

Un Peuple - Un But - Une Foi

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR,
DES UNIVERSITÉS ET DES CENTRES UNIVERSITAIRES RÉGIONAUX ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ GASTON BERGER DE SAINT-LOUIS



U.F.R DES SCIENCES APPLIQUÉES ET TECHNOLOGIES

Section Informatique

Mémoire de Master Recherche

Option : Informatique

présenté par

Djibril MBOUP

Structure et dynamique des réseaux complexes aléatoires

Sous la direction de Dr. Cherif DIALLO
Enseignant-chercheur à l'Université Gaston Berger

Année Universitaire 2012-2013

Remerciements

Nous remercions d'abord et avant tout ALLAH LE TOUT PUISSANT sans l'assistance de qui ce travail n'arriverait jamais à son terme;
Gloire à son Prophète Seydina Mouhamed, que la Paix soit sur lui et sur ses compagnons;
À notre vénéré guide Cheikh Ahmadou Bamba Khadimoul Rassoul.

Mes prochains remerciements s'adressent à mon encadrant Dr. Cherif Diallo d'avoir accepté d'encadrer ce travail, sans oublier M. Baldé pour son soutient et sa disponibilité. Je vous remercie infiniment pour votre générosité, pour vos précieux conseils méthodologiques. Je remercies tous les membres du jury pour l'intérêt qu'ils ont porté à mon travail et je leur exprime ma profonde gratitude.

Je tiens à remercier toute ma famille particulièrement à ma mère pour votre soutien inlassable sur tous les points. Sans vos prodigieux conseils, je n'arriverai pas à ce stade.
Merci à tous les professeurs de l'U.F.R Sciences Appliquées et technologies qui ont contribué à ma formation de la première année jusqu'en Master;
À tous mes camarades de promotion, particulièrement ceux du M2itic;
À tous mes amis;
Aux membres du Dahira Mafaatihoul Bichri de l'Université Gaston Berger de Saint-Louis.

Table des matières

Remerciements **i**

Introduction **1**

1 Préliminaire **3**
1.1 Historique des réseaux complexes . . . 3
1.2 Qu'est-ce qu'un réseau ? . . . 4
1.3 Pourquoi complexe ? . . . 5
1.4 Les réseaux réels . . . 5
1.4.1 Réseaux sociaux . . . 6
1.4.2 Réseaux d'informations . . . 6
1.4.3 Réseaux technologiques . . . 7
1.4.4 Réseaux biologiques . . . 8

2 Définitions et concepts de la théorie des graphes **9**
2.1 Graphe . . . 10
2.2 Degré d'un nœud . . . 11
2.3 Étude de la Connexité . . . 12
2.4 Plus court chemin (*shortest path*), diamètre . . . 13
2.5 Densité . . . 13

3 Propriétés des réseaux complexes **14**
3.1 L'effet petit-monde (*small-world effect*) . . . 14
3.2 Coefficient de *Clustering* . . . 16
3.3 Degré de distribution . . . 17
3.4 Corrélation de degré et Assortativité . . . 19
3.5 Centralité . . . 21
3.5.1 Centralité de degré (*degree centrality*) . . . 21
3.5.2 Centralité de proximité (*closeness centrality*) . . . 21
3.5.3 Centralité intermédiaire (*betweenness centrality*) . . . 22
3.6 Structure de communauté . . . 22

4 Modèles de réseaux **24**
4.1 Modèle d'Erdös et de Rényi (*random graphs*) 24
4.2 Modèle de Watts & Strogatz (*small-world*) 27
4.3 Modèle de Barabási et Albert (*scale-free*) 30

5 Dynamique des réseaux complexes **35**
5.1 Analyse des phénomènes dynamiques 35
5.2 Robustesse et vulnérabilité du réseau 36
5.3 Modélisation de la dynamique dans les réseaux mobiles 37
5.4 Simulation 40
5.4.1 Étude de la connectivité du graphe dans le temps 41
5.4.2 Peut-on classer le graphe obtenu par rapport aux modèles existants 43

Conclusion **44**

Bibliographie **45**

Table des figures

1.1 1.a) Réseau de contact sexuel du lycée de Jefferson From P.S. Bearman et al. AJS, 110, 44 (2004). 1.b) Représentation d'une chaine alimentaire des animaux de l'Arctique [10] 5
1.2 A gauche :le réseau de citation des papiers scientifiques, à droite : le réseau World Wide Web [20] 7
1.3 Distribution de degrés des nœuds observé dans certains réseaux réels. Ces distributions obéissent tous à une loi puissance (*power-law*). Extrait de [20] 8

2.1 Ponts de Königsberg 9
2.2 (a) Graphe non orienté, (b) graphe orienté, (c) graphe pondérés 10
2.3 Graphe biparti 11
2.4 Matrice d'adjacence d'un graphe 12
2.5 Graphes complets K_3, K_4, K_5 13

3.1 Six degrés de séparation 14
3.2 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 clusterings locaux est de $C_{1,2,3,4,5} = \{0, 0, 1, 1, \frac{1}{6}\}$, ce qui donnent la moyenne $C^2 = \frac{13}{30}$ 17
3.3 La distribution des degrés $P(k)$: nous avons la loi puissance en rouge et loi de poisson en bleu 18
3.4 Exemple du degré de corrélation d'un sommet i 20
3.5 Exemple de réseau montrant trois structures de communauté 23

4.1 Exemple du processus de création du modèle de ER Random Graph $G^{ER}_{(N,p)}$, extrait dans Réf [5] 25
4.2 Procédure de recablage de Watts et Strogatz. D'après [27] 28
4.3 Caractéristiques de la moyenne de longueurs l et du coefficient de clustering C dans le modèle de WS 29
4.4 Variante du modèle *small-world* de Newman, Moore et Watts. Extrait de [16] 30
4.5 Degré de distribution générée par le modèle de Barabási-Albert sur un réseau de N nœuds avec $m = 3$ 32
4.6 Graphe construit suivant le modèle d'attachement préférentiel d'Albert et Barabàsi 32

5.1 Attaque ciblé, le réseau n'est plus connexe 36
5.2 Visualisation de $G_t(V_0, E_t)$ aux instants $t = 10, t = 30, t = 50$ 40
5.3 En haut: densité du graphe des graphes obtenus en fonction du temps, en bas: nombre de composantes connexes des graphes obtenus en fonction du temps 42
5.4 En haut: le coefficient de clustering des graphes obtenus en fonction du temps, en bas : l'efficience globale des graphes obtenus en fonction du temps 42
5.5 Les degrés de distribution obtenus à $t = 50$ 43

Liste des tableaux

3.1 Exemple de mesures de corrélation de nœuds de réseaux réels extrait de Newman [19] où n indique le nombre de nœud et r l'assortativité 20

4.1 Comparaison entre les propriétés statistiques d'un graphe aléatoire et d'un réseau réel; avec N le nombre de nœuds, $\langle k \rangle$ le degré moyen, l la longueur moyenne des plus courts chemins, l_{rand} la longueur moyenne des plus courts chemins dans un modèle ER du même graphe, C le coefficient de clustering de l'eq 5.4, C_{rand} le coefficient de clustering dans un modèle ER du même graphe. Extrait de [5] 26

Introduction

Depuis très longtemps, l'étude des réseaux a constitué une grande importance dans les disciplines qu'elles soient scientifiques ou non. Cette étude a connu un essor surtout dans les domaines des sciences sociales, de la physique, de l'informatique et de la biologie moléculaire. Un réseau constitue une représentation abstraite des entités ou des éléments d'un domaine sous forme de *nœuds* ou points et leur relation ou interaction sous forme de liens aussi appelés *arêtes*. Plus connu dans le domaine des mathématiques sous le nom de *graphe*, un réseau permet, par exemple dans un réseau social, la représentation de la propagation des rumeurs dont les nœuds seraient les personnes et les liens leurs différentes relations ou bien Internet où les machines et les serveurs seraient les nœuds et les liens les câbles ou l'air.

Vers les années 1990, nous assistons à une explosion d'activités de recherche autour des réseaux dits « complexes » (structures irrégulières, complexes et évoluant dynamiquement dans le temps [9]) qui sont en général le résultat d'une évolution décentralisée et non planifiée. En effet, l'étude des réseaux complexes a été encouragée du fait de l'apparition récente 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 y ont introduit de nouveaux concepts pour mieux mimer sa structure et son évolution dynamique. Après plusieurs modèles proposés, la question concernant l'évolution dynamique des réseaux complexes reste encore un thème d'actualité dans le domaine de la recherche. C'est pour cette raison que nous essayons d'étudier, d'analyser et de comprendre les deux aspects que sont la structure et la dynamique dans les réseaux mobiles.

Ce rapport s'articule autour de cinq chapitres. Dans un premier temps, nous allons faire une présentation sur les réseaux complexes afin d'expliquer leurs origines et où est-ce que nous pouvons les trouver.
Dans le second chapitre nous présentons quelques notions de la théorie des graphes. Le

troisième chapitre est consacré aux propriétés essentielles pour la compréhension et l'analyse des réseaux complexes.
Nous allons abordés, dans le chapitre quatre, les différents modèles de graphes qui font que nous parlons de réseaux complexes : les *random graph*, les *small-world* et les *scale-free*.
Dans le chapitre cinq, nous parlerons de l'aspect dynamique d'un tel système et l'approche proposée pour la modélisation de la mobilité. Enfin, l'aspect technique sera étudié avec la simulation d'un réseau mobile. Nous essayons d'en tirer la réponse des questions suivantes : Quelle type de graphe a-t-on à l'instant t? Est-ce que la classe du graphe change dans le temps? Quels sont les paramètres caractéristiques de ce graphe qui décrivent de manière exhaustive son évolution?

Chapitre 1

Préliminaire

1.1 Historique des réseaux complexes

L'étude des réseaux complexes a connu une grande réputation dans la communauté scientifique durant cette dernière décennie. Les réseaux complexes font partie de ce qu'on appelle les *« systèmes complexes »*. Cependant, nous ne pourrons pas parler de réseaux complexes sans commencer par 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. Selon Newman dans [21], 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 de ses parties est plus que leurs différents comportements. Le comportement collectif est, parfois, appelé comportement « émergent ». 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. Nous pouvons citer comme exemples de ces systèmes : les systèmes de matière condensée, l'écosystème, les marchés financiers, le cerveau, le système immunitaire, le trafic routier, Internet etc. Des progrès substantiels qui ont permis à la compréhension quantitative des systèmes complexes, ont commencé à être enregistrés dans les années 1980 surtout dans le domaine de la physique.

La théorie des systèmes complexes peut être divisée entre deux approches [21]. La première concerne l'étude et la création de modèles mathématiques qui ne représentent pas exactement le comportement 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 qui peuvent être mis en simulation sur ordinateur. C'est dans ce cadre que les réseaux sont venus comme un élément incontournable dans la modélisation des systèmes complexes.

Un nouveau mouvement de recherche sur l'étude des réseaux complexes est né vers la fin des années 1990. De florissants travaux ont abouti à la découverte par Watts et Strogatz des graphes petit-monde (*small-world*), en 1998, et par Barabási et Albert des graphes sans échelles (*scale-free*), en 1999. Ces travaux constituent, aujourd'hui, un élément fondamental dans l'étude et la compréhension des réseaux complexes. Après ces pionniers, de nombreuses revues d'articles et de livres vont apparaitre dans la littérature des réseaux complexes avec des thèmes parlant aussi bien sur la structure que sur la dynamique de ces réseaux. D'ailleurs Dorogovtsev et Mendes dans leur article intitulé *Evolution of networks* expliquent comment est organisé la structure des *scale-free*, le rôle du mécanisme d'attachement préférentiel, la topologie et les propriétés structurelles des réseaux qui évoluent. Albert et Barabási revient sur le même sujet que Dorogovtsev et Mendes dans *Statistical mechanics of complex networks*. L'article de Newman : *Structure and Function of Complex Networks*, nous donne en détails les propriétés des réseaux et les différents modèles de réseaux. Ainsi, beaucoup d'autres chercheurs ont participé à l'amélioration du domaine et au perfectionnement des outils utilisés.

1.2 Qu'est-ce qu'un réseau ?

Le terme « réseau » peut avoir plusieurs définitions dans le domaine ou le contexte où nous nous plaçons. En effet, le terme réseau est utilisé pour représenter les systèmes tels que les réseaux sociaux, les circuits électriques, les réseaux économiques, les composants chimiques, les réseaux neuraux, les systèmes de transports, la propagation d'épidémie, les réseaux métaboliques, la chaine alimentaire (*food web*), Internet, le World Wide Web etc. Pour récapituler, le terme réseau désigne, en général, tout système admis, dans la représentation des mathématiques abstraites, comme un graphe dont les nœuds identifient les éléments du système et un ensemble de liens (arêtes) représentant la présence de relations ou d'interactions entre les éléments de ce système [3]. Par exemple dans la figure 1.a, on considère un réseau social de contact sexuel dont les nœuds représentent les individus (garçons et filles) et les liens leurs relations. De même, la figure 1.b indique un réseau de chaine alimentaire des animaux de l'Arctique dont les nœuds représentent les espèces et les liens la relation entre proie et prédateur.

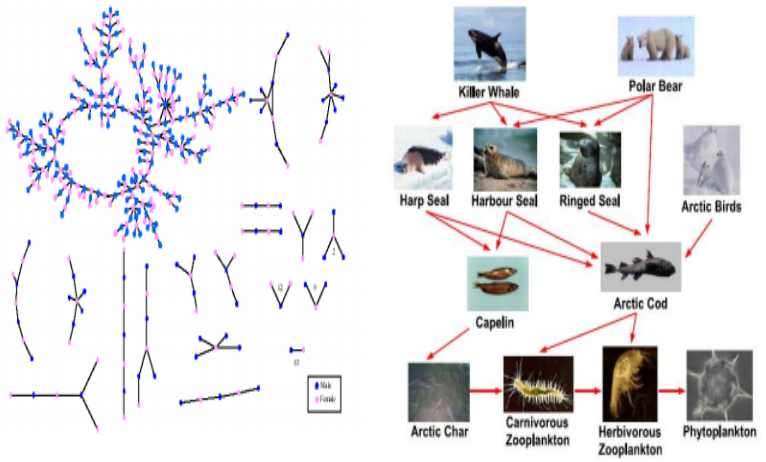


FIGURE 1.1 – 1.a) Réseau de contact sexuel du lycée de Jefferson From P.S. Bearman et al. AJS, 110, 44 (2004). 1.b) Représentation d'une chaine alimentaire des animaux de l'Arctique [10]

1.3 Pourquoi complexe ?

Après avoir défini le terme réseau, répondre à la question : pourquoi un réseau est-il complexe ? revient tout simplement à dire qu'il appartient à la grande famille des systèmes complexes. Ainsi, nous avons rencontré plusieurs définitions de réseau complexe dans la littérature, mais nous n'en retenons que quelques. Selon Newman un réseau complexe est donc 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 [20]. Dans un exposé intitulé *Applications of Complex Networks* [10] Buldú définit un réseau complexe comme un réseau avec des caractéristiques topologiques non triviales, avec un mode de connexion entre leurs éléments qui ne sont ni régulièrs ni purement aléatoires. Enfin, pour Boccaletti un réseau complexe est un réseau dont la structure est irrégulière complexe et qui évolue dynamiquement dans le temps [9].

1.4 Les réseaux réels

Les réseaux réels sont les réseaux du monde réel. Ces réseaux ont été étudiés par les chercheurs de différentes disciplines ces dernières années. Dans cette partie nous adoptons la catégorisation faite par Newman dans [20]. Ce qui nous amène à diviser ces réseaux en quatre catégories : *les réseaux sociaux*, les *réseaux d'informations*, les *réseaux techno-*

logiques et les *réseaux biologiques*.

1.4.1 Réseaux sociaux

Un réseau social est défini formellement comme un ensemble d'individus ou d'entités sociales qui sont liés à travers des interactions entre eux [9]. Ces interactions peuvent être de nature diverse : amicales, 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. La sociométrie a aussi contribué à l'essor de l'analyse des réseaux sociaux. L'apport empirique de la sociométrie est dû en partie à l'œuvre de Moreno, qui est perçu comme un des précurseurs de l'analyse des réseaux et de la psychologie sociale. L'expérience du psychologue social américain en 1967, Stanley Milgram, va annoncer ce qu'on appelle l'effet « petit-monde » qui sera détaillé plus tard. En outre, beaucoup de concepts fondamentaux et de méthodes utilisés récemment par les physiciens dans l'analyse des réseaux complexes ont leurs origines dans la sociométrie. Néanmoins, l'étude des réseaux sociaux souffre souvent de problèmes d'imprécisions, de subjectivité, et d'échantillons de petite taille [20]. 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 analysés par un certain nombre d'auteurs [2] [5] [4]. D'autres exemples de réseaux sociaux ont été étudiés par la communauté scientifique : les réseaux de co-auteurs dans une académie, les réseaux d'appel téléphonique, les réseaux d'email, les réseaux de contact sexuel, etc.

1.4.2 Réseaux d'informations

Notre seconde catégorie de réseau est les réseaux d'informations. 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 leur 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 B est cité dans A (figure 1.2). 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 existaient déjà. Redner (1998), en étudiant la distribution des citations de 783339 papiers catalogués par *the Institute of Scientifc Information* et de 24296 publiés

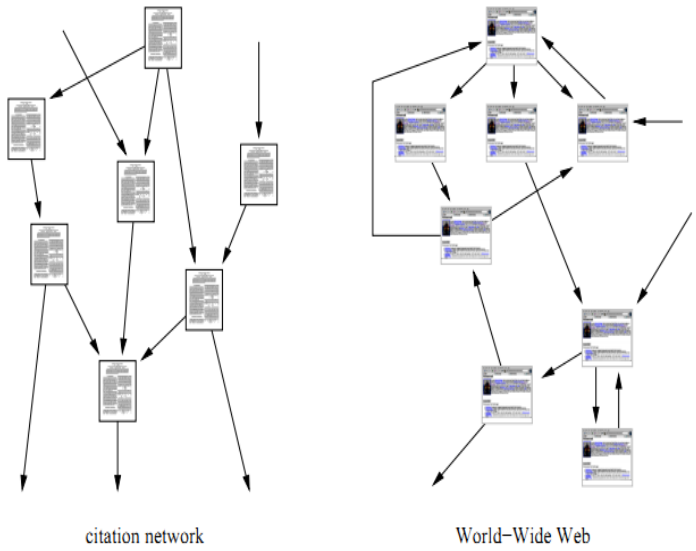


FIGURE 1.2 – A gauche :le réseau de citation des papiers scientifiques, à droite : le réseau World Wide Web [20]

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 ($p_k \sim k^{-\lambda}$) [5].

Un autre type de réseau d'informations est le World Wide Web avec ses pages web (contenant des informations) et ses hyperliens (à ne pas confondre avec le réseau Internet qui est le réseau physique reliant les ordinateurs du monde entier entre eux). Ce réseau contient aujourd'hui des milliards de nœuds. 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'apparence des boucles [20]. Les études sur le réseau WWW a commencé depuis la fin des années 1990 avec Albert, Jeong, Barabási (1999) et Kumar et *al.* (1999).

1.4.3 Réseaux technologiques

Un réseau technologique est un réseau créé par l'homme principalement pour la distribution d'un service ou d'énergie. Le premier réseau étudié dans cette catégorie est la topologie du réseau électrique du *Western States Power Grid* aux États Unis [27]. C'est un réseau de haute tension (*power grid*) qui permet la distribution électrique entre les États de l'Ouest. D'autres réseaux de transport ont été étudiés, y compris le réseau aérien, le trafic routier, le trafic ferroviaire.

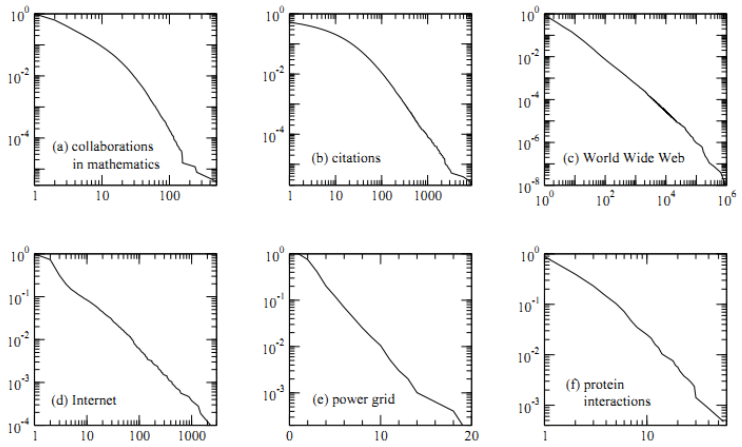


FIGURE 1.3 – Distribution de degrés des nœuds observé dans certains réseaux réels. Ces distributions obéissent tous à une loi puissance (*power-law*). Extrait de [20]

Il y a un autre réseau technologique qui attire plus notre attention: Internet. Internet est un réseau de machines et d'autres appareils de télécommunication qui sont interconnectés entre eux par des câbles ou l'air. La topologie d'Internet est étudiée en deux niveaux différents : au niveau des routeurs où les routeurs 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 représenté par un seul nœud et il existe une arête si deux domaines sont reliés par un routeur [5]. 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 [17].

1.4.4 Réseaux biologiques

Enfin la dernière catégorie est constituée des réseaux biologiques. Ils concernent tous les réseaux liés aux vivants. L'exemple le plus classique parmi les réseaux biologiques 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 de cette réaction agit sur les *substrats*. En outre, il existe d'autres classes importantes de réseaux biologiques qui ont toujours intéressé la communauté: les réseaux d'interaction entre protéines, les réseaux de régulateur génétique, les réseaux neuraux, les réseaux de chaine alimentaire ou *food web* (figure: 1.a) etc.

Chapitre 2

Définitions et concepts de la théorie des graphes

Historiquement, la théorie des graphes est née en 1936 quand Léonard Euler démontra qu'il est impossible de traverser chacun des sept ponts de la ville de Königsberg (aujourd'hui Kaliningrad) une fois exactement et de revenir au point de départ. Les ponts enjambent les bras du fleuve Pregel qui coulent de part et d'autre de l'île de Kneiphoff. Euler simplifie ce problème en représentant cette situation à l'aide d'un « dessin » où les sommets représentent les terres et les arêtes les ponts comme le montre la figure 2.1. Il démontre que, pour que cette traversée soit possible il faut que le nombre de ponts soit pair, or le nombre de ponts de Königsberg est impair.

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 chimie (modélisation de structures), la biologie (génome), les sciences sociales (modélisation des relations) ou en vue d'applications industrielles (problème du voyageur de commerce). En plus de son existence purement mathématique, le graphe est aussi une structure puissante de représentation de problèmes informatiques.

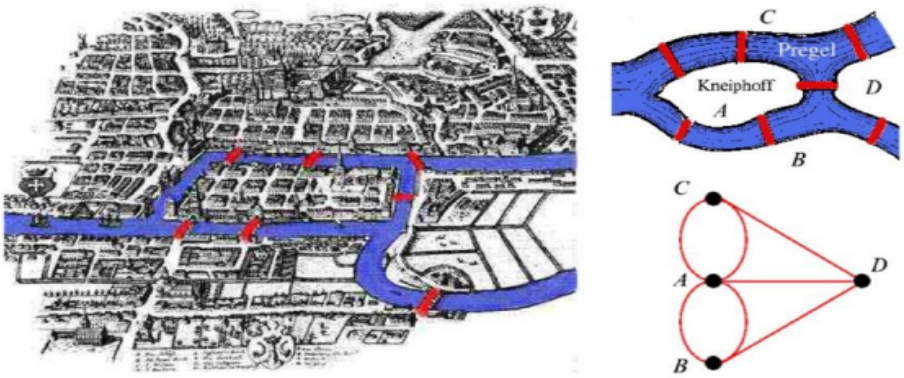


Figure 2.1 – Ponts de Königsberg

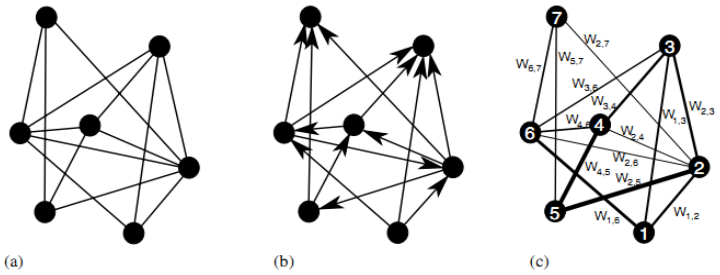


FIGURE 2.2 – (a) Graphe non orienté, (b) graphe orienté, (c) graphe pondérés

2.1 Graphe

Un *graphe* G est 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 graphes.

Une arête e du graphe est une paire $e = (v_i, v_j)$ de sommets. Les sommets v_i et v_j sont les extrémités de l'arête. On dit que e est incidente en v_i et en v_j, et que v_j est voisin de (ou adjacent à) v_i (et vice versa). L'arête (v_i, v_i) est appelé *boucle*

Cette définition correspond au graphe **non orienté** (*undirected graph*).
Si les arêtes sont orientés, on dit que le graphe est **orienté** (*directed graph*). Dans ce cas, on parle d'arc $e = (v_i, v_j)$, v_i est l'extrémité initiale, v_j l'extrémité finale (ou bien origine et destination).
Si les arêtes (ou arcs) sont quantifiées par une valeur ou poids, on parle de **graphe pondéré** (*weighted graph*).
La figure 2.2 donne en illustration des exemples de graphes orientés, non orientés et pondérés.

Un multigraphe est un graphe pour lequel il peut exister plusieurs arêtes entre deux sommets.

Un graphe $G = (V, E)$ est **simple** :
- s'il n'est pas un multigraphe
- s'il n a pas de boucle.

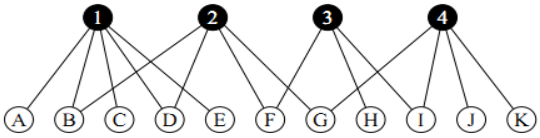


FIGURE 2.3 – Graphe biparti

Un graphe biparti (*bipartite graph*) 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 (voir 2.3).

La matrice d'adjacence (*adjacency matrix*) du graphe $G = (V, E)$ est une représentation matricielle qui est exactement équivalente au graphe.
Soit $A = (a_{ij})$, la matrice d'adjacence de G de dimension $n \times n$ telle que

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

La figure 2.4 montre un exemple de graphe et sa matrice d'adjacence correspondante.

2.2 Degré d'un nœud

Le degré d'un nœud i est le nombre de liens adjacents à ce nœud. Le degré d'un nœud i noté par k_i est donné par l'expression :

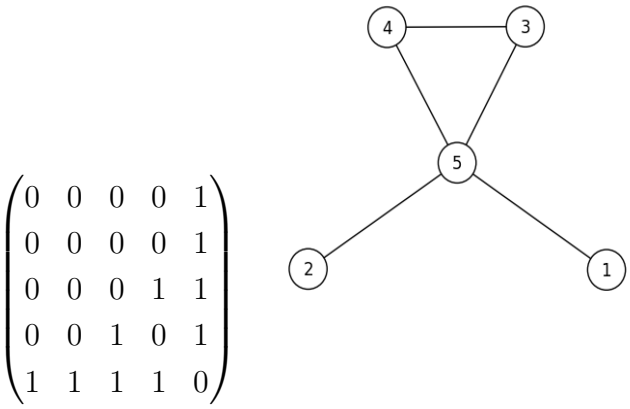
$$k_i = \sum_{j \in V} a_{ij} \text{ ; avec } a_{ij} \in A \tag{2.1}$$

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

$$k_i^{in} = \sum_{j \in V} a_{ij} \tag{2.2}$$

Le degré entrant (ou intérieur) k_i^{out} est le nombre d'arcs ayant i comme extrémité finale.

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



$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

FIGURE 2.4 – Matrice d'adjacence d'un graphe

2.3 Étude de la Connexité

Chaine et cycle

Une *chaine* 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'encadre dans la séquence. Un *cycle* est une chaine dont les extrémités coïncident.

Une chaine est *élémentaire* si tous les sommets qui la composent sont tous distincts. Une chaine est *simple* si toutes les arêtes qui la composent sont tous distinctes.

Chemin et circuit

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 chemin est *élémentaire* si tous les sommets qui le composent sont tous distincts. Un chemin est *simple* si tous les arcs qui le composent sont tous distincts.

Connexité

Un graphe $G = (V, E)$ est *connexe* si $\forall i, j \in V$ il existe une chaine entre i et j.
Une *composante connexe* est un sous-ensemble de sommets qui est connexe. Un graphe est connexe s'il comporte une composante connexe maximale et une seule.

Forte connexité

Un graphe $G = (V, E)$ est *fortement connexe* si $\forall i, j \in V$ il existe un chemin entre i et j.
Une *composante fortement connexe* (cfc) est un sous-ensemble de sommets tel qu'il existe

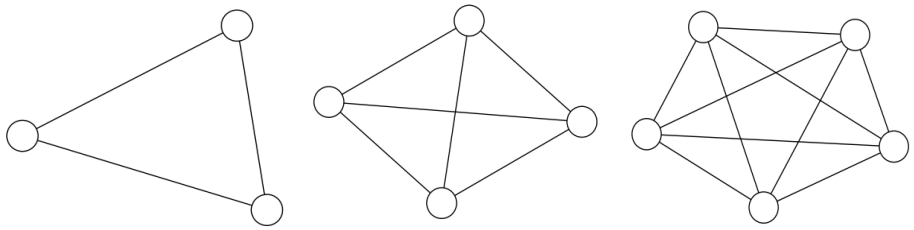


FIGURE 2.5 – Graphes complets K_3, K_4, K_5

un chemin entre deux sommets quelconques. Une cfc maximale (cfcm) est un ensemble maximal de cfc. Les différentes cfcm définissent une partition de V. Un graphe est fortement connexe s'il comporte une seule cfcm.

Remarque: Graphes complets

Les *graphes complets* sont les graphes dont tous les sommets sont « reliés » 2 à 2. Ce sont des graphes qui sont toujours connexes (voir 2.5). Si G est un graphe complet avec un nombre de nœuds égal à N alors le nombre de liens $K = N(N-1)/2$. G est noté K_N. Un *clique* est un sous graphe complet.

2.4 Plus court chemin (*shortest path*), diamètre

Plus court chemin (*shortest path*)

Le *plus court chemin* d_{ij}, ou encore appelé la *distance géodésique*, entre i et j est la distance minimale entre tous les chemins qui connecte i et j.

Diamètre

Le *diamètre* d'un graphe G est le maximum entre tous les *plus courts chemins* du graphe.

$$Diam(G) = \max(d_{ij}) \tag{2.4}$$

2.5 Densité

La *densité* d'un graphe est le coefficient entre nombre d'arêtes existantes et le nombre d'arêtes possibles.

$$densite = \frac{\text{nbre d'arêtes existantes}}{\text{nbre d'arêtes possibles}} \tag{2.5}$$

Chapitre 3

Propriétés des réseaux complexes

Les concepts de théorie des graphes étudiés dans la Section 2 nous permettent d'avoir une vue sur les réseaux complexes mais ils ne suffisent pas pour faire l'analyse de tels réseaux. Il a fallu aux chercheurs de ce domaine d'inventer d'autres moyens plus adaptés pour compléter l'étude fondée sur la théorie des graphes. La structure d'un réseau complexe est un peu particulière à cause de sa grande dimension et des nombreuses interactions qui influent sur sa caractéristique. Dans cette partie nous allons définir un certain nombre de propriétés qui nous permettra de mieux comprendre la topologie d'un réseau complexe.

3.1 L'effet petit-monde (*small-world effect*)

L'effet petit-monde est né de l'expérience du psychologue social américain Stanley Milgram. 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 donnent 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

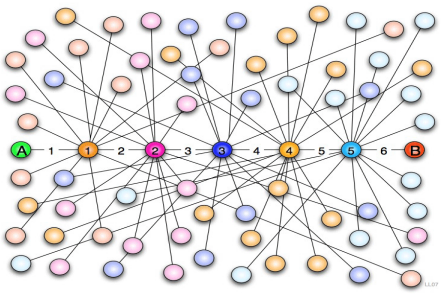


Figure 3.1 – Six degrés de séparation

personnes qu'ils jugeaient le plus susceptible de connaitre 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 [29]. Dans la littérature, ce concept fait souvent référence à la moyenne des longueurs (*Average path length*).

Formellement, 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 :

$$l = \frac{1}{N(N-1)} \sum_{\substack{i,j \\ i \neq i}} d_{ij} \tag{3.1}$$

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} [14].

$$E_{glob}(G) = \frac{1}{N(N-1)} \sum_{\substack{i,j \\ i \neq i}} \frac{1}{d_{ij}} \tag{3.2}$$

Ainsi, tout couple de nœuds appartenant à des composantes déconnectés du graphe contribue à la sommation dans $E_{glob}(G)$ à une valeur égale à zero. Nous avons, par définition, $0 \leq E_{glob}(G) \leq 1$. En outre, il a été montré par Latora et Marchiori que sa grande valeur permet de caractériser l'effet petit-monde [14].

La moyenne des longueurs est une donnée très utile dans l'analyse des réseaux complexes. On a découvert qu'elle présente une valeur faible aussi bien dans les réseaux aléatoires que les dans les réseaux réels (voir le chapitre 4).

3.2 Coefficient de *Clustering*

Le coefficient de *« clustering »*, encore appelé *Transitivité* [9], 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 [29]. 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 [20].
Dans la littérature, deux définitions du coefficient de clustering sont utilisés : 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 :

$$C^1 = \frac{3 \times \text{nombre de triangles dans G}}{\text{nombre de triplets connectés dans G}} \tag{3.3}$$

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 [20] :

$$C^1 = \frac{6 \times \text{nombre de triangles dans G}}{\text{nombre de chemins de longueur deux}} \tag{3.4}$$

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.

$$C_i = \frac{\text{nombre de triangles connectés à } i}{\text{nombre de triplets centrés sur } i} \tag{3.5}$$

Cette définition est donnée par Watts et Strogatz dans [27]. 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 3.5 autrement .

$$C_i = \frac{2 \times e_i}{k_i(k_i - 1)} \tag{3.6}$$

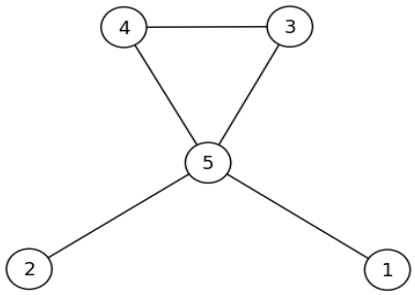


FIGURE 3.2 – Illustration du coefficient de *clustering*: $C^1 = 3 \times \frac{1}{8} = \frac{3}{8}$ car nous avons un triangle et huit triplets alors que l'ensemble des clusterings locaux est de $C_{1,2,3,4,5} = \{0, 0, 1, 1, \frac{1}{6}\}$, ce qui donnent la moyenne $C^2 = \frac{13}{30}$

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.

$$C^2 = \frac{1}{N} \sum_i C_i \tag{3.7}$$

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 5.4. Mais les détails de cette différence sont expliqués dans [20]. D'autres extensions du coefficient de clustering sont fait dans les graphes pondérés, par exemple dans l'article de Tore Opsahl et Pietro Panzarasa [26]. Une autre mesure alternative des propriétés de clustering d'un graphe G est *l'efficience locale* définie par:

$$E_{loc} = \frac{1}{N} \sum_{i \in V} E(G_i) \tag{3.8}$$

où $E(G_i)$ est l'efficience de G_i évalué par la formule 3.2.

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 [29].

3.3 Degré de distribution

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

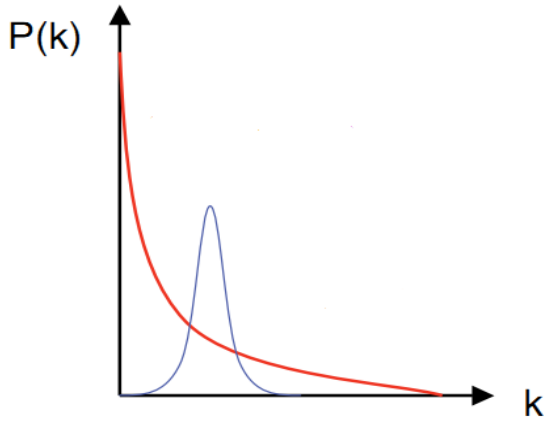


FIGURE 3.3 – La distribution des degrés $P(k)$: nous avons la loi puissance en rouge et loi de poisson en bleu

k sur le nombre total de nœud d'un graphe. Le terme p_k signifie également la probabilité qu'un nœud choisi de manière aléatoire est 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* (voir la figure 5.5). Le degré de distribution $P(k)$ satisfait la condition de normalisation $\sum_{k=0}^{\infty} P(k) = 1$.

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 est défini comme étant:

$$\langle k^n \rangle = \sum_k k^n P(k) \tag{3.9}$$

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 de connectivité 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 [9].

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 ($p_k \sim k^{-\lambda}$) [20]. 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*.

3.4 Corrélation de degré et 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 [9]:

$$kP(k^{'}|k) = k^{'}P(k|k^{'}) \tag{3.10}$$

Si un graphe est non corrélé, $P(k^{'}|k)$ ne dépend plus de k. D'après la normalisation et l'équation 3.10, on peut obtenir facilement la relation suivante :

$$P(k^{'}|k) = \frac{k^{'}P(k^{'})}{\langle k \rangle} \tag{3.11}$$

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:

$$k_{nn,i} = \frac{1}{k_i} \sum_{j=1}^{N} a_{ij} k_j \tag{3.12}$$

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'eq. 3.12, nous pouvons définir $k_{nn}(k)$ ainsi

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i/k_i=k} k_{nn,i}\,; \tag{3.13}$$

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 :

$$k_{nn}(k) = \sum_{k} k^{'}P(k^{'}|k) \tag{3.14}$$

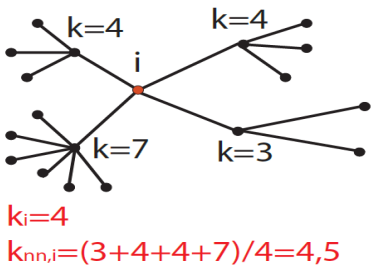


FIGURE 3.4 – Exemple du degré de corrélation d'un sommet i

	network	n	r
real-world networks	physics coauthorship	52 909	0.363
	biology coauthorship	1 520 251	0.127
	mathematics coauthorship	253 339	0.120
	film actor collaborations	449 913	0.208
	company directors	7 673	7 673
	Internet	10 697	-0.189
	World-Wide Web	269 504	-0.065
	protein interactions	2 115	-0.156
	neural network	307	-0.163
	food web	92	-0.276

TABLE 3.1 – Exemple de mesures de corrélation de nœuds de réseaux réels extrait de Newman [19] où n indique le nombre de nœud et r l'assortativité

Si le graphe n'est pas corrélé alors la formule 3.14 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 réseau sociaux (popularité). En revanche, un graphe est *« disassortative »* si $k_{nn}(k)$ décroit 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.

3.5 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 apport à d'autres [11]. 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*).

3.5.1 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

$$C_D(i) = \frac{1}{N} \sum_j a_{ij} \tag{3.15}$$

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 [23].

3.5.2 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 :

$$C_C(i) = \frac{N-1}{\sum_{j \in V \setminus \{i\}} dist(i,j)} \tag{3.16}$$

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 [11]. 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.

3.5.3 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 [11]. Elle est définie par la relation suivante:

$$C_B(i) = \sum_{j,k \in V, j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \tag{3.17}$$

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 i à j.

3.6 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 diverses 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. 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 apport au reste du réseau (Girvan Newman, 2002). Cette définition met l'accent sur un élément essentiel

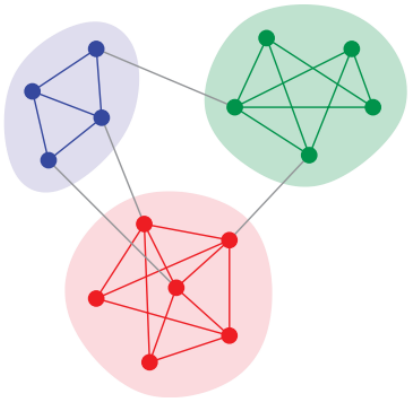


FIGURE 3.5 – Exemple de réseau montrant trois structures de communauté

qui est la forte densité de connexion des nœuds internes comparée aux nœuds externes comme l'indique la figure 3.5. 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 [11].

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 celles qui maximisent la modularité Q [1] (qualité d'une partition) introduite par Girvan et Newman [15].

Dans cette partie, nous pouvons nous limiter là pour la compréhension du reste de l'exposé mais il existe d'autres propriétés qui sont étudiées par la communauté des réseaux complexes. Parmi ces propriétés, nous pouvons citer le spectre d'un graphe, les motifs de Milo et *al.*, la navigabilité de Kleinberg etc.

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

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

Chapitre 4

Modèles de réseaux

Le graphe était et demeure, depuis très longtemps, un outil puissant dans la modélisation des réseaux du monde réel. Différentes sortes de modèles ont été proposées par la communauté pour mieux analyser les propriétés statistiques de ces réseaux. Les trois grandes classes de modèle les plus utilisés sont: le modèle d'Erdös et de Rényi (*random graphs*), le modèle de Watts & Strogatz (*small-world*) et le modèle de d'Albert et Barabási (*scale-free*)

4.1 Modèle d'Erdös et de Rényi (*random graphs*)

L'étude des graphes aléatoires ou *random graphs* a été initiée par Erdös et de Rényi en 1959 dans leur premier article intitulé *On random graphs*. Dans cet article, 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 4.1. 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 $G^{ER}_{(N,E)}$. 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 :

$$\langle k \rangle = \frac{N(N-1)p}{N} = (N-1)p \simeq Np \tag{4.1}$$

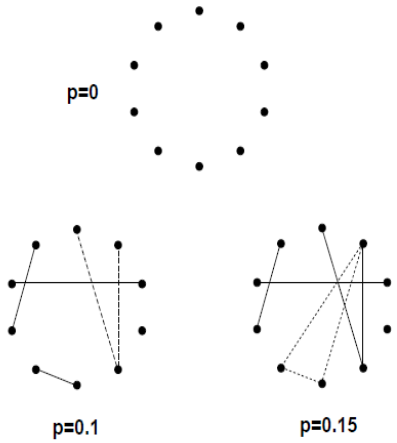


FIGURE 4.1 – Exemple du processus de création du modèle de ER Random Graph $G^{ER}_{(N,p)}$, extrait dans Réf [5]

Erdös et de 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 trouvé plus tard par Bollobás (1981) [5]. Dans un *random graph*, la probabilité qu'un nœud i a un degré $k = k_i$ suit une loi binomiale de paramètres $N-1$ et p.

$$P(k = k_i) = C^k_{N-1} p^k (1-p)^{N-1-k} \tag{4.2}$$

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 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

$$P(k = k_i) \simeq \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle} \tag{4.3}$$

C'est la raison pour laquelle cette distribution homogène est souvent appelé *Poisson Random Graph*. Pourtant, cette distribution est tout à fait différente de celle des réseaux réels dont la distribution suit une loi puissance.

L'étude des ER Random Graph révèle une propriété intéressante qui est le phénomène des valeurs critiques. Erdös et de Rényi prouvent qu'en fonction d'une valeur seuil de probabilité $p_c = \frac{1}{N}$, correspondant au degré moyen critique $\langle k \rangle_c = 1$:

- Si $p < p_c$, alors il est sûr avec une forte probabilité que le graphe n'a pas de composante de taille supérieure au plus de $O(\ln N)$ et aucune composante n'a un nombre de cycles supérieur à un;

Network	N	$\langle k \rangle$	l	l_{rand}	C	C_{rand}	Référence
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic 1999
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts, Strogatz 1998
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook et al. 2001
Power grid	4 941	2.67	18.7	12.4	0.08	0.005	Watts, Strogatz 1998
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts, Strogatz 1998

TABLE 4.1 – Comparaison entre les propriétés statistiques d'un graphe aléatoire et d'un réseau réel; avec N le nombre de nœuds, $\langle k \rangle$ le degré moyen, l la longueur moyenne des plus courts chemins, l_{rand} la longueur moyenne des plus courts chemins dans un modèle ER du même graphe, C le coefficient de clustering de l'eq 5.4, C_{rand} le coefficient de clustering dans un modèle ER du même graphe. Extrait de [5]

- Si $p = p_c$, alors la structure du graphe change brutalement et le graphe devient soudainement composé d'une composante géante ayant une taille de $O(N^{2/3})$;
- Si $p > p_c$, alors le graphe a sûrement une composante de taille $O(N)$ avec un nombre $O(N)$ de cycles et la taille des autres composantes est en $O(\ln N)$ avec au plus un cycle;

Les graphes aléatoires reproduisent bien l'une des principales caractéristiques des réseaux réels discutée dans la section 3.1, 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 (voir Tableau 4.1). Ce modèle présente aussi un degré de distribution de type *"Poissonniène"* qui est contraire à celui des réseaux réels. En plus, il n y a pas de corrélation entre le degré des nœuds adjacents ni de structure de communauté [20].

Modèle configurationnel

Ce modèle, introduit par Bender and Canfield (1978), est une généralisation des *ER Randoms Graphs*. Il permet de générer un graphe aléatoire avec un degré de distribution contrôlé (pas forcement Poissonniène) sans autre corrélation non souhaitée. Les améliorations qu'apporte ce modèle permettent de mieux satisfaire la représentations des réseaux

réels. D'ailleurs, la formalisation de ce modèle ne sera faite plus tard que par Molloy et Reed (1995).
La méthode de construction consiste à prendre une sequence de degré c'est-à-dire un ensemble $D = \{k_1, k_2, ..., k_N\}$ tel que $\sum_i k_i = 2E$; où E est le nombre de liens. En fait, cet ensemble sera choisi de manière à ce que la fraction de nœuds ayant comme degré k tende vers la distribution de degrés désirée $P(k)$ au fur et à mesure que N augmente. Une fois que chaque nœud possède sa séquence de degré, la méthode pour générer le graphe est la suivante : on donne à chaque nœud i un nombre k_i de "points de connexion" (un point de connexion correspond à l'extrémité d'une arête), et on choisit des paires de ces points de connexion de manière aléatoire uniforme pour joindre les arêtes. Quand tous les points de connexion ont été utilisés, le graphe résultant est un membre choisi de façon aléatoire parmi l'ensemble des graphes possédant la séquence du degré désiré [15].

Malgré la distribution hétérogène des degrés (c'est-à-dire on peut faire de sorte que la distribution des degrés suit une loi puissance), ce modèle présente toujours un coefficient de clustering faible mais aussi il ne possède pas de structure de communauté.

4.2 Modèle de Watts & Strogatz (*small-world*)

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*) [27]. 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 recablage des liens selon une probabilité p (voir figure 4.2). 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 triviale.

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 $C(p)$

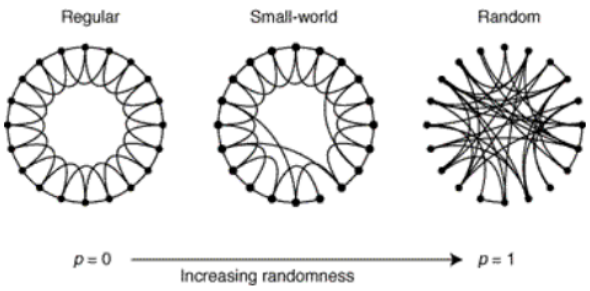


FIGURE 4.2 – Procédure de recablage de Watts et Strogatz. D'après [27]

et de la moyenne des longueurs $l(p)$ en fonction de la probabilité de recablage 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 p donné (voir figure 4.3) [27]. 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 replacées de manière aléatoire, sur un intervalle $0 \ll p \ll 1$, 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 présentent dans les réseaux réels [5].

Les propriétés statistiques de ce modèle ont, très tôt, intéressé les scientifiques. À propos du coefficient de clustering, Barrat et Weigt (2000) ont montré qu'il peut être obtenu facilement par la relation suivante :

$$C(p) \simeq C(0)(1-p)^3 = \frac{(3m-3)}{(4m-2)}(1-p)^3 \tag{4.4}$$

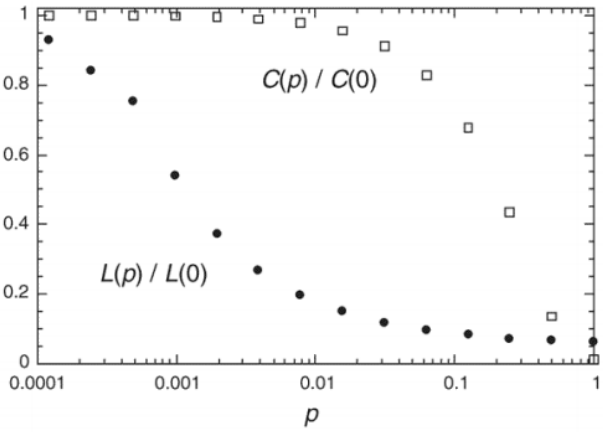


FIGURE 4.3 – Caractéristiques de la moyenne de longueurs l et du coefficient de clustering C dans le modèle de WS

En ce qui concerne la longueur moyenne des plus courts chemins l, il n'y a pas encore une valeur exacte trouvée mais des solutions approximatives ont été faites. Barthélémy et Amaral (1999) conjecturent que l satisfait la relation suivante:

$$l(p) = \xi g(N/\xi) \tag{4.5}$$

avec

$$g(x) \sim \begin{cases} x & \text{si} \quad x \ll 1 \\ \ln(x) & \text{si} \quad x \gg 1 \end{cases}$$

où ξ est une longueur de corrélation qui dépend de p et $g(x)$ est une fonction qui ne dépend seulement que de la dimension du système et de la géométrie du réseau mais pas de N, ξ ou p [20].

La distribution des degrés du modèle de Watts et Strogatz ne correspond pas à la plupart de celle des réseaux du monde réels, mais ce n'est pas surprenant car ce n'était pas l'objectif initial de ce modèle [20]. Cette distribution, selon Barrat et Weigt (2000), peut être écrite de la forme suivante:

$$P(k) = \sum_{i=0}^{min(k-m,m)} \binom{m}{i} (1-p)^i p^{1-i} \frac{(pm)^{k-m-i}}{(k-m-i)!} e^{-pm} \tag{4.6}$$

Malgré les nombreuses avancées qu'il apporte dans l'étude des réseaux complexes, le modèle original proposé par Watts et Strogatz présente quelques difficultés. Le fait qu'une seule des extrémités d'une arête soit redirigée, qu'aucun nœud ne puisse être connecté

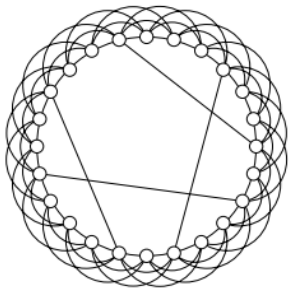


FIGURE 4.4 – Variante du modèle *small-world* de Newman, Moore et Watts. Extrait de [16]

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 [16]. Ce modèle reproduit le modèle de WS à la seule différence qu'ici les arêtes ne sont plus replacées mais de nouvelles arêtes servant de « raccourcis » sont ajoutées entre deux nœuds choisis suivant une probabilité p (voir figure 4.4). De ce fait, d'intéressants résultats concernant le coefficient de clustering $C(p)$ et la moyenne des longueurs $l(p)$ émergent du modèle de Newman, Moore et Watts [16].

$$C(p) = \frac{3(k-1)}{2(2k-1) + 4kp(p+2)} \tag{4.7}$$

$$l(p) = \frac{N}{k} f(Nkp) \tag{4.8}$$

avec

$$f(x) = \frac{1}{2\sqrt{x^2+2x}} \tanh \sqrt{\frac{x}{x+2}}$$

4.3 Modèle de Barabási et Albert (*scale-free*)

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. Il découvrit étonnamment que, 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.

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 [20].

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*. 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 nœuds 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.

$$\prod_{j \to i} = \frac{k_i}{\sum_l k_l} \tag{4.9}$$

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 4.5). Selon les auteurs, l'origine de la loi puissance est essentiellement due à l'attachement préférentiel qui entraine la présence d'un faible groupe de nœuds de fortes connexions appelés « hubs ». Cette faible présence de « hubs » a pour conséquence la résistance au panne mais laisse le réseau vulnérable aux attaques ciblées.
Par ailleurs, la distribution du modèle de Barabási-Albert peut être obtenue analytiquement par différentes méthodes : *" mean-field approximation "*(Barabási et Albert, 1999; Barabási, Albert et Jeong, 1999), *" rate equation "* (Krapivsky, Redner et Leyvraz, 2000), *" master equation "* (Dorogovtsev, Mendes et Samukhin, 2000).

Concernant la moyenne des longueurs l (*Average path length*) de ce modèle, des expériences ont montré qu'elle est plus faible dans les *scale-free* que dans les *random graphs* [5]. l est donné par la relation suivante (Bollobás et Riordan,2003; Cohenand et Havlin,

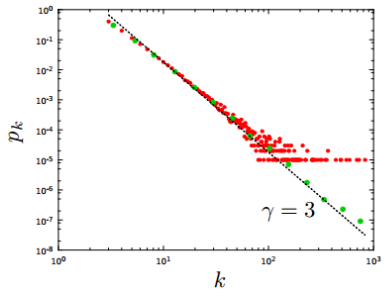


FIGURE 4.5 – Degré de distribution générée par le modèle de Barabási-Albert sur un réseau de N nœuds avec $m = 3$

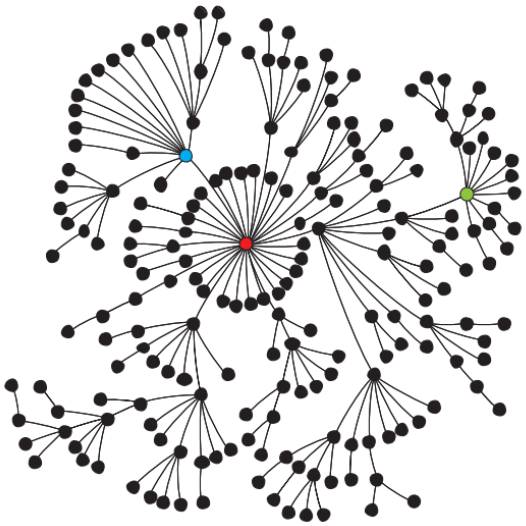


FIGURE 4.6 – Graphe construit suivant le modèle d'attachement préférentiel d'Albert et Barabàsi

2003):

$$l \sim \frac{logN}{loglogN} \tag{4.10}$$

Le coefficient de clustering du modèle de BA est calculé par Klemm et Eguiluz (2002) et prouvé par Bollobás and Riordan (2003).

$$C = \frac{m-1}{8}\frac{(\ln N)^2}{N} \tag{4.11}$$

Le modèle de Barabási-Albert est l'un des modèles de réseau évolutif et dynamique (ajout de nouveaux nœuds) qui présentent une caractéristique longtemps observée dans les réseaux réels: la distribution des degrés en loi puissance. Cette caractéristique découle

principalement de l'attachement préférentiel.

Ces dernières années, le modèle de Barabási-Albert a beaucoup attiré l'attention dans la littérature, ce qui fait qu'un grand nombre de recherches ont été mené dans le sens où des modifications ou des extensions ont été apportées sur ce modèle. Ceux-ci altèrent le comportement du modèle original ou le rendent un peu plus réaliste vis-à-vis des réseaux réels, car même si ce modèle parait pertinent, nous obtenons un modèle de réseau sous forme d'arbre (voir figure 4.6), alors que les réseaux réels n'arborent presque jamais cette forme.
Parmi les variantes du modèle de Barabási-Albert, Krapivsky et Redner considèrent la version orienté de ce modèle. Ils ont étudié l'âge a d'un nœud i, défini comme le nombre de nœuds ajouter 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. La méthode est basée sur l'algorithme itérative suivante : a) m nouveau nœuds sont ajoutés avec une probabilité p; b) m arêtes sont recablées 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) 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.

En dehors des modèles de bases évoqués ci-dessus, plusieurs modèles ou variantes ont été proposés. N'empêche, vu la complexité des réseaux réels, aucun de ces modèles ne prend en compte toutes les propriétés d'un tel système. Appréhender l'aspect topologique et l'aspect dynamique dans un modèle reste encore un défi à relever. Pour le moment,

certaines simulations donnent des résultats plus ou moins pertinentes dans un cadre bien défini, qui peuvent être étendus dans d'autres domaines.

Chapitre 5

Dynamique des réseaux complexes

La plupart des réseaux complexes sont le siège de phénomènes dynamiques. Par exemple, Internet et le World Wide Web voient, chaque jour, leurs nœuds croître exponentiellement et supportent, en plus, un flux continu d'informations; les réseaux sociaux sont le siège de formation d'opinions ou de propagation d'informations, ainsi que d'épidémies; les réseaux d'infrastructures de communication tels que les téléphones mobiles ou les réseaux *ad hoc* permettent de collecter des milliers d'informations par jour. La mise en évidence du caractère hétérogène (distribution des nœuds du réseau en *loi puissance*) de la topologie des réseaux a donc naturellement mené à la question de l'influence de telles caractéristiques sur ces phénomènes dynamiques [7]. En effet, la principale problématique est de proposer des modèles qui permettront de modéliser la dynamique locale et globale d'un réseau complexe

5.1 Analyse des phénomènes dynamiques

La dynamique d'un système correspond à son évolution dans le temps. Nous pouvons résumer les phénomènes dynamiques d'un réseau complexe en trois grandes catégories: l'évolution dynamique du réseaux (état *intrinsèque* des nœuds et des liens), la dynamique sur le réseau et la dynamique inter-agissant entre la topologie du réseau et les processus qui s'y évoluent [12].
D'une part, la topologie des réseaux réels est loin d'être fixée, le nombre de nœuds et de liens change ensemble en fonction des propriétés locales et globale du système. En particulier, les principes évolutifs du réseau sont les ingrédients nécessaires pour expliquer certaines propriétés particulières du réseau. Par exemple, l'attachement préférentiel est un principe nécessaire pour comprendre l'émergence de l'hétérogénéité des degrés dans les réseaux comme Internet et le World Wide Web [5].
D'autres part, les réseaux sont des structures où se déroulent des processus dynamiques, il est donc intéressant d'étudier le comportement des modèles de systèmes dynamiques évoluant sur ces réseaux. Beaucoup d'entre eux, comme la propagation de rumeurs ou

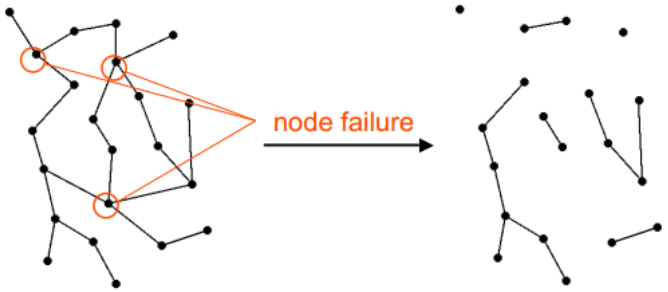


FIGURE 5.1 – Attaque ciblé, le réseau n'est plus connexe

d'épidémies, la diffusion, les algorithmes de routage, les oscillations, etc, ont des applications directes dans l'étude des phénomènes dynamiques observés sur les réseaux réels. Par exemple, nous pouvons étudier pour prédire, contrôler ou optimiser la propagation d'une épidémie ou d'un feux de forêt. La marche aléatoire, les modèles de la physique statistique, la théorie de la percolation (c'est-à-dire l'étude du comportement d'un graphe lorsqu'on retire une fraction de nœuds ou de liens au hasard) fournissent des informations plus générales qui peuvent être utilisées pour construire un cadre théorique commun dans lequel les différentes propriétés des processus dynamiques sur les réseaux peuvent être analysées et expliquées.
Enfin, la troisième situation, caractérisée par l'interaction entre la dynamique « des réseaux » et la dynamique « sur les réseaux » est plus compliquée et ce genre de problèmes a été récemment considéré par la communauté des réseaux complexes. D'ailleurs, ce cas a été d'habitude négligé en raison des différentes échelles temporelles des deux types de dynamiques. Nous pouvons supposé que les propriétés structurelles d'un réseau évoluent avec un temps τ_T alors qu'un processus qui se déroulent sur le réseau avec un temps τ_D; la situation mentionnée ci-dessus correspond au cas où $\tau_T \simeq \tau_D$. Lorsque $\tau_T \gg \tau_D$, l'évolution des processus dynamiques est étudiée sur des réseaux à structure topologique « trempée » tandis que le cas où $\tau_D \gg \tau_T$ signifie que l'évolution temporelle des processus sur le réseau est négligée par rapport à celle de la structure du réseau lui-même.

5.2 Robustesse et vulnérabilité du réseau

Il a été compris tout d'abord empiriquement que la structure très hétérogène des réseaux complexes implique une grande résistance aux pannes, c'est-à-dire à un phénomène où des sites deviennent inopérants et sont donc retirés du réseau de manière aléatoire, mais en même temps une grande fragilité devant des attaques ciblées sur les sites mieux

connectés. Pour un réseau hétérogène diverses études se sont intéressées à l'évolution de la taille de la plus grande composante connexe en fonction de la fraction de nœuds qui sont supprimés. Si des nœuds sont retirés au hasard, le réseau peut rester connexe mais la dégradation de la taille du réseau survivant entraîne, par exemple, la diminution de la capacité de transmission. Ceci est dû au faite que de telles pannes touchent typiquement des sites faiblement connectés (car ce sont les plus nombreux), plutôt périphériques; l'intégrité du réseau est dans ce cas peu remise en question. Au contraire, l'élimination d'une fraction des sites les plus connectés conduit très rapidement la désintégration du réseau en petites composantes isolées (voir figure 5.1). En effet, ces sites de fort degré mettent en communication de nombreuses parties du réseau et leur suppression a des conséquences beaucoup plus importantes. Il est possible de comprendre quantitativement cette phénoménologie par une approche physique statistique, en traduisant en terme de problème de percolation. Les fluctuations de connectivité apparaissent clairement comme responsables de cette ambivalence entre *robustesse* et *vulnérabilité* des réseaux complexes.

5.3 Modélisation de la dynamique dans les réseaux mobiles

Modéliser un réseau dynamique n'est pas une tâche facile. Parce qu'il faut prendre en compte les différentes catégories de phénomènes dynamiques précités dans la section 5.1. C'est pour cela, dans notre cas nous ne prenons en compte que l'évolution dynamique du réseau. En effet, notre objectif est de pouvoir créer des graphes dynamiques représentant des modèles de réseaux mobiles et d'en déduire leur nature et leur comportement dans le temps.

Un réseau mobile peut être représenté par un graphe dynamique dont les nœuds sont les stations mobiles et les liens la "connexion" entre deux stations. Contrairement à un réseau statique, la structure topologique d'un réseau mobile change en fonction du temps. Ainsi, un réseau mobile peut être vu de manière abstraite comme une séquence de graphes $G_t(V_t, E_t)$ dans le temps. Comme les stations mobiles sont en mouvement, il faut à chaque instant t étudier la connectivité du réseau. Dans notre étude, nous considérons qu'il existe un lien entre deux nœuds du réseau si la distance qui les sépare est inférieure à une certaine valeur R.

Mobilité

Nous ne pouvons pas parler de réseau mobile sans parler de la notion de mobilité. La mobilité consiste à mettre en mouvement dans un environnement connu l'ensemble des stations mobiles constituant ce réseau. En effet, à un instant t deux stations peuvent être en communication et à un instant $t + \delta t$, la communication peut être rompue à cause de leur déplacement. Pour simuler la mobilité dans les réseaux mobiles, il y a un certain nombre de modèles de mobilité qui sont utilisés:

Random Walk. Ce modèle fut proposé pour simuler le mouvement imprévisible de particules en physique. Les physiciens y font d'ailleurs référence en parlant de « mouvement brownien ». C'est un modèle itératif qui fait évoluer toutes les stations à chaque étape. À son tour, une station choisit une direction et une vitesse de manière aléatoire et se rend à sa nouvelle destination. La distribution du choix de l'angle (entre 0 et 2π) et de la vitesse (entre 0 et v_{vmax}) suit une distribution gaussienne [22].

Random Waypoint (RWP). Ce modèle proposé par J. BROCH et *al.* (1998) est devenu le principal modèle de mobilité utilisé pour les simulations dans les publications académiques, en particulier pour tester des algorithmes de routage. Le principe dérive du Random Walk. Itérativement les stations se déplacent avec une vitesse constante choisie aléatoirement de manière uniforme, vers une destination elle aussi choisie aléatoirement dans l'espace de simulation. Une fois à destination, la station attend pendant une durée (un nombre d'étapes) choisie entre 0 et w_{max} [22]. A propos de ce modèle, Bettstetter (2002) propose la fonction suivante décrivant la densité de probabilité de présence de stations mobiles.

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

Gauss-Markov. Ce modèle propose, pour calculer la vitesse au temps t , de s'appuyer sur la vitesse au temps $t - 1$ mais aussi sur la vitesse moyenne de la station considérée et sur une valeur aléatoire uniforme. Un paramètre α permet de donner plus ou moins d'importance à ces trois composantes pour calculer la nouvelle vitesse de la station. Les auteurs notent que si le paramètre α est réglé à zéro, alors la formule ne tient plus compte de la vitesse à $t - 1$ et le modèle est équivalent au Random Walk. Si au contraire α vaut 1, la vitesse de la station est constante tout au long de la simulation et le modèle se rapproche des mécanismes de diffusion en physique. Tout réglage de α entre 0 et 1 mène à un comportement plus réaliste que le Random Walk évitant les cas extrêmes où la vitesse est très rapide à l'étape i et subitement très lente à l'étape $i + 1$ [22]. Plus

précisément, la valeur de la vitesse et la direction au n-ième intervalle de temps sont calculées en fonction des valeurs de la vitesse et de la direction du (n-1)-ième intervalle de temps et d'une variable aléatoire, en utilisant les équations suivantes:

$$s_n = \alpha.s_{n-1} + (1-\alpha).\overline{s} + \sqrt{(1-\alpha^2)s_{n-1}}$$

$$d_n = \alpha.d_{n-1} + (1-\alpha).\overline{d} + \sqrt{(1-\alpha^2)d_{n-1}}$$

où s_n et d_n sont les nouvelles vitesse et direction du n?ud mobile à l'instant n; $\overline{s}$ et $\overline{d}$ sont des constantes représentant la valeur moyenne de la vitesse de la direction quand n tend vers l'infini; et $_{n-1}$ et d_{n-1} sont des variables aléatoires d'une distribution de Gauss. La position d'un nœud mobile est donnée par les équations suivantes :

$$x_n = x_{n-1} + s_{n-1}\cos d_{n-1}$$

$$y_n = y_{n-1} + s_{n-1}\cos d_{n-1}$$

où (x_n, y_n) et (x_{n-1}, y_{n-1}) sont respectivement les coordonnées x et y de la position du nœud au n-ième et (n-1)-ième intervalle.

La liste n'est pas exhaustive car notre ambition est de présenter quelques modèles de mobilité de manière à discuter la façon dont ils influent sur la nature du graphe.

Méthodologie

Un réseau dynamique peut être modéliser par un ensemble de « *snapshot* » de graphe statique $G_t(V_t, E_t)$ à des instants t avec $t = \{1, 2, ..., N\}$. Il faut d'abord préciser que le nombre de nœuds sera fixe tout au long de l'étude, donc $V_t = V_0$. Ainsi, le but du travail est de modéliser un graphe dynamique permettant de représenter un ensemble d'utilisateurs mobiles interconnectés entre eux. Considérons un espace euclidien $E(x, y)$ où chaque nœud sera représenté par ses coordonnées (x, y). A chaque instant t, $G_t(V_0, E_t)$ sera obtenu en exécutant les étapes suivantes

- Les nœuds sont générés de manière aléatoire dans l'espace de simulation selon l'un des algorithmes de mobilité suivante: *Random Walk*, *Random Waypoint (RWP)* ou *Gauss-Markov*.
- Deux nœuds sont interconnectés si la distance (distance euclidien) qui les sépare est inférieure à une certaine valeur R.

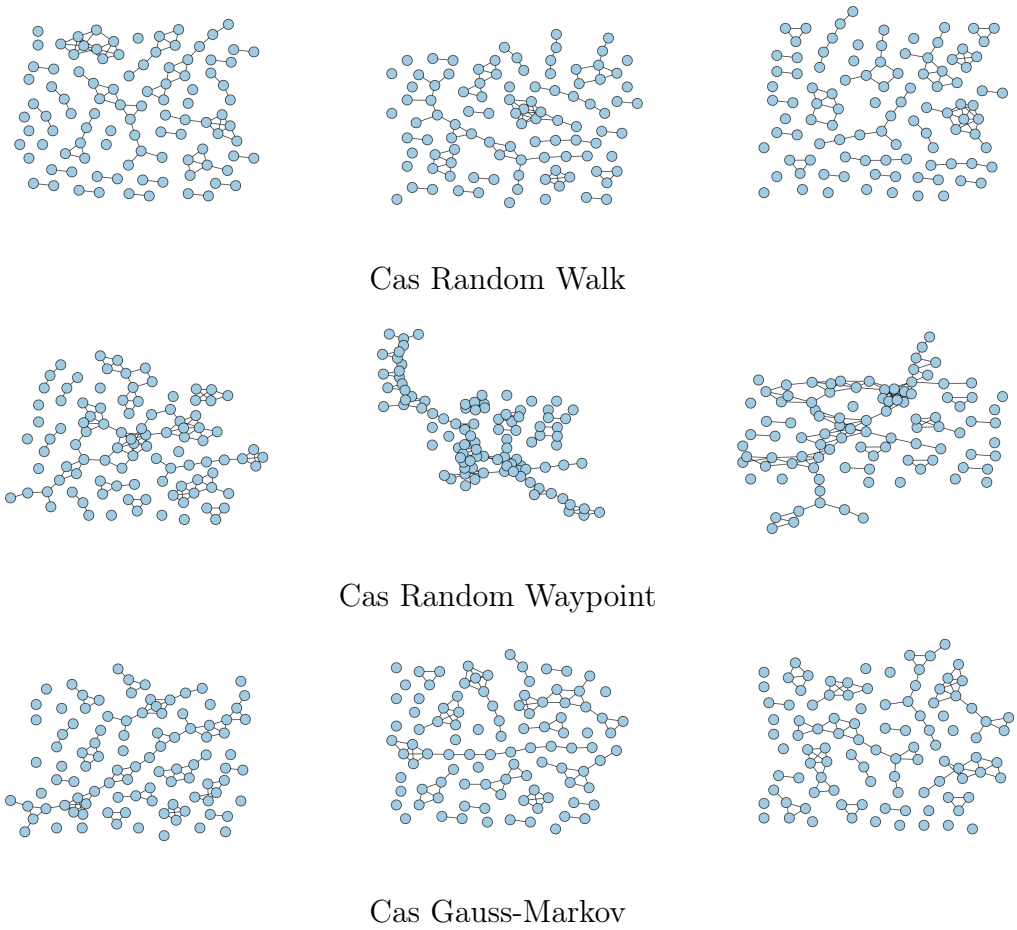


FIGURE 5.2 – Visualisation de $G_t(V_0, E_t)$ aux instants $t = 10, t = 30, t = 50$

Pour comprendre et décrire la nature de ce graphe, nous avons besoins d'étudier le comportement dans le temps des propriétés suivantes : la densité, le nombre de composante connexe N_{Comp}, la distribution des degrés $P(k)$, le coefficient de clustering C, la moyenne des longueurs l ou l'efficience globale E_{glob}.

5.4 Simulation

Dans cette partie, nous allons abordé la partie technique. Pour simuler les réseaux complexes, nous avons fait appel au langage Python. Python est un langage de programmation puissant très sollicité dans le calcul scientifique. Il fournit de nombreux modules, libres et gratuits, qui sont essentiellement destinés à la simulation de problème scientifique et de leur visualisation. Parmi ces modules, nous nous intéressons à ceux qui sont

destinés à la représentation des réseaux et de leur visualisation:

- **Networkx** est un module Python qui permet de créer, de manipuler et d'étudier la structure, le dynamique et les fonctions des réseaux complexes. En plus, il dispose des algorithmes précis de calcul des propriétés des graphes, ensuite il offre des fonctions de génération de graphes (graphes complets, *Erdös Rényi (ER) Random Graph*, *Watts Strogatz Graph*, *scale-free*, etc.) [6].
- **Matplotlib** est un excellent module de visualisation 2D intégré dans le langage Python. Il propose des interfaces comme **pylab** et **pyplot** qui s'inspirent fortement des fonctions d'affichage de MATLAB et fournissent un ensemble de fonctions très simples permettant d'afficher des courbes, des images, ou des champs de vecteurs à partir des tableaux **numpy** (module python permettant d'utiliser efficacement les tableaux et les matrices) [13].
- **pymobility** est un module développé par André Panisso. Il fournit un ensemble de fonctions permettant de simuler certains modèles de mobilité existants (Random Walk, RWP, Gauss-Markov, Truncated Levy Walk). Chaque fonction retourne la position des nœuds à chaque instant $t + \delta t$. Ce module diminue vraiment la peine d'implémenter les algorithmes de mobilité.

Rappelons d'abords que, pour simuler l'évolution dynamique d'un c'est tout à fait naturel de prendre en compte la dynamique des nœuds (c'est-à-dire des nœuds apparaissent et disparaissent au fil du temps) et la dynamique des liens (de même la création et la suppression de liens). Mais, nous nous concentrons sur la dynamique des liens car les algorithmes de mobilité utilisés prennent en entrée le nombre de nœuds fixe. Dans toute l'expérience, nous travaillons avec un nombre de nœuds égal à 100.

5.4.1 Étude de la connectivité du graphe dans le temps

D'après la figure 5.3, la première observation faite sur les résultats des simulations est la faible densité du graphe et sa fluctuation au cours du temps causées par des événements propres à l'expérimentation tels les types de mobilité utilisés (voir). Nous remarquons que dans tous les modèles de mobilité choisis, la densité ne dépasse pas 5 %. Ceci a une forte conséquence sur la connexité du graphe comme le montre la figure 5.3. Si nous nous referons au cas *Random Walk*, le nombre de composantes connexes peut aller jusqu'à 44. Cependant, calculer la moyenne des longueurs $l(t)$ (*Average path length*) ne sera pas possible, donc dans le reste de la simulation, pour étudier l'effet petit-monde, nous allons considérer l'efficience globale E_{glob}.

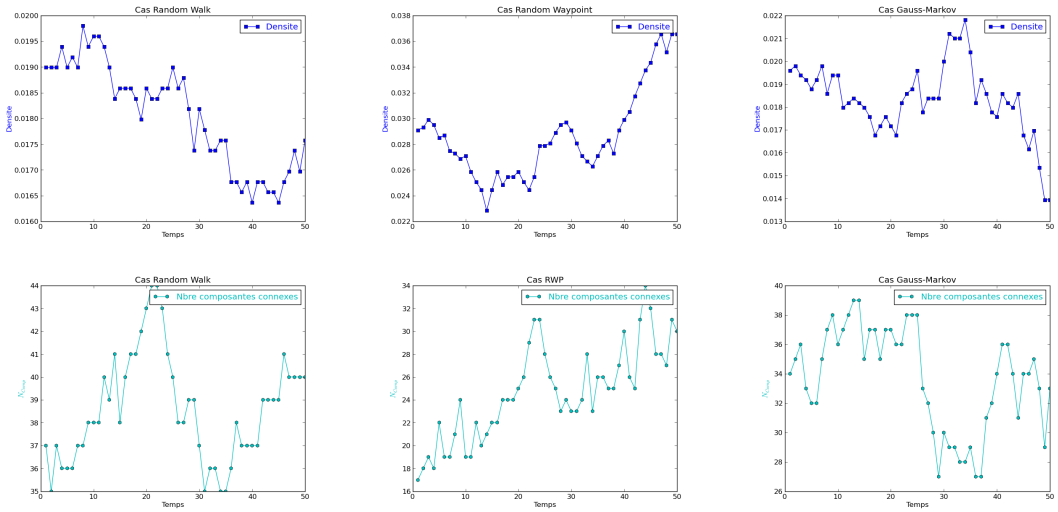


FIGURE 5.3 – En haut: densité du graphe des graphes obtenus en fonction du temps, en bas: nombre de composantes connexes des graphes obtenus en fonction du temps

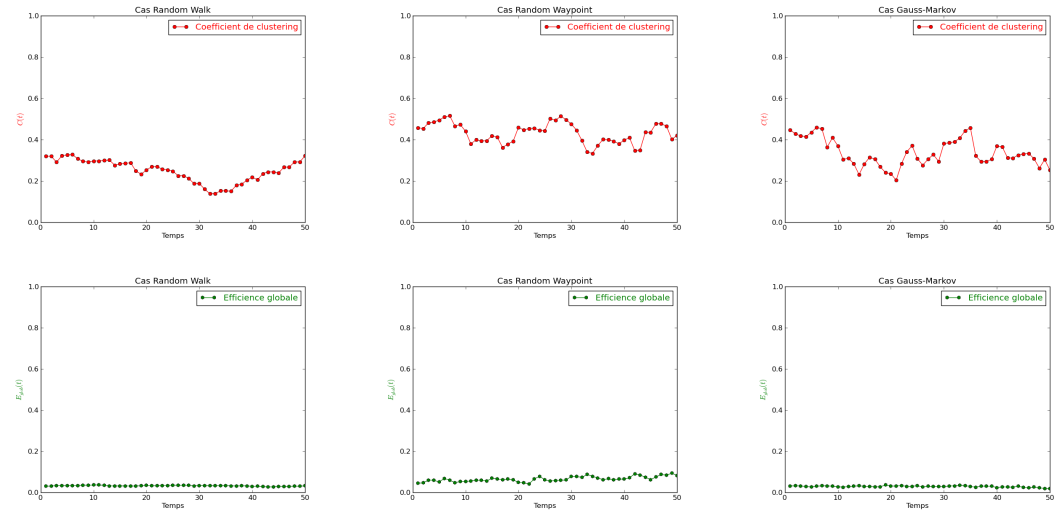


FIGURE 5.4 – En haut: le coefficient de clustering des graphes obtenus en fonction du temps, en bas : l'efficience globale des graphes obtenus en fonction du temps

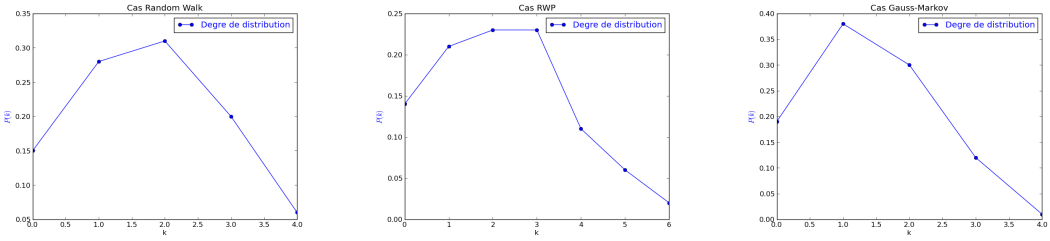


FIGURE 5.5 – Les degrés de distribution obtenus à $t = 50$

5.4.2 Peut-on classer le graphe obtenu par rapport aux modèles existants

Dans notre contexte, classer un graphe signifie: est ce que le graphe est petit-monde (*small-world*), sans échelle (*scale-free*), les deux ou autre? Les résultats empiriques obtenus montrent, d'abords, que le coefficient de clustering est considérable dans tous les modèles de mobilité utilisés. La figure 5.4 montre que cette valeur est d'autant plus important dans le cas RWP que dans les cas Random Walk et Gauss-Markov. Malgré sa structure non connexe, il y a forte fréquence de *triads* comme observé dans les réseaux sociaux. Ce phénomène est une caractéristique des graphes petit-monde.

Par contre, les mesures obtenues sur l'efficience globale ne sont pas importantes, elles oscillent entre 0.02 et 0.1. Elles sont presque nulles dans les cas Random Walk et Gauss-Markov. Ceci ne nous permet pas de conclure que les graphes obtenus présentent des effets petit-monde. Nous constatons que cette faible valeur de E_{glob} est dû probablement à la faible connexité des graphes. Ce phénomène se conserve toute la durée de la simulation.

Après plusieurs *run*, les degrés de distribution des graphes dynamiques obtenus sont tous de la forme de ceux observés dans les graphes aléatoires. La figure 5.5 en est une parfaite illustration. Cette distribution n'est pas hétérogène comme la plupart des réseaux réels, elle s'explique par la distribution que suit l'emplacement des nœuds déterminé par les algorithmes de mobilité considérés. En plus, le phénomène d'attachement préférentiel n'est pas prise en compte l'or de la création des liens.
Par conséquent, les graphes obtenus peuvent être classés dans la catégorie des graphes aléatoires avec la particularité d'avoir un coefficient de clustering non négligeable.

Conclusion

La modélisation des réseaux dits complexes est un problème délicat du fait de la complexité qui peut intervenir sur sa structure et sa dynamique. Cependant, beaucoup d'efforts ont été réalisés ces derniers années.

Dans ce rapport nous avons tenté d'apporter notre modeste contribution à la modélisation d'un ensemble d'utilisateurs mobiles. Ces utilisateurs sont interconnectés entre-eux si la distance qui les sépare est inférieur à une certaine distance. Nous avons parvenu à mettre en simulation ce type de réseau qui fournit un certain nombre de résultats intéressantes. En effet, la structure de notre modèle dépend essentiellement de la mobilité du réseau. C'est pour cela, nous souhaitons l'améliorer en y ajoutant d'autres propriétés comme l'attachement préférentiel.

Dans les perspective, une prochaine étape sera d'appliquer notre modélisation à de très grands réseaux et d'en estimer ainsi la limite, en terme de complexité et d'analyse. En plus, nous projetons valider ce modèle à partir des données réels.

Bibliographie

[1] P. Borgnat A. Scherrer and al. Description and simulation of dynamic mobility networks. *Elsevier*, 2008.

[2] L. A. Adamic and B. A. Huberman. Power-law distribution of the world wide web. *Science 287*, 2000. 6

[3] Marc Barthelemy Alain Barrat and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008. 4

[4] R. Albert and A. L. Barabási. Dynamics of complex systems: Scaling laws for the period of boolean networks. *Phys. Rev. Lett. 84*, 2000. 6

[5] R. Albert and A. L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 2002. iv, vi, 6, 7, 8, 25, 26, 28, 31, 35

[6] Daniel A. Schult Aric A. Hagberg and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, August 2008. 41

[7] Alaint Barrat. Les reseaux complexes et physique statistique. *Archicube*, 2010. 35

[8] Gil Zussman Berk Birand, Murtaza Zafer and Kang-Won Lee. Dynamic graph properties of mobile networks under levy walk mobility. *X*, 20XX.

[9] S. Boccaletti and al. Complex networks: Structure and dynamics. *Physics*, 2006. 1, 5, 6, 16, 18, 19

[10] Javier M. Buldú. *Applications of Complex Networks*, 2012. iv, 5

[11] Nacim Fateh CHIKHI. *Calcul de centralité et identification de structures de communautés dans les graphes de documents*. PhD thesis, UNIVERSITÉ DE TOULOUSE, 2010. 21, 22, 23

[12] Luca DALL'ASTA. *Phénomènes dynamiques sur les réseaux complexes*. PhD thesis, UNIVERSITÉ PARIS 11, 2006. 35

[13] John D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, May-Jun 2007. 41

[14] Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *PHYSICAL REVIEW LETTERS*, 2001. 15

[15] Vincent LEVORATO. *Contributions à la Modélisation des Réseaux Complexes : Prétopologie et Applications*. PhD thesis, Université Paris 8, 2008. 23, 27

[16] C. Moore M. E. J. Newman and D. J. Watts. Mean-feld solution of the small-world network model. *Physical Review Letters*, 2000. iv, 30

[17] P. Faloutsos et C. Faloutsos M. Faloutsos. On power-law relationships of the internet topology. *In ACM SIGCOMM Computer Communication Review*, 1999. 8

[18] Thang N. Dinh Nam P. Nguyen, Yilin Shen and al. The evolution of overlapping communities in dynamic mobile networks. *IEEE TRANSACTION ON MOBILE COMPUTING*, 20XX.

[19] M. E. J. Newman. Assortative mixing in networks. *arXiv:cond-mat/0205405v1*, 2002. vi, 20

[20] M. E. J. Newman. The structure and function of complex networks. *[cond-mat.stat-mech]*, 2003. iv, 5, 6, 7, 8, 16, 17, 18, 26, 29, 31

[21] M. E. J. Newman. Complex systems: A survey. *Am. J. Phys.*, 2011. 3

[22] Yoann PIGNÉ. *Modélisation et Traitement Décentralisé des Graphes Dynamiques - Application aux Réseaux Mobiles AD HOC*. PhD thesis, UNIVERSITÉ DU HAVRE, 2008. 38

[23] Massoud SEIFI. *Cœurs stables de communautés dans les graphes de terrain*. PhD thesis, Université Pierre et Marie Curie, 2012. 21

[24] Chita R. Das Sunho Lim, Chansu Yu. Clustered mobility model for scale-free wireless networks. *X*, 20XX.

[25] Lionel TABOURIER. *Méthode de comparaison des topologies de graphes complexes. Applications aux réseaux sociaux.* PhD thesis, UNIVERSITÉ PIERRE ET MARIE CURIE, 2010.

[26] Pietro Panzarasa Tore Opsahl. Clustering in weighted networks. *Preprint submitted to Social Networks*, 2009. 17

[27] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 1998. iv, 7, 16, 27, 28

[28] Wikipedia. Analyse des reseaux sociaux, 2013.

[29] Faraz ZAIDI. *Analyse, Structure et Organisation des Réseaux Complexes*. PhD thesis, UNIVERSITE BORDEAUX I, 2010. 15, 16, 17