilité d'être infecté dépend de l'interaction avec les individus infectieux et du facteur de transmissibilité de l'agent pathogène. Cette interaction peut être, dans plusieurs cas de maladie, un contact direct avec l'agent infectieux ou indirect dans d'autres cas. Dans le cas des maladies où l'agent pathogène peut se suspendre dans l'air comme la grippe, la proximité (distance) et le voisinage avec les nœuds qui sont infectieux peuvent faciliter la dispersion rapide du virus. En effet, dans ce mode de transmission, les particules sont plus petites ($< 5\mu m$) et peuvent rester en suspension dans l'air pendant de longues périodes [SPL15]. Par conséquent, la portée de la transmission par voie aérienne est

Algorithm 1 Spatial weighted graph

INPUT : $G(t)$ - Graph at time t,
N - Network size,
P - $\{(i,(x,y)) : i = 1...N\}$,
MAX_X, MAX_Y - $MAX_X \times MAX_Y$ grid size,
t - Time step
OUTPUT : $G(t+1)$

if $P = \emptyset$ **then**
  for each u in $G(t)$ **do**
    $(x,y) \leftarrow random_position(MAX_X, MAX_Y)$ ▷ give a random position in the grid
    $G(t).update_nodes(u,(x,y))$
  end for
else
  for each couple $(i,(x_i,y_i))$ in P **do**
    $G(t).update_nodes(i,(x_i,y_i))$
  end for
end if
$L \leftarrow getListNodesWithposition(G(t))$ ▷ This method return a set of couple $(i,(x_i,y_i))$
while $L \neq \emptyset$ **do**
  $(u,(x_1,y_1)) \leftarrow L.pop()$
  for each couple $(v,(x_2,y_2))$ in L **do**
    $d \leftarrow euclidean_distance((x_1,y_1),(x_2,y_2))$ ▷ Compute euclidean distance
    if $d \leq d_{th}$ **then**
      $weight \leftarrow d$
      $G(t).update_edges(u,v,weight)$
    end if
  end for
end while
$G(t+1) \leftarrow G(t)$

Algorithm 2 Temporal network

INPUT : $:mob$ - Mobility model N - Network size,
MAX_X, MAX_Y - $MAX_X \times MAX_Y$ grid size,
T - max time step
OUTPUT : : $\mathcal{G}$ - $\{G(t1),...,G(T)\}$

$t \leftarrow t_0$
$\mathcal{G} \leftarrow List[]$
while $t \leq T$ **do**
  if $t = t_0$ **then**
    $G(t) < -create_new_graph(N)$
    $P \leftarrow \{\}$
  else
    $P \leftarrow getNodesPositionsFromMobility(mob)$ ▷ Give postions of nodes, $\{(i,(x,y)) : i = 1...N\}$, from mobility model
  end if
  $G(t+1) \leftarrow spatial_weighted_graph(G(t),n,P,d_{th},t)$
  insert $G(t+1)$ into $\mathcal{G}$
  $t \leftarrow t+1$
end while

TABLE 4.1: Paramètres de simulation

Paramètres	**Valeurs**	**Description**
MAX_X,MAX_Y	100, 100	Grille de simulation de taille $MAX_X \times MAX_Y$
N	[100, 200, 500, 1000]	Nombre de nœuds
T	200	Valeur maximale du sequence des pas de temps $\{1,2,3,\ldots,T\}$
d_{th}	4 m	Distance maximale pour que deux individus soient considérer en contacts
wt_{max}	5. s	Durée maximale de pause de la mobilité *RWP*
$[v_{min},v_{max}]$	0.1, 1 (m/s)	L'intervalle de vitesse des nœuds suivant *RWP*
v_{mean}	1	Vitesse moyenne du mobile dans la mobilité *GM* (μ_v)
μ_θ	1	La variance du paramètre de direction dans *GM*
α_{GM}	0.2	Paramètre de calibrage du modèle *GM* pour faire varier le caractère aléatoire
fl_{exp}	-2.6	Exposant de la loi puissance permettant de déterminer la distance des *flights* dans *TLW*
wt_{exp}	-1.8	Exposant de la loi puissance permettant de déterminer la durée des pauses dans le modèle *TLW*
fl_{max}	50 m	Distance maximale d'un *flight* dans la mobilité *TLW*
wt_{max}	100. s	Durée maximale du temps de pause de *TLW*
α_{STEPS}	2.5	Exposant de la loi de puissance de la distribution permettant de choisir une distance d à partir d'une zone Z_i dans *STEPS*
zone_size	10 m	Taille d'une zone Z_i carrée
rwp_speed	[0.83, 1,38] (m/s)	Vitesses min et max des nœuds suivant RWP dans une zone dans le modèle STEPS
rwp_pause_time	[1, 5]	Intervalles du temps de pause *STEPS*
zone_speed	[0.83, 3,6] (m/s)	Vitesses min max des nœuds pour les déplacements inter-zones
zone_time	[20, 30] s	Intervalle de temps représentant la durée d'un agent mobile sur une zone Z_i

considérablement plus étendue que celle des gouttelettes, l'infection peut se produire pendant des heures après que la personne infectée ait occupé une pièce et le taux de transmission par voie aérienne dépend fortement de l'environnement. La prise en compte de ces facteurs nous conduit à intégrer le mouvement des individus et leurs interactions dans le processus de propagation. C'est dans cette optique que nous faisons appel aux modèles de mobilité afin de pouvoir générer des graphes temporels qui mettront en évidence ces éléments. Dans un premier temps, nous allons effectuer un zoom sur la diffusion spatiale des nœuds en étudiant leur *déplacement quadratique moyen* (*Mean Squared Displacement, MSD*) considérant les différents de type de mobilité. Ensuite, nous allons effectuer une analyse statistique des caractéristiques structurelles et topologiques de ces graphes temporels dans le but de comprendre l'évolution de leur dynamique interne (modèle épidémiologique SIS dans notre cas).

4.4 *Déplacement Quadratique Moyen* (MSD)

Comme nous l'avons évoqué dans la partie 3.3.1.2, le déplacement quadratique moyen (MSD) permet de quantifier la dispersion des individus dans l'espace par rapport à leur origine. Nous rappelons que le MSD permet souvent de distinguer trois types diffusions (*normal-*

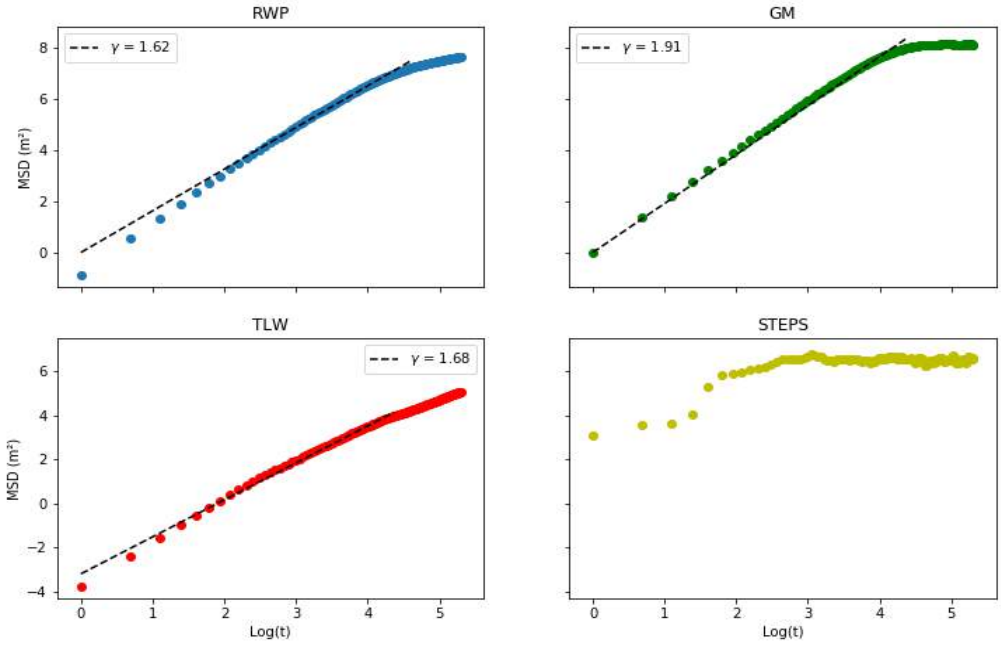


FIGURE 4.2: *Déplacement Quadratique Moyen* (MSD) des différentes modèles de mobilités utilisés

diffusion, *sub-diffusion* et *super-diffusion*) caractérisée par une fonction puissance dépendant du temps, $\boldsymbol{MSD(t) \propto t^{\gamma}}$. Dans le cas d'une diffusion normale, le MSD est une fonction linéaire telle que $\gamma = 1$. La super-diffusion, comme nous l'avons constaté dans la plupart des traces humains, est un processus plus rapide que la diffusion normale, elle est caractérisée par une valeur de $\gamma > 1$. Enfin, la sub-diffusion ou la diffusion ultra-faible se déduit par un exposant $\gamma < 1$. Avec les paramètres de simulations décrites dans le tableau 4.7, nous avons étudié la nature de la diffusion en utilisant la formule de MSD de la section 3.3.1.2. Dans la figure 4.2, nous constatons la nature de super-diffusion sur les trajectoires *RWP*, *GM* et *TLW*. En utilisant une échelle logarithmique, nous avons pu remarquer une linéarité du MSD(t) révélant des valeurs de $\gamma > 1$. Par contre, les résultats obtenus sur le modèle *STEP* ne nous permettent pas de déduire le comportement de la diffusion des agents mobiles. La Figure 4.2 montre en même temps les résultats de γ que nous avons trouvé (RWP :$\gamma = 1,62$, GM :$\gamma = 1,98$, TLW :$\gamma = 1,68$). D'autres études similaires montrent que ces valeurs peuvent être influencées par les paramètres de mobilité utilisés. Par exemple, Kim et *al.* montrent qu'il existe une forte corrélation entre la direction du mobile θ_n et le paramètre d'échelle α_{GM} sur la valeur de γ dans le modèle *GM* [KLE10b]. L'exposant fl_{exp} de la distribution des pas de déplacement (*flight-length*) du modèle TLW permet aussi de contrôler cette valeur [KLE10b].

4.5 Analyse topologique des propriétés des réseaux de contacts temporels obtenus

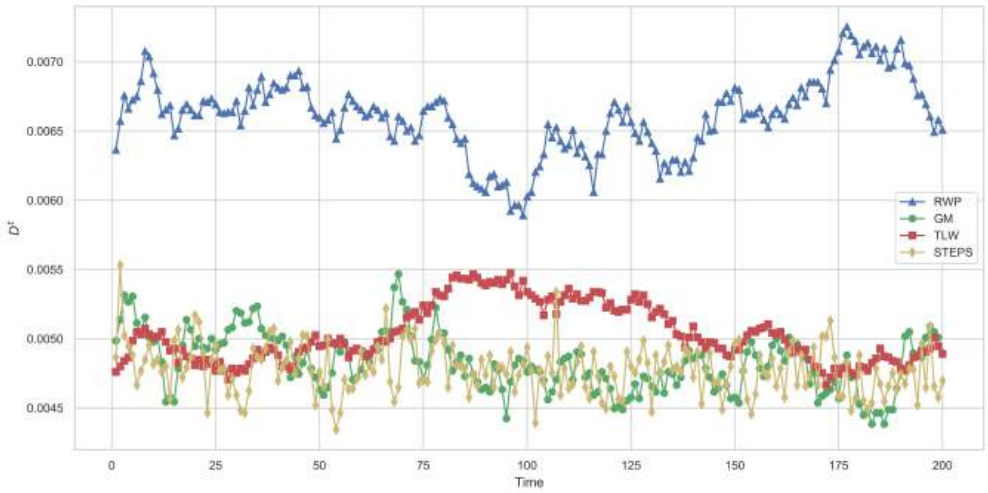
Nous avons généré plusieurs traces de contacts dynamiques à partir de différents modèles de mobilité, afin de mieux comprendre les caractéristiques des réseaux obtenus. Pour chaque type de mobilité, nous avons remarqué que les paramètres utilisés et la taille ont une influence considérable sur le mouvement des agents et leurs interactions. Ce qui motive l'étude de certaines propriétés permettant d'analyser la nature des réseaux temporels obtenus. En comparaison avec les caractéristiques des graphes temporels aléatoires générés, nous avons étudié simultanément quelques réseaux de contacts de traces humaines. Ces traces ont été récoltées dans cadre du projet *Sociopattern* avec des badges RFID tels que deux individus sont considérés en contact s'ils sont face à face durant un intervalle de **20** secondes. Nous avons choisis deux *datasets* : le premier contient des données de contacts entre les élèves et les enseignants dans une école primaire [GBC14], le deuxième est une liste de contacts dynamiques entre des étudiants d'une classe préparatoire collectée pendant 5 jours [MFB15].

4.5.1 Densité et composantes connexes

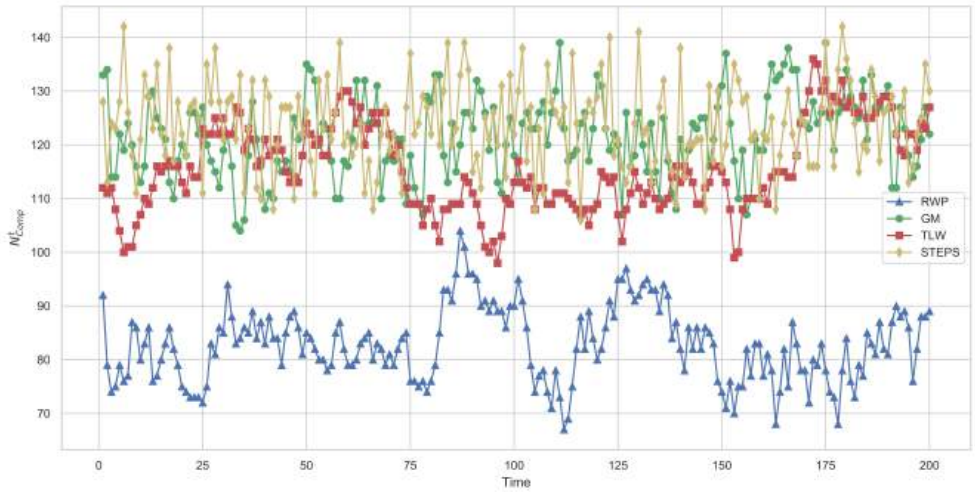
Rappelons d'abord que la densité rencontrée dans les graphes statiques permet d'avoir une vue sur la structure globale du graphe, c'est à dire si le graphe est essentiellement épars ou dense. Dans les réseaux temporels, la notion de densité peut être importante si l'on considère qu'un réseau temporel est une succession de *snapshots* de graphes statiques. La densité devient dès lors un élément pertinent pour étudier la variation de la structure d'un graphe dynamique qui dépend des facteurs de mobilité. Cette mesure de densité sera notée $\boldsymbol{D^t}$ dans le reste du document. La première observation faite sur les résultats obtenus de la Figure 4.3 montre que tous les graphes temporels générés sont éparses et présentent chacun un nombre de composants connexes, $\boldsymbol{N^t_{Comp}}$, très élevé (voir Figure 4.3b). Ce qui nous montre qu'il existe une forte déconnexion dans les graphes temporels obtenus. Dans la figure 4.3a, nous observons une densité plus forte et un nombre faible de composants dans le réseau temporel généré par *RWP*, $\mathscr{G}_{RWP}$, par rapport aux autres graphes générés. Par conséquent, nous pouvons en déduire que le graphe $\mathscr{G}_{RWP}$ est plus connexe que $\mathscr{G}_{GM}$, $\mathscr{G}_{TLW}$ et $\mathscr{G}_{STEPS}$. En général, cette connectivité importante peut être justifiée du fait que les agents mobiles suivant le modèle *RWP* tendent à se concentrer souvent au milieu de l'espace de simulation au fil du temps comme le montre Bettstetter dans [BRS03].

Nous avons aussi constaté dans la Figure 4.4 qu'il existe une forte corrélation entre la densité $\boldsymbol{D^t}$ et le nombre de composantes connexes $\boldsymbol{N^t_{Comp}}$ dans chaque TVG généré. Le coefficient de corrélation de Pearson [2] calculé démontre qu'il existe une corrélation négative entre ces deux variables dans les différents réseaux temporels $\mathscr{G}_{RWP}$, $\mathscr{G}_{GM}$, $\mathscr{G}_{TLW}$, $\mathscr{G}_{STEPS}$ (voir Tableau 4.2). Plus la densité augmente, plus le nombre de composantes devient faible avec $\boldsymbol{N^t_{Comp}} \longrightarrow \boldsymbol{1}$. Dans $\mathscr{G}_{TLW}$, nous remarquons un degré de corrélation plus significatif ($|\boldsymbol{Pearson_{D^t,N^t_{Comp}}}| > 0,5$).

2. $r_{pb} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum(x-m_x)(y-m_y)}{\sqrt{\sum(x-m_x)^2(y-m_y)^2}}$



(a) Densité $\boldsymbol{D^t}$



(b) Nombre de composantes $\boldsymbol{N^t_{Comp}}$

FIGURE 4.3: Évolution temporelle de la densité et du nombre de composantes connexes

En outre, nous nous sommes aussi intéressés au comportement de la densité, $\boldsymbol{D^t}$ des graphes temporels générés par rapport à l'évolution de la taille $\boldsymbol{N}$. Si nous considérons un environnement fermé sans obstacle, il est évident que le nombre d'interactions par individu augmente en fonction de la taille de la population. Pourtant, malgré l'augmentation de la taille, nous observons une fluctuation de la densité $\boldsymbol{D^t}$ autour d'une valeur moyenne dans tous les TVGs. Par exemple dans la Figure 4.5, lorsque $\boldsymbol{N}$ égal à 100 nœuds nous avons remarqué dans le graphe $\mathcal{G}_{\boldsymbol{RWP}}$ une évolution non homogène de la densité $\boldsymbol{D^t}$ entre les valeurs 0.5% à 0.10% et d'une stabilité lorsque la taille du graphe augmente à partir de 500 nœuds avec un pourcentage de 0.7% . Ce même phénomène est constaté sur les graphes temporels $\mathcal{G}_{\boldsymbol{GM}}$ et $\mathcal{G}_{\boldsymbol{STEPS}}$ avec une densité moyenne de 0.5% (Figure 4.6 et 4.8). Par contre, dans $\mathcal{G}_{\boldsymbol{TLW}}$, nous avons noté une instabilité de l'évolution de la densité en fonction du nombre de nœuds (Figure 4.7). Les Figures 4.9a et 4.9b résument l'évolution de la moyenne et de la variance

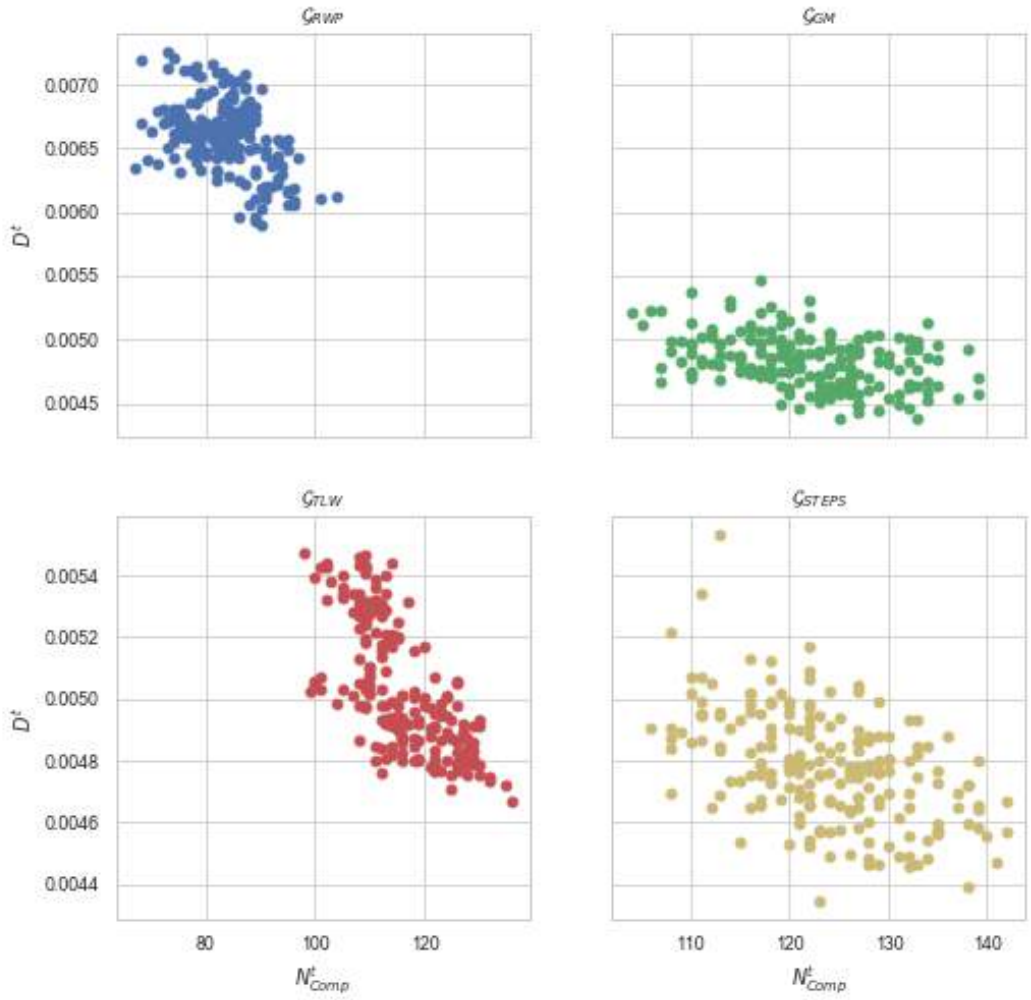


FIGURE 4.4: Corrélation entre la densité et le nombre de composantes connexes

Réseau temporel	$Pearson_{D^t,N^t_{Comp}}$	p-value
$\mathcal{G}_{RWP}$	-0.449	2.41 e-11
$\mathcal{G}_{GM}$	-0.392	8.94 e-9
$\mathcal{G}_{TLW}$	-0.694	4.27 e-30
$\mathcal{G}_{STEPS}$	-0.493	1.12 e-13

TABLE 4.2: Mesures de corrélation entre D^t et N^t_{Comp}

(Écart-type) de D^t dans tous les réseaux temporels qui ont été générés à partir de modèles de mobilité *RWP*, *GM*, *TLW* et *STEPS*. Malgré l'observation d'une moyenne de densité D^t presque constante, nous constatons une décroissance de la variance en fonction de la taille du réseau. Ainsi, plus la taille augmente, plus la densité évolue de manière homogène. La densité des réseaux générés de taille faible évolue progressivement avec une large variance. Cependant, cette variance est plus faible dans les graphes $\mathcal{G}_{TLW}$ et $\mathcal{G}_{STEPS}$ (voir Figure 4.9b).

Contrairement aux réseaux de contacts aléatoires, la densité de contacts réels est marquée

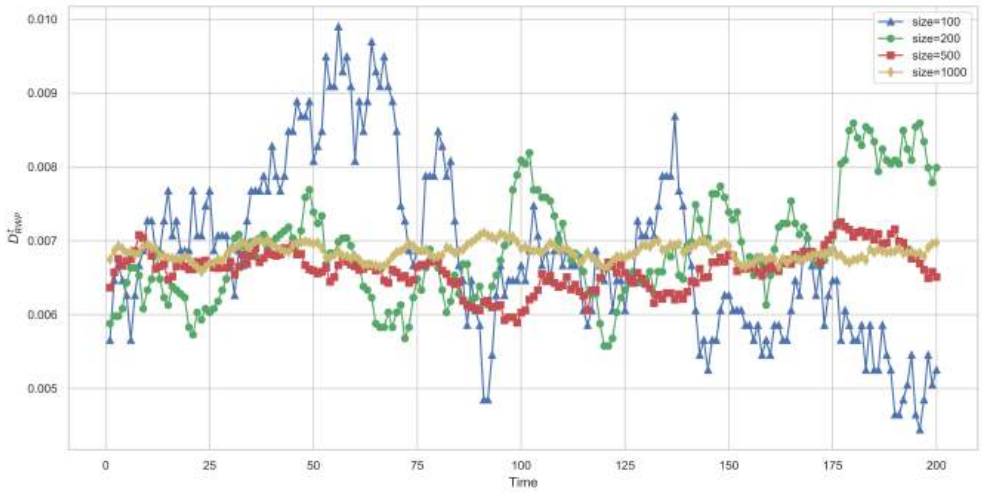


FIGURE 4.5: Évolution de $\boldsymbol{D^t}$ dans le graphe $\mathscr{G}_{\boldsymbol{RWP}}$

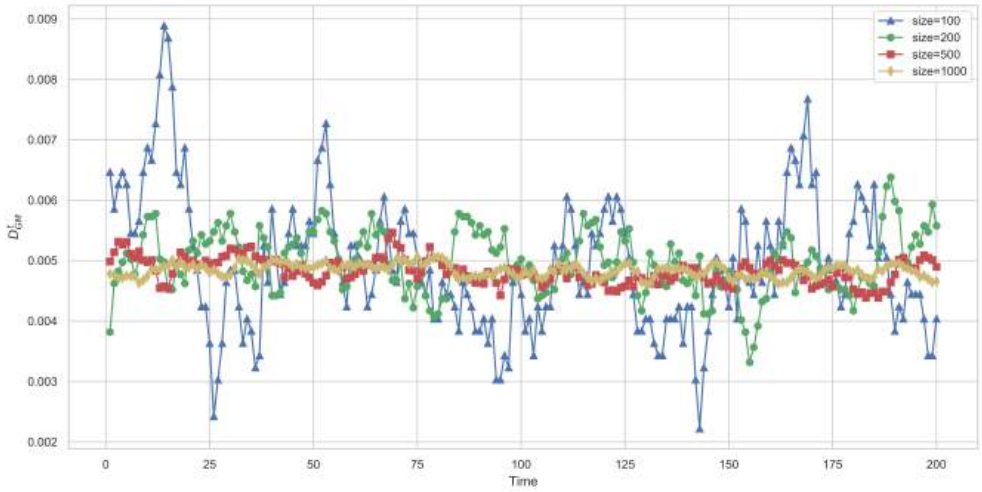


FIGURE 4.6: Évolution de $\boldsymbol{D^t}$ dans le graphe $\mathscr{G}_{\boldsymbol{GM}}$

par la présence des motifs réguliers présentant des pics durant les heures de pauses (10h - 12h), des ruptures marquant les heures de descente, d'une périodicité des fréquences des contacts comme le montrent les données de contacts de la Figure 4.10.

4.5.2 Chemin temporel et accessibilité

La notion d'accessibilité dans les réseaux temporels est cruciale dans la mesure où elle permet de savoir comment une épidémie se propage dans le réseau, quelle est la rapidité du processus de propagation. Dans la section 2.5.6.1, nous avons discuté en détail de la différence qui existe entre le chemin temporel et le chemin dans les réseaux statiques. Nous avons aussi montré que la condition d'accessibilité *(reachability)* dans les réseaux temporels dépendait surtout de l'ordre chronologique des contacts entre les individus (*time-dependent*).

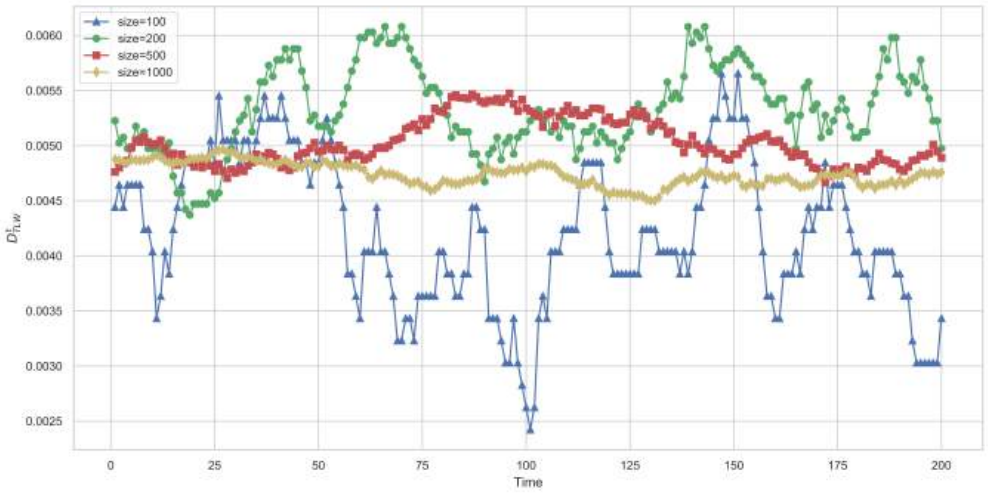


FIGURE 4.7: Évolution de D^t dans le graphe $\mathcal{G}_{TLW}$

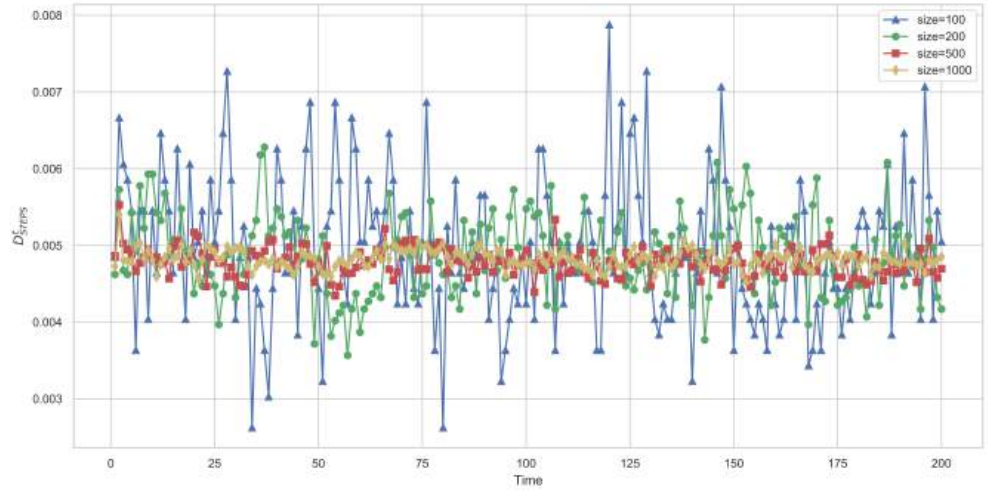
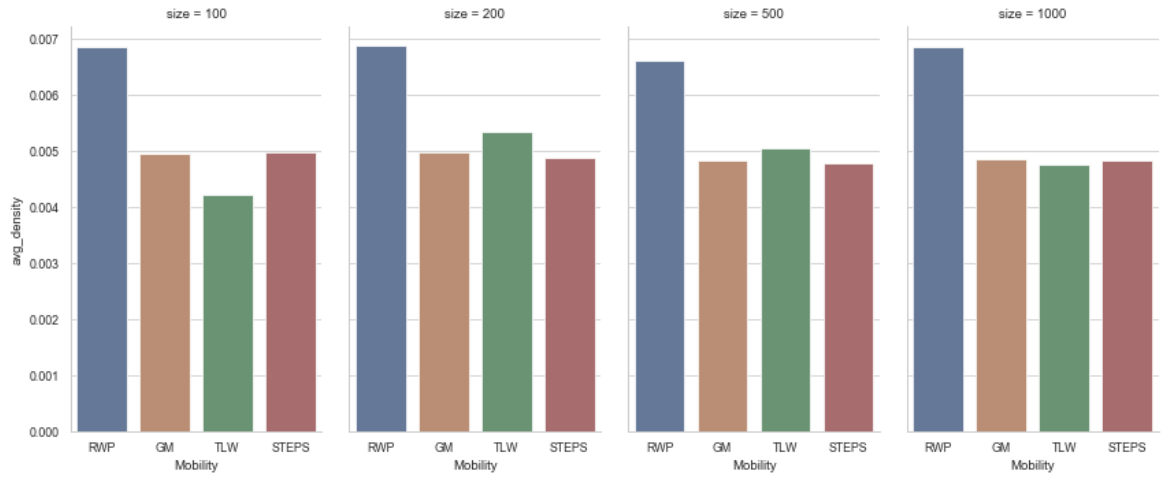


FIGURE 4.8: Évolution de D^t dans le graphe $\mathcal{G}_{STEPS}$

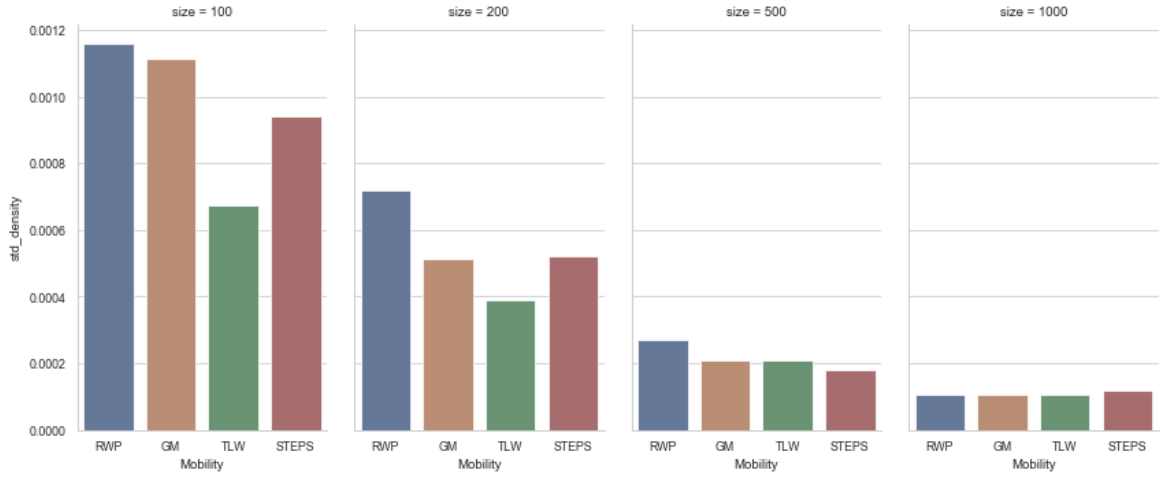
Contrairement aux réseaux statiques, l'accessibilité temporelle est mesurée en unité de temps et est équivalente à la durée estimée pour atteindre les nœuds du réseau. Dans [Hol05], Holme définit la notion de ratio d'accessibilité *(reachability ratio)* et de temps d'accessibilité *(reachability time)* comme étant successivement le pourcentage d'arêtes reliées par un chemin temporel et la durée moyenne de tous les chemins temporels. Dans cette partie, nous allons utiliser la définition de *Reachability Latency* proposée par Thompson et *al.* [TBF17] qui n'est rien d'autre que la durée moyenne pour atteindre une portion r de nœuds dans le graphe temporel.

$$R_r = \frac{1}{TN} \sum_t \sum_i d^t_{i,k} \tag{4.1}$$

où T représente la valeur maximale d'une séquence de *timestep*, $\{1,2,3,\ldots,T\}$, d^t_i est un



(a) Moyenne



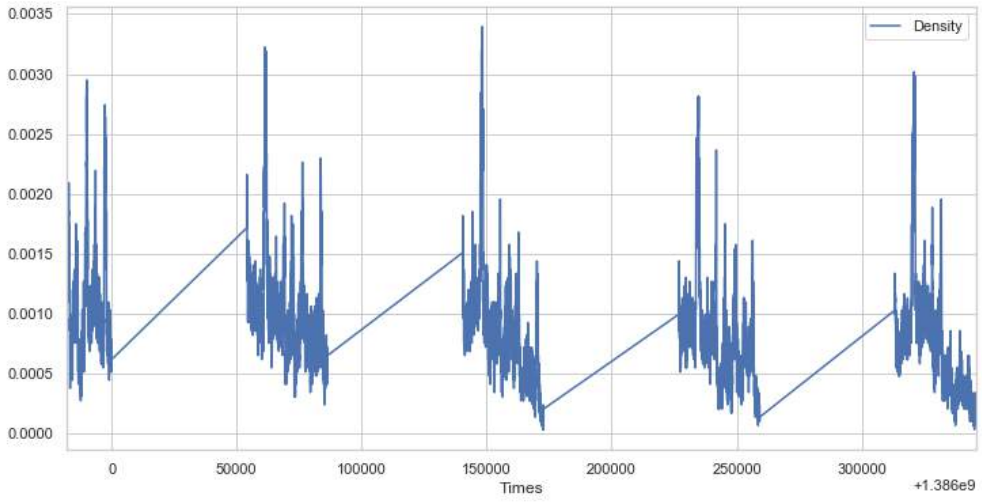
(b) Ecart-type

FIGURE 4.9: Moyenne et Écart-type de la densité D^t des graphes temporels en fonction de la taille N

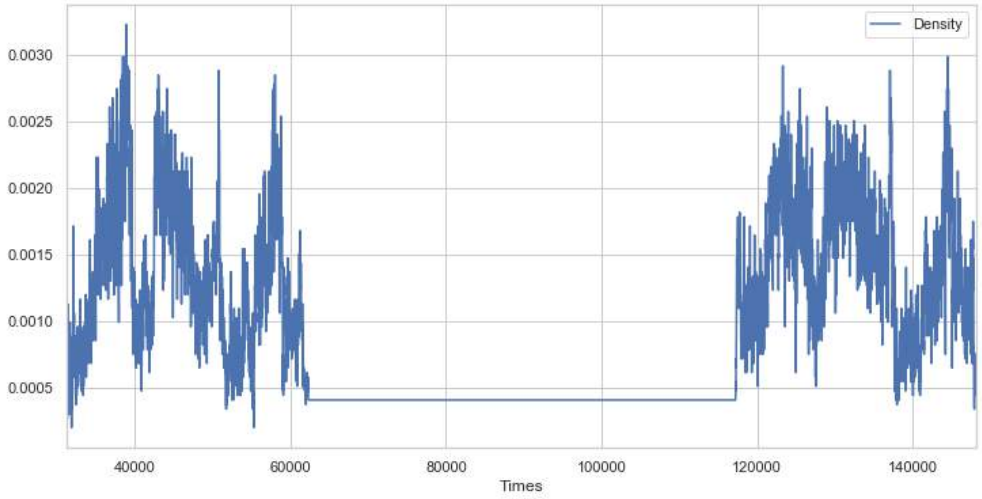
vecteur ordonné de taille N contenant les plus courts chemins temporels vers le nœud i à l'instant t. La valeur k représente le $\lfloor rN \rceil$ ième élément du vecteur, qui est la partie entière du produit entre la fraction r de nœuds pouvant être atteinte et le nombre de nœuds N du réseau temporel. Avec un ratio $r = 1$, la formule 4.1 devient le diamètre temporel du réseau [TBF17]

$$R_d = \frac{1}{TN} \sum_t \sum_i maxjd_{i,j}^t \tag{4.2}$$

Par conséquent, cette dernière formule nous renseigne sur l'accessibilité globale des nœuds du réseau. Nous avons calculé la *Reachability Latency* pour $r = 1$ c'est à dire les diamètres temporels des différents graphes temporels selon l'évolution de leur taille N (Voir Figure 4.11). Les résultats obtenus montrent que la mobilité joue un rôle très important dans l'accessibilité aux nœuds du réseaux. Dans 4.11, les nœuds suivant la mobilité STEPS présentent le plus faible diamètre temporel comparé aux autres graphes générés. Par exemple, dans le modèle STEPS 100% des nœuds peuvent être atteints sur un temps $R_d = 45,33$ lorsque le nombre de nœuds est égal à 200 comparé aux nœuds qui se déplacent en RWP où $R_d = 96,57$. En plus, nous avons aussi remarqué que le coefficient de *Reachability Latency* décroit en fonction de la



(a) Densité de contacts des étudiants d'une classe préparatoire (*High School*, HS) collecté pendant 5 jours [MFB15]



(b) Densité du graphe temporel de contacts obtenu à partir des données provenant des travaux de [GBC14] (*Primary School*, PS)

FIGURE 4.10: Densité temporelle des graphes de contacts réels

taille des individus en interaction dans certains cas de mobilité ($\mathcal{G}_{RWP}$, $\mathcal{G}_{GM}$, $\mathcal{G}_{STEPS}$). En revanche, nous constatons de manière étrange, dans le cas des nœuds suivant *TLW*, que la valeur de R_d augmente lorsque $N = 200$ et $N = 500$ au lieu de diminuer.

4.5.3 Coefficient de corrélation temporelle et l'efficience temporelle

Comme nous l'avons constaté dans la plupart des réseaux réels, l'effet "*petit monde*" est aussi d'actualité dans les graphes temporels. Cet phénomène est quantifié par une importante valeur du coefficient de clustering et une faible valeur de la moyenne des longueurs (*Average path lenght*) ou de l'Efficience globale dans les réseaux complexes statiques. Dans les réseaux temporels, le concept de clustering peut être remplacé selon Tang et *al.* par le coefficient de

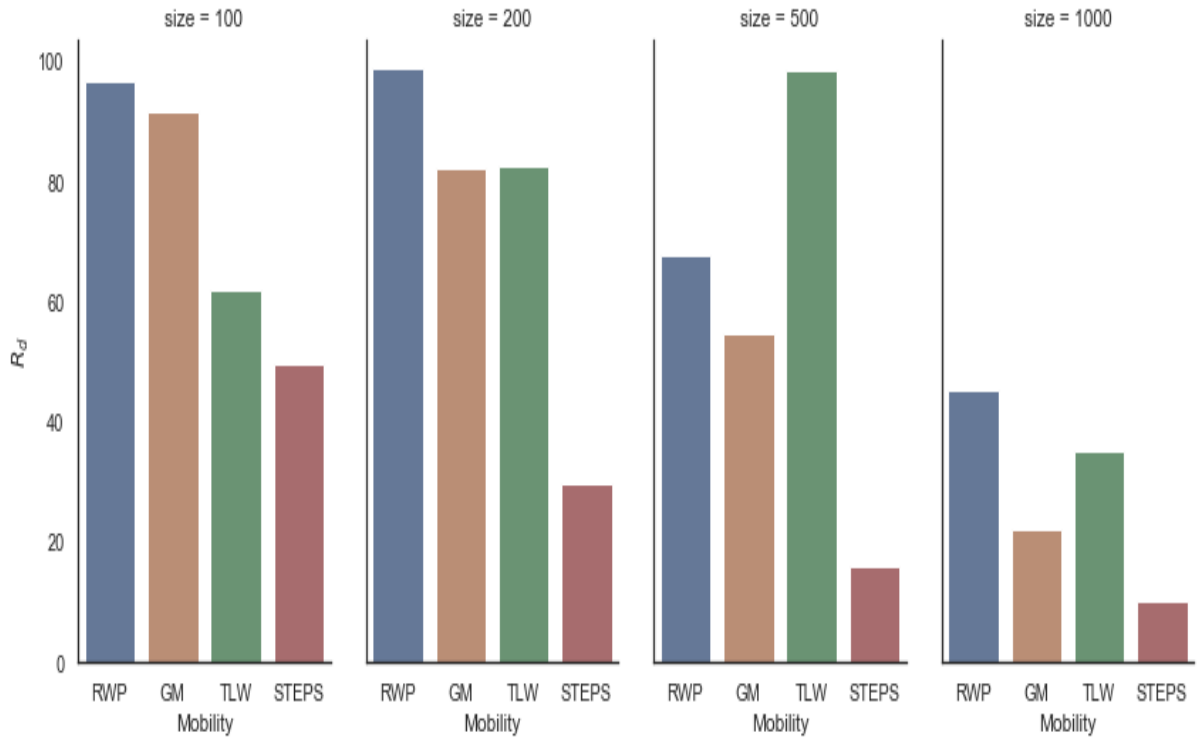


FIGURE 4.11: Reachability Latency

corrélation temporel *(Temporal Correlation Coefficient)* noté $\mathscr{C}$. Il nous permet de comprendre comment les nœuds sont groupés dans le réseau temporel et l'efficience temporelle *(Temporal Efficiency)* $\mathscr{E}$ permet de capturer la longueur de la moyenne des plus courts chemins temporels. Le Tableau 4.3 montre les mesures de coefficient de corrélation temporel et l'efficience temporelle dans les graphes aléatoires générés en comparaison avec celles obtenues dans les graphes de contacts des données de Sociopattern.

Nous constatons dans l'histogramme 4.12 correspondant aux mesures du Tableau 4.3 que l'effet small-world est bien présent dans les graphes $\mathscr{G}_{RWP}$, $\mathscr{G}_{GM}$, $\mathscr{G}_{TLW}$ avec d'importantes valeurs du coefficient de corrélation temporelle ($\mathscr{C}_{RWP} = 0.835$, $\mathscr{C}_{GM} = 0.715$, $\mathscr{C}_{TLW} = 0.873$) et de faibles valeurs de l'efficience temporelle ($\mathscr{E}_{RWP} = 0.047$, $\mathscr{E}_{GM} = 0.033$, $\mathscr{E}_{TLW} = 0.014$). Par contre, dans le graphe de contacts dynamiques générés par le modèle de STEPS, nous avons un faible coefficient de corrélation et une valeur significative de l'efficience temporelle ($\mathscr{C}_{STEPS} = 0.351$, $\mathscr{E}_{STEPS} = 0.111$).

Comparées aux données temporels générés, les données de Sociopattern, PS (*Primairy School*) et HS (*High School*), révèlent que de très faibles valeurs du coefficient de corrélation temporelle $\mathscr{C}$ et l'efficience temporelle $\mathscr{E}$ (voir Tableau 4.3). Ce constat est valable dans d'autres jeux de données réelles [BSK16]. Cependant, Tang *(al.)*, en étudiant l'effet "*petit monde*" dans [Tan+10a], montrent que les données de contacts récoltées dans *INFOCOM'06* présentent d'importantes valeurs de $\mathscr{C}$ et de $\mathscr{E}$. Par exemple, avec les données de contacts récoltées dans des conférences qui ont pu se dérouler durant quatre jours les auteurs ont trouvés respectivement les résultats suivants : $\mathscr{C}_{d1} = 0.80$ et $\mathscr{E}_{d1} = 0.192$; $\mathscr{C}_{d2} = 0.78$ et $\mathscr{E}_{d2} = 0.293$; $\mathscr{C}_{d3} = 0.81$ et $\mathscr{E}_{d3} = 0.134$; $\mathscr{C}_{d4} = 0.83$ et $\mathscr{E}_{d4} = 0.019$ [Tan+10a].

TABLE 4.3: Mesures de coefficient de corrélation temporelle $\mathscr{C}$ et l'efficience temporelle $\mathscr{E}$

TVG	$\mathscr{C}$	$\mathscr{E}$	$\|V\|$
$\mathscr{G}_{RWP}$	0.835	0.047	500
$\mathscr{G}_{GM}$	0.715	0.033	500
$\mathscr{G}_{TLW}$	0.873	0.014	500
$\mathscr{G}_{STEPS}$	0.351	0.111	500
$\mathscr{G}_{PS}$	0.089	0.003	242
$\mathscr{G}_{HS}$	0.082	0.001	327

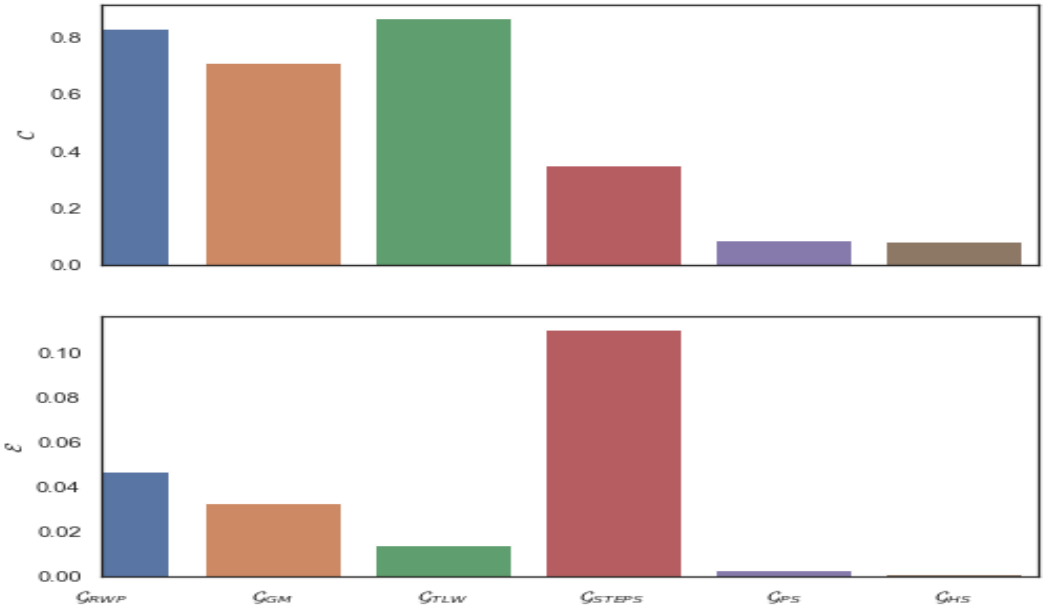


FIGURE 4.12: L'effet *petit monde*'dans les graphes temporels

4.5.4 Les mesures de centralité dans les réseaux temporels

Nous savons intuitivement que tous les individus n'ont pas le même impact sur la transmission d'une maladie ou la diffusion d'une information dans un réseau. Certains individus ont un rôle central du fait de leurs importantes connectivités (appelés *'hubs'*) ou bien selon leur capacité à faciliter l'accessibilité ou à faire passer l'information aux autres nœuds. Les mesures de centralité rencontrées dans la section 2.5.6.3 nous permettent de capturer l'importance de certains nœuds dans le processus de propagation dans les réseaux temporels.

4.5.4.1 Centralité temporelle de degré

En mesurant la centralité de degré dans les TVGs, nous constatons que les données obtenues sont concentrées autour d'une valeur moyenne ressemblant à une distribution gaussienne. Dans la figure 4.13, nous avons utilisé la méthode du *Kernel Density Estimation* (KDE), pour estimer la distribution de la centralité temporelle de degré. Le test de normalité D'Agostino

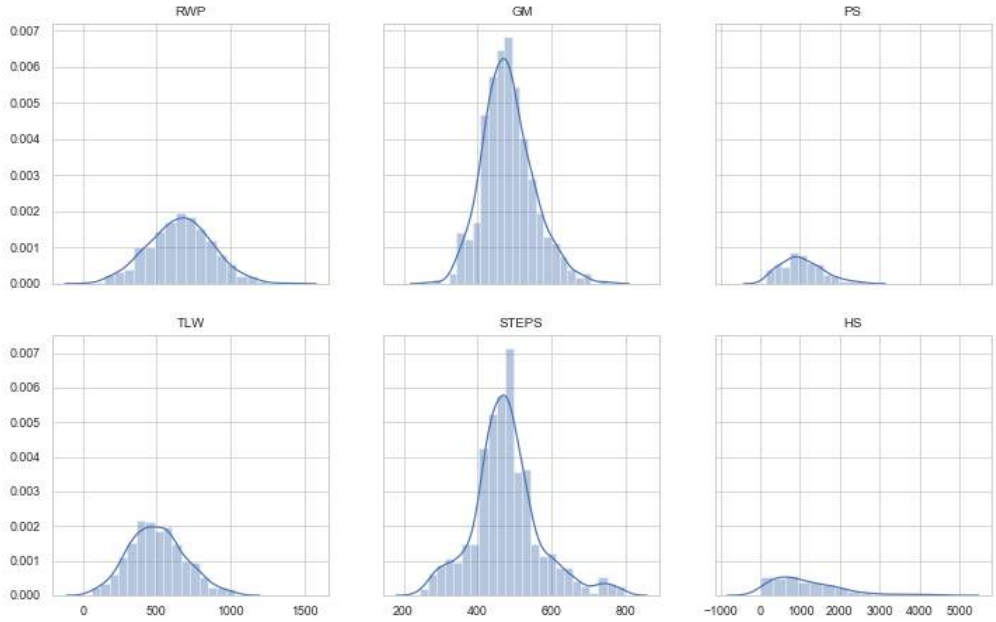


FIGURE 4.13: Centralité de temporelle de degré

et Pearson [3] utilisé nous montre que seules les données de $\mathcal{G}_{RWP}$ et $\mathcal{G}_{TLW}$ ont une distribution gaussienne avec $p-value = 0.890$ et $p-value = 0.067$ respectivement. Pour la distribution de centralité des graphes $\mathcal{G}_{GM}$ et $\mathcal{G}_{STEPS}$, nous avons effectué un *fitting* avec plusieurs fonctions de distribution de probabilité. Les résultats obtenus sont détaillés dans l'Appendix (Voir 6).

Comme nous l'avons constaté dans la Figure 4.13, la plupart des nœuds des graphes de contacts réels ($\mathcal{G}_{PS}$ et $\mathcal{G}_{HS}$) ont une centralité temporelle de degré très faible comparée à celle des graphes temporels aléatoires. Ce constat montre que les graphes aléatoires générés ont une capacité de diffusion plus rapide que les réseaux réels car les nœuds de fortes centralité joue le rôle de "super-diffuseur" (*super spreader*). Ce rôle de super-diffuseur de certains nœuds apparait plus en détails lorsqu'on étudie la distribution des paires de contacts c'est à dire le nombre fois qu'un individu $\boldsymbol{A}$ rencontre un autre individu $\boldsymbol{B}$ (Voir section 4.5.5).

4.5.4.2 La centralité temporel de proximité

Avec la méthode de KDE (*Kernel Density*), nous pouvons déduire intuitivement la nature de la distribution des mesures de centralité temporelle de proximité obtenus (Voir Figure 4.14). En utilisant le test de normalité, les distributions des données de $\mathcal{G}_{GM}$, $\mathcal{G}_{TLW}$ et $\mathcal{G}_{STEPS}$ ressemblent très bien à des distributions gaussiennes de moyennes $\mu_{GM} = 0.0369$, $\mu_{TLW} = 0.0185$ et $\mu_{STEPS} = 0.115$. Le *fitting* fait sur la courbe représentant les mesures de centralité de proximité obtenues sur graphe $\mathcal{G}_{RWP}$ montre que la densité de probabilité (2) ne suit pas la loi normale. De même, les mesures de centralité obtenues sur les graphes temporels de Sociopettern ne sont pas normalement distribuées.

3. **https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html**.

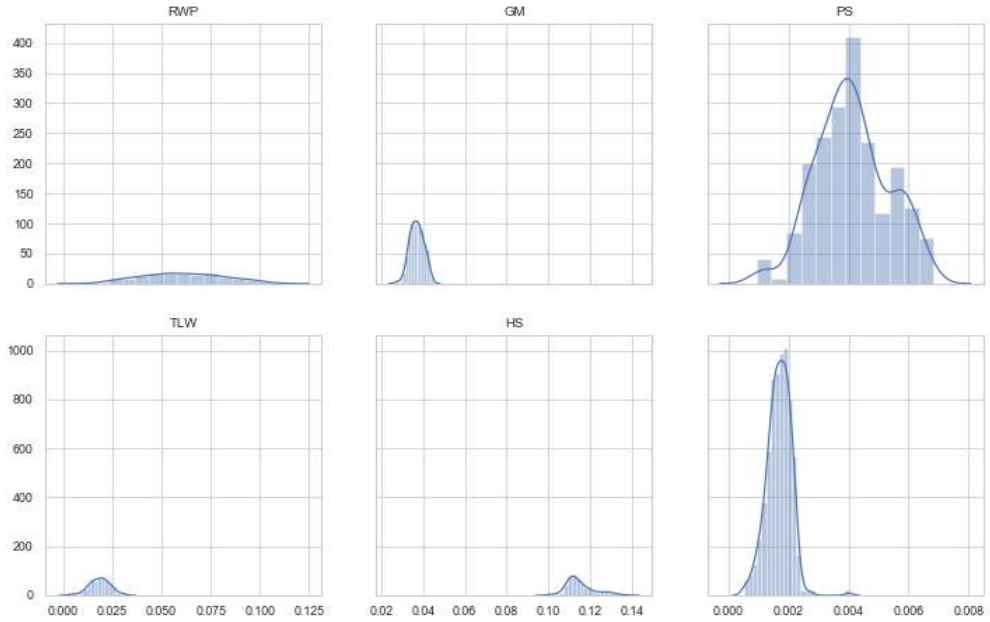


FIGURE 4.14: centralité temporelle de proximité

TABLE 4.4: Mesures de corrélation entre la centralité temporelle de degré (TCD) et la centralité temporelle de proximité (TCP)

Réseau temporel	Pearson(TCD, TCP)	p-value	$\|V_t\|$
$\mathcal{G}_{RWP}$	0.855	1.669e-144	500
$\mathcal{G}_{GM}$	0.449	2.756e-26	500
$\mathcal{G}_{TLW}$	0.741	2.252e-88	500
$\mathcal{G}_{STEPS}$	0.726	3.605e-83	500
$\mathcal{G}_{PS}$	0.678	6.121e-34	242
$\mathcal{G}_{HS}$	0.576	2.202e-30	327

Nous avons étudié curieusement la relation qui existe entre la centralité de degré et la centralité de proximité. Ainsi nous observons dans la figure 4.15 une forte corrélation entre ces deux variables dans les graphe temporels générés. Les coefficients de corrélation calculés sur les différents TVGs prouvent bien ce constat (voir Tableau 4.4). En bref, nous pouvons dire que les nœuds de forte centralité temporelle de degré ont tendance à faire passer plus de plus courts chemins. En d'autres termes, les nœuds de fort contact ont tendance à faire passer plus d'informations dans le réseau.

4.5.5 Occurrence des paires de contacts

Les mesures de centralité nous permettent de quantifier l'influence d'un nœud dans le réseau c'est à dire la capacité d'un nœud à diffuser rapidement l'information dans le réseau. Récemment, il a été démontré que les mesures de centralité ne révèlent pas toujours les super-diffuseurs (super spreaders) [Madotto2016]. Dans un processus de propagation, nous pensons qu'en plus du rôle de super-diffuseur de certains nœuds il faut surtout prendre

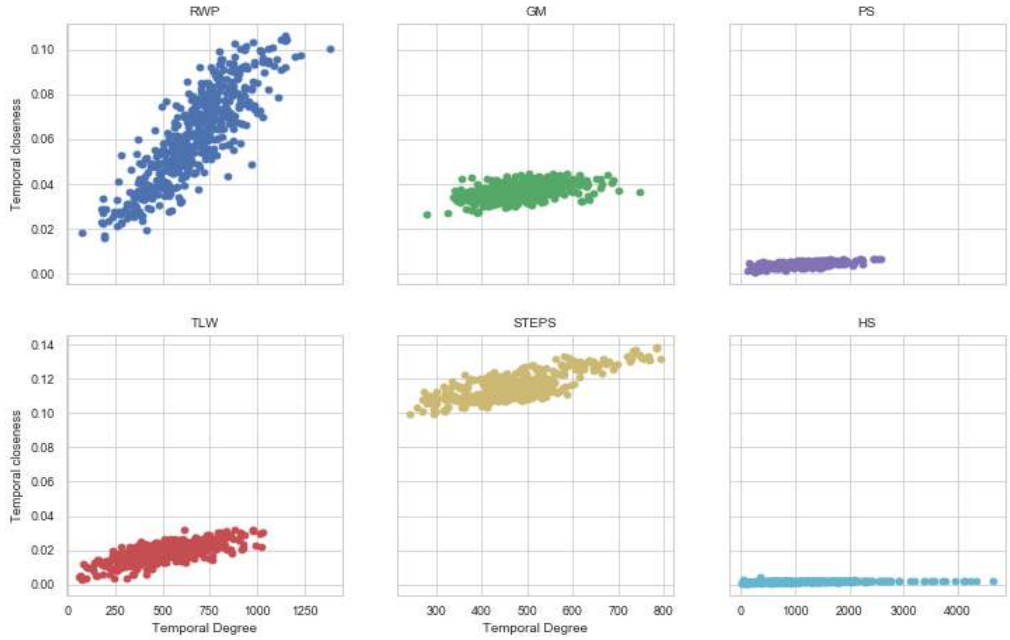


FIGURE 4.15: Corrélation entre la centralité de degré et la centralité de proximité

compte la fréquence des contacts entre les paires de nœuds c'est-à-dire dire paires contacts de grandes occurrences. Ces liens de grande occurrence peuvent révéler des informations très intéressantes par exemple dans le cas des réseaux sociaux temporels, les individus à forte occurrence de contacts peuvent représenter des relations sociales telle que les relations amicales, relations parentales, les relations professionnelles etc. Par conséquent, la probabilité d'infection augmente entre les individus qui entreprennent ces genres de contacts. Ainsi, nous avons étudié cette occurrence dans les graphes temporels générés ainsi que les graphes dynamiques obtenus à partir de traces. La Figure 4.16 nous montre de manière surprenante que la distribution des occurrences de tous les graphes temporels sont de type queue lourde (*heavy tailed*) c'est-à-dire si nous avons un coefficient d'aplatissement, C_k [4], qui est supérieur à celui de la loi normale (pour laquelle $C_k = 3$). Dans cette étude, nous avons utilisé la librairie Python *powerlaw* pour effectuer le *fitting* des distributions obtenues [ABP14]. Cette librairie permet d'analyser les distributions *heavy tailed*, notamment la loi puissance(*Power Law, PL*), la loi de puissance tronquée (*Truncated Power Law, TPL*), la loi exponentielle (*Exp*) ou avec une coupure exponentielle (*Stretched Exponential, S Exp*) ou la loi log-normal. Le tableau 4.5 donne les expressions mathématiques de ces distributions. Les résultats obtenus dans la Figure 4.16 montrent que les occurrences des paires de contacts réels sont toujours de loi puissance (avec une puissance $\alpha = 1.46$ pour PS et $\alpha = 1.43$ pour PS), tandis que cette loi puissance n'est pas toujours vérifier dans les données synthétiques comme dans le cas des individus suivant le modèle *STEPS* et le modèle *TLW*. En effet, cette fameuse loi de puissance explique que la plupart des individus entretiennent de fortes relations avec un nombre très limité qui peut déboucher sur différentes interprétations selon le contexte du graphe de contact étudié. Par

4. $E\left[\frac{X-\mu_X}{\sigma_X}\right] > 3$; avec X une variable aléatoire

TABLE 4.5: Quelques distributions à queue lourde (*heavy tailed*)

Nom	Expression de $p(x)$	Abbréviation
Loi puissance (*Power Law*)	$Cx^{-\alpha}$	*PL*
Loi de puissance tronquée (*Truncated Power Law*)	$Cx^{-\alpha}e^{-\lambda x}$	*TPL*
Exponentielle étirée (*Stretched Exponential*)	$Cx^{\beta-1}e^{-\lambda x^{\beta}}$	*S Exp*
Log-Normal	$\frac{1}{x\sigma\sqrt{2\pi}}\exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right]$	

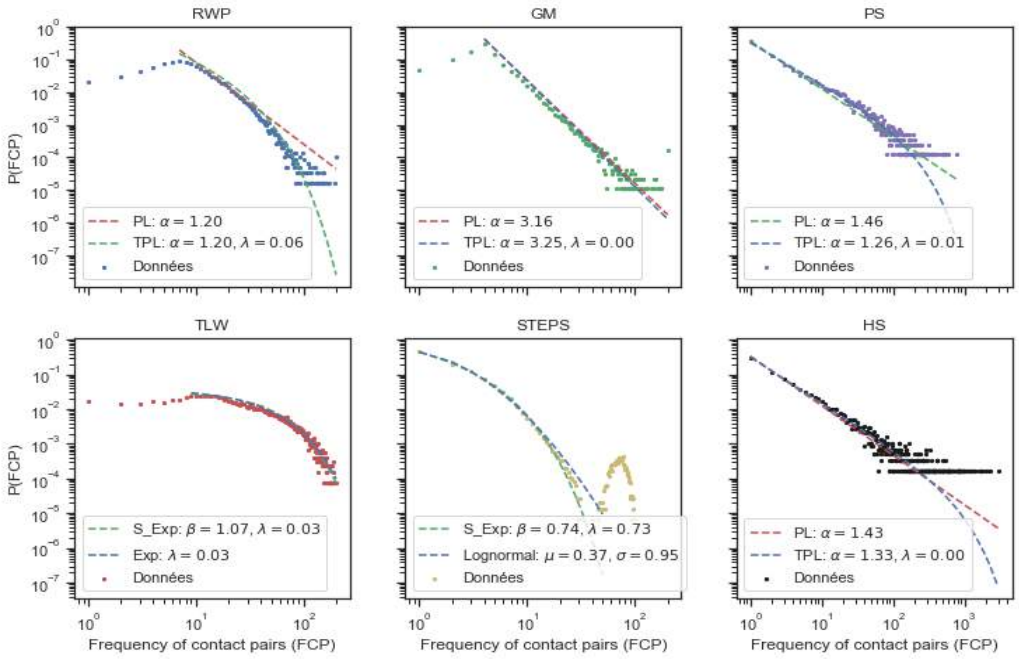


FIGURE 4.16: Distribution des occurrences de paires de contacts sur une échelle logarithmique

exemple dans le cas des données de Socioppatern, il peut s'agir des contacts entre deux élèves ou étudiants qui sont amis ou qui s'assoient toujours ensemble ou bien entre étudiants et professeurs.

4.5.6 Durée de contact (*Contact Time*) et Durée d'inter-contact (*Inter-Contact-Time*)

De même que la fréquence des paires de contacts, la durée de contacts pourrait constituer un élément essentiel dans le processus de contamination car elle augmente la probabilité d'infection. C'est la raison pour laquelle nous avons essayé de comprendre son comportement dans les différents types de graphes étudiés. Comme nous l'avons remarqué précédemment dans la fréquences de contacts des individus, la loi de puissance apparaît encore dans la durée des contacts entre les agents (voir figure 4.17). Dans la figure 4.17, tous les graphes temporels basés sur des données de contacts réels décrivent exactement cette même caractéristique. Par ailleurs, ces résultats sont similaires à ceux produits dans plusieurs études faites sur d'autres données de contacts réels [CLF07 ; AS09 ; Hon+08]. Cependant, nous observons que la distribution de la durée de contacts des données issues de la mobilité synthétique sont plus

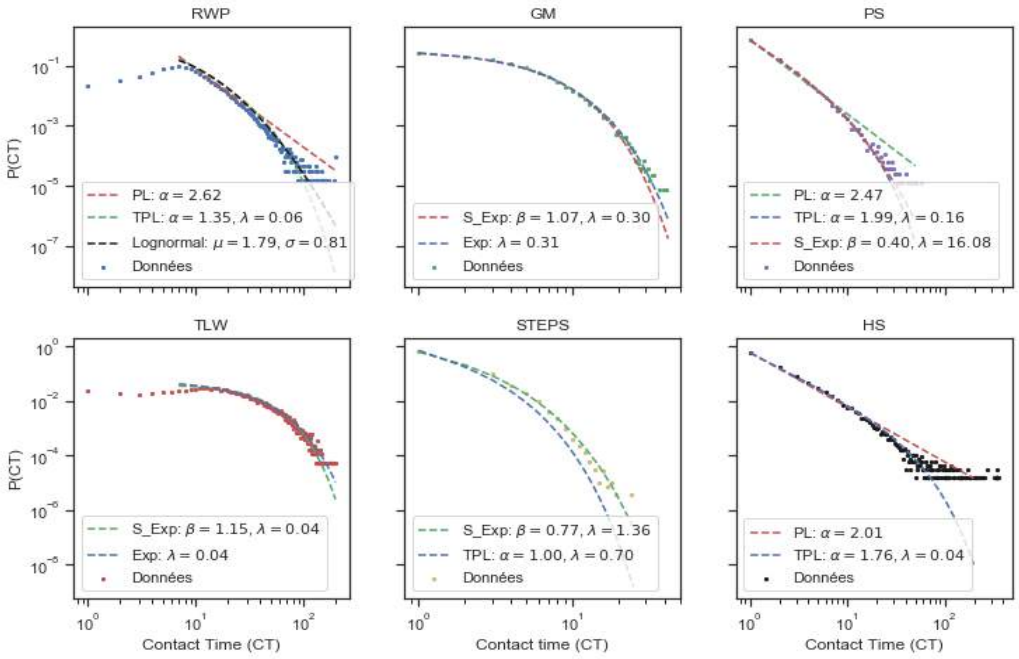


FIGURE 4.17: Distribution de la durée des contact sur une échelle log-log

proches de la loi *Stretched Exponential* et de *Log-normal* que la loi puissance (voir Figure 4.17).

La plupart des réseaux réels révèle un phénomène naturel appelé *'bursty'* [GB06 ; Kar+12 ; Váz+06 ; Bar+18]. Comme nous l'avons détaillé dans la section 2.5.6.4, il est caractérisé par une longue période d'inactivité des connexions suivie souvent d'une augmentation intermittente des connexions. Ce phénomène dépend du temps (*time-dependent*) et est caractérisé par une distribution de la durée d'inter-contacts ou d'inter-événements en loi puissance dans les réseaux temporels. Dans cette partie, nous avons étudié le phénomène de *bursty* sur les graphes temporels obtenus à partir de données synthétiques. Avec une échelle logarithmique, les résultats obtenus dans la figure 4.18 montrent que la distribution de la durée d'inter-contact des graphes $\mathcal{G}_{HS}$,$\mathcal{G}_{PS}$,$\mathcal{G}_{GM}$ et $\mathcal{G}_{STEPS}$ peut être assimilée à une loi de puissance (PL) ou une loi de puissance tronquée (TPL). Par contre, la distribution observée supr les TVGs $\mathcal{G}_{RWP}$ et $\mathcal{G}_{TLW}$ sont plus proches de la loi log-normal (voir 4.18).

De manière alternative, la nature de *burstiness* peut être facilement déduite à partir de la mesure du coefficient de burstiness B ou du coefficient de variation locale LV. Ainsi, nous avons calculé ces deux coefficients dans les paires de contacts de chaque graphe temporel. Rappelons d'abord qu'un phénomène de *bursty* est caractérisé par un coefficient $B \longrightarrow 1$ ou une valeur de LV comprise entre 1 et 3 (voir section 2.5.6.4). La Figure 4.19 confirme une phénomène de bursty dans $\mathcal{G}_{STEPS}$, $\mathcal{G}_{HS}$ et $\mathcal{G}_{PS}$ avec une concentration des points ayant un coefficient $B > 0$ et $1 < LV \leq 3$. Dans les autres cas, si nous interprétons les résultats sur l'axe LV, nous constatons que la plupart des mesures obtenues sur les graphes $\mathcal{G}_{RWP}$, $\mathcal{G}_{GM}$ et $\mathcal{G}_{TLW}$ sont très proches de 0 montrant une situation qui semble périodique. Cependant, nous distinguons plusieurs valeurs du coefficient de *bursty* qui sont positifs. Les résultats obtenus

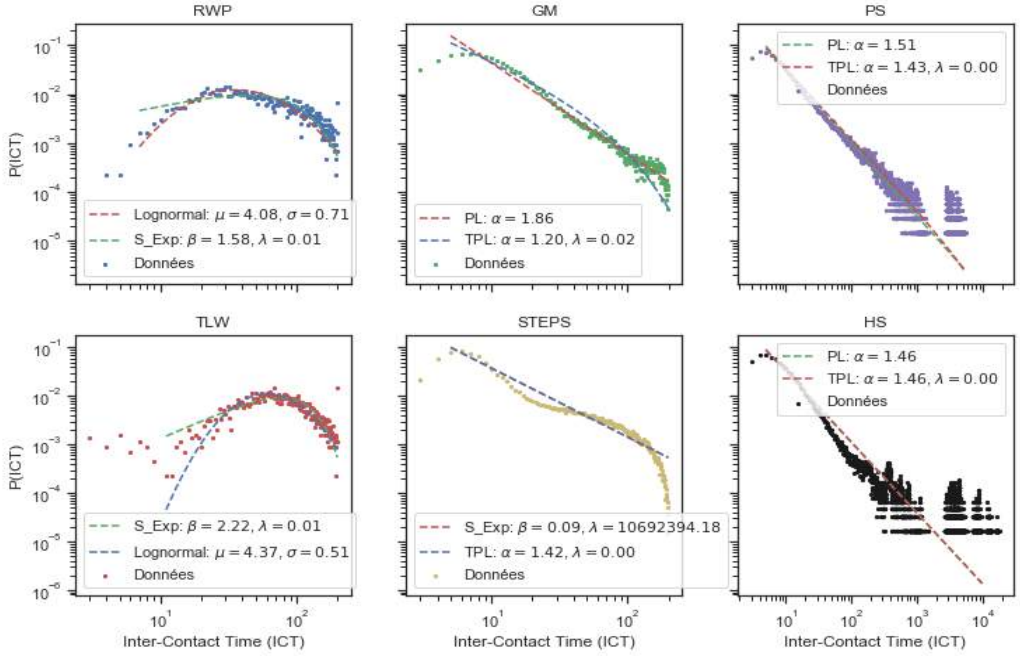


FIGURE 4.18: Distribution de la durée d'inter-contact sur une échelle log-log

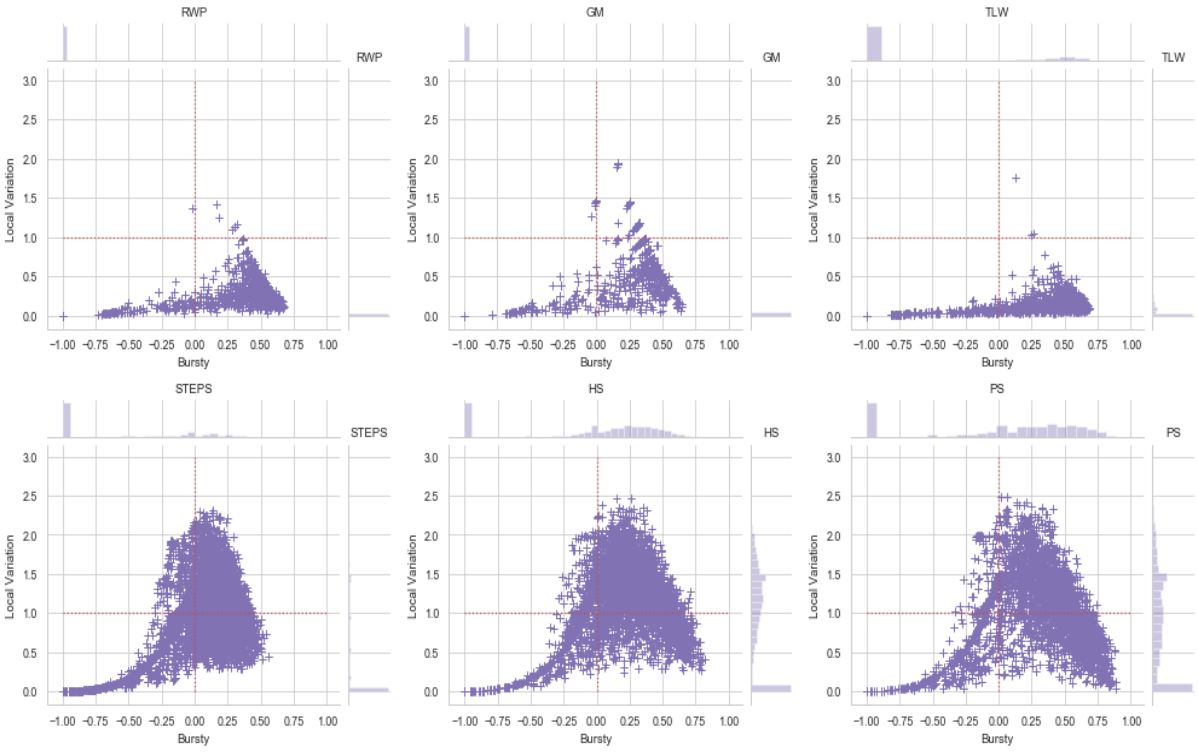


FIGURE 4.19: Mesures du coefficient de Burstiness $\boldsymbol{B}$ en fonction de la variation locale $\boldsymbol{LV}$

sur $\mathscr{G}_{RWP}$, $\mathscr{G}_{GM}$ et $\mathscr{G}_{TLW}$ ne nous permettent pas de conclure sur la nature du phénomène.

TABLE 4.6: Mesures de Fluctuabilité et Volatilité

Réseau temporel	Fluctuabilité	Volatilité
$\mathcal{G}_{RWP}$	0.0872	0.0011
$\mathcal{G}_{GM}$	0.1922	0.0019
$\mathcal{G}_{TLW}$	0.028	0.0003
$\mathcal{G}_{STEPS}$	0.3193	0.0056
$\mathcal{G}_{PS}$	0.0661	0.0017
$\mathcal{G}_{HS}$	0.0308	0.0003

4.5.7 Fluctuabilité et Volatilité

Nous savons que la fluctuabilité permet de mesurer la diversité des connexions temporelles tandis que la volatilité quantifie le taux de changement des liens en fonction de l'ordre temporel. Dans un contexte de réseau de contacts temporels influencé surtout par la mobilité les mesures de fluctuabilité et de volatilité nous permettent de comprendre la variabilité et le changement rapide des liens temporels. Les résultats obtenus montrent de faibles valeurs de fluctuabilité et de volatilité (voir Tableau 4.6) indiquant la répétition de certaines connexions temporelles (non unicité des liens). Cette caractéristique était déjà montrée avec l'apparition de la loi puissance dans les occurrences des paires de contacts. En plus, nous observons avec les faibles valeurs de volatilité un changement lent de la connectivité temporel. Ce dernier résultat peut être utiliser pour comprendre la rapidité du processus de propagation dans les réseaux de contacts temporels générés.

4.6 Modèles épidémiques

Pour étudier le phénomène de propagation dans un réseau de contact temporel nous allons adopter dans notre approche le modèle épidémiologique SIS (*Susceptible-Infected-Susceptible*) qui est largement utilisé dans la communauté scientifique. Ce modèle cyclique est très simple à mettre en place et permet de répondre à plusieurs questions de recherches. En plus, les résultats sont faciles à étendre dans d'autres modèles épidémiques similaires (SIR, SEIR etc.). L'approche analytique du modèle a été développée dans la Section 3. Avec les données de contacts temporels, nous avons développé deux modèles de propagation SIS utilisant l'impact de la distance et du voisinage dans le processus de contamination. Dans cette partie, nous avons restreint notre champ d'étude à trois différents modèles de mobilité suivants : *RWP*, *GM* et *TLW*.

4.6.1 Modèle 1 : Description

Dans le contexte de l'approche compartimentale, la plupart des modèles épidémiologiques considère de manière probabiliste le facteur de contamination en tant que coefficient représentant le taux d'infection [BBV08 ; Pas+15]. De plus, nous supposons souvent que ce taux d'infection garde la même valeur dans tout le processus de propagation. En réalité, il n'est pas évident de croire que la probabilité d'être infecté par une maladie pourrait être homogène sur tous les individus d'une population. Pour la plupart des maladies, la proximité des distances et la durée de contact entre deux individus font parties des principaux facteurs du processus

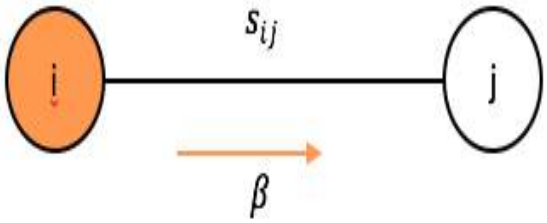


FIGURE 4.20: Le nouveau taux de transmissibilité $\bar{\beta}_{ij}$ dépendant du taux β et de la force du lien s_{ij}. $\bar{\beta}_{ij} = \beta * s_{ij}$

de contamination. Dans cette partie, nous tenterons de mettre en évidence la relation qui existe entre la mobilité et le processus de propagation en définissant un nouveau paramètre de transmissibilité (*spreading rate*) en fonction de la proximité.

Pour étudier l'évolution du processus épidémique présenté ci-dessus, nous considérons le même type de graphe temporel de contact que précédemment $\mathcal{G} = \{G_1, G_2, \ldots, G_T\}$ où $Gt = (V, Et)$ est généré à partir des modèles de mobilité suivants : *RWP*, *GM* et *TLW*. Nous supposons aussi que la taille du réseau temporel est constituée de 1000 nœuds qui se déplacent dans un espace de simulation de $100m \times 100m$ selon un type de mobilité défini. Les nœuds sont temporairement localisés par les traces de leurs positions de coordonnées (x, y) dans un espace euclidien. En plus, deux nœuds i et j sont considérés en contact si la distance d_{ij} est inférieure à un seuil d_{th}. Maintenant, nous supposons deux assertions :

— Premièrement, nous considérons que les graphes temporels $\mathcal{G}$ générés sont non-valués (*unweighted temporal network*) et que le processus de propagation dépend seulement des paramètres β et μ.

— Deuxièmement, nous générons des graphes temporels pondérés par un coefficient s_{ij} qui dépend de la distance de proximité d_{ij}. Nous définissons s_{ij} comme étant le rapport entre d_{th} et d_{ij}.

$$s_{ij} = \frac{d_{th}}{d_{ij}} \tag{4.3}$$

Nous supposons que notre paramètre de transmissibilité β dépend de la proximité et change en fonction de la force du lien (i, j) dans le modèle (voir figure 4.20) et devient :

$$\bar{\beta}_{ij} = \beta * s_{ij} \tag{4.4}$$

Tous les paramètres de simulations utilisés ont été répertoriés dans le Tableau .

4.6.2 Résultats

Dans le but de tester les hypothèses évoquées en début de chapitre, nous avons effectué des simulations du modèle épidémique SIS sur les différents graphes temporels générés à partir des modèles de mobilité *RWP*, *GM* et *TLW* ($\mathcal{G}_{RWP}$, $\mathcal{G}_{GM}$ et $\mathcal{G}_{TLW}$). Dans le premier cas, le taux

TABLE 4.7: Paramètres de Simulation

Paramètres	**Valeurs**	**Description**
MAX_X, MAX_Y	100, 100	Grille de simulation de taille $MAX_X \times MAX_Y$
N	1000	Nombre de nœuds
T	100	Valeur maximale du séquence des pas de temps $\{1,2,3,\ldots,T\}$
d_{th}	4 m	Distance maximale pour que deux individus soient considérés en contacts
wt_{max}	5. s	Durée maximale de pause de la mobilité *RWP*
$[v_{min}, v_{max}]$	0.1, 1 (m/s)	L'intervalle de vitesse des noeuds suivant *RWP*
v_{mean}	1	Vitesse moyenne du mobile dans la mobilité *GM* (μ_v)
μ_θ	1	La variance du paramètre de direction dans *GM*
α_{GM}	0.2	Paramètre de calibrage du modèle *GM* pour faire varier le caractère aléatoire
fl_{exp}	-2.6	Exposant de la loi puissance permettant de déterminer la distance des *flights* dans *TLW*
wt_{exp}	-1.8	Exposant de la loi puissance permettant de déterminer la durée des pauses dans le modèle *TLW*
fl_{max}	50 m	Distance maximale d'un *flight* dans la mobilité *TLW*
wt_{max}	100. s	Durée maximale du temps de pause de *TLW*
ρ	1%	Proportion d'individus initialement infectés
β_0	0.2	Taux de transmissibilité (*spreading rate*)
μ	0.5	Taux de guérison (*recovery rate*)

de transmissibilité β est considéré fixe tout au long de l'expérience. La première observation faite sur les résultats de la Figure 4.21 montre de manière globale que la densité d'infection sur tous les graphes atteint un seuil endémique faible par rapport à la deuxième expérience (voir Figure 4.23). Nous constatons, en plus, que le taux de prévalence $i(t)$ du graphe $\mathcal{G}_{RWP}$ évolue plus vite que dans les autres graphes et atteint à l'état endémique la valeur constante de 44% à partir de $t = 20$. Dans $\mathcal{G}_{GM}$ et $\mathcal{G}_{TLW}$, les densités d'infection évoluent plus lentement et atteignent les mêmes valeurs seuils de 30% à partir de $t = 20$. Comme nous l'avons constaté précédemment, ici le processus de propagation dépend des propriétés structurelles du réseau temporel surtout de la connectivité qui est entièrement une résultante de la mobilité. La Figure 4.22 résume en moyenne le taux de prévalence dans chaque type de graphe.

Dans notre seconde expérience, le taux $\bar{\beta}_{ij}$ dépend d'un coefficient β et la force du lien s_{ij} ($\bar{\beta}_{ij} = \beta * s_{ij}$). Il faut noté que si $d_{ij} = d_{th}$ alors nous revenons au premier cas c'est-à-dire $\bar{\beta}_{ij} = \beta$. Comme β et d_{th} sont fixes la probabilité de transmission dépend plus de la proximité des individus. Nous remarquons d'abords dans la Figure 4.21 que la densité d'infection $i(t)$ augmente plus rapidement que dans notre premier cas. En plus, la valeur seuil des nœuds infectés à l'état endémique est supérieure à 50%. Nous constatons toujours que les nœuds qui suivent le *RWP* ont un taux de prévalence toujours supérieur à ceux des autres graphes temporels. Par conséquent, nous pouvons noter que le coefficient $\bar{\beta}_{ij}$ influence beaucoup la probabilité d'infection dans le réseau. Et ce fait découle de plus de l'effet de la proximité entre les individus qui est une conséquence immédiate de la mobilité.

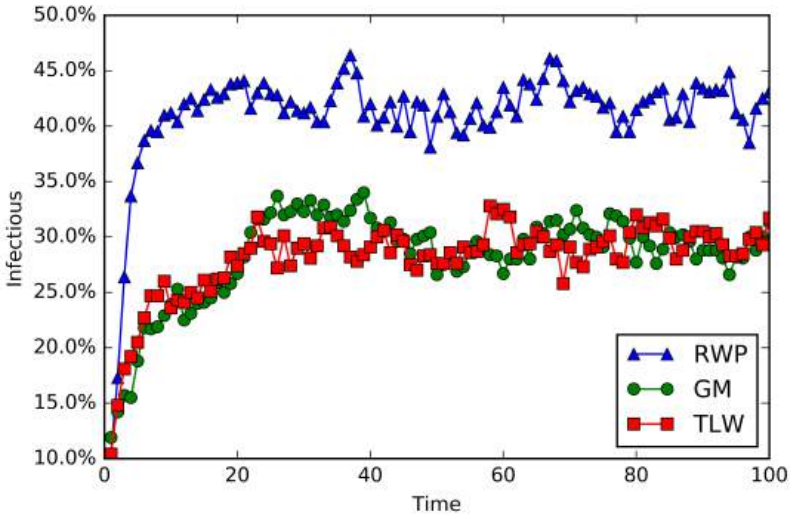


FIGURE 4.21: Taux de prévalence obtenus sur les différents TVG $\mathcal{G}_{RWP}$,$\mathcal{G}_{GM}$,$\mathcal{G}_{TLW}$

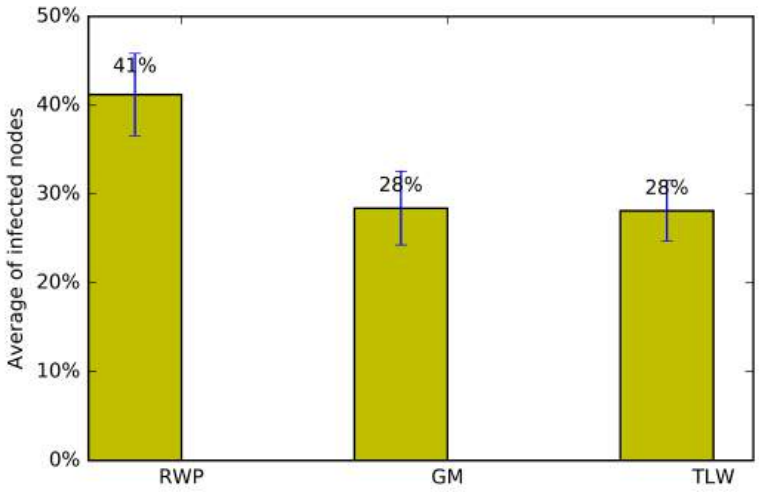


FIGURE 4.22: Comparaison entre les moyenne totales de nœuds infectés

4.6.3 Modèle 2 : Description

Dans cette partie, nous étudions la dynamique de la propagation épidémique dans le réseau temporel généré par des modèles de mobilité synthétiques en tenant en compte, en plus de la proximité, du voisinage. Contrairement au premier modèle, nous supposons que la probabilité pour un nœud sain d'être infecté dépend de son voisinage avec les nœuds infectés et de la distance qui les sépare de ces derniers. En considérant les mêmes types de mobilité que le modèle 4.6.1, nous avons gardé le même 'environnement de simulation (voir Tableau 4.6.1). Dans ce modèle, nous avons généré les graphes temporels $\mathcal{G}_{RWP}$, $\mathcal{G}_{GM}$ et $\mathcal{G}_{TLW}$ dont chaque lien est pondéré par une force s_{ij}. D'abord, nous définissons le poids w_{ij} comme étant le rapport entre d_{th} et d_{ij} :

$$w_{ij} = \frac{d_{ij}}{d_{th}} \tag{4.5}$$

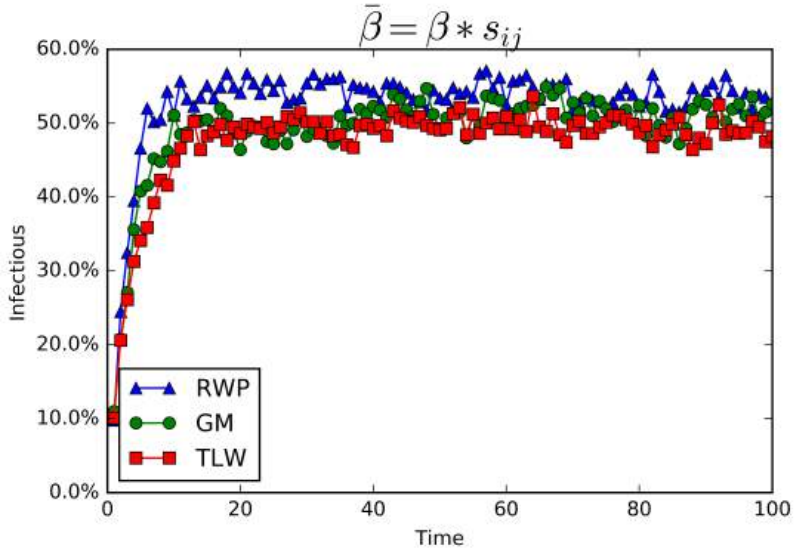


FIGURE 4.23: Taux de prévalence obtenus sur $\mathcal{G}_{RWP}, \mathcal{G}_{GM}, \mathcal{G}_{TLW}$ lorsque β est remplacé par le nouveau coefficient $\bar{\beta}$

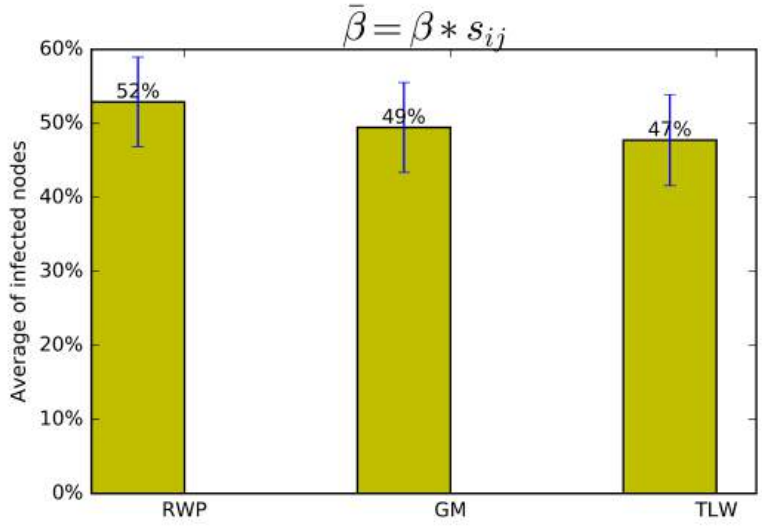


FIGURE 4.24: Comparaison des moyennes totales des nœuds infectés lorsque β devient $\bar{\beta}$ sur les différents TVG utilisés

Comme,

$$0 < d_{ij} \leq d_{th}$$

Alors,

$$0 < w_{ij} \leq 1$$

En utilisant le poids w_{ij}, nous allons définir la force de la proximité, s_{ij}, entre deux nœuds i et j, par la formule suivante :

$$s_{ij} = \frac{1}{(w_{ij})^{\alpha}} \tag{4.6}$$

avec un exposant α représentant un paramètre d'échelle. Par conséquent, avec différentes valeurs de α, nous distinguons plusieurs cas selon les types de mobilité.

— Si $\alpha = 0$ alors $s_{ij} = 1$ ne dépend plus de la distance de proximité.

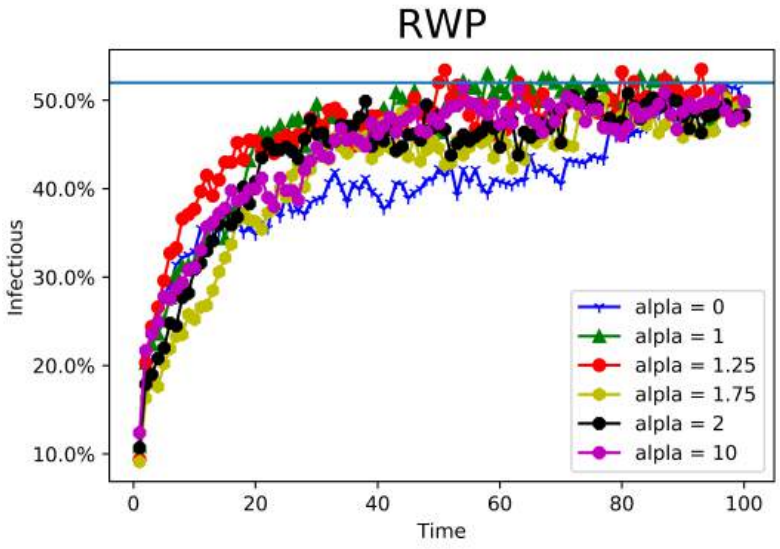


FIGURE 4.25: Taux de prévalence dans $\mathcal{G}_{RWP}$ en fonction du paramètre α

— Si $\alpha = 1$, nous retrouvons la même valeur de s_{ij} défini dans le modèle précédent.
— Si $\alpha > 1$, la force du lien s_{ij} augmente l'effet de proximité dans l'interaction entre les nœuds.

Ensuite, à partir de s_{ij} nous avons défini p_{ij} comme étant la probabilité pour que le nœud j soit infecté par son voisin infectieux i. Cette probabilité est donnée par la relation suivante :

$$p_{ij} = \frac{s_{ij}}{\sum_{l \in \Gamma(i)} s_{il}} \tag{4.7}$$

où $\Gamma(i)$ est l'ensemble des voisins du nœud i et s_{ij} est la force entre i et j. Il faut noter que les valeurs de p_{ij} sont contenues évidemment dans l'intervalle $[0,1]$. Si pour tout voisin j de i, la force du lien $s_{ij} = 1$ alors la probabilité dépend du nombre de voisins de i et devient $p_{ij} = 1/k$ avec k le degré de nœud i. Par conséquent, l'équation 4.7 intègre l'effet de la proximité et l'interaction d'un nœud infecté avec son voisinage dans la probabilité de transmission de maladie dans le réseau temporel.

4.6.4 Résultats

Les résultats obtenus dans les Figures 4.25 4.26 4.27 montrent que la densité des nœuds infectés augmente plus rapidement dans $\mathcal{G}_{GM}$ que dans $\mathcal{G}_{RWP}$ et $\mathcal{G}_{TLW}$. Nous remarquons aussi que $\mathcal{G}_{GM}$ atteint plus vite son seuil endémique (72% à l'instant t = 12). Les seuils endémiques obtenus dans $\mathcal{G}_{TLW}$ et $\mathcal{G}_{RWP}$ sont respectivement 61% et 52%. Dans cette expérience, la moyenne du taux de prévalence dans le graphe $\mathcal{G}_{TLW}$ est la plus faible. En effet, ces résultats sont complètement différents de ce qu'on avait constaté dans 4.6.2 où on avait le pourcentage le plus élevé de nœuds infectés dans $\mathcal{G}_{RWP}$.

Nous avons étudié ce modèle SIS en fonction des différentes valeurs de α. Dans $\mathcal{G}_{GM}$, nous ne sentons pas l'influence du paramètre α, puisque toutes les courbes de densité se superposent et convergent très rapidement vers la même valeur seuil. Dans les TVGs $\mathcal{G}_{TLW}$ et $\mathcal{G}_{RWP}$, pour certaines valeurs de α, la probabilité d'infection p_{ij} augmente. Cependant, elles convergent vers le même seuil endémique.

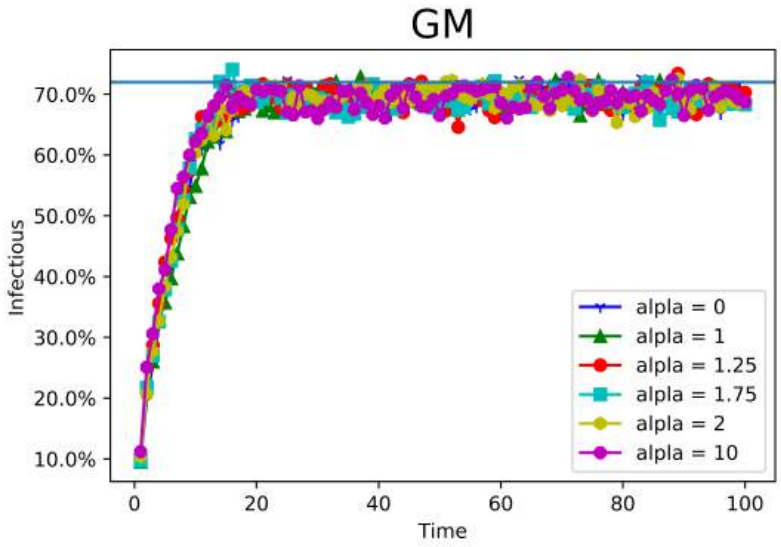


FIGURE 4.26: Taux de prévalence dans $\mathscr{G}_{GM}$ en fonction du paramètre α

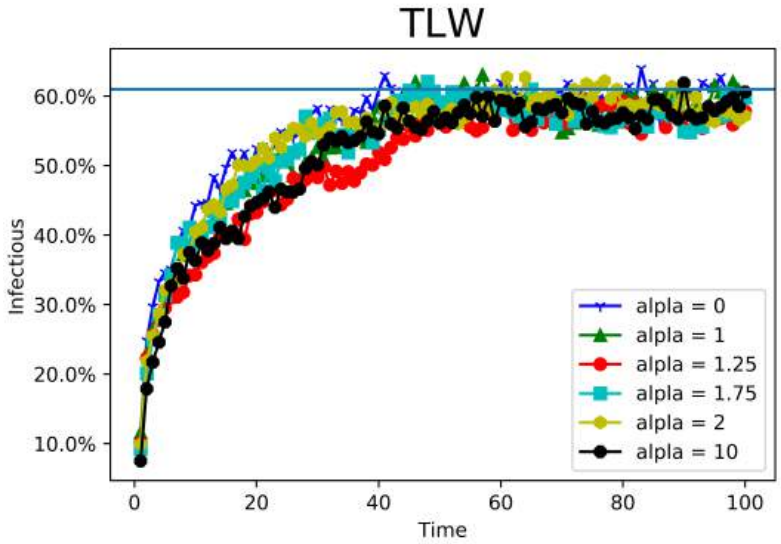


FIGURE 4.27: Taux de prévalence dans $\mathscr{G}_{TLW}$ en fonction du paramètre α

4.7 Discussion

Dans le contexte du processus de propagation d'épidémie, les modèles mathématiques utilisés supposent que la probabilité d'être infecté dépend d'un taux de transmissibilité (*spreading rate*) qui ne reflète pas exactement la réalité. En nous appuyant sur le modèle analytique SIS, nous avons dans un premier temps redéfini le coefficient β par un nouveau coefficient $\tilde{\beta}$ qui dépend de la proximité entre les nœuds supposés être en contact. Dans la seconde expérience, nous avons élaboré un modèle SIS dont la probabilité de transmission intègre la distance qui sépare le nœud infecté avec son voisinage. Ces résultats peuvent être interprétés par les différentes caractéristiques découlant de la dynamique externe du réseau. En faisant une analyse statistique, nous avons pu comprendre que les phénomènes de propagation peuvent être influencés par plusieurs facteurs tels que le MSD, la densité du réseau, la connexité, l'accessibilité, la corrélation et l'efficience temporelle, la centralité des

nœuds, la fréquence des paires de contacts, le phénomène de "bursty". Les résultats de l'analyse topologique montrent que toutes les graphes temporels qui ont été générés (*RWP*, *GM*, *TLW* et *STEPS*) présentent une évolution de la densité très faible et de forte déconnexions malgré une importante connexité dans $\mathcal{G}_{RWP}$. Si nous nous référons des résultats $\boldsymbol{R}_d$ obtenus dans des différentes TVGs, $\mathcal{G}_{STEPS}$ et $\mathcal{G}_{GM}$ font passer plus rapidement l'information dans le réseau que $\mathcal{G}_{RWP}$ et $\mathcal{G}_{GM}$. De même, les résultats par rapport à la fluctuation et la volatilité montrent qu'un important changement de connexions dans les graphes $\mathcal{G}_{GM}$ $\mathcal{G}_{GM}$. Ces deux caractéristiques peuvent expliquer la rapide dispersion des fractions de nœuds infectés due à la mobilité des agents. L'effet petit-monde est aussi constaté dans les TVGs. La centralité et l'occurrence des liens permettent de détecter l'influence de certains individus dans le processus de propagation. En bref, ces deux expériences mettent en évidence l'impact de la mobilité sur le processus de propagation dans la mesure où la structure topologique des graphes de contacts temporels dépend entièrement de la mobilité.

Chapitre 5

Conclusion & Perspectives

L'épidémiologie fait partie des thématiques les plus étudiées parmi les phénomènes dynamiques qui se produisent dans les réseaux complexes. Son but consiste à aider les services de santé publique de mettre en place des politiques stratégiques permettant de surveiller, de contrôler ou de stopper la prolifération des maladies épidémiques. Ce qui justifie la place importante qu'elle occupe dans les réseaux complexes. Plusieurs modèles théoriques ont été développés pour comprendre leurs comportements. Cependant, la plupart d'entre eux ne prennent pas en compte certaines caractéristiques liées à la dynamique topologique découlant de la mobilité des individus. Après une étude approfondie sur l'état de l'art, nous avons, d'une part, étudié les caractéristiques topologiques des réseaux de contacts temporels générés à partir de différents modèles de mobilité, d'autres part, nous avons proposé un modèle épidémique, SIS, permettant d'étudier la dynamique interne en intégrant l'impact de la force des liens de contacts comme la distance de proximité entre les individus et leur voisinage avec les nœuds infectés dans le processus de contamination. Pour atteindre notre objectif nous avons généré quatre différents types de traces basés essentiellement sur des modèles synthétiques de mobilité RWP, GM, TLW et STEPS afin d'étudier l'impact de la mobilité sur les phénomènes dynamiques. L'étude des caractéristiques des traces obtenus révèlent que la dynamique topologique influence profondément sur les processus dynamique. De manière globale, nous avons pu constater que les réseaux de contacts temporels aléatoires sont des réseaux qui sont fortement déconnectés où nous distinguons parfois des caractéristiques petit-monde (une fort coefficient de corrélation temporel et une faible mesure d'efficience temporel). Nous avons aussi constaté que la durée des contacts et le nombre de contacts entre deux paires de nœuds suivent une loi puissance montrant que la plupart des individus passent plus de temps avec un nombre très limité d'individus. De plus, nous remarquons aussi de faibles mesures de fluctuabilité et de volatilité montrant la répétition de certaines connexions sur les TVGs. Contrairement aux réseaux de contacts réels, le phénomène de *bursty* n'est pas très fréquent dans les traces synthétiques.

En dehors de caractéristiques rencontrées, il existe d'autres propriétés découlant de la mobilité humaine qui doivent être prise en compte dans l'étude l'évolution de la dynamique.

Par exemple, la nature des contacts entre les individus, la régularité de certains évènements (par exemple les navettes), les pics de contacts durant la journée particulièrement aux heures de pause et dans les POI (*Points of Interets*), la diminution périodique de la fréquence des contacts durant les nuits, la fréquence de visite de certains lieux etc. Toutes ces propriétés peuvent expliquer le comportement de la propagation d'un phénomène épidémique dans une population donnée.

Étant donné que, les données de mobilité synthétique ne suffisent pas pour décrire exactement le comportement de la mobilité humaine alors nous espérons étendre les travaux effectués dans cette thèse sur des données de contact basées sur une mobilité réelle, de dégager l'approche analytique qui soutient nos modèles et de prédire analytiquement le seuil épidémique. En plus, il s'avère aussi très pertinent d'ajuster le modèle en y intégrant la durée de contact et le nombre individus infectieux antérieurement rencontré dans le facteur de contamination.

Bibliographie

[AS12] W. Quattrociocchi A. Casteigts P. Flocchini et N. Santoro.
"Time-Varying Graphs and Dynamic Networks".
In : *arxiv:1012.0009v3* (2012).

[AS09] Muhammad Abdulla et Robert Simon.
"The Impact of Intercontact Time within Opportunistic Networks: Protocol Implications and Mobility Models".
In : (jan. 2009).

[AH00] L. A. Adamic et B. A. Huberman.
"Power-law distribution of the world wide web".
In : *Science* (2000).

[AB02] R. Albert et A.-L. Barabási.
"Statistical mechanics of complex networks".
In : *Reviews of Modern Physics* 74 (jan. 2002), p. 47-97.
eprint : `cond-mat/0106096`.

[AJB99] Réka Albert, Hawoong Jeong et Albert-László Barabási.
"Diameter of the World-Wide Web".
In : *Nature* 401 (sept. 1999), 130 EP -.

[AM18] Alberto Aleta et Yamir Moreno.
"Multilayer Networks in a Nutshell".
In : *arXiv e-prints*, arXiv:1804.03488 (avr. 2018), arXiv:1804.03488.
arXiv : `1804.03488 [physics.soc-ph]`.

[ABP14] Jeff Alstott, Ed Bullmore et Dietmar Plenz.
"powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions".
In : *PLOS ONE* 9.1 (jan. 2014), p. 1-11.

[Alv+16] Hamidreza Alvari et al.
"Identifying Community Structures in Dynamic Networks".
In : *CoRR* abs/1609.02622 (2016).
arXiv : `1609.02622`.

[Ama+00] L. A. N. Amaral et al.
"Classes of small-world networks".
In : *Proceedings of the National Academy of Sciences* 97.21 (2000), p. 11149-11152.
eprint : `http://www.pnas.org/content/97/21/11149.full.pdf`.

[AM91] Roy M. Anderson et Robert M. May.

Infectious diseases of humans : dynamics and control / Roy M. Anderson and Robert M. May. English.
Oxford University Press Oxford ; New York, 1991,
viii, 757 p. :

[AL00] William Andreopoulos et Dirk Labudde.
"In Protein Purification and Analysis I: Methods and Applications".
In :
ISBN:978-1-477555-057, 0.
Chap. Protein-Protein Interaction Network.

[Bal+09] Duygu Balcan et al.
"Multiscale Mobility Networks and the Spatial Spreading of Infectious Diseases".
In : *Proceedings of the National Academy of Sciences of the United States of America* 106.51 (2009), p. 21484-21489.

[BA99] Albert L. Barabási et Réka Albert.
"Emergence of Scaling in Random Networks".
In : *Science* (1999).

[BP16] Albert-László Barabási et Márton Pósfai.
Network science.
Cambridge : Cambridge University Press, 2016.

[Bar+18] Hugo Barbosa et al.
"Human mobility: Models and applications".
In : *Physics Reports* 734 (2018). Human mobility: Models and applications, p. 1-74.

[BBV08] Alain Barrat, Marc Barthlemy et Alessandro Vespignani.
Dynamical Processes on Complex Networks.
1st.
New York, NY, USA : Cambridge University Press, 2008.

[Bar11] M. Barthélemy.
"Spatial networks".
In : *physrep* 499 (fév. 2011), p. 1-101.
arXiv : `1010.0302 [cond-mat.stat-mech]`.

[Bar16] Jeff Bartlett.
"Mathematical Modeling of the 2014/2015 Ebola Epidemic in West Africa".
In : 9 (jan. 2016).

[Bas+16] Partha Basuchowdhuri et al.
"Detecting Community Structures in Social Networks by Graph Sparsification".
In : *Proceedings of the 3rd IKDD Conference on Data Science, 2016.*
CODS '16.
Pune, India : ACM, 2016,
5:1-5:9.

[BC78] E. A. Bender et E. R. Canfield.

"The asymptotic number of labelled graphs with given degree sequences".
In : *J. Comb. Theory A* (1978).

[Ber60] Daniel Bernoulli.
"Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de Finoculation pour la prévenir."
In : *Mém. Math. Phys. Acad. Roy. Sei. Paris* (1760), p. 1-45.

[BRS03] C. Bettstetter, G. Resta et P. Santi.
"The node distribution of the random waypoint mobility model for wireless ad hoc networks".
In : *IEEE Transactions on Mobile Computing* 2.3 (juil. 2003), p. 257-269.

[BW02] Christian Bettstetter et Christian Wagner.
"The Spatial Node Distribution of the Random Waypoint Mobility Model".
In : *Mobile Ad-Hoc Netzwerke, 1. Deutscher Workshop ÜBer Mobile Ad-Hoc Netzwerke WMAN 2002.*
GI, 2002,
P. 41-58.

[Blo+12] Vincent Blondel et al.
"Data for Development: the D4D Challenge on Mobile Phone Data".
In : (sept. 2012).

[Boc+06] S. Boccaletti et al.
"Complex networks: Structure and dynamics".
In : *Physics Reports* 424.4 (2006), p. 175-308.

[Boc+14] S. Boccaletti et al.
"The structure and dynamics of multilayer networks".
In : *Physics Reports* 544.1 (2014). The structure and dynamics of multilayer networks, p. 1-122.

[BPV03] M. Boguñá, R. Pastor-Satorras et A. Vespignani.
"Epidemic Spreading in Complex Networks with˜Degree Correlations".
In : *Statistical Mechanics of Complex Networks.*
Sous la dir. de R. Pastor-Satorras, M. Rubi et A. Diaz-Guilera.
T. 625.
Lecture Notes in Physics, Berlin Springer Verlag.
2003,
P. 127-147.
eprint : `cond-mat/0301149`.

[Bol81] Béla Bollobás.
"Degree sequences of random graphs".
In : *Discrete Mathematics* 33.1 (1981), p. 1-19.

[BR00] S. Bornholdt et T. Rohlf.
"Topological Evolution of Dynamical Networks: Global Criticality from Local Dynamics".

In : *Physical Review Letters* 84 (juin 2000), p. 6114-6117.
eprint : cond-mat/0003215.

[Bov15] Alexandre Bovet.
"An Introduction to Non-diffusive Transport Models".
In : *ArXiv e-prints* (août 2015).

[BHG06] D. Brockmann, L. Hufnagel et T. Geisel.
"The scaling laws of human travel".
In : *Nature* 439 (jan. 2006), 462 EP -.

[Bro+00] Andrei Broder et al.
"Graph Structure in the Web".
In : *Comput. Netw.* 33.1-6 (juin 2000), p. 309-320.

[BK12] Piotr Bródka et Przemyslaw Kazienko.
"Multi-layered Social Networks".
In : *CoRR* abs/1212.2425 (2012).
arXiv : 1212.2425.

[Bul12] J. M. Buldú.
Applications of Complex Networks.
2012.

[BSK16] Kathrin Büttner, Jennifer Salau et Joachim Krieter.
"Temporal correlation coefficient for directed networks".
In : *SpringerPlus* 5.1 (2016), p. 1198.

[CV07] Guido Caldarelli et Alessandro Vespignani.
Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science.
River Edge, NJ, USA : World Scientific Publishing Co., Inc., 2007.

[Cas+16] Gastone C Castellani et al.
"Systems medicine of inflammaging".
In : *Briefings in Bioinformatics* 17.3 (mai 2016), p. 527-540.

[CP10] C. Castellano et R. Pastor-Satorras.
"Thresholds for Epidemic Spreading in Networks".
In : *Physical Review Letters* 105.21, 218701 (nov. 2010), p. 218701.
arXiv : 1010.1646 [cond-mat.stat-mech].

[Cha+08] Deepayan Chakrabarti et al.
"Epidemic Thresholds in Real Networks".
In : *ACM Trans. Inform. Syst. Secur.* (2008).

[Cha+17] Vivek Charu et al.
"Human mobility and the spatial transmission of influenza in the United States".
In : *PLoS Computational Biology* 13.2 (fév. 2017). Sous la dir. de Marcel Salathé, e1005382.

[Che+15] Hocine Cherifi et al.

"Overlapping Community Structure in Co-authorship Networks: a Case Study".
In : *CoRR* abs/1504.00247 (2015).
arXiv : 1504.00247.

[Che+18] Hocine Cherifi et al.
"Multilayer Network Model of Movie Script".
In : *CoRR* abs/1812.05718 (2018).
arXiv : 1812.05718.

[CHI10] Nacim Fateh CHIKHI.
"Calcul de centralité et identification de structures de communautés dans les graphes de documents".
Thèse de doct. Université de Toulouse, 2010.

[Chi+10] Camelia Chira et al.
"Complex Systems and Cellular Automata Models in the Study of Complexity".
In : *Studia Universitatis Babes-Bolyai : Series Informatica* LV (déc. 2010).

[Cho+16] Gerardo Chowell et al.
"Mathematical models to characterize early epidemic growth: A review".
In : *Physics of Life Reviews* 18 (2016), p. 66-97.

[Col14] Ewan Colman.
"Structure and dynamics of evolving complex networks".
Thèse de doct. School of Information Systems, Computing et Mathematics Brunel University, 2014.

[CLF07] Vania Conan, Jérémie Leguay et Timur Friedman.
"The heterogeneity of inter-contact time distributions: its importance for routing in delay tolerant networks".
working paper or preprint.
Jan. 2007.

[CVS11] Eduardo Cuestas, Mario Vilaró et Pablo Catalán Serra.
"[Spatial and temporal spread predictability of influenza A H1N1 epidemic in Argentina by the percolation method]."
In : *Revista Argentina de microbiologia* 43 3 (2011), p. 186-90.

[CFK17] Marjan Cugmas, Anuška Ferligoj et Luka Kronegger.
"Scientific co-authorship networks".
In : *arXiv e-prints*, arXiv:1711.00770 (nov. 2017), arXiv:1711.00770.
arXiv : 1711.00770 [stat.AP].

[CZL13] Jing Cui, Yi-Qing Zhang et Xiang Li.
"On the clustering coefficients of temporal networks and epidemic dynamics".
In :
Mai 2013,
P. 2299-2302.

[D B10] S. A. Hill snd D. Braha.
"Dynamic model of time-dependent complex networks".

In : *PHYSICAL REVIEW* (2010).

[DP15] Eric H. Davidson et Isabelle S. Peter.
"Chapter 2 - Gene Regulatory Networks".
In : *Genomic Control Process*.
Sous la dir. d'Eric H. Davidson et Isabelle S. Peter.
Oxford : Academic Press, 2015,
P. 41-77.

[De +13] M. De Domenico et al.
"Mathematical Formulation of Multilayer Networks".
In : *Physical Review X* 3.4, 041022 (oct. 2013), p. 041022.
arXiv : 1307.4977 [physics.soc-ph].

[Die93] Klaus Dietz.
"The Estimation of the Basic Reproduction Number for Infectious Diseases".
In : *Statistical Methods in Medical Research 2(1)* (1993).

[DG10] A.-L. Do et T. Gross.
"Contact processes and moment closure on adaptive networks".
In : *ArXiv e-prints* (nov. 2010).
arXiv : 1011.2689 [physics.soc-ph].

[DMS00] S. N. Dorogovtsev, J. F. F. Mendes et A. N. Samukhin.
"Structure of Growing Networks with Preferential Linking".
In : *Phys. Rev. Lett. 85, 4633* (2000).

[DA05] Jordi Duch et Alex Arenas.
"Community detection in complex networks using extremal optimization".
In : *pre* 72, 027104 (août 2005), p. 027104.
arXiv : cond-mat/0501368 [cond-mat.dis-nn].

[Ein05] A. Einstein.
"INVESTIGATIONS ON THE THEORY OF THE BROWNIAN MOVEMENT".
In : *Ann. der Physik* (1905). Sous la dir. de R. Fürth.

[EKB17] Rasha Elhesha, Tamer Kahveci et Benjamin Baiser.
"Motif centrality in food web networks".
In : *Journal of Complex Networks* 5.4 (2017), p. 641-664.
eprint : /oup/backfile/content_public/journal/comnet/5/4/10.1093_comnet_cnw032/2/cnw032.pdf.

[ER59] P. Erdös et A. Renyi.
"On Random Graphs".
In : *Publ. Math. Debrecen* (1959).

[FFF99] Michalis Faloutsos, Petros Faloutsos et Christos Faloutsos.
"On Power-law Relationships of the Internet Topology".
In : *SIGCOMM Comput. Commun. Rev.* 29.4 (août 1999), p. 251-262.

[Fen+16] Yun Feng et al.

"Epidemic spreading on weighted networks with adaptive topology based on infective information".
In : *Physica A: Statistical Mechanics and its Applications* 463 (2016), p. 493-502.

[For10] Santo Fortunato.
"Community detection in graphs".
In : *physrep* 486.3-5 (fév. 2010), p. 75-174.
arXiv : `0906.0612 [physics.soc-ph]`.

[Fre78] Linton C. Freeman.
"Centrality in social networks conceptual clarification".
In : *Social Networks* (1978), p. 215.

[Fri+09] Mark D. Fricker et al.
"Adaptive Biological Networks".
In :
Adaptive Networks: Theory, Models and Applications.
Sous la dir. de Thilo Gross et Hiroki Sayama.
Berlin, Heidelberg : Springer Berlin Heidelberg, 2009,
P. 51-70.

[Fun+09] Sebastian Funk et al.
"The spread of awareness and its impact on epidemic outbreaks".
In : *Proceedings of the National Academy of Sciences* 106.16 (2009), p. 6872-6877.
eprint : `http://www.pnas.org/content/106/16/6872.full.pdf`.

[GV05] I. Farkas G. Palla I. Derenyi et Tamas Vicsek.
"Uncovering the overlapping community structure of complex networks in nature and society".
In : *Nature 435* (2005).

[GB15] Riccardo Gallotti et Marc Barthelemy.
"The multilayer temporal network of public transport in Great Britain".
In : *Scientific Data* 2 (jan. 2015), 140056 EP -.

[Gao+18] Lei Gao et al.
"Effects of individual popularity on information spreading in complex networks".
In : *Physica A: Statistical Mechanics and its Applications* 489.Supplement C (2018), p. 32-39.

[GBC14] Valerio Gemmetto, Alain Barrat et Ciro Cattuto.
"Mitigation of infectious disease at school: targeted class closure vs school closure."
en.
In : *BMC infectious diseases* 14.1 (déc. 2014), p. 695.

[GN02] M. Girvan et M. E. J. Newman.
"Community structure in social and biological networks".
In : *Proceedings of the National Academy of Sciences* 99.12 (2002), p. 7821-7826.
eprint : `http://www.pnas.org/content/99/12/7821.full.pdf`.

[Gle18] Gleamviz.

Global Epidemic and Mobility Model.
Juil. 2018.
URL : http://http://www.gleamviz.org/.

[GB06] Kwang-Il Goh et Albert-Laszlo Barabasi.
"Burstiness and Memory in Complex Systems".
In : *arXiv e-prints*, physics/0610233 (oct. 2006), physics/0610233.
arXiv : physics/0610233 [physics.data-an].

[GAE14] S. R. Goldberg, H. Anthony et T. S. Evans.
"Modelling Citation Networks".
In : *ArXiv e-prints* (août 2014).
arXiv : 1408.2970 [physics.soc-ph].

[GHB08] Marta C. González, César A. Hidalgo et Albert-László Barabási.
"Understanding individual human mobility patterns".
In : *Nature* 453 (juin 2008), 779 EP -.

[Gra09] Andrzej Grabowski.
"Opinion formation in a social network: The role of human activity".
In : *Physica A: Statistical Mechanics and its Applications* 388.6 (2009), p. 961-966.

[GDB06] Thilo Gross, Carlos J. Dommar D'Lima et Bernd Blasius.
"Epidemic Dynamics on an Adaptive Network".
In : *Phys. Rev. Lett.* 96 (20 mai 2006), p. 208701.

[GS09] Thilo Gross et Hiroki Sayama.
Adaptive Networks: Theory, Models and Applications.
1st.
Springer Publishing Company, Incorporated, 2009.

[Gui+03] R. Guimerà et al.
"Self-similar community structure in a network of human interactions".
In : *Phys. Rev. E* 68 (6 déc. 2003), p. 065103.

[GN05] Roger Guimerà et Lués A Nunes Amaral.
"Functional cartography of complex metabolic networks".
In : *Nature* 433.7028 (fév. 2005), p. 895-900.

[HG97] F. Harary et G. Gupta.
"Dynamic Graph Models".
In : *Mathl. Comput. Modelling* (1997).

[HA13] Matthew Hartfield et Samuel Alizon.
"Introducing the Outbreak Threshold in Epidemiology".
In : *PLoS Pathogens* 9.6 (juin 2013), e1003277.

[HB15] Stefan Heinz et Hakima Bessaih.
Stochastic Equations for Complex Systems: Theoretical and Computational Topics.
Jan. 2015.

[HSD13] Morris W. Hirsch, Stephen Smale et Robert L. Devaney.

Differential Equations, Dynamical Systems, and an Introduction to Chaos. English.
Elsevier Inc., 2013.

[Hol15] P. Holme.
"Modern temporal network theory: a colloquium".
In : *European Physical Journal B* 88, 234 (sept. 2015), p. 234.
arXiv : `1508.01303 [physics.soc-ph]`.

[HN06] P. Holme et M. E. J. Newman.
"Nonequilibrium phase transition in the coevolution of networks and opinions".
In : 74.5, 056108 (nov. 2006), p. 056108.
eprint : `physics/0603023`.

[HS12] P. Holme et J. Saramäki.
"Temporal networks".
In : *Phys. Rep. 519* 519 (oct. 2012), p. 97-125.
arXiv : `1108.1780 [nlin.AO]`.

[Hol05] Petter Holme.
"Network reachability of real-world contact sequences".
In : 71, 046119 (avr. 2005), p. 046119.
arXiv : `cond-mat/0410313 [q-bio.OT]`.

[HS13] Petter Holme et Jari Saramäki.
"Temporal Networks as a Modeling Framework".
In :
Temporal Networks.
Sous la dir. de Petter Holme et Jari Saramäki.
Berlin, Heidelberg : Springer Berlin Heidelberg, 2013,
P. 1-14.

[Hon+08] Seongik Hong et al.
"Routing Performance Analysis of Human-Driven Delay Tolerant Networks Using the Truncated Levy Walk Model".
In :
Jan. 2008,
P. 25-32.

[Hon+16] Sheng Hong et al.
"Epidemic spreading model of complex dynamical network with the heterogeneity of nodes".
In : *International Journal of Systems Science* 47.11 (2016), p. 2745-2752.

[HXS15] Changjun Hu, Wenwen Xu et Peng Shi.
"Information Diffusion in Online Social Networks: Models, Methods and Applications".
In :
Web-Age Information Management: WAIM 2015 International Workshops: HENA, HRSUNE, Qingdao, China, June 8-10, 2015, Revised Selected Papers.

Sous la dir. de Xiaokui Xiao et Zhenjie Zhang.
Cham : Springer International Publishing, 2015,
P. 65-76.

[HLV06] E. Hyytia, P. Lassila et J. Virtamo.
"Spatial node distribution of the random waypoint mobility model with applications".
In : *IEEE Transactions on Mobile Computing* 5.6 (juin 2006), p. 680-694.

[JN03] M. Liu J. Yoon et B. Noble.
"Random waypoint considered harmful".
In : *Proceedings of INFOCOM'03* (2003).

[Jeo+00] H. Jeong et al.
"The large-scale organization of metabolic networks".
In : *Nature* 407 (oct. 2000), 651 EP -.

[Jeo+01] H. Jeong et al.
"Lethality and centrality in protein networks".
In : *Nature* 411 (mai 2001), 41 EP -.

[Jia+17] Shixiong Jiang et al.
"Human mobility in space from three modes of public transportation".
In : *Physica A: Statistical Mechanics and its Applications* 483 (2017), p. 227-238.

[JM96] David B. Johnson et David A. Maltz.
"Dynamic Source Routing in Ad Hoc Wireless Networks".
In :
Mobile Computing.
Sous la dir. de Tomasz Imielinski et Henry F. Korth.
Boston, MA : Springer US, 1996,
P. 153-181.

[Jur+15] Raja Jurdak et al.
"Understanding Human Mobility from Twitter".
In : *PLOS ONE* 10.7 (juil. 2015), p. 1-16.

[Kan+15] Chaogui Kang et al.
"A Generalized Radiation Model for Human Mobility: Spatial Scale, Searching Direction and Trip Constraint".
In : *PLOS ONE* 10.11 (nov. 2015), p. 1-11.

[KBV10] T. Karagiannis, J. Le Boudec et M. Vojnovic.
"Power Law and Exponential Decay of Intercontact Times between Mobile Devices".
In : *IEEE Transactions on Mobile Computing* 9.10 (oct. 2010), p. 1377-1390.

[Kar+12] Márton Karsai et al.
"Universal features of correlated bursty behaviour".
In : *Scientific Reports* 2, 397 (mai 2012), p. 397.
arXiv : `1111.7235 [physics.soc-ph]`.

[KR07] M.J. Keeling et P. Rohani.
Modeling Infectious Diseases in Humans and Animals.
Princeton University Press, 2007.

[KM27] W. O. Kermack et A. G. McKendrick.
"A contribution to the mathematical theory of epidemics".
In : *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 115.772 (1927), p. 700-721.
eprint : `http://rspa.royalsocietypublishing.org/content/115/772/700.full.pdf`.

[KS18] Hyobin Kim et Hiroki Sayama.
"Robustness and Evolvability of Multilayer Gene Regulatory Networks".
In :
Jan. 2018,
P. 546-547.

[KS17] Minjun Kim et Hiroki Sayama.
"Predicting stock market movements using network science: an information theoretic approach".
In : *Applied Network Science* 2.1 (oct. 2017), p. 35.

[KLE10a] S. Kim, C. Lee et D. Y. Eun.
"Superdiffusive Behavior of Mobile Nodes and Its Impact on Routing Protocol Performance".
In : *IEEE Transactions on Mobile Computing* 9.2 (fév. 2010), p. 288-304.

[KLE10b] S. Kim, C. Lee et D. Y. Eun.
"Superdiffusive Behavior of Mobile Nodes and Its Impact on Routing Protocol Performance".
In : *IEEE Transactions on Mobile Computing* 9.2 (fév. 2010), p. 288-304.

[KPY16] Yup Kim, Seokjong Park et Soon-Hyung Yook.
"The origin of the criticality in meme popularity distribution on complex networks".
In : *Scientific Reports* 6 (mar. 2016), 23484 EP -.

[Kiv+14] Mikko Kivelä et al.
"Multilayer networks".
In : *Journal of Complex Networks* 2.3 (2014), p. 203-271.
eprint : `/oup/backfile/content_public/journal/comnet/2/3/10.1093_comnet_cnu016/2/cnu016.pdf`.

[Kra+16] Andrew M Kramer et al.
"Spatial spread of the West Africa Ebola epidemic".
In : *Royal Society Open Science* 3.8 (août 2016), p. 160294.

[KR01] P. L. Krapivsky et S. Redner.
"Organization of growing random networks".
In : *Phys. Rev. E 63* (2001).

[KSH10] J. Kroc, P. Sloot et A. Hoekstra.
Simulating Complex Systems by Cellular Automata.
Juin 2010.

[Kum+99] Ravi Kumar et al.
"Extracting Large-Scale Knowledge Base from the Web".
In : *Proceedings of the 9th ACM Symposium on Principles of Database Systems* (1999).

[Kup13] M. N. Kuperman.
"Invited review: Epidemics on social networks".
In : *ArXiv e-prints* (déc. 2013).
arXiv : 1312.3838.

[LM01] V. Latora et M. Marchiori.
"Efficient Behavior of Small-World Networks".
In : *PHYSICAL REVIEW LETTERS* (2001).

[LNR17] V. Latora, V. Nicosia et Giovanni Russo.
Complex Networks Principles, Methods and Applications.
Cambridge University Press, 2017.

[LM05] C. Li et P. K. Maini.
"An evolving network model with community structure".
In : *JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL* (2005).

[LW07] Xiang Li et Xiaofan Wang.
"On the stability of epidemic spreading in small-world networks: how prompt the recovery should be?"
In : *International Journal of Systems Science* 38.5 (2007), p. 401-411.
eprint : https://doi.org/10.1080/00207720701276333.

[LH03] B. Liang et Z. J. Haas.
"Predictive distance-based mobility management for multidimensional PCS networks".
In : *IEEE/ACM Transactions on Networking* 11.5 (oct. 2003), p. 718-732.

[Lon+15] András London et al.
"Complex network analysis of public transportation networks: A comprehensive study".
In :
Juin 2015.

[Lu+16] Jinhu Lu et al.
Complex Systems and Networks: Dynamics, Controls and Applications.
Jan. 2016.

[MW00] C. Moore M. E. J. Newman et D. J. Watts.
"Mean-field solution of the small-world network model".
In : *Physical Review Letters* (2000).

[M K14] N. Perra & Vespignani M. Karsai.
"Time varying networks and the weakness of strong ties".
In : *Sci. Rep. 4, 4001; DOI:10.1038/srep04001* (2014).

[M M09] C. Mascolo M. Musolesi.
"Mobility Models for Systems Evaluation: A Survey".
In : *Springer Berlin Heidelberg* (2009).

[M15] Smith A. F. M.
"George Edward Pelham Box. 10 October 1919 —28 March 2013".
In : *Biographical Memoirs of Fellows of the Royal Society* 61 (fév. 2015), p. 23-37.

[ML16a] Andrea Madotto et Jiming Liu.
"Super-Spreader Identification Using Meta-Centrality".
In : *Scientific Reports* 6 (déc. 2016), 38994 EP -.

[MFB15] Rossana Mastrandrea, Julie Fournet et Alain Barrat.
"Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys".
In : *PLoS ONE* 10.9 (sept. 2015), e0136497.

[ML16b] N. Masuda et R. Lambiotte.
A Guide to Temporal Networks.
Sous la dir. d'UK) Henrik Jeldtoft Jensen (Imperial College London.
Series on Complexity Science, 2016.

[MH17] Naoki Masuda et Petter Holme.
Temporal Network Epidemiology.
Jan. 2017.

[MDL16] Djibril Mboup, Cherif Diallo et Moussa Lo.
"Structural and dynamical analysis of complex networks".
In : *The 2nd Information Technology and Mechatronics Engineering Conference (ITOEC 2016).*
Atlantis Press, mai 2016.

[MDL17a] Djibril Mboup, Cherif Diallo et Moussa Lo.
"Impact of distance-proximity in spreading process in mobile networks".
In : *2017 Global Information Infrastructure and Networking Symposium, GIIS 2017, Saint Pierre, France, October 25-27, 2017.*
2017,
P. 71-74.

[MDL17b] Djibril Mboup, Cherif Diallo et Moussa Lo.
"Mobility and Epidemic Process in Temporal Networks".
In : *Proceedings of the Second International Conference on Advanced Wireless Information, Data, and Communication Technologies, AWICT 2017, Paris, France, November 13-14, 2017.*
2017,
5:1-5:7.

[Mil67] Stanley Milgram.
"The Small World Problem".
In : *Psychology Today* 2 (1967), p. 60-67.

[Mil+02] R. Milo et al.
"Network Motifs: Simple Building Blocks of Complex Networks".
In : *Science* 298.5594 (2002), p. 824-827.
eprint : `http://science.sciencemag.org/content/298/5594/824.full.pdf`.

[Mon+14] Yves-Alexandre de Montjoye et al.
"D4D-Senegal: The Second Mobile Phone Data for Development Challenge".
In : (juil. 2014).

[MS02] Jose M. Montoya et Ricard V. Solé.
"Small World Patterns in Food Webs".
In : *Journal of Theoretical Biology* 214.3 (2002), p. 405-412.

[MN00] C. Moore et M. E. J. Newman.
"Epidemics and percolation in small-world networks".
In : *pre* 61 (mai 2000), p. 5678-5682.
eprint : `cond-mat/9911492`.

[Nef13] Nikolai Nefedov.
"Analysis of Communities Evolution in Dynamic Social Networks".
In :
Complex Networks IV: Proceedings of the 4th Workshop on Complex Networks CompleNet 2013.
Sous la dir. de Gourab Ghoshal, Julia Poncela-Casasnovas et Robert Tolksdorf.
Berlin, Heidelberg : Springer Berlin Heidelberg, 2013,
P. 39-46.

[Nek+07] M. Nekovee et al.
"Theory of rumour spreading in complex social networks".
In : *Physica A Statistical Mechanics and its Applications* 374 (jan. 2007), p. 457-470.
arXiv : `0807.1458 [physics.soc-ph]`.

[New01] M. E. J. Newman.
"Clustering and preferential attachment in growing networks".
In : *Phys. Rev. E* 64 (2 juil. 2001), p. 025102.

[New03] M. E. J. Newman.
"The Structure and Function of Complex Networks".
In : *SIAM Review* 45.2 (2003), p. 167-256.
eprint : `https://doi.org/10.1137/S003614450342480`.

[New11] M. E. J. Newman.
"Resource Letter CS-1: Complex Systems".
In : *American Journal of Physics* 79 (août 2011), p. 800-810.

arXiv : 1112.1440 [cond-mat.stat-mech].

[NG03] M. E. J. Newman et M. Girvan.
"Finding and evaluating community structure in networks".
In : *arXiv:cond-mat/0308217v1 [cond-mat.stat-mech]* (2003).

[Ngu+11] Anh Dung Nguyen et al.
"STEPS - An Approach for Human Mobility Modeling".
In : *NETWORKING 2011.*
Sous la dir. de Jordi Domingo-Pascual et al.
Berlin, Heidelberg : Springer Berlin Heidelberg, 2011,
P. 254-265.

[Nic+12] V. Nicosia et al.
"Components in time-varying graphs".
In : *Chaos* 22.2, 023101 (juin 2012), p. 023101.
arXiv : 1106.2134 [physics.soc-ph].

[Nic+13] V. Nicosia et al.
"Graph Metrics for Temporal Networks".
In :
Temporal Networks, Understanding Complex Systems. ISBN 978-3-642-36460-0. Springer-Verlag Berlin Heidelberg, 2013, p. 15.
Sous la dir. de P. Holme et J. Saramäki.
2013,
P. 15.

[Nou+12] Anastasios Noulas et al.
"A Tale of Many Cities: Universal Patterns in Human Urban Mobility".
In : *PLOS ONE* 7.5 (mai 2012), p. 1-10.

[PA13] Giuliano Andrea Pagani et Marco Aiello.
"The Power GridÂ as a complex network: A survey".
In : *Physica A: Statistical Mechanics and its Applications* 392.11 (2013), p. 2688-2700.

[PS11] Raj Kumar Pan et Jari Saramäki.
"Path lengths, correlations, and centrality in temporal networks".
In : *Physical review. E, Statistical, nonlinear, and soft matter physics* 84 (juil. 2011), p. 016105.

[Pan14] André Panisson.
pymobility v0.1 - python implementation of mobility models.
Mai 2014.

[PS16] Luca Pappalardo et Filippo Simini.
"Data-driven generation of spatio-temporal routines in human mobility".
In : *arXiv e-prints*, arXiv:1607.05952 (juil. 2016), arXiv:1607.05952.
arXiv : 1607.05952 [cs.SI].

[PV01a] R. Pastor-Satorras et A. Vespignani.

"Epidemic dynamics and endemic states in complex networks".
In : *pre* 63.6, 066117 (juin 2001), p. 066117.
eprint : `cond-mat/0102028`.

[PV01b] Romualdo Pastor-Satorras et Alessandro Vespignani.
"Epidemic Spreading in Scale-Free Networks".
In : *Phys. Rev. Lett.* 86 (14 avr. 2001), p. 3200-3203.

[PV02] Romualdo Pastor-Satorras et Alessandro Vespignani.
"Epidemic dynamics in finite size scale-free networks".
In : *Phys. Rev. E* 65 (3 mar. 2002), p. 035108.

[PV04] Romualdo Pastor-Satorras et Alessandro Vespignani.
Evolution and Structure of the Internet: A Statistical Physics Approach.
New York, NY, USA : Cambridge University Press, 2004.

[Pas+15] Romualdo Pastor-Satorras et al.
"Epidemic processes in complex networks".
In : *Rev. Mod. Phys.* 87 (3 août 2015), p. 925-979.

[PE08] J. L. Payne et M. J. Eppstei.
"Evolutionary Dynamics on Scale-Free Interaction Networks".
In : *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION* (2008).

[PSW12] I. Pestov, H. Sayama et C. Wong.
"Modeling discrete distributed heterogeneous systems".
In : *Int. Conf. on Modeling, Simulation and Visualization Methods, MSV.*
2012,
P. 7.

[Pod+01] J. Podani et al.
"Comparable system-level organization of Archaea and Eukaryotes".
In : *Nature Genetics* 29 (août 2001), 54 EP -.

[PT17] Anton V. Proskurnikov et Roberto Tempo.
"A tutorial on modeling and analysis of dynamic social networks. Part I".
In : *Annual Reviews in Control* 43.Supplement C (2017), p. 65-79.

[R A99] A. L. Barabási (1999) R. Albert H. Jeong.
"Diameter of the World-Wide Web".
In : *Nature 401* (1999).

[Red98] Redner, S.
"How popular is your paper? An empirical study of the citation distribution".
In : *Eur. Phys. J. B* 4.2 (1998), p. 131-134.

[Rhe+11] I. Rhee et al.
"On the Levy-Walk Nature of Human Mobility".
In : *IEEE/ACM Transactions on Networking* 19.3 (juin 2011), p. 630-643.

[RLH10] Luis E C Rocha, Fredrik Liljeros et Petter Holme.
"Information dynamics shape the sexual networks of Internet-mediated prostitution".

In : *Proceedings of the National Academy of Sciences of the United States of America* 107.13 (mar. 2010), p. 5706-5711.

[Roy11a] Radhika Ranjan Roy.
"Gravity Mobility".
In :
Handbook of Mobile Ad Hoc Networks for Mobility Models.
Boston, MA : Springer US, 2011,
P. 443-482.

[Roy11b] Radhika Ranjan Roy.
"Random Gauss Markov Mobility".
In :
Handbook of Mobile Ad Hoc Networks for Mobility Models.
Boston, MA : Springer US, 2011,
P. 311-344.

[RSM18] Stefania Rubrichi, Zbigniew Smoreda et Mirco Musolesi.
"A comparison of spatial-based targeted disease mitigation strategies using mobile phone data".
In : *EPJ Data Science* 7.1 (juin 2018), p. 17.

[Sah+19] F. D. Sahneh et al.
"Contact Adaption During Epidemics: A Multilayer Network Formulation Approach".
In : *IEEE Transactions on Network Science and Engineering* 6.1 (jan. 2019), p. 16-30.

[Sal+17] Kankoé Sallah et al.
"Mathematical models for predicting human mobility in the context of infectious disease spread: introducing the impedance model."
In : *Int J Health Geogr* 16.1 (2017), p. 42.

[Sap+18] Anna Sapienza et al.
"Estimating the outcome of spreading processes on networks with incomplete information: A dimensionality reduction approach".
In : *Phys. Rev. E* 98 (1 juil. 2018), p. 012317.

[Say+13] Hiroki Sayama et al.
"Modeling complex systems with adaptive networks".
In : *Computers & Mathematics with Applications* 65.10 (2013). Grasping Complexity, p. 1645-1664.

[SEI12] Massoud SEIFI.
"Cœurs stables de communautés dans les graphes de terrain".
Thèse de doct. Université Pierre et Marie Curie, 2012.

[SSL16] Vedran Sekara, Arkadiusz Stopczynski et Sune Lehmann.
"Fundamental structures of dynamic social networks".
In : *Proceedings of the National Academy of Sciences* 113.36 (2016), p. 9977-9982.

eprint : http://www.pnas.org/content/113/36/9977.full.pdf.

[SKS10] G. Shabbir, H. Khan et M. A. Sadiq.
"A note on Exact solution of SIR and SIS epidemic models".
In : *ArXiv e-prints* (déc. 2010).
arXiv : 1012.5035 [math.CA].

[SMK05] Shigeru Shinomoto, Keiji Miura et Shinsuke Koyama.
"A measure of local variation of inter-spike intervals".
In : *Biosystems* 79.1 (2005). Selection of Papers from the 5th International Workshop on Neural Coding, p. 67-72.

[Sie14] Robert Siegfried.
Modeling and simulation of complex systems: A framework for efficient agent-based modeling and simulation.
Oct. 2014,
P. 1-227.

[Sim+12] Filippo Simini et al.
"A universal model for mobility and migration patterns".
In : *Nature* 484 (fév. 2012), 96 EP -.

[SM06] E. Soedarmadji et R. J. McEliece.
"A Dynamic Graph Algorithm for the Highly Dynamic Network Problem".
In : *PERCOMW?06* (2006).

[Sol65] Derek J. de Solla Price.
"Networks of scientific papers".
In : *Science* (1965).

[Son+10] Chaoming Song et al.
"Modelling the scaling properties of human mobility".
In : *Nature Physics* 6 (sept. 2010), 818 EP -.

[SPL15] A. Stopczynski, A. S. Pentland et S. Lehmann.
"Physical Proximity and Spreading in Dynamic Social Networks".
In : *ArXiv e-prints* (sept. 2015).
arXiv : 1509.06530 [physics.soc-ph].

[Str15] Steven H. (Steven Henry) author Strogatz.
Nonlinear dynamics and chaos : with applications to physics, biology, chemistry, and engineering.
Second edition. Boulder, CO : Westview Press, a member of the Perseus Books Group, [2015], 2015.

[Sun+14] Ye Sun et al.
"Epidemic spreading on weighted complex networks".
In : *Physics Letters A* 378 (jan. 2014), p. 635-640.
arXiv : 1308.4014 [physics.soc-ph].

[SBS98] Helena Szczerbicka, Matthias Becker et Michael Syrjakow.

"Genetic Algorithms: a Tool for Modelling, Simulation, and Optimization of Complex Systems."
In : *Cybernetics and Systems* 29 (oct. 1998), p. 639-659.

[T O09] P. Panzarasa T. Opsahl.
"Clustering in weighted networks".
In : *Preprint submitted to Social Networks* (2009).

[Tan+10a] J. Tang et al.
"Small-world behavior in time-varying graphs".
In : *pre* 81, 055101 (mai 2010), p. 055101.
arXiv : `0909.1712 [physics.soc-ph]`.

[Tan+15] Jinjun Tang et al.
"Uncovering urban human mobility from large scale taxi GPS data".
In : *Physica A: Statistical Mechanics and its Applications* 438 (2015), p. 140-153.

[Tan+10b] John Tang et al.
"Analysing Information Flows and Key Mediators Through Temporal Centrality Metrics".
In : *Proceedings of the 3rd Workshop on Social Network Systems.*
SNS '10.
Paris, France : ACM, 2010,
3:1-3:6.

[Tel+11] Qawi K Telesford et al.
"The ubiquity of small-world networks".
In : *Brain connectivity* 1.5 (déc. 2011), p. 367-375.

[TBF17] William Hedley Thompson, Per Brantefors et Peter Fransson.
"From static to temporal network theory: Applications to functional brain connectivity".
In : *Network Neuroscience* 1.2 (2017), p. 69-99.
eprint : `https://doi.org/10.1162/NETN_a_00011`.

[Tiz+14] Michele Tizzoni et al.
"On the Use of Human Mobility Proxies for Modeling Epidemics".
In : *PLOS Computational Biology* 10.7 (juil. 2014), p. 1-15.

[Val+11] Nicholas C. Valler et al.
"Epidemic Spread in Mobile Ad Hoc Networks: Determining the Tipping Point".
In :
NETWORKING 2011: 10th International IFIP TC 6 Networking Conference, Valencia, Spain, May 9-13, 2011, Proceedings, Part I.
Sous la dir. de Jordi Domingo-Pascual et al.
Berlin, Heidelberg : Springer Berlin Heidelberg, 2011,
P. 266-280.

[Van+13] Philippe Vanhems et al.

"Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors".
In : *PLoS ONE* 8.9 (sept. 2013), e73970.

[Váz+06] Alexei Vázquez et al.
"Modeling bursts and heavy tails in human dynamics".
In : *pre* 73, 036127 (mar. 2006), p. 036127.
arXiv : `physics/0510117 [physics.soc-ph]`.

[VG15] Christian L. Vestergaard et Mathieu Génois.
"Temporal Gillespie Algorithm: Fast Simulation of Contagion Processes on Time-Varying Networks".
In : *PLOS Computational Biology* 11.10 (oct. 2015), p. 1-28.

[Wan+17] W. Wang et al.
"Unification of theoretical approaches for epidemic spreading on complex networks".
In : *Reports on Progress in Physics* 80.3, 036603 (mar. 2017), p. 036603.
arXiv : `1612.04216 [physics.soc-ph]`.

[WS98] Duncan J. Watts et Steven H. Strogatz.
"Collective dynamics of 'small-world' networks".
In : *Nature* (1998).

[WM00] Richard J. Williams et Neo D. Martinez.
"Simple rules yield complex food webs".
In : *Nature* 404 (mar. 2000), 180 EP -.

[WNW] LAM WK, ZHONG NS et TAN WC.
"Overview on SARS in Asia and the World".
In : *Respirology* 8.s1 (), S2-S5.
eprint : `https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1440-1843.2003.00516.x`.

[YS13] Junichi Yamanoi et Hiroki Sayama.
"Post-merger Cultural Integration from a Social Network Perspective: A Computational Modeling Approach".
In : *Comput. Math. Organ. Theory* 19.4 (déc. 2013), p. 516-537.

[Zac77] Wayne W. Zachary.
"An Information Flow Model for Conflict and Fission in Small Groups".
In : *Journal of Anthropological Research* 33.4 (1977), p. 452-473.

[ZAI10] Faraz ZAIDI.
"Analyse, Structure et Organisation des Réseaux Complexes".
Thèse de doct. UNIVERSITE BORDEAUX I, 2010.

[Zha+16] Jian-Xiong Zhang et al.
"Identifying a set of influential spreaders in complex networks".
In : *Scientific Reports* 6 (juin 2016), 27823 EP -.

[ZHA+15] ZHAOYANG ZHANG et al.
"Modeling Epidemics Spreading on Social Contact Networks".
In : *IEEE transactions on emerging topics in computing* 3.3 (sept. 2015), p. 410-419.

[Zha+15] Kai Zhao et al.
"Explaining the power-law distribution of human mobility through transportation modality decomposition".
In : *Scientific Reports* 5 (mar. 2015), 9136 EP -.

[Zhe+09] Yu Zheng et al.
"Mining Interesting Locations and Travel Sequences from GPS Trajectories".
In : *Proceedings of the 18th International Conference on World Wide Web*. WWW '09.
Madrid, Spain : ACM, 2009,
P. 791-800.

[ZK06] Changsong Zhou et Jürgen Kurths.
"Dynamical Weights and Enhanced Synchronization in Adaptive Complex Networks".
In : *Phys. Rev. Lett.* 96 (16 avr. 2006), p. 164102.

[ZX14] Yinzuo Zhou et Yingjie Xia.
"Epidemic spreading on weighted adaptive networks".
In : *Physica A: Statistical Mechanics and its Applications* 399 (2014), p. 16-23.

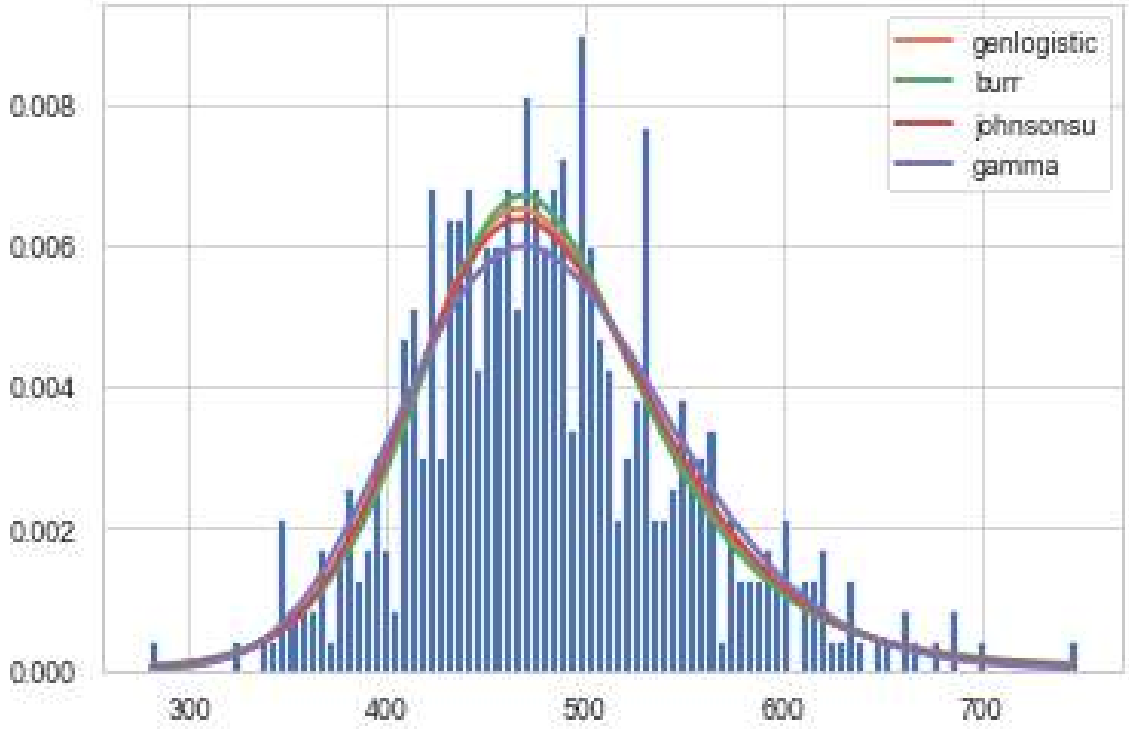
[Zip46] George Kingsley Zipf.
"The P1 P2/D hypothesis: On the intercity movement of persons".
In : *American sociological review* 11.6 (déc. 1946), p. 677-686.

Appendix A

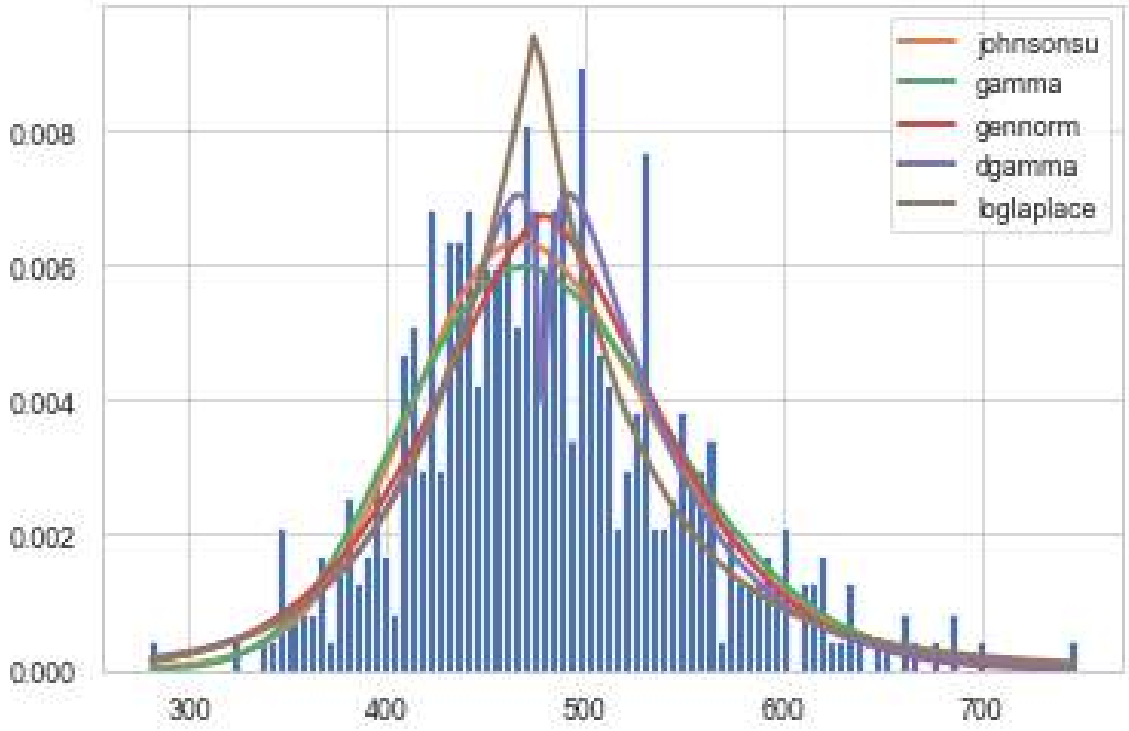
Dans cette partie, nous donnons le *fitting* de la distribution de probabilité de la centralité de degré et de proximité en utilisant la librairie Fitter [1] écrite en Python. Ce package utilise un ensemble de distribution pour *fitter* les données.

1. https ://pypi.org/project/fitter/

??



(a) La distribution de la centralité de degré des noeuds suivant le modèle GM



(b) La distribution de la centralité de degré des noeuds suivant le modèle STEPS

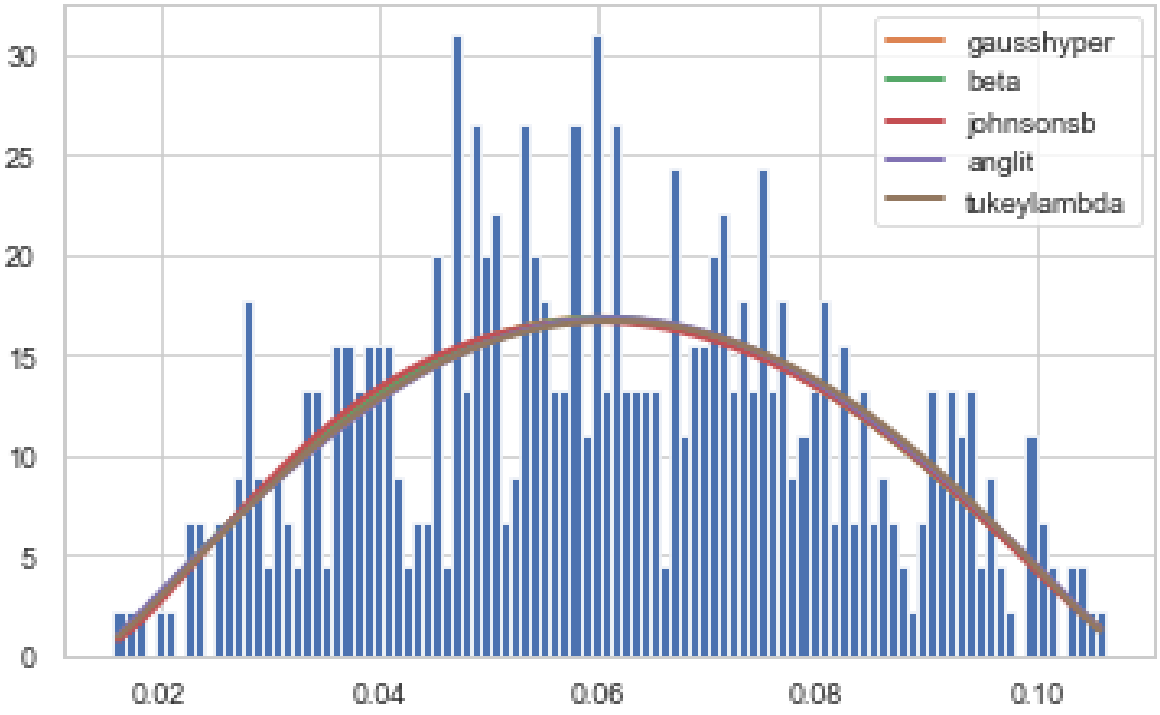


FIGURE 2: La distribution de la centralité de proximité des noeuds suivant le modèle RWP