



HAL
open science

Etude de la robustesse des graphes sociaux émergents

Slimane Lemmouchi

► **To cite this version:**

Slimane Lemmouchi. Etude de la robustesse des graphes sociaux émergents. Autre [cs.OH]. Université Claude Bernard - Lyon I, 2012. Français. NNT : 2012LYO10098 . tel-00944441

HAL Id: tel-00944441

<https://theses.hal.science/tel-00944441>

Submitted on 10 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Etude de la Robustesse des Graphes Sociaux Emergents

THÈSE

présentée et soutenue publiquement le 26 juin 2012

pour l'obtention du

Doctorat de l'Université Claude Bernard Lyon 1, UCBL
(Spécialité Informatique)

par

Slimane Lemmouchi

Directeurs de thèse : Hamamache Kheddouci & Mohammed Haddad

Composition du jury

<i>Rapporteurs :</i>	M. Raphaël Couturier	Professeur, Université de Franche-Comté (France)
	M. Olivier Togni	MCF - HDR, Université de Bourgogne (France)
<i>Examineurs :</i>	M. Hocine Chérifi	Professeur, Université de Bourgogne (France)
	M. Stéphane Lallich	Professeur, Université Lumière Lyon 2 (France)
<i>Directeur de thèse :</i>	M. Hamamache Kheddouci	Professeur, Université Claude Bernard Lyon 1
<i>Co-directeur de thèse :</i>	M. Mohammed Haddad	MCF, Université Claude Bernard Lyon 1

Mis en page avec la classe thloria.

Remerciements

Presque quatre années viennent de s'écouler depuis le début de ma thèse. Quatre années pendant lesquelles Mon Directeur de thèse, Le Professeur Hamamache Kheddouci, avec tout son professionnalisme qui sied à son rang et son enthousiasme redoutablement communicatif, a su me diriger, m'orienter, m'avoir à moult reprises, poussé quand j'étais peu convaincu de certaines pistes de recherche à avancer et à mener à terme mes travaux de recherche. Qu'il trouve ici toute ma gratitude pour tous ses conseils prodigués et les rencontres qu'il a su catalyser.

A Mohammed Haddad, Mon Co-Directeur de thèse. Avec ses connaissances pluridisciplinaires avérées conjuguées à sa disponibilité en tout temps et à tout moment, parfois même au détriment de certains de ses impératifs, Mohammed était toujours là pour m'apporter toute son aide si précieuse, ses conseils et ses orientations étaient toujours opportunes, à même de me guider pour venir à bout de ma thèse. Pour tout cela, je ne saurais le remercier à sa juste valeur, mais je lui dis tout simplement "MERCI".

J'adresse mes vifs remerciements à M. Olivier Togni, MCF-HDR à l'Université de Bourgogne et à M. Raphaël Couturier, Professeur à l'Université Franche-Comté, pour m'avoir fait l'insigne honneur d'être rapporteurs de mon mémoire de thèse et surtout d'avoir accepté de juger mon travail dans des délais très court malgré leurs emplois de temps plus que saturés. Leurs remarques et leurs conseils constructifs m'ont été d'une grande aide.

Mes remerciements vont également à l'adresse de M. Stéphane Lallich, Professeur à l'Université Lumière Lyon 2 et M. Hocine Chérifi, Professeur à l'Université de Bourgogne, d'avoir accepté de faire partie de mon Jury de thèse.

Un grand merci à tous les membres de notre Laboratoire et plus particulièrement à Eric Duchêne et Brice Effantin pour avoir pris la peine de me corriger lacunes et incompréhensions dans certaines parties de mon manuscrit de thèse. Qu'ils trouvent ici toute ma reconnaissance pour leurs aides si précieuses.

Enfin, un grand merci à tous les membres de ma famille pour m'avoir supporté et surtout soutenu pendant ces 4 années.

Table des matières

Remerciements	i
Table des figures	vii
Liste des tableaux	ix
Introduction générale	5

Chapitre 1

Les communautés dans les réseaux sociaux

1.1	Introduction	12
1.2	Notions de théorie des graphes	12
1.2.1	Définitions	13
1.2.2	Terminologie	13
1.3	Utilisations des graphes dans les réseaux sociaux	14
1.4	Qu'est ce qu'un réseau social	14
1.4.1	Origines des réseaux sociaux	15
1.4.2	Succès des réseaux sociaux	16
1.4.3	Exemples de réseaux sociaux	16
1.4.4	Propriétés des réseaux sociaux	17
1.5	Les réseaux complexes : Robustesse et Communautés	18
1.5.1	Définitions	18
1.6	Méthodes de détection des communautés	20
1.6.1	Les approches agglomératives	21
1.6.2	Les approches séparatives	23
1.7	Exemples de réseaux sociaux et leurs communautés d'intérêts	24
1.7.1	Le graphe Zachary [116]	24

1.7.2	Le graphe de pages web [91]	25
1.7.3	Le graphe de collaboration entre scientifiques [47]	26
1.7.4	Le graphe football [47, 94]	26
1.7.5	Autres exemples de réseaux	26
1.8	Conclusion	28

Chapitre 2

Réseaux d'interactions sociales

2.1	Introduction	32
2.2	Quelques exemples de réseaux d'interactions sociales	32
2.2.1	Les réseaux d'acteurs	32
2.2.2	Les réseaux de citations	33
2.2.3	Les réseaux de co-auteurs	33
2.2.4	Les réseaux de connaissances	33
2.2.5	Les réseaux d'appels téléphoniques	33
2.3	Propriétés communes aux réseaux d'interactions	33
2.3.1	Distribution des degrés en loi de puissance	34
2.3.2	Petit diamètre et distance moyenne faible	34
2.3.3	Densité faible	35
2.3.4	Coefficient de clustering fort	35
2.4	Modélisation des réseaux d'interactions	36
2.4.1	Modèle aléatoire d'Erdős et Rényi	36
2.4.2	Modélisation des graphes "scale free"	39
2.4.3	Modèle petits mondes (" <i>Small world</i> ") :	42
2.5	Conclusion	43

Chapitre 3

Etude des structures sociales émergées

3.1	Introduction	46
3.2	Plateforme d'émergence de structures sociales	46
3.3	Description de la plateforme	49
3.4	Etude des structures émergées	49
3.4.1	Propriétés des structures émergées	49
3.4.1.1	Longueur moyenne des chemins (Average path length " <i>APL</i> ") :	50

3.4.1.2	Coefficient de Clustering ou Transitivité	51
3.4.1.3	Distribution de degrés	52
3.4.1.4	Discussion	53
3.4.2	Classification des structures de communautés émergées	55
3.4.2.1	Corrélation entre réseaux initiaux et structures de communautés	57
3.4.2.2	Impact des paramètres R , S et T sur la construction des communautés	58
3.5	Conclusion	60

Chapitre 4

Robustesse des communautés dans les réseaux de communication

4.1	Introduction	62
4.2	Objectifs de notre étude	62
4.3	Perturbations et Robustesse	63
4.3.1	Robustesse d'un système complexe	64
4.3.2	Mesures de la robustesse dans les réseaux complexes	65
4.3.3	Robustesse et communautés	65
4.3.4	Métrique pour la robustesse des structures des communautés	66
4.3.5	Positionnement et motivation de notre étude	66
4.3.6	Impact des paramètres R , S , T sur les structures de graphes	67
4.4	Etude de la robustesse des graphes émergés	67
4.4.1	Processus de perturbation du réseau	67
4.4.2	Évaluation de la robustesse	68
4.4.2.1	Asect global :	70
4.4.2.2	Aspect local	70
4.5	Analyse de la robustesse	71
4.5.1	Analyse de l'aspect global	72
4.5.2	Aspect local	73
4.6	Conclusions	80

Chapitre 5

Analyse des positions des nœuds importants

5.1	Introduction	82
5.2	Détermination des nœuds importants	82

5.2.1	Centralité de degré (<i>Degree Centrality</i>)	83
5.2.2	Centralité de proximité (<i>Closeness Centrality</i>)	84
5.2.3	Centralité d'intermédiarité (<i>Betweenness Centrality</i>)	84
5.3	Evaluation de la robustesse des nœuds importants	85
5.4	Analyse de la robustesse	86
5.4.1	Effets de la perturbation sur la population des nœuds importants .	86
5.4.1.1	Selon la centralité d'intermédiarité (<i>betweenness centrality</i>)	86
5.4.1.2	Selon la centralité de proximité (<i>closeness centrality</i>) . . .	88
5.4.1.3	Selon la centralité de degrés <i>degree centrality</i>	90
5.4.2	Robustesse des différentes configurations (<i>selon les proportions R, S, T</i>)	90
5.4.2.1	Les effets de l'intensité des perturbations	90
5.4.3	Impact moyen sur les différentes perturbations	91
5.5	Conclusion	92
	Conclusion générale	99
	Bibliographie	103

Table des figures

1.1	Un réseau avec trois communautés. Chaque communauté est dense en liens. Très peu de liens relient les communautés	21
1.2	(a) Un graphe à deux communautés, (b) Le dendrogramme résultant de l'application d'un algorithme agglomératif.	22
1.3	Exemples de réseaux sociaux.	25
2.1	Exemple d'un graphe aléatoire	37
2.2	Evolution du coefficient de clustering d'un graphe aléatoire	38
2.3	Distribution de degrés d'un graphe aléatoire	38
2.4	Un exemple de réseau scale-free qui s'accroît selon le principe "rich gets richer"	40
2.5	Distribution de degrés des graphes scale-free	40
2.6	Un graphe de 130 nœuds généré avec le modèle de Barabasi-Albert. Les 5 nœuds (en rouge) sont en contact avec les 60% des autres nœuds (en vert)	41
2.7	Modèle de Watts et Strogatz. A : $p=0$, uniquement les nœuds voisins proches sont connectés. Le coefficient de clustering et la distance moyenne entre n'importe quels deux nœuds sont grands. B : $p < 1$ mais $p > 0$, peu de courts chemins sont introduits dans les réseau, ce qui réduit la distance moyenne. C : p est grand et le réseau est équivalent à un graphe aléatoire. Dans ce cas, la distance moyenne et le coefficient de clustering sont trop faibles.	43
3.1	Architecture de la plateforme	48
3.2	Emergence du graphe d'interactions	50
3.3	APL moyen des graphes émergés de taille 1500 nœuds	51
3.4	Distribution de degrés d'un graphe émergé G' de 1500 nodes	53
3.5	Un graphe émergé avec $\gamma = 2.54$	54
3.6	Les interactions dans une communauté	56
3.7	(a) Les classes de graphes, (b) Classification structurelle	58

3.8	Influence de R, S, T sur les structures de communautés	60
4.1	Evaluation de la robustesse des communautés	68
4.2	Impacts moyens sur les serveurs et les ressources : selon l'intensité de la perturbation	74
4.3	Impacts moyens sur les serveurs (a) et sur les ressources (b) : selon les configurations	75
4.4	Classement des impacts moyens pour toutes les perturbations	76
4.5	Moyenne des impacts pour toutes les perturbations	77
4.6	Perturbations appliquées sur les serveurs	78
4.7	Perturbations appliquées sur les ressources	79
5.1	Evolution des proportions des nœuds avec une centralité d'intermédiarité élevée	87
5.2	Evolution des proportions des nœuds avec une centralité de proximité élevée	93
5.3	Evolution des proportions des nœuds avec une centralité de degré élevée . .	94
5.4	Effets sur les nœuds à fort centralité de proximité	95
5.5	Effets sur les nœuds à fort centralité de proximité	96
5.6	Effets sur les nœuds à fort centralité de degrés	97
5.7	Impact moyen : Intermédiarité et Proximité	98

Liste des tableaux

1.1	Récapitulatif des algorithmes de détection de communautés.	29
1.2	Exemples de réseaux sociaux connus dans la littérature.	30
2.1	Distribution des degrés de quelques réseaux d'interactions	34
2.2	Le degré moyen et la distance moyenne	34
2.3	Nombre de sommets, de liens et densité de quelques réseaux d'interaction .	35
2.4	Le CC de quelques réseaux d'interaction	36
3.1	Valeurs de l'APL	51
3.2	Valeurs moyennes du coefficient de clustering	52
3.3	Valeurs de γ pour un graphe de 1500 nœuds	53
3.4	Correspondance entre R, S, T et les structures de communautés. (F : Faible, M : Moyen, E : Elevé)	59
4.1	Perturbations appliquées aux serveurs : variations du <i>Coefficient de clustering</i> , <i>APL</i> , <i>Modularité</i> , et <i>Distribution de degrés</i>	72
4.2	Perturbations appliquées aux ressources : variations du <i>Coefficient de clustering</i> , <i>APL</i> , <i>Modularité</i> , et <i>Distribution de degrés</i>	73

Résumé

Les réseaux sont présents dans pratiquement tous les aspects de la vie. Le monde qui nous entoure comporte énormément de réseaux. Par exemple, les réseaux de communication constitués de téléphones, les réseaux électriques, les réseaux d'ordinateurs, le réseau des lignes aériennes, ... *etc*, sont autant de réseaux importants dans la vie de chaque jour. Le cadre mathématique des réseaux est bien approprié pour décrire plusieurs systèmes composés d'un grand nombre d'entités qui interagissent entre elles. Chaque entité est représentée par un nœud du réseau et chaque interaction par un lien entre deux nœuds. Il est donc possible de modéliser ces réseaux par des graphes. Pour la plupart de ces réseaux, la difficulté provient principalement du grand nombre d'entités, ainsi que de la façon dont elles sont interconnectées. Une approche naturelle pour simplifier de tels systèmes consiste donc à réduire leur taille. Cette simplification n'est pas faite aléatoirement, mais de telle façon à ce que les nœuds de la même composante aient plus de liens entre eux qu'avec les autres composantes. Ces groupes de nœuds ou composantes sont appelés *communautés d'intérêt*.

Notre thèse se positionne dans le domaine de l'étude des graphes sociaux. Elle s'intéresse principalement à l'étude de la robustesse des structures sociales émergentes dans les réseaux d'interactions. L'aspect de la robustesse des réseaux constitue un challenge très important pour comprendre leur fonctionnement, le comportement des entités les constituant et surtout pour comprendre les interactions qui peuvent se produire entre elles, permettant l'émergence de certains comportements qui n'étaient pas du tout prévisibles au préalable. Actuellement, les études de la robustesse des réseaux qui existent dans la littérature traitent cet aspect du point de vue purement structurel, *i.e.* toutes les perturbations sont appliquées soit sur les nœuds, soit sur les arêtes du graphe. Pour ce qui est de notre étude, nous nous sommes intéressés à définir une nouvelle stratégie qui se base sur des perturbations appliquées sur les paramètres qui permettent l'émergence des graphes sociaux dans les réseaux d'interaction. Cette façon d'aborder l'aspect robustesse des graphes constitue une nouvelle manière d'évaluer et de quantifier les changements qui peuvent intervenir dans les structures de ces graphes.

Notre travail est scindé en 3 grandes parties. La première partie s'intéresse à l'étude des graphes sociaux émergents dans les réseaux de communication. Nous modéliserons ces réseaux par des graphes et nous les caractériserons par un ensemble de paramètres influents sur l'émergence. Nous extrairons le réseau social issu de la communication et

nous étudierons les propriétés des communautés qui en émergent. Nous énumérerons et étudierons les types possibles de ces structures de communautés que nous classerons selon certaines de leurs propriétés structurelles. Nous proposerons une corrélation directe entre les paramètres et les types de structures de communautés émergées.

Dans la deuxième partie, nous proposons un modèle pour étudier la robustesse des structures sociales émergées des échanges entre entités dans les réseaux d'interaction. Nous étudions cette robustesse selon deux aspects : global et local. L'aspect global s'intéresse à étudier les propriétés du graphe social dans son ensemble telles que : la présence de triangles, la longueur moyenne des chemins ou la loi de distribution des degrés. Quant à l'aspect local, celui-ci s'intéresse plus particulièrement aux propriétés des communautés telles que : leur qualité, la classe de graphe qui leur correspond ou encore leur composition. Nous définissons dans cette partie les configurations les plus robustes face à des perturbations définies.

En complément aux deux précédentes parties, nous nous sommes intéressés également à l'étude des positions des nœuds importants dans le réseau, *i.e.* de voir comment certains nœuds tels que : les nœuds ayant des degrés élevés (hubs) ou les nœuds se trouvant à proximité d'un ou de plusieurs hubs, influent sur l'organisation du graphe social émergé ainsi que sa robustesse. Nous donnerons aussi les configurations des structures émergées les plus robustes ainsi qu'un classement de celles-ci face aux perturbations introduites.

Mots-clés: Communautés d'intérêt, Robustesse, Réseaux complexes, Emergence.

Abstract

Networks are present in virtually all aspects of life. The world surrounding us includes to many networks. For example, communication networks constituted of phones, electrical networks, computers networks, aerial lines network,? etc, are such important networks in our daily life. The mathematical framework of networks is well appropriated to describe different systems composed of many entities interacting with each other. Each entity is represented by a network node and each interaction by a link between two nodes. Therefore, it is possible to model these networks by graphs. For most of these networks, the difficulty comes mainly from the large number of entities and the way they are interconnected. A natural approach to simplify such systems is therefore to reduce their size. This simplification is not made randomly, but in such a way that the nodes of the same component would have more connections between themselves than with other components. These groups of nodes or components are called communities of interest.

Our thesis is positioned in the field of social graphs study. It is mainly interested in studying the robustness of social structures emerging in interaction networks. The aspect of networks robustness is a very important challenge to understand their functioning, the behavior of the constituting entities and especially to understand the interactions that may occur between them, allowing the emergence of certain behaviors that were not predictable at all in advance. Currently, studies of networks robustness that exist in the literature treat this aspect from a purely structural point of view, ie, all perturbations are applied either on nodes or on the edges of the graph. In terms of our study, we focused on defining a new strategy based on perturbations applied on the parameters that allow the emergence of social graphs in interaction networks. This way to approach the robustness appearance of the graphs is a new way to assess and quantify the changes that may occur in the structures of these graphs.

Our work is split into three major parts. The first part concerns the study of emerging social graphs in communications networks. We will model these networks by graphs and will characterize them by a set of parameters influencing emergence. We will extract the social network from the communication and will study the properties of emerging communities. We list and study the possible types of community structures that we classify according to certain of their structural properties. We will suggest a direct correlation between the parameters and types of emerged community structures.

In the second part, we will propose a model to study the robustness of emerged social structures from the exchanges between entities in interaction networks. We study this robustness according to two aspects : global and local. The global aspect is interested in studying the properties of the social graph as a whole such as : the presence of tri-

angles, the average path length or the degrees distribution law. As for the local aspect, it is particularly interested in the properties of communities such as : quality, the class of corresponding graphs or their composition. We define in this part the most robust configurations facing defined disturbances.

In addition to the two preceding parts, we also were interested in studying the positions of important nodes in the network, ie, to see the way some nodes such as : nodes with high degrees (hubs) or nodes in the proximity of one or more hubs, affect the organization of emerged social graph as well as its robustness. We will also give the configurations of the strongest emerged structures and their ranking facing the introduced disturbances.

Keywords: Communities of interest, Robustness, Complex Networks, Emergence.

Introduction générale

Le monde qui nous entoure comporte énormément de réseaux. Les réseaux de communication constitués de téléphones, les réseaux électriques, les réseaux d'ordinateurs, le réseau des lignes aériennes, ... *etc*, sont autant de réseaux importants dans la vie de chaque jour. Le symbolique réseau des pages HTML et des liens - le World Wide Web (WWW)- est un réseau virtuel que nous utilisons quotidiennement. La société est aussi semblable à un réseau. Le réseau des relations d'amitié entre les individus, les relations de travail, les intérêts communs ou le réseau des relations commerciales entre les individus et les sociétés sont également des exemples de réseaux sociaux et économiques. Les pays et les villes sont connectés par des réseaux de routes ou de lignes aériennes.

Les systèmes composés d'un grand nombre d'entités qui interagissent entre elles peuvent être représentés par des graphes où chaque entité est représentée par un nœud du réseau et chaque interaction par un lien entre deux nœuds. Par exemple, il est possible de représenter graphiquement les interactions entre protéines d'une même cellule ou les connections entre neurones sous forme d'un réseau. Finalement, les réseaux sont fréquemment utilisés en sciences sociales pour analyser les collaborations entre différents scientifiques ou les contacts humains. Les graphes représentent les propriétés topologiques essentielles en traitant le réseau comme une collection des nœuds et des arêtes. Pour la plupart de ces réseaux, la complexité provient principalement du grand nombre d'entités ainsi que de la façon avec laquelle elles sont interconnectées. Une approche naturelle pour simplifier de tels systèmes consiste donc à réduire leur taille.

Un graphe G , selon sa définition mathématique, est une paire d'ensembles (V, E) , où V est l'ensemble des nœuds du graphe G (représentant les individus ou entités du réseau), et E l'ensemble des arêtes entre les nœuds du graphe G (représentant les liens ou relations entre les individus du réseau). Pour une meilleure analyse et compréhension du graphe représentant le réseau, celui-ci est généralement décomposé en sous-groupes de nœuds. Cette décomposition n'est pas faite aléatoirement, mais de manière à ce que les nœuds de la même composante aient plus de liens entre eux qu'avec le reste du réseau. Ces groupes de nœuds ou composantes sont appelés *communautés d'intérêt*. Une communauté peut

être définie, d'une manière structurelle, comme un ensemble de nœuds qui ont plus de liens entre eux qu'avec les autres nœuds du graphe, et d'une manière sémantique, comme un ensemble de nœuds partageant les mêmes centres d'intérêt.

La plupart des travaux qui existent dans le domaine des communautés se scindent en deux grandes thématiques : la détection de communautés et l'analyse de communautés. La détection de communautés consiste à trouver les groupes d'intérêt dans un réseau donné, sans connaître à priori ni la taille ni le nombre des communautés. De nombreux travaux sont proposés pour la détection de communautés dans les réseaux [41]. Quant à l'analyse des communautés, elle s'intéresse beaucoup plus à l'examen des structures mêmes de ces communautés et de leur étude par le biais de certaines propriétés théoriques telles que : la densité, le distribution de degrés, *etc.*

Notre thèse s'intéresse à l'étude de la robustesse des structures de graphes sociaux qui émergent des échanges entre entités dans les réseaux d'interactions. Ainsi, deux aspects sont considérés dans notre étude. Le premier aspect concerne l'étude des structures de communautés ou groupes d'intérêt émergés et, le deuxième aspect concerne l'étude des propriétés de certains nœuds considérés comme importants dans le graphe.

Pour le premier aspect, nous nous intéressons à la classification des structures émergées selon leur environnement d'émergence, c'est-à-dire, selon les paramètres ayant concourus à leur apparition. Dans cette partie, nous avons proposé une plateforme, validée par des expérimentations, qui permet l'émergence de ces structures. Nous avons conçu par la suite l'étude de la corrélation directe entre les paramètres du réseau et les types de structures sociales émergées. Cette classification nous a permis de déterminer pour quelles valeurs de ces paramètres, nous passons d'une classe de graphe à une autre. Nous nous sommes intéressés à l'étude de l'impact des perturbations sur les communautés émergées. Pour cela, nous avons aussi défini également un modèle que nous avons validé aussi par des expérimentations. Les perturbations que nous introduisons sur les communautés ne concernent pas leur aspect structurel, comme c'est étudié dans la plupart des travaux existants dans la littérature, mais considère beaucoup plus les changements intervenus dans le contenu du réseau, c'est à dire, nous nous intéressons à l'étude de l'influence de la distribution des ressources et des fournisseurs sur la formation des communautés. Nous introduisons des perturbations sur les paramètres qui font émerger ces communautés. Cette étude est complétée dans le second aspect par une analyse des propriétés de certains nœuds importants tels que : les nœuds ayant des degrés élevés et qui deviennent influents dans l'organisation des communautés ainsi que les nœuds intermédiaires entre plusieurs

nœuds importants et que leur disparition pourrait faire basculer la communauté considérée d'un type de classe à un autre type.

Cette thèse est organisée comme suit :

- **Chapitre 1** : *Les communautés dans les réseaux sociaux*. Dans ce chapitre, nous donnerons quelques définitions et la terminologie. Nous aborderons les réseaux et leur modélisation en graphes, la définition structurelle et sémantique des communautés et nous donnerons des exemples de réseaux sociaux ainsi que leurs communautés d'intérêt correspondantes. Nous parlerons aussi de partitionnement de graphes, des méthodes de partitionnement existantes.
- **Chapitre 2** : *Réseaux d'interactions sociales*. Dans ce chapitre, nous donnons une définition des réseaux sociaux d'interactions ainsi qu'une présentation détaillée des trois modèles de réseaux d'interactions les plus connus, à savoir : le modèle "aléatoire", le modèle "scale-free" et le modèle "small world". Nous définissons également les propriétés structurelles de chacun de ces réseaux d'interactions que nous illustrons avec des exemples.
- **Chapitre 3** : *Etude des structures sociales émergées*. Ce chapitre traite de l'émergence des structures sociales dans des réseaux de communication. Nous modéliserons les réseaux de communication par les graphes et nous les caractériserons par un ensemble de paramètres (taille du réseau, densité du réseau, nombre de ressources dans le réseau, nombre de fournisseurs dans le réseau, etc). Nous extrairons un réseau social issu de la communication et nous étudierons les propriétés des communautés émergées. Nous énumérerons et étudierons les types possibles de structures de communautés émergées, et nous les classerons selon certaines propriétés structurelles. Nous proposerons une corrélation entre les paramètres du réseau initial et les types de structures de communautés émergées.
- **Chapitre 4** : *Robustesse des communautés dans les réseaux de communication*. Dans ce chapitre, nous avons conçu un modèle pour étudier la robustesse des structures sociales émergées des échanges entre entités dans les réseaux d'interaction. Nous étudions cette robustesse selon deux aspects : global et local. L'aspect global s'intéresse à étudier les propriétés du graphe social dans son ensemble telles que : la présence de triangles, la longueur moyenne des chemins ou la loi de distribution des degrés. Quant à l'aspect local, celui-ci s'intéresse plus particulièrement aux propriétés des communautés telles que : leur qualité, la classe de graphe qui leur correspond ou encore leur composition. Nous définissons dans ce chapitre les configurations les plus robustes face à des perturbations définies.
- **Chapitre 5** : *Analyse des positions des nœuds importants*. En complément à l'étude

réalisée dans le chapitre 4, nous nous intéresserons dans cette partie à l'étude des propriétés des nœuds importants dans le réseau, de voir comment ces nœuds importants, tels que les nœuds ayant des degrés élevés (hubs) ou les nœuds se trouvant à proximité d'un ou plusieurs hubs, influent sur l'organisation du graphe social émergées ainsi que sa robustesse. Nous donnerons aussi les configurations des structures émergées les plus robustes ainsi qu'un classement de celles-ci face aux perturbations introduites

- **Conclusion générale :** Nous terminerons notre thèse par une conclusion, où nous rappellerons nos apports et contributions. Nous citerons aussi nos travaux en cours et nos perspectives.

Les travaux réalisés pendant ma thèse ont abouti à plusieurs résultats concrétisés par des publications internationales et nationale.

1. Papiers acceptés :

- **S. Lemmouchi**, M. Haddad and H. Kheddouci. *Robustness Study of Emerged Social Communities from Exchanges in Networks*. Computer Communication Journal, 2012 (Accepté avec révision mineure) [75].
- **S. Lemmouchi**, M. Haddad and H. Kheddouci : *Study of robustness of community emerged from exchanges in networks communication*. MEDES 2011, pp. 189-196, San Francisco (USA), November 2011. [74].
- **S. Lemmouchi**, M. Haddad and H. Kheddouci. *Robustness Evaluation Framework*. In the proceedings of , 4th Workshop on Software Engineering Challenges for Ubiquitous and Pervasive Computing (UPC'2010), Berlin (Germany), July 2010 [73].
- **S. Lemmouchi**, M. Haddad and H. Kheddouci. *Robustesse de communautés émergées des graphes sociaux issus des réseaux de communication*. In the proceedings of EGC 2010 (Atelier Web Social), pp 41-52, Hamamat (Tunisie), Janvier 2010. [72]
- **S. Lemmouchi**, M. Haddad and H. Kheddouci. *Robustness of emerged community in social network*. MEDES 2009, pp 477-479, Lyon (France), October 2009. [71].

2. Papiers en cours de préparation :

- **S. Lemmouchi**, M. Haddad and H. Kheddouci. *Robustness Study of Important Nodes in Social Networks*. 2012.

Chapitre 1

Les communautés dans les réseaux sociaux

Sommaire

1.1	Introduction	12
1.2	Notions de théorie des graphes	12
1.2.1	Définitions	13
1.2.2	Terminologie	13
1.3	Utilisations des graphes dans les réseaux sociaux	14
1.4	Qu'est ce qu'un réseau social	14
1.4.1	Origines des réseaux sociaux	15
1.4.2	Succès des réseaux sociaux	16
1.4.3	Exemples de réseaux sociaux	16
1.4.4	Propriétés des réseaux sociaux	17
1.5	Les réseaux complexes : Robustesse et Communautés	18
1.5.1	Définitions	18
1.6	Méthodes de détection des communautés	20
1.6.1	Les approches agglomératives	21
1.6.2	Les approches séparatives	23
1.7	Exemples de réseaux sociaux et leurs communautés d'intérêts	24
1.7.1	Le graphe Zachary [116]	24
1.7.2	Le graphe de pages web [91]	25
1.7.3	Le graphe de collaboration entre scientifiques [47]	26
1.7.4	Le graphe football [47, 94]	26
1.7.5	Autres exemples de réseaux	26
1.8	Conclusion	28

1.1 Introduction

L'existence dans les réseaux d'interaction de zones plus densément connectées que d'autres dans les graphes découle souvent de la présence dans ce genre de graphes de structures de "communautés". Ce type de structure intervient dans de nombreux réseaux d'interactions et apporte de l'information sur leur organisation. Ces structures de communautés peuvent avoir des interprétations différentes selon le réseau considéré (réseaux sociaux, réseaux biologiques, réseaux d'information, réseaux de télécommunication ou autres). La connaissance de ces structures de communautés permet de comprendre le fonctionnement de ces réseaux et d'appréhender leur performance. Par exemple en sociologie, les réseaux de connaissances, de travail, ou encore d'amitié, sont caractérisés par des groupes d'individus fortement connectés entre eux, représentant les communautés d'intérêts. En informatique, les réseaux physiques et logiques d'ordinateurs, les graphes du web (graphes de forums, graphes de pages web, *etc*) peuvent être décomposés en communautés. Dans ce chapitre nous donnerons les définitions et la terminologie utilisées dans cette thèse. Nous nous intéressons à la définition des concepts généraux des réseaux sociaux et des communautés d'intérêts ainsi que leurs principaux apports dans les réseaux. Nous citerons également les techniques qui permettent la détection de ces communautés et nous donnerons quelques exemples des réseaux les plus connus et leurs communautés d'intérêts.

1.2 Notions de théorie des graphes

Les graphes sont utilisés entre autres pour décrire mathématiquement des concepts de réseaux. Ils représentent les propriétés topologiques essentielles d'un réseau en le réduisant à un ensemble de nœuds et de liens. Par exemple, dans les réseaux d'ordinateurs tels qu'Internet, les ordinateurs peuvent être représentés par des nœuds et les câbles les reliant par des liens. Dans le réseaux WWW, les nœuds sont des pages HTML et les arêtes représentent les liens hypertextes entre ces pages. C'est un concept simple et puissant à la fois. Parce qu'avec cette simplicité, ce concept prend en considération les systèmes complexes en utilisant les mêmes outils et méthodes mathématiques et dans la plupart des cas, les propriétés de ces réseaux sont identiques. La théorie des graphes a vraisemblablement connu ses premières applications avec la formalisation du problème des sept ponts de Königsberg, par Leonard Euler en 1736. Dans les années 60, Erdős et A. Rényi (ER), deux mathématiciens, ont introduit un nouveau concept, combinant la théorie des graphes et les outils de la théorie probabiliste, pour la considération de familles de graphes plutôt que des graphes spécifiques [34, 35]. Enfin, avec l'avancée de l'informatique et la disponibilité des données à grande échelle ainsi que des outils pour leurs études, il a été

mis en évidence que la théorie classique ne pouvait décrire avec précision plusieurs réseaux réels. Les travaux de Barabasi *et al.* [10] et de Faloutsos *et al.* [39] ont prouvé que la distribution des liens dans ce type de réseaux n'est pas complètement aléatoire et ne peut être décrite par la théorie d'ER. Ces travaux ont permis d'expliquer comment certains phénomènes tels que la propagation des virus peuvent survivre dans Internet pour longtemps.

1.2.1 Définitions

Un graphe $G = (V, E)$ est un ensemble de sommets $V = \{v_1, v_2, \dots, v_n\}$, un ensemble d'arêtes $E = \{e_1, e_2, \dots, e_m\}$ et une fonction d'incidence qui associe deux sommets chaque arête $F : E \rightarrow V \times V$, où chaque arête est notée par $e = uv$. L'ordre du graphe G désigne le nombre de sommets de G et est noté par $|V|$. Le nombre d'arêtes du graphe G est appelé la *taille* de G .

Un sous-graphe de G est noté $G' = (V', E')$, où $V' \subseteq V$, $E' \subseteq E$.

1.2.2 Terminologie

Dans cette section, nous définissons l'ensemble des termes utilisés dans cette thèse. Pour plus de détails sur les notions de graphes, se rapporter au livre de Claude Berge [13].

- **Un nœud** est l'unité de base d'un réseau. Dans un réseau social on parle d'**individus**.
- **Une arête** est une connexion entre deux nœuds.
- Une arête est **orientée** si elle est utilisée dans une seule direction. On parle d'**arc** lorsque l'arête est orientée.
- Une arête est **pondérée** lorsqu'on lui attribue un poids.
- **Un graphe** est défini par un ensemble de nœuds et un ensemble d'arêtes.
- **Un graphe orienté** désigne un graphe avec des arêtes orientées (des arcs).
- **Un graphe pondéré** désigne un graphe avec des arêtes pondérées. Des étiquettes ou des labels peuvent être associés aux nœuds et/ou des arêtes d'un graphe. On parlera alors d'un graphe étiqueté.
- **Le degré** d'un nœud est le nombre d'arêtes qui lui sont incidentes.
- **Un chemin** est une séquence d'arêtes qui relie deux nœuds.
- **Un chemin orienté** est une séquence d'arêtes qui relie deux nœuds en respectant l'orientation du parcours à chaque arête.
- **Un cycle** est un chemin fermé (un chemin dont les extrémités sont reliées).
- **La distance** entre deux nœuds d'un graphe G est la longueur d'un plus court chemin entre cette paire de nœuds du graphe.
- **Le diamètre** d'un graphe est la distance maximum dans le graphe.

- Un graphe est **complet** lorsqu'il existe une arête entre toute paire de nœuds.
- Un graphe est dit **connexe** lorsqu'il existe un chemin entre toute paire de nœuds.
- La **densité** d'un graphe est la proportion d'arêtes existantes dans le graphe (nombre d'arêtes dans le graphe divisé par le nombre d'arêtes maximum qui peuvent exister).
- Le **coefficient de clustering d'un nœud** est la probabilité que deux voisins d'un nœud soient aussi voisins entre eux.
- Le **coefficient de clustering d'un graphe** est la moyenne des coefficients de clustering de tous ses nœuds.
- La **distribution des degrés** d'un graphe G est l'association de chaque entier k au nombre de nœuds de G ayant un degré égale à k .
- La **matrice d'adjacence** est une représentation particulière d'un graphe. L'élément a_{ij} de cette matrice représente le nombre d'arêtes ou le poids de l'arête liant le nœud i au nœud j .
- La **matrice diagonale** d'un graphe G est une matrice où tous ses éléments, or la diagonale, sont nuls. Les éléments de la diagonale sont les degrés des nœuds de G .

1.3 Utilisations des graphes dans les réseaux sociaux

Les graphes non orientés sont adaptés pour les réseaux sociaux avec des relations non orientées. Les graphes orientés sont adaptés pour représenter des relations non symétriques comme les réseaux de confiance par exemple. Les graphes pondérés sont adaptés aux réseaux sociaux qui contiennent différents niveaux d'intensités dans les relations. Les graphes étiquetés permettent de représenter différents types de relations. Les graphes multipartis sont adaptés pour des réseaux sociaux incluant différents types de ressources manipulées par les acteurs et qui sont le support des interactions. Des exemples de graphes seront donnés plus loin.

Pour ce qui est notre étude dans cette thèse, nous utilisons les graphes simples non orientés, non pondérés et non étiquetés.

1.4 Qu'est ce qu'un réseau social

Les réseaux sociaux sont formés de groupes d'individus (entités) connectés par des liens sociaux. Les relations entre les individus dépendent du contexte. Elles peuvent être des relations d'amitié dans le cas d'un réseau de connaissances, des relations de citations dans un réseau de documents, des liens de connexions physiques ou logiques dans un réseau d'ordinateurs, etc. Les réseaux sociaux ont suscité l'intérêt des chercheurs dans des

domaines de recherches multiples. On en trouve plusieurs applications : propagation et recherche d'information dans le web [61, 109, 117], recherche de ressources dans les réseaux pair-à-pair [20, 23, 33, 99, 110, 115], sécurité [78], biologie [1, 47, 101], *etc.*

Stanley Milgram a été l'un des premiers chercheurs à avoir étudié les réseaux sociaux [81]. Les réseaux sociaux sont caractérisés par deux propriétés : ce sont des graphes dits *petit monde* [81] et qui ont un coefficient de clustering élevé [112]. Les graphes petit monde sont des graphes où chaque nœud peut être relié à n'importe quel autre nœud dans le graphe avec une petite chaîne de relations sociales. Ceci donna naissance au concept de "*six degrés de séparation*" (1967). Ce concept suggère que deux personnes, choisies au hasard parmi les citoyens américains, sont reliées en moyenne par une chaîne de six relations.

Les réseaux sociaux ont des structures spécifiques. Ils sont caractérisés par certaines zones denses. Les individus de ces zones denses ont plus de liens entre eux qu'avec le reste du graphe social [42, 86]. Ces zones denses sont dites "*communautés d'intérêt*".

Pour une meilleure compréhension du réseau, celui-ci est généralement décomposé en communautés. La décomposition n'est pas faite d'une manière aléatoire, mais de telle sorte que les nœuds de la même composante aient au moins autant de liens entre eux qu'avec les autres nœuds appartenant à d'autres composantes. Du point de vue sémantique, une communauté est un ensemble de nœuds qui partagent les mêmes centres d'intérêt ou ayant le même profil.

1.4.1 Origines des réseaux sociaux

Les réseaux sociaux sont nés de l'idée de l'existence d'une structure déterminée dans laquelle les individus se connaissent entre eux, directement ou indirectement. La notion des "*six degrés de séparation*" a propulsé l'idée que tout le monde peut être connecté par des connaissances communes.

La plupart des chercheurs sur les réseaux sociaux en ligne se concentrent sur les interactions et les échanges afin d'obtenir un emploi, partager des histoires avec d'autres personnes ou échanger des fichiers, *etc.* Pourtant, pendant les dernières décennies, les chercheurs spécialisés dans les sciences du comportement ont étudié les réseaux sociaux de toutes sortes - interaction "*offline*" (face à face, lettres, téléphones, *etc.*) ainsi que les interactions "*en ligne*" pour déterminer comment les réseaux sociaux sont développés et maintenus et comment ils affectent nos vies.

J. L. Moreno a étudié les réseaux sociaux au début des années 1930 [84]. Son objectif était de visualiser graphiquement un réseau social. Les personnes sont représentées par des points et les relations entre ces personnes par des flèches. Cette façon de représenter est désignée par le terme "*sociogramme*" et fut un premier outil d'identification rapide des caractéristiques d'un réseau social. Moreno a introduit le concept "*étoile*" pour désigner les personnes ayant le plus de relations dans un réseau social, en référence à l'étoile formée par un point et ses connections.

Les mathématiciens ont rapidement fait le rapprochement entre les représentations sociogrammes et la théorie des graphes au sens mathématique. Une évolution de la représentation des réseaux sociaux est passée en revue dans [102]. Au milieu du vingtième siècle, Cartwright et Harary furent les premiers à avoir appliqué la théorie des graphes à l'analyse des réseaux sociaux. Le graphe est devenu par la suite la représentation adoptée par toutes les sciences manipulant l'analyse des réseaux sociaux, dont la sociologie, les mathématiques et l'informatique.

1.4.2 Succès des réseaux sociaux

Les réseaux sociaux les plus populaires poussent leurs membres, à travers le marketing viral du comportement naturel humain, à parler entre eux sur les produits et les services. L'idée derrière l'emploi des réseaux sociaux est de voir un retour sur investissement suffisant sur l'implication de certaines mesures - par exemple, le temps, l'énergie ou l'argent - pour continuer à utiliser un service. Le succès des réseaux sociaux ne signifie pas forcément "*profitable*". Alors que de nombreux sites de réseaux sociaux sont plus ou moins rentables, d'autres, comme Wikipédia, l'encyclopédie libre en ligne, cherchent à servir la société en stockant le maximum de connaissances humaines accessibles à tous et de façon gratuite.

1.4.3 Exemples de réseaux sociaux

Différents réseaux sociaux mettent l'accent sur différents aspects des interactions humaines. Par exemple, MySpace, Facebook et YouTube montrent les interactions qui peuvent se produire au sein des réseaux sociaux centrés sur l'objectif principal de relier les gens en ligne. Chacun de ces réseaux sociaux offre des outils qui permettent aux utilisateurs de s'exprimer et d'interagir avec les autres utilisateurs de différentes manières, en choisissant de se concentrer sur différents domaines de l'interaction sociale. Bien que chaque réseau exprime la sociabilité d'une manière différente, ils sont tous construits au-

tour de l'idée de connecter les individus en leur permettant d'interagir. A titre d'exemple, facebook (www.facebook.com), développé en 2004 par un étudiant de l'Université de Harvard (USA), a été exclusivement réservé aux étudiants de cette université. Après deux années, il a été généralisé à toutes les universités. Facebook est un réseau pair à pair de relations basé sur un réseau social qui permet aux utilisateurs de créer des profils personnels décrivant leurs vies réelles et d'établir des liens avec d'autres utilisateurs. En plus des informations de base tels que le nom, l'âge, les études, *etc*, les utilisateurs peuvent ajouter d'autres informations tels que des livres favoris, des films ou des photos.

1.4.4 Propriétés des réseaux sociaux

Les réseaux sociaux possèdent des propriétés intéressantes. Ils influencent nos vies sans que nous soyons conscients des implications qu'ils soulèvent : Comment un virus se propage et infecte les individus ? Comment un sujet de recherche est devenu d'actualité ? Et tant d'autres questions qui suscitent l'intérêt et nous exhortent à trouver des modèles qui pourraient fournir de puissants outils à même d'améliorer, par exemple, notre système de santé. Pedro Domingos [30] étudie la façon avec laquelle on peut exploiter les réseaux sociaux à des fins de marketing en utilisant notre richesse de données. C'est extraordinaire de voir comment de telles techniques pourraient transformer les stratégies de commercialisation qui sont en cours. Peter Mika [80] considère comme une forme particulière d'influence, la façon dont les individus sont d'accord sur la terminologie et les implications de ce phénomène pour la construction des ontologies et du Web sémantique.

La principale caractéristique d'un réseau social est "*l'effet petit monde*" qui découle de l'expérience de Stanley Milgram (1960) [81]. Le plus court chemin entre deux sommets dans un réseau social de taille n est de l'ordre de $\log(n)$. Dans [85], Newman a montré que les individus d'un même réseau social possèdent la faculté de trouver facilement ces plus courts chemins.

Une autre caractéristique aussi importante est le "*Coefficient de clustering*" élevé dans un réseau social. En effet, les individus dans un tel réseau ont tendance se socialiser en se regroupant en *Communautés*. Si un sommet v_1 est connecté à un autre sommet v_2 qui est lui même connecté à un sommet v_3 , alors il y a une forte probabilité pour que v_1 soit aussi connecté à v_3 . En d'autres termes, la façon avec laquelle les sommets dans un réseau social favorisent l'émergence de structures de graphe de type "*Triangle*". Ces communautés possèdent une forte densité locale et une faible densité globale.

La troisième caractéristique des réseaux sociaux est la *distribution des degrés* qui suit une loi de puissance de type $p(k) \sim k^{-\gamma}$, c'est à dire, plus on considère un degré

élevé, plus le nombre de sommets ayant ce degré dans un même réseau est faible. γ est un paramètre compris entre 2 et 3.

1.5 Les réseaux complexes : Robustesse et Communautés

De nombreux systèmes sont constitués d'un grand nombre d'entités en interaction. Ces systèmes sont dit "Complexes" lorsque, même si les interactions entre ces entités sont bien comprises, il émerge un comportement global qui n'était pas prévisible à partir des règles d'interaction. *Un système est informellement dit complexe si le résultat final n'est pas prédictible directement en connaissant les règles.* Cette limitation vient de l'impossibilité de modéliser le système par des équations solvables et prédictives. Ce qui est primordial est le nombre de paramètres ainsi que le fait que chacun de ces paramètres peut avoir une influence essentielle sur le comportement du modèle. Pour prévoir ce comportement, il est nécessaire de tous les prendre en compte, ce qui revient à effectuer une simulation du système étudié.

1.5.1 Définitions

Il n'existe pas de définition précise de ce qui est un système complexe. Néanmoins, nous pouvons considérer deux définitions d'un système complexe selon sa constitution et selon son comportement.

En effet, du point de vue de sa constitution, un système complexe est d'un grand nombre fini d'entités, possédant un comportement global structuré qui émerge et dont le graphe d'interaction est non trivial, c'est à dire que ce n'est pas simplement tout le monde qui interagit avec tout le monde (il y a au moins des liens privilégiés). On note que ce type de réseau est souvent en lui-même intéressant car il présente des caractéristiques telles que : le petit monde, invariance d'échelle, communautés, ...

Du point de vue de son comportement, un système complexe, dans lequel on retrouve une auto-organisation et une émergence des propriétés ou de structures cohérentes ainsi qu'une apparition de motifs, présente également une robustesse locale et une fragilité à moyenne échelle : si un élément est affecté par un événement extérieur, ses voisins le sont aussi. Il s'ensuit que le système est souvent plus robuste à une petite perturbation locale. Bien cibler cette perturbation est cependant très difficile. Par exemple, les virus sont un parfait exemple : avec une dizaine de gènes, un virus est capable de modifier

profondément un organisme de plusieurs dizaines de milliers de gènes, et ce même pour une faible concentration au départ.

L'aspect de la robustesse des systèmes complexes constitue un challenge très important pour comprendre leur fonctionnement, le comportement des entités les constituant et surtout comprendre les interactions qui peuvent se produire entre elles, permettant l'émergence de certains comportements qui n'étaient pas du tout prévisibles au préalable.

En théorie des graphes, un réseau complexe est défini comme un ensemble de n nœuds - représentant les entités du système à étudier - connectés entre eux par un ensemble de m arêtes. Les nœuds sont représentés par des sommets et les arêtes par des liens. Nous adopterons cette notation tout au long de notre travail.

Un réseau complexe est dûment caractérisé par la connaissance de chaque nœud et de chaque arête. Ces réseaux sont souvent décrits par leur matrice d'adjacence A , où A_{ij} représente la connection du nœud i au nœud j . $A_{ij} \neq 0$ indiquent la présence d'une arête.

Les réseaux complexes ont été étudiés dans différents domaines. Plusieurs systèmes dans la nature peuvent être décrits par des modèles de systèmes complexes dont leurs structures consistent en un ensemble de nœuds connectés par des arêtes, par exemple, Internet est un réseau de routeurs et de serveurs, le WWW est un réseau de pages web, *etc.*

Le cadre mathématique des réseaux complexes s'est avéré bien approprié pour décrire plusieurs systèmes composés d'un grand nombre d'unités qui interagissent entre elles. Chaque unité est représentée par un nœud du réseau et chaque interaction par un lien entre deux nœuds.

Pour la plupart de ces systèmes, la complexité provient principalement du grand nombre d'entités le composant, ainsi que de la façon dont elles sont interconnectées. Nous nous intéressons ici aux définitions des communautés. Dans la littérature, on ne trouve pas une définition précise des communautés et il n'existe pas de définition formelle consensuelle d'une structure de communautés dans un réseau d'interactions, mais, la plupart des chercheurs dans ce domaine se restreignent à donner des définitions plutôt structurelles et sémantiques. Ce problème de non formalisation de la définition des communautés dans les réseaux a été soulevé dans plusieurs articles [41,42,67,97]. Cependant, un consensus sur ce qu'est une communauté, repris notamment dans la plupart des publications scientifiques, stipule qu'une *communauté est un groupe de nœuds qui ont plus de liens entre eux qu'avec le reste du graphe* [41,42,67,86,91,97].

Lorsqu'on s'intéresse au problème d'extraction de la structure de communautés, il est

nécessaire d'avoir une définition formelle de celle-ci ; or il en existe plusieurs. C'est pourquoi, les algorithmes d'extraction de communautés doivent être capable de détecter ces communautés selon le contexte dans lequel elles existent.

Dans notre travail, nous ne nous intéressons pas aux algorithmes qui procèdent à la formation des communautés dans les graphes, mais plutôt à ceux qui nous permettent de les extraire. En effet, les communautés que nous considérons émergent par elles-mêmes et aucune supposition n'est faite *a priori*.

Nous allons donner ici deux définitions des communautés, l'une sémantique et l'autre structurelle.

Définition sémantique

Une communauté est un ensemble de nœuds qui partagent les mêmes centres d'intérêt ou ayant le même profil.

Cette définition est intuitive dans le sens où les individus ont tendance à bâtir des liens et à se mettre dans le même groupe avec les individus qui ont les mêmes intérêts. Les travaux de la partie analyse de communautés se basent sur cette définition pour comprendre et expliquer le comportement des réseaux étudiés.

Définition structurelle

Une communauté est un ensemble de nœuds fortement liés entre eux et faiblement liés avec les autres nœuds du graphe.

Comme le montre la Figure 1.1, une communauté peut être vue comme des groupes de nœuds où chaque groupe est dense en liens et a peu de liens avec les autres groupes. Dans la section suivante, les approches proposées s'appuient sur cette définition structurelle pour identifier les communautés dans les réseaux.

1.6 Méthodes de détection des communautés

Comme mentionné plus haut, la notion de communautés dans les graphes n'a pas de définition formelle. Cependant, l'existence de zones plus densément connectées que d'autres est le résultat d'une présence de structures de graphes dont les nœuds se sont regroupés en communautés du fait de leur ressemblance ou de leurs intérêts communs. Cette ressemblance ou ce partage d'intérêt peut avoir des interprétations différentes selon

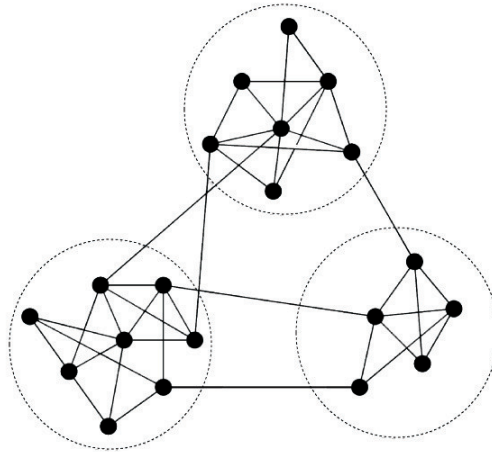


FIGURE 1.1 – Un réseau avec trois communautés. Chaque communauté est dense en liens. Très peu de liens relie les communautés

la nature et le type du réseau d'interaction considéré (réseaux sociaux, réseaux biologiques, réseaux de communications, *etc*).

A partir de cette définition informelle de ce qu'est une communauté, plusieurs méthodes, approches et algorithmes, permettant la détection ou l'extraction des communautés, ont été définis dans la littérature. Le but commun de ces méthodes, approches et algorithmes est de trouver une partition des sommets qui soient en adéquation avec la définition de la communauté.

D'une manière générale, nous pouvons classer toutes les méthodes, approches et algorithmes en deux grandes familles d'approches. Il s'agit des approches *agglomératives* et *séparatives*.

1.6.1 Les approches agglomératives

Le principe global de ces approches consiste à regrouper les nœuds itérativement en communautés. Au début, les nœuds sont considérés comme si chacun d'eux constituait une communauté à part, *i.e.* il y a autant de communautés que de nœuds. Les communautés sont réunies deux à deux jusqu'à avoir une grande communauté représentant l'ensemble des nœuds du graphe. Il faut souligner qu'à chaque étape de regroupement de deux communautés, une métrique (fonction de qualité) est calculée et le partitionnement ayant la plus haute valeur de la métrique considérée représente le meilleur partitionnement du graphe en communautés. Le résultat de ces approches est un dendrogramme reprenant

l'historique des jointures de communautés à chaque étape.

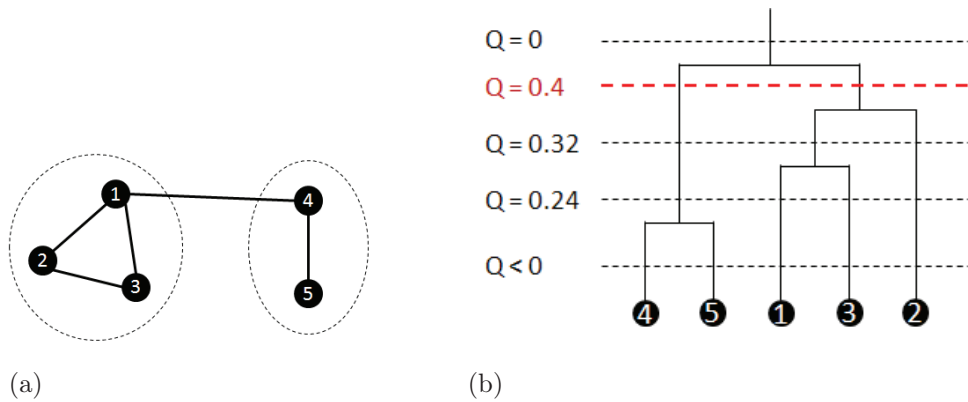


FIGURE 1.2 – (a) Un graphe à deux communautés, (b) Le dendrogramme résultant de l'application d'un algorithme agglomératif.

Parmi les algorithmes les plus populaires dans cette famille d'approches, on peut citer :

- **L'algorithme glouton de Newman** [87] : Dans cet algorithme, on démarre avec un ensemble de nœuds isolés considérés comme une communauté. Les communautés sont jointes deux à deux jusqu'à avoir une grande communauté représentant le graphe entier. La métrique utilisée dans cet algorithme est la *modularité* qui est calculée à chaque jointure de deux communautés i et j :

$$\Delta Q = 2(e_{ij} - a_i^2)$$

où, e_{ij} est la proportion d'arêtes à l'intérieur des communautés et a_i la proportion d'arêtes attendue dans le graphe aléatoire de G . Dans l'exemple de la Figure 1.2, la modularité est maximale au niveau 3, ce qui donne un partitionnement du graphe de la figure 1.2 (a) en deux communautés.

- **L'algorithme de Donetti *et al.*** [31] Les auteurs de cet article utilisent les vecteurs propres de la matrice Laplacienne L du graphe pour mesurer les similarités entre les nœuds, qui est considérée comme le score des arêtes (Pour rappel, $L = A - D$, où A est la matrice d'adjacence du graphe et D la matrice diagonale).
- **L'algorithme de Zhou *et al.*** [?] : Basé sur la marche aléatoire pour mesurer la distance entre nœuds, cet algorithme utilise cette distance comme score pour la jointure des communautés en commençant d'abord par la jointure des communautés ayant la plus petite distance.

1.6.2 Les approches séparatives

Contrairement à la familles des approches agglomératives, les approches séparatives scindent le graphe en plusieurs communautés en retirant progressivement les arêtes reliant les communautés distinctes. Les arêtes sont retirées une à une faisant apparaître des composantes connexes du graphe qui sont identifiées à des communautés. Ce processus est réitéré jusqu'au retrait de toutes les arêtes. Les méthodes existantes diffèrent par la façon de choisir les arêtes à retirer. Ci-dessous, nous donnons quelques unes des méthodes les plus connues dans la littérature de la classe des approches séparatives.

- **L’intermédiarité de Newman et Girvan [91]** : Dans cet article, les auteurs utilisent la mesure dite *betweenness* pour détecter les arêtes inter-communautés à supprimer en premier pour trouver les communautés. Cette mesure donne des valeurs fortes aux arêtes qui ont une extrémité dans une communauté i et une autre extrémité dans une autre communauté j , et des valeurs faibles aux arêtes d’une même communauté. Cette *betweenness* est calculée comme étant le nombre des plus courts chemins du graphe qui passent par cette arête. De la même façon, l’algorithme proposé calcule la *betweenness* de chaque arête et supprime celle ayant la plus grande valeur. Ce processus est réitéré jusqu’à ce qu’on ait autant de communautés que de nœuds dans le graphe. Si cette méthode ainsi proposée donne de bons résultats, il n’en demeure pas moins que son inconvénient majeur est le coût de calcul. En effet, pour chaque suppression d’une arête, le score de toutes les arêtes est affecté et ainsi, celui-ci sera recalculé, ce qui est relativement coûteux. Le complexité de l’algorithme est de l’ordre de $O(m^2.n)$.
- **L’algorithme de Radicchi [97]** : Radicchi *et al.* ont défini une autre métrique pour détecter les arêtes inter-communautés à supprimer. Les auteurs s’inspirent de la notion du coefficient de clustering pour donner des poids aux arêtes. La fonction de score d’une arête est calculée en divisant le nombre de triangles construits par cette arête sur le nombre maximum de triangles possibles. Le poids d’une arête (i, j) est donné par l’équation suivante :

$$C_{i,j}^{(3)} = \frac{Z_{i,j}^{(3)}}{\min [(k_i - 1), (k_j - 1)]}$$

où $Z_{i,j}^{(3)}$ est le nombre de triangles construits par l’arête (i, j) et $\min [(k_i - 1), (k_j - 1)]$ est le nombre maximum de triangles qui peuvent être construits par l’arête (i, j) . Les auteurs proposent aussi d’utiliser des cycles de taille k , mais en pratique seuls des cycles de taille 3 et 4 sont utilisés. L’avantage de cet algorithme est que le calcul des poids des arêtes est local, et lors de la suppression d’une arête, le score des arêtes d’autres localités n’est pas affecté (seul le score des arêtes voisines sera recalculé),

alors, les autres ne seront pas recalculés. L'algorithme est moins coûteux en temps, mais ne donne pas de résultats assez satisfaisants. La complexité de l'algorithme est de $O(m^2)$.

- **L'algorithme de Fortunato** [42] : Le principe de cet algorithme est le même que toutes les approches séparatives. La métrique utilisée par Fortunato *et al.* pour identifier les arêtes inter-communautés est la centralité d'information. La centralité d'information d'une arête est la diminution de l'efficacité du réseau due à la suppression de cette arête. L'efficacité d'une arête $e(i, j)$ est définie comme l'inverse de la distance entre le nœud i et le nœud j . Pour toute arête, on calcule l'efficacité puis la diminution du réseau due à la suppression de cette arête, et on affecte cette diminution au score de l'arête. Le calcul de l'efficacité du réseau est coûteux en temps. La complexité de cet algorithme est de $O(m^3.n)$.

En plus de ces deux grandes familles d'algorithmes pour la détection de structures de communautés dans les réseaux, il existe d'autres approches telles que les approches basées sur l'optimisation de la modularité [26, 51, 90] et les approches dédiées à des types de réseaux spécifiques (réseaux dits "signés", réseaux mobiles) [7, 8, 108].

Le Tableau 1.1 synthétise les méthodes les plus connues de détection de communautés dans les réseaux, vues dans ce chapitre. Nous rappelons la classe et la complexité de chacun des algorithmes présentés.

1.7 Exemples de réseaux sociaux et leurs communautés d'intérêts

Dans la littérature, il existe énormément de réseaux sociaux. Fortunato *et al.* [43] a donné plusieurs exemples de ces réseaux sociaux et dans cette section, nous citons uniquement les réseaux sociaux les plus connus.

1.7.1 Le graphe Zachary [116]

Le réseau de Zachary (figure 1.3 (a)) est un réseau social des membres d'un club de karaté de l'université San Francisco aux États Unis. Le club de karaté compte 78 membres. Zachary a fait une étude sur les membres du club qui ont des relations d'amitié en dehors du club. Parmi les 78 membres du club, seuls 34 ont des relations d'amitié. Le réseau d'amitié issu de ce club est représenté dans la figure 1.3 (a). Il est constitué de 34 nœuds représentant les membres du club et 78 liens représentant les amitiés entre les membres. Ce réseau social contient deux communautés réelles, instructeurs et administrateurs du

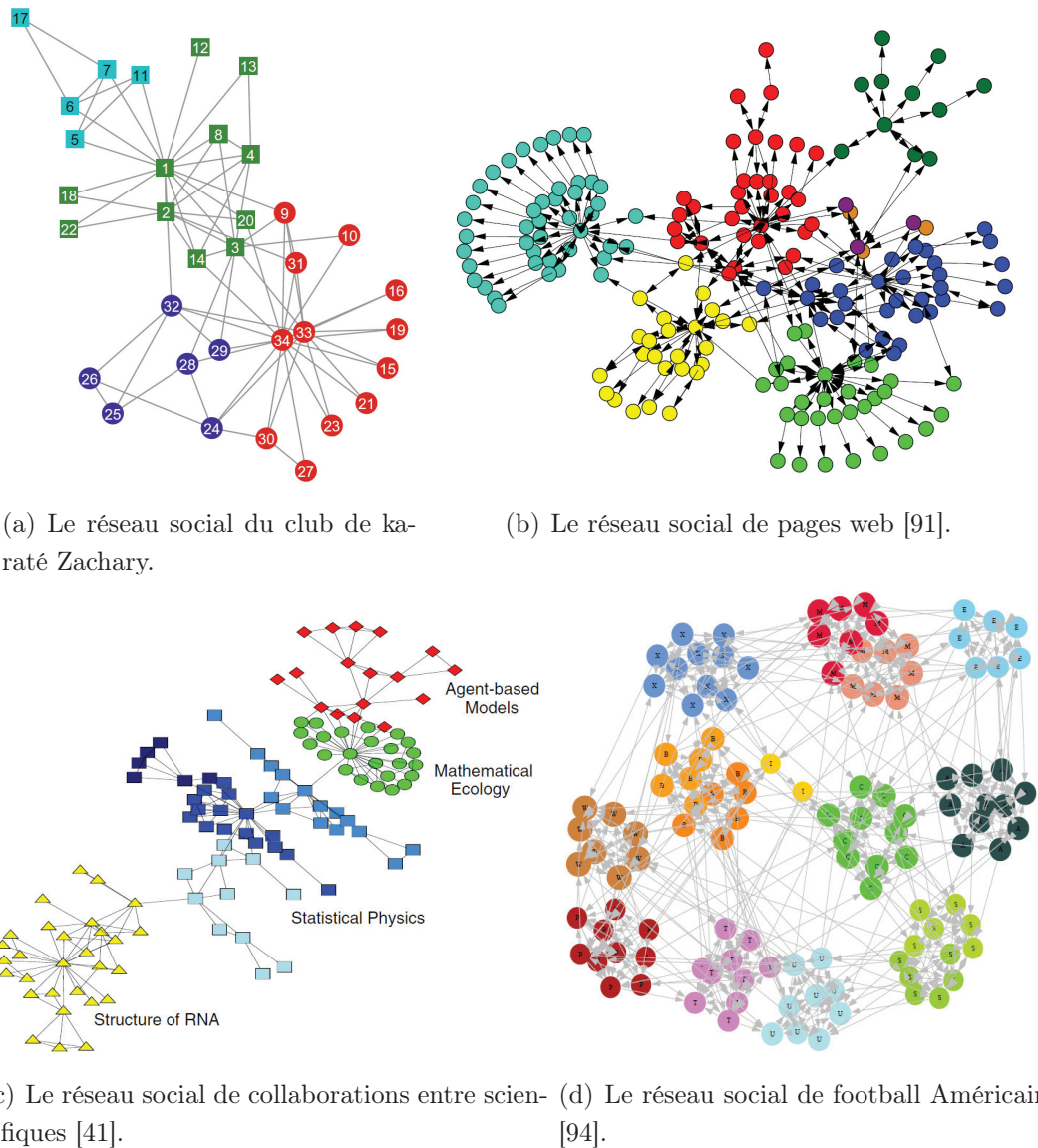


FIGURE 1.3 – Exemples de réseaux sociaux.

club (les deux communautés sont représentées dans la figure 1.3 (a) par les formes, carrés et cercles).

1.7.2 Le graphe de pages web [91]

Ce graphe est un réseau social d'indexation de pages web. Les nœuds de ce graphe représentent les pages web et les arcs représentent les hyperliens d'une page à une autre. Dans ces graphes, les liens sont orientés (généralement, si une page A indexe une autre page B , le contraire n'est pas systématiquement vrai - la page B peut ne pas indexer

la page A). Le groupe de pages qui ont beaucoup de liens entre elles, peut être une communauté de pages web qui traitent un même sujet. Rassembler les pages du même sujet en communautés, peut faciliter et accélérer le processus d’indexation et de réponse aux requêtes.

1.7.3 Le graphe de collaboration entre scientifiques [47]

Ce réseau est un graphe de collaboration de scientifiques de différentes disciplines de l’institut Santa Fe dans le nouveau Mexique aux USA. Le graphe contient 271 nœuds représentant les scientifiques de l’institut durant les deux années 1999 et 2000. Un lien est mis entre deux nœuds (scientifiques) s’ils sont co-auteurs d’au moins un article de recherche durant ces deux années.

La figure 1.3 (c) montre le réseau de collaboration entre les scientifiques de l’institut Santa Fe, où les formes représentent les différentes disciplines de recherche (chaque communauté représente l’ensemble des chercheurs d’une discipline donnée).

1.7.4 Le graphe football [47, 94]

Le graphe *football* est un réseau de relations entre les équipes de football américain de division I entre des universités aux USA. Les nœuds du réseau représentent les équipes (étiquetés par les noms des collèges) et les liens représentent les matchs entre équipes (un lien est mis entre deux équipes si elles se rencontrent durant la saison). Les équipes sont divisées en conférences de 8 à 12 équipes chacune. Il existe plus de matchs entre les équipes de la même conférence qu’avec les autres. Une équipe joue en moyenne sept matchs dans sa conférence et quatre avec des équipes d’autres conférences (une équipe joue autour de 75% de ses matchs dans sa conférence). Les communautés réelles de ce réseau sont les groupes d’équipes (nœuds) de mêmes conférences.

La figure 1.3 (d) montre un exemple de réseau de football de 115 équipes regroupées en 12 conférences de la saison 2004 [94].

1.7.5 Autres exemples de réseaux

Il existe plusieurs réseaux et benchmark issus de domaines multiples (biologie, web, sociologie, etc). La plupart des méthodes de détection de communautés proposées dans la littérature sont testées sur ces graphes pour valider leur performances en temps et en précision. La performance en temps est la capacité de l’algorithme à détecter les communautés dans les grands réseaux (algorithmes de faible complexité). Tandis que la performance en précision est la capacité de l’algorithme à détecter les bonnes communautés du réseau.

En plus des quatre graphes sociaux présentés ci-dessus, la table 1.2 énumère des exemples de réseaux les plus utilisés parmi les nombreux qui existent. Nous donnons le nombre de nœuds, le nombre d'arêtes et une brève description de chaque réseau.

1.8 Conclusion

Dans ce chapitre, nous avons donné les définitions et la terminologie utilisées dans cette thèse. Nous avons aussi introduit les concepts de réseaux sociaux, de réseaux complexes et de communautés dans les réseaux. Nous avons donné les définitions des communautés d'intérêt et cité les apports et importances de ces dernières dans la compréhension du fonctionnement des réseaux. Nous avons terminé le chapitre en donnant des exemples de réseaux connus et leur communautés d'intérêt.

La détection de communautés dans les réseaux est un problème bien abordé et traité dans la littérature. Cet engouement montre l'importance de la connaissance de ces modules dans les réseaux. Nous consacrerons tout un chapitre pour énumérer et classier les différentes structures de communautés qui émergent dans les réseaux sociaux d'interaction. Mais avant cela, il nous a paru important de définir d'abord ces réseaux, de définir quelques unes de leurs propriétés et de citer les modélisation des réseaux sociaux d'interaction les plus utilisées. C'est d'ailleurs l'objet du chapitre suivant.

TABLE 1.1 – Récapitulatif des algorithmes de détection de communautés.

<i>Algorithme</i>	<i>Type de l'algorithme/ Classe de graphes</i>	<i>Complexité</i>	<i>Référence</i>
Intermédiarité (Newman et Girvan)	Séparatif / Graphes simples	$O(m^2.n)$	[91]
Centralité d'information (Fortunato <i>et al.</i>)	Séparatif/ Graphes simples	$O(m^3.n)$	[43]
Coefficient de clustering (Radicchi <i>et al.</i>)	Séparatif/ Graphes simples	$O(m^2)$	[97]
Glouton (Newman)	Agglomératif/ Graphes simples	$O((m + n).n)$	[88]
Sémantique (Donetti <i>et al.</i>)	Agglomératif/ Graphes simples	$O(n^3)$	[31]
Proximité (Zhou <i>et al.</i>)	Agglomératif/ Graphes simples	$O(n^3)$	[119]
Marches aléatoires (Pons et Latapy)	Agglomératif/ Graphes simples	$O(mn^2)$	[96]
Spectral (Newman)	Optimisation de la modularité/ Graphes simples	$O(n^2.log n)$	[90]
Extrême (Duch et Arenas)	Optimisation de la modularité/ Graphes simples	$O(n^2.log^2 n)$	[26]
Bissection (Fiedler)	Partitionnement de graphes	$O(n^3)$	[40]
KL (Kernighan et Lin)	Partitionnement de graphes	$O(n^3)$	[62]
spectrale (Leicht et Newman)	Graphes orientés	$O(n^2.log n)$	[68]
Réduction de taille (Arenas <i>et al.</i>)	Graphes orientés et pondérés	$O(n^2.log n)$	[7]
Motifs (Arenas <i>et al.</i>)	Graphes avec motifs	dépend des paramètres	[8]
Mod^+ et Mod^- (Gómez <i>et al.</i>)	Graphes pondérés	dépend des paramètres	[49]
Capocci (Capocci <i>et al.</i>)	Graphes simples	$O(n^2)$	[19]
Chevauchement (Palla <i>et al.</i>)	Graphes simples	$O(exp(n))$	[93]

TABLE 1.2 – Exemples de réseaux sociaux connus dans la littérature.

<i>Graphe</i>	<i>nœuds</i>	<i>Arêtes</i>	<i>Description</i>	<i>Référence</i>
Zachary	34	78	Un réseau social d'amitié entre 34 membres d'un club de karaté d'une université Américaine	[116]
Cortex	55	564	Réseau de connexions entre les corticales dans le cerveau d'un chat	[101]
Dolphins	62	159	Un réseau social des associations fréquentes entre 62 dauphins dans une communauté vivant à Doubtful Sound, Nouvelle-Zélande.	[70]
Adjnoun	112	425	Un réseau de la contiguïté des adjectifs et des noms communs dans le roman David Copperfield de Charles Dickens	[89]
Football	115	613	Un réseau de football américain entre les collèges de la division I-A	[47]
Elec s208	122	189	Réseau de circuits électroniques	[82]
Jazz	198	2742	Réseau de musiciens du Jazz	[48]
Elec s420	252	399	Réseau de circuits électroniques	[82]
Neurons	279	2287	Réseau de connexion de neurones de <i>C. Elegans</i>	[1]
Transcription	423	519	Réseau d'interaction de protéines	[1]
Elec s838	512	819	Réseau de circuits électroniques	[82]
E-mail	1133	5451	Réseau d'e-mails échangés entre le personnel de l'université Rovira i Virgili	[52]
Aéroport	3618	14142	Réseau de plus grands aéroports connectés à travers le monde	[53]
PGP	10680	24316	Réseau d'utilisateurs de l'algorithme PGP pour les transactions d'information sécurisée	[15]

Chapitre 2

Réseaux d'interactions sociales

Sommaire

2.1	Introduction	32
2.2	Quelques exemples de réseaux d'interactions sociales	32
2.2.1	Les réseaux d'acteurs	32
2.2.2	Les réseaux de citations	33
2.2.3	Les réseaux de co-auteurs	33
2.2.4	Les réseaux de connaissances	33
2.2.5	Les réseaux d'appels téléphoniques	33
2.3	Propriétés communes aux réseaux d'interactions	33
2.3.1	Distribution des degrés en loi de puissance	34
2.3.2	Petit diamètre et distance moyenne faible	34
2.3.3	Densité faible	35
2.3.4	Coefficient de clustering fort	35
2.4	Modélisation des réseaux d'interactions	36
2.4.1	Modèle aléatoire d'Erdős et Rényi	36
2.4.2	Modélisation des graphes "scale free"	39
2.4.3	Modèle petits mondes (" <i>Small world</i> ") :	42
2.5	Conclusion	43

2.1 Introduction

Depuis un peu plus de deux décennies, date à laquelle le Web fut créé, celui-ci ne cesse de croître atteignant aujourd'hui des dizaines de milliards de pages Web. Ce réseau n'est pas le seul réseau de cette importance. On retrouve de nombreux réseaux d'interactions similaires dans différents domaines. Les études menées sur ce type de réseaux montrent qu'en les modélisant, nous pouvons observer des propriétés communes et ce, en dépit de leurs origines diverses. La connaissance des propriétés de ces réseaux d'interactions est nécessaire pour prévoir leurs évolutions et déterminer leurs capacités à résister aux différents phénomènes ou tout simplement comprendre leur nature. En effet, il est très important de comprendre si un tel réseau est robuste ou non à des attaques ou à des pannes ou encore comment un virus pourrait se propager dans un réseau social.

Comme mentionné ci-dessus, notre étude concerne la robustesse des structures de graphes qui émergent des échanges entre les entités d'un grand réseau d'interactions, mais avant d'aborder cette étude, il nous est apparu nécessaire de donner plus de détails sur ce type de réseaux pour mieux comprendre leur fonctionnement.

Il existe beaucoup de grands réseaux utilisés dans différents domaines. Nous citons à titre d'exemple trois grandes familles : les réseaux liés au Web, les réseaux utilisés en biologie et les réseaux sociaux. Dans notre étude, nous nous intéressons à cette dernière catégorie des réseaux. Il existe, également, beaucoup de réseaux qui sont de nature sociale tels que : les réseaux de citations, les réseaux d'acteurs, les réseaux de co-auteurs ou les réseaux de connaissances. Dans ce chapitre, nous allons aussi discuter quelques modèles décrivant le processus de génération de réseaux complexes et de leurs propriétés.

Ce chapitre est structuré comme suit : nous citons quelques exemples des réseaux d'interactions sociales. Nous donnons par la suite les propriétés communes à ces réseaux et nous terminons le chapitre en citant les modèles de réseaux d'interactions les plus connus.

2.2 Quelques exemples de réseaux d'interactions sociales

2.2.1 Les réseaux d'acteurs

Le graphe des acteurs est un graphe dont les sommets sont des acteurs et deux acteurs sont reliés s'ils ont joué ensemble dans un film [10] [14]. En 2000, la taille du graphe d'acteurs étudié était de 449913 sommets [92].

2.2.2 Les réseaux de citations

Le graphe des citations est un graphe dont les sommets sont des publications scientifiques et deux publications u et v sont liées par l'arc (u, v) si la publication u a cité en référence la publication v . Dans [98], des études ont été réalisées sur un graphe de 783.339 sommets issu du catalogue de l'*Institute for Scientific Information* et un graphe de 24.296 sommets issu des publications de la revue *PhysicalReviewD*.

2.2.3 Les réseaux de co-auteurs

Le graphe de co-auteurs est un graphe dont les sommets sont des auteurs scientifiques et deux auteurs sont reliés s'ils ont une publication commune [9,91,92]. Barabasi *et al.* [9] ont utilisé des données issues des mathématiques (70975 auteurs et 70901 articles) et des neurosciences (209293 auteurs et 210750 articles) entre 1991 et 1998.

2.2.4 Les réseaux de connaissances

Le graphe de connaissance est un graphe dont les sommets sont des personnes et deux personnes sont liées si elles se connaissent [81]. Bien entendu, une telle relation ne peut être définie de façon très formelle et le graphe ne peut être totalement construit.

2.2.5 Les réseaux d'appels téléphoniques

Le graphe des appels téléphoniques est un graphe orienté dont les sommets représentent des numéros de téléphones et un arc signale qu'un numéro a au moins une fois appelé un autre [3].

2.3 Propriétés communes aux réseaux d'interactions

Comme mentionné ci-dessus, l'étude des réseaux d'interactions a révélé que ceux-ci partagent quelques propriétés intéressantes. En effet, plusieurs propriétés sont définies dans la littérature tels que : la distribution des degrés en loi de puissance, la distance moyenne faible ou le petit diamètre, le fort coefficient de clustering, la densité faible, la composante connexe géante et le nombre élevé de bicliques.

Graphe	Taille	Densité moyenne	γ	Référence
WWW_1	325729	4.51	2.45	[5]
WWW_2	$4 \cdot 10^7$	7	2.38	[66]
WWW_3	$2 \cdot 10^8$	7.5	2.72	[18]
Routeur	150000	2.66	2.4	[50]
Appels	$53 \cdot 10^6$	3.16	2.1	[3]
Citations	783339	8.57	3	[98]
Acteurs	212250	28.78	2.3	[10]

TABLE 2.1 – Distribution des degrés de quelques réseaux d'interactions

Graphe	Taille	Degré moyen	Distance moyenne	Référence
Routeur	150000	2.66	11	[50]
Acteurs	212250	61	3.65	[112]
Co-auteurs, neuro.	209293	11.5	6	[9]
Co-auteurs, math.	70975	3.9	9.5	[9]

TABLE 2.2 – Le degré moyen et la distance moyenne

2.3.1 Distribution des degrés en loi de puissance

La distribution des degrés d'un réseau d'interactions suit une loi de puissance de type : $p(k) \sim k^{-\gamma}$ où, k désigne le degré, γ est l'exposant de la loi de puissance. En pratique, la valeur de γ est comprise entre 2 et 3. Le Tableau 2.1 donne, à titre indicatif, la valeur de γ pour quelques types de réseaux d'interactions.

Une distribution des degrés en loi de puissance signifie qu'il existe beaucoup de sommets de faibles degrés et très peu de sommets de forts degrés. γ représente la vitesse de décroissance de la courbe des degrés. Plus γ est grand et plus la probabilité d'obtenir des sommets de forts degrés est petite. Le degré moyen dans un graphe dont la distribution des degrés suit une loi de puissance n'est pas significatif car l'écart-type est très important. Ce type de graphes sont dits *graphes à invariance d'échelle (scale-free graphs)*

2.3.2 Petit diamètre et distance moyenne faible

Contrairement à leurs tailles, les réseaux d'interactions possèdent des petits diamètres. En effet, des études menées sur certains réseaux ont mis à l'évidence cette caractéristique et que le diamètre est de l'ordre de $(\log(n))$ où n est la taille du graphe. Le Tableau 2.2 donne quelques valeurs pour la distance moyenne de quelques graphes.

	Internet	Web	Acteurs	Co-auteurs	Co-occurrence	Protéines
V	75885	325729	392340	16401	9297	2113
E	357317	1090108	15038083	29552	392066	2203
densité(G)	$1.2e^{-4}$	$2.1e^{-5}$	$1.9e^{-4}$	$2.2e^{-4}$	$9.1e^{-3}$	$9.9e^{-4}$

TABLE 2.3 – Nombre de sommets, de liens et densité de quelques réseaux d'interaction

Le Tableau 2.2 montre que les réseaux d'interactions ont une distance moyenne très faible au regard de leurs tailles. Dans la littérature, la distance moyenne des réseaux d'interactions n'est pas souvent donnée mais elle est souvent d'un facteur de 2 à 5 de la distance moyenne pour les graphes non orientés.

Border *et al.* [18] ont montré empiriquement que le graphe du Web a un diamètre très petit par rapport à sa taille. Sur un graphe de 200 millions de pages Web, le diamètre est égal à 500 alors que la distance moyenne est seulement de 16, c'est à dire, si l'on choisit aléatoirement deux pages Web, *si un chemin existe entre elles*, alors la distance moyenne est de 16 dans le cas où le graphe est orienté et 6 dans le cas non orienté. Cette caractéristique trouve son explication dans le fait que la recherche dans les pages Web se fait généralement en largeur.

Dans son expérience, le sociologue *Stanley Milgram* [81] a mis en évidence que deux personnes quelconques sont séparés par une chaîne de longueur au plus 6.

2.3.3 Densité faible

La densité pour un graphe est la proportion d'arêtes dans le graphe. Si n est la taille du graphe G et m le nombre de liens qui existent entre ses nœuds, alors :

$$\text{Densité}(G) = \frac{2m}{n(n-1)}$$

La densité des réseaux d'interactions est très peu élevée (de l'ordre de $\frac{1}{n}$). En effet, le nombre d'arêtes dans un graphe d'interactions est du même ordre que son nombre de sommets. Le Tableau 2.3 donne la densité de quelques graphes d'interaction.

2.3.4 Coefficient de clustering fort

Introduit par Watts et Strogatz [112], le *Coefficient de clustering* mesure la probabilité que le voisinage d'un sommet soit dense. Le Tableau 3.2 donne le coefficient de clustering

Graphe	Taille	Degré moyen	CC	Référence
Sites Web	153127	35.21	0.1078	[2]
Acteurs	225226	61	0.79	[112]
Co-auteurs, neuro.	209293	11.5	0.76	[9]
Co-auteurs, math.	70975	3.9	0.59	[9]

TABLE 2.4 – Le CC de quelques réseaux d'interaction

de certains réseaux d'interaction.

Du tableau 3.2, nous constatons que les réseaux d'interaction possèdent un fort coefficient de clustering malgré que leurs densités sont assez faibles. En résumé, nous pouvons dire que les réseaux d'interaction sont denses localement et densément faibles globalement.

2.4 Modélisation des réseaux d'interactions

Modéliser un réseau d'interaction revient à construire des graphes avec des propriétés similaires aux réseaux réels. L'intérêt d'une telle modélisation est de pouvoir effectuer des simulations de pannes, d'attaques, de propagation et d'autres événements qui peuvent survenir sur les réseaux réels. Dans ce qui suit, nous allons essayer de donner quelques présentations de différents modèles de réseaux d'interactions.

Deux techniques sont principalement utilisées pour la génération de réseaux d'interactions. La première consiste à expliquer comment les propriétés communes aux réseaux d'interactions émergent et la seconde technique consiste à utiliser une propriété commune aux réseaux d'interactions et de faire émerger les autres.

2.4.1 Modèle aléatoire d'Erdős et Rényi

Entre 1950 et 1960, P. Erdős et A. Rényi (ER) ont étudié et proposé les premiers modèles de réseaux d'interactions appelés les graphes aléatoires [34] [35] [36]. Le principe de ce modèle consiste à générer un graphe à n sommets et m arêtes. Les arêtes entre les sommets sont choisies d'une manière aléatoire (cf. Figure 2.1). Dans ces études, les auteurs ont montré que lorsque la taille du graphe est grande, la distribution des degrés d'un graphe $G_{n,p}$, où n est la taille du graphe et p la probabilité avec laquelle le lien entre deux sommets est établi, suit une loi de Poisson [34–36], tel que : $p(k) \sim e^{-\lambda} \frac{\lambda^k}{k!}$.

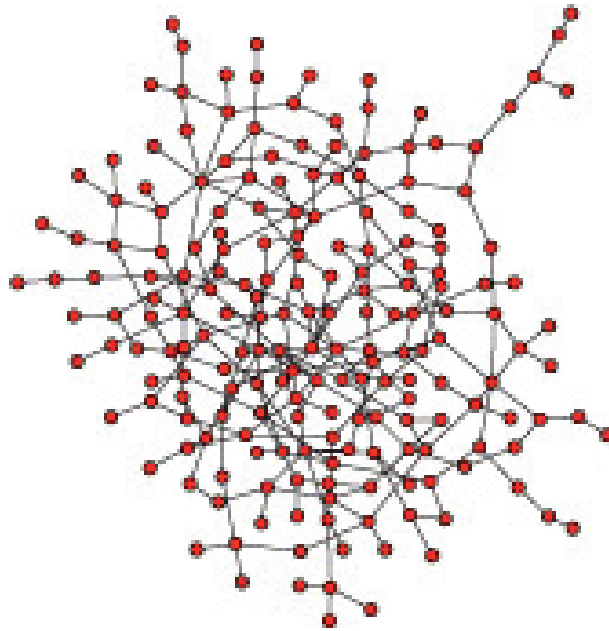


FIGURE 2.1 – Exemple d'un graphe aléatoire

Une telle distribution signifie que les degrés des sommets sont proches du degré moyen *degré moyen* = pn .

Les résultats de ces études montrent que la distribution des degrés des $G_{n,p}$ est différente d'une loi de puissance et que le coefficient de clustering est plus fort dans les réseaux d'interactions que dans les $G_{n,p}$. Par contre, on note deux points en commun entre les $G_{n,p}$ et les réseaux d'interactions : une distance moyenne faible et un diamètre petit.

En outre, il a été aussi démontré que les $G_{n,p}$ ne peuvent être utilisés pour modéliser les réseaux d'interactions. De plus, les $G_{n,p}$ sont des graphes statiques : le nombre de sommets est fixe et ne change plus au cours du temps. Or, il existe des réseaux d'interactions tel que le Web, où le nombre de pages Web et le nombre de liens augmentent chaque jour [69]. En d'autres termes, la modélisation d'Erdős et Rényi ne prend pas en compte la dynamique des réseaux d'interactions.

Propriétés des graphes aléatoires Le coefficient de clustering d'un graphe aléatoire est donné par le nombre de liens entre les voisins d'un nœuds divisé par le nombre maximal des liens possibles. Du moment que la probabilité p est uniforme pour tout le réseau, alors le coefficient de clustering moyen C est simplement donné par p . Étant donné que tous les nœuds dans un graphe aléatoire ont pratiquement un degré proche du degré moyen du graphe, alors le coefficient de clustering décroît avec la taille du graphe n . De la figure 2.2, nous constatons que le coefficient de clustering est indépendant des degrés des nœuds.

Une autre propriété importante d'un graphe aléatoire est sa distance moyenne et son diamètre qui sont inversement proportionnels à sa taille n . Les deux quantités sont représentées à l'échelle $\ln(n)$ [16]. Ce qui indique que ce type de graphe est caractérisé par la propriété *small-world*.

Par ailleurs, dans ce type de réseaux, la distribution des degrés des nœuds suivent une loi de Poisson (cf. Figure 2.3), qui indique que la plupart des nœuds ont approximativement le même nombre de liens.

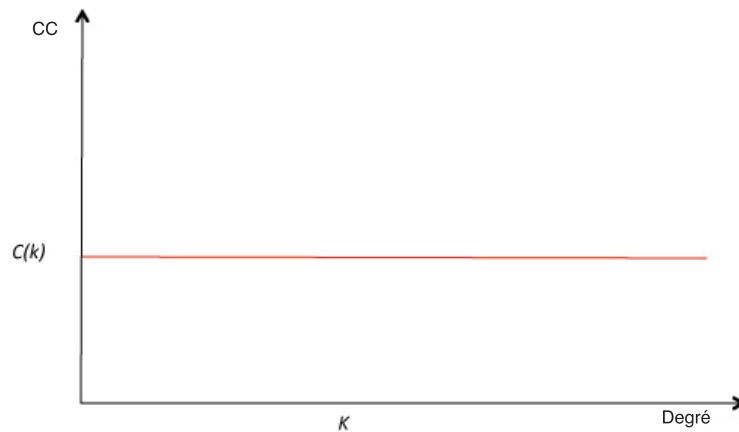


FIGURE 2.2 – Evolution du coefficient de clustering d'un graphe aléatoire

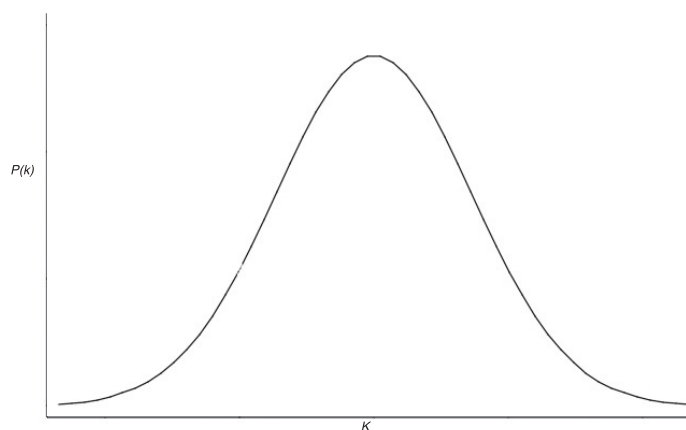


FIGURE 2.3 – Distribution de degrés d'un graphe aléatoire

2.4.2 Modélisation des graphes "scale free"

Comme mentionné plus haut, les graphes $G_{n,p}$ possèdent une distribution de degrés qui suit la loi de Poisson, alors que celle des réseaux d'interactions suit une loi de puissance. Ainsi, plusieurs modèles aléatoires ont été proposés pour générer des graphes aléatoires avec des degrés en loi de puissance.

Le comportement "scale-free" a des conséquences importantes sur la topologie du réseau parce qu'il implique une grande probabilité de l'existence de nœuds de degrés élevés qui agissent en tant que "hubs" dans le réseau.

Les premiers modèles traitant de la particularité "scale-free" ont été discutés longuement dans [28,104,113]. Plus récemment, Barabasi et Albert (BA) ont introduit un modèle ayant un intérêt important pour les réseaux [10].

Modèle à base d'attachement préférentiel Le modèle de Barabasi-Albert pour la génération de graphes d'interaction est un modèle basé sur le concept de l'**attachement préférentiel** où un sommet a une forte probabilité d'être connecté à un nœud de degré élevé [10]. En d'autres termes, un sommet choisi à une étape t a une grande probabilité d'être choisi à l'étape $t + 1$.

Le modèle de Barabasi-Albert ne regroupe pas toutes les propriétés communes des réseaux d'interactions. En effet, les graphes générés ont un taux de regroupement quasi nul et très peu de bicliques.

Le modèle de Barabasi-Albert est basé sur deux hypothèses simples concernant l'évolution du réseau : La *croissance* (growth) et l'*attachement préférentiel* (preferential attachment).

- **Croissance** (growth) : les nouveaux nœuds sont ajoutés au réseau où chaque nouveau nœud est connectée à m nœuds existants.
- **Attachement préférentiel** (preferential attachment) : C'est le principe même du modèle. Chaque nouveau nœud est connecté aux nœuds existants avec une probabilité proportionnelle à leurs degrés.

Considérons l'évolution du réseau à un temps t , où pour chaque intervalle de temps, un nouveau nœud est ajouté au réseau et est connecté aux m nœuds existants avec une probabilité $\Pi(i) = \frac{k_i}{\sum_j k_j}$ (principe du "*rich gets richer*"). La Figure 2.4 explique ce principe.

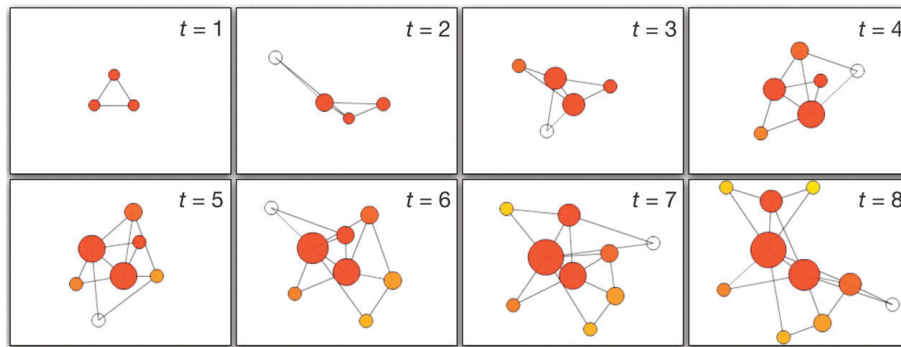


FIGURE 2.4 – Un exemple de réseau scale-free qui s'accroît selon le principe "rich gets richer"

Le modèle de Barabasi-Albert est caractérisé par la production d'une distribution de degrés suivant une loi de puissance de type $p(k) \sim k^{-\gamma}$, avec $2 \leq \gamma \leq 3$. (cf. Figure 2.5)

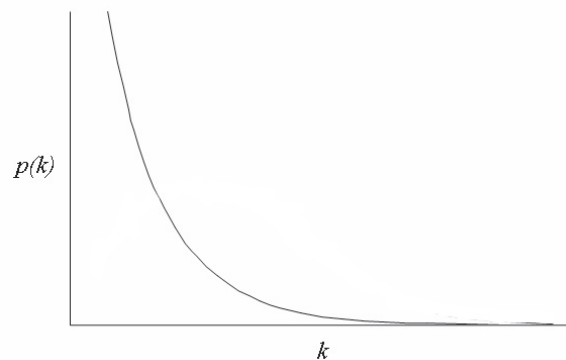


FIGURE 2.5 – Distribution de degrés des graphes scale-free

Par exemple, la distribution de degrés du graphe de la Figure 2.6 avec le modèle de Barabasi-Albert est caractérisé par un exposant $\gamma = 3$.

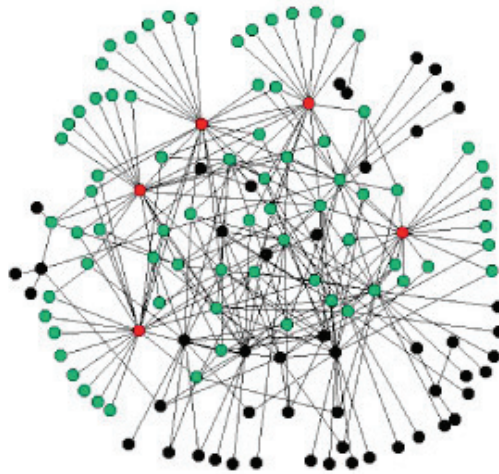


FIGURE 2.6 – Un graphe de 130 nœuds généré avec le modèle de Barabasi-Albert. Les 5 nœuds (en rouge) sont en contact avec les 60% des autres nœuds (en vert)

De nombreux autres modèles basés sur l'attachement préférentiel ont été introduits afin de proposer des processus de construction plus réalistes ou de permettre une analyse formelle des propriétés des graphes engendrés. Pour plus de détails, nous suggérons de lire [9, 32, 38, 106].

Distribution des degrés fixés : Dans [3], les auteurs proposent une méthode basée sur l'appariement de demi-arêtes où le degré de chaque sommet du graphe est fixé selon une distribution des degrés en loi de puissance d'exposant donné et chaque sommet reçoit autant de demi-arêtes que son degré. Ensuite, les demi-arêtes sont reliées entre elles de manière aléatoire et indépendante. Ainsi, la distribution des degrés dépend de deux paramètres α et β tel que :

$$|\{v \mid d(v) = x\}| = y = \frac{e^\alpha}{x^\beta}$$

où, α représente le logarithme du nombre de sommets de degré 1 et β l'exposant de la loi de puissance qui représente la pente de la courbe ou la vitesse de la décroissance des degrés. Le degré maximum du graphe est $e^{\frac{\alpha}{\beta}}$. Par définition, la distribution des degrés suit une loi de puissance de sommets de paramètre β . Le graphe généré par cette méthode peut contenir des boucles et des arêtes multiples mais pour des graphes de grande taille, le nombre de boucles et d'arêtes multiples est négligeable.

Modèle de copie : Ce modèle est basé sur le comportement de l'utilisateur sur le Web. Lorsqu'un utilisateur crée une page web, celle-ci va être interconnectée au reste du Web par des liens hypertextes. Pour cela, l'utilisateur choisit des pages sur le Web (des prototypes). Un prototype est une page web populaire et donc possédant de nombreux liens entrants. Le choix des prototypes n'est pas aléatoire mais plutôt en relation avec le contenu de la nouvelle page Web. Ainsi, l'utilisateur copie des liens hypertextes des prototypes dans la nouvelle page Web. Ce modèle à base de copie permet de construire des graphes riches en bicliques.

Kumar *et al.* [65] proposent deux modèles de copie : le modèle linéaire et le modèle exponentiel. Ces deux modèles génèrent des graphes aléatoires riches en bicliques et avec une distribution des degrés en loi de puissance. Dans le modèle linéaire, à chaque étape un seul sommet est ajouté, or dans le Web plusieurs pages Web sont ajoutées en même temps. Le graphe exponentiel permet d'ajouter à chaque étape un nombre de sommets qui dépend de la taille du graphe à un instant t .

2.4.3 Modèle petits mondes ("*Small world*") :

Un graphe **petit monde** (*small world*) est caractérisé par une distance moyenne faible et un fort coefficient de clustering. Il existe plusieurs modèles générant des graphes avec une distance moyenne faible et d'autres générant des graphes avec un fort coefficient de clustering, mais il n'existe que très peu de modèles regroupant les deux propriétés.

Watts et Strogatz [112] proposent une méthode pour générer des graphes petits mondes (cf. Figure 2.7). Partant d'un anneau régulier à n sommets où chaque sommet est relié à ses $2k$ plus proches voisins (k voisins de chaque côté). Le coefficient de clustering d'un sommet u de l'anneau régulier est assez important : $C(u) = \frac{3(k-2)}{4(k-1)}$ et la distance moyenne, dans un anneau régulier, est elle aussi très élevée.

L'idée de Watts et Strogatz est de modifier suffisamment l'anneau régulier en déplaçant les arêtes afin de diminuer la distance entre les sommets. Concrètement, pour chaque sommet et pour chaque arête, selon une probabilité p , l'arête est redirigée vers un autre sommet choisi d'une manière aléatoire et uniforme. Avec une probabilité $1 - p$, l'arête est gardée.

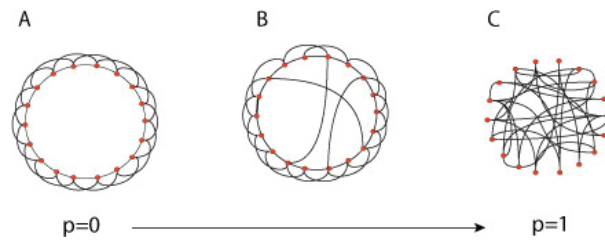


FIGURE 2.7 – Modèle de Watts et Strogatz. **A** : $p=0$, uniquement les nœuds voisins proches sont connectés. Le coefficient de clustering et la distance moyenne entre n'importe quels deux nœuds sont grands. **B** : $p < 1$ mais $p > 0$, peu de courts chemins sont introduits dans le réseau, ce qui réduit la distance moyenne. **C** : p est grand et le réseau est équivalent à un graphe aléatoire. Dans ce cas, la distance moyenne et le coefficient de clustering sont trop faibles.

Deux autres méthodes ont été proposées par Latapy *et al.* [96]. Ces auteurs ont proposé deux modèles aléatoires pour les réseaux pouvant être représentés par des graphes bipartis. A titre d'exemple, le graphe des acteurs est un graphe dont les sommets sont des acteurs de cinéma et dont deux sommets sont reliés entre eux si deux acteurs ont joué dans un même film. Ce graphe peut-être vu comme un graphe biparti $G = (V_1, V_2, E)$ dans lequel V_1 est l'ensemble des films et V_2 l'ensemble des acteurs. E représente l'ensemble des liens entre les sommets de V_1 et de V_2 . Un lien est établi quand un acteur a joué dans un film. Les réseaux de co-auteurs et co-occurrence sont également des réseaux décomposables en plusieurs bipartis. Les auteurs observent que les trois réseaux réels (acteurs, co-auteurs et co-occurrence) ont une propriété commune : la présence de lois de puissance dans les distributions de degrés de V_2 . Les distributions des degrés de V_1 sont de deux types : en loi de Poisson pour "Co-occurrence" et "Co-auteurs" et, en loi de puissance, pour "Acteurs".

2.5 Conclusion

Au cours des dernières années, des progrès dans les réseaux complexes ont permis de découvrir des similitudes importantes entre ces divers systèmes tels que l'Internet, les réseaux cellulaires neuronaux, les systèmes métaboliques, *etc.*

Aujourd'hui, les réseaux sont de plus en plus intégrés et interconnectés pour plus d'informations, de l'énergie, de transport ou de commerce. La nature critique de ces réseaux suscite des inquiétudes sur le risque et l'impact des défaillances du système, et il est impératif de mieux comprendre le fonctionnement de ces réseaux complexes. Pour cela, une étude approfondie de ces systèmes est nécessaire qui permettra de mieux comprendre comment sont structurés ces réseaux complexes. Dans ce chapitre, nous avons donné les

modèles de graphes les plus utilisés et cité quelques exemples de réseaux d'interactions sociales. Nous avons vu que le modèle de Watts et Strogatz génère des graphes petits mondes mais avec une distribution des degrés différente d'une loi de puissance. A l'inverse, le modèle de Barabasi-Albert génère un graphe avec une distribution des degrés en loi de puissance.

Dans le chapitre suivant, nous allons étudier les graphes sociaux que nous obtenons avec notre modèle selon les trois propriétés (loi de puissance, diamètre et coefficient de clustering) définies précédemment. Nous étudierons les propriétés mathématiques de ces graphes sociaux et nous proposerons une classification de ceux-ci.

Chapitre 3

Etude des structures sociales émergées

Sommaire

3.1	Introduction	46
3.2	Plateforme d'émergence de structures sociales	46
3.3	Description de la plateforme	49
3.4	Etude des structures émergées	49
	3.4.1 Propriétés des structures émergées	49
	3.4.2 Classification des structures de communautés émergées	55
3.5	Conclusion	60

3.1 Introduction

L'étude des structures sociales est un challenge important dans l'analyse des réseaux. Plusieurs méthodes et modèles ont été proposés dans la littérature pour déterminer d'une manière statistique la signification des structures découvertes. Par ailleurs, la plupart des analyses existantes considèrent uniquement l'aspect structurel des communautés émergées. Dans notre travail, nous nous intéressons à l'étude de la robustesse de ces structures émergées dans les réseaux d'interaction. Plus précisément, nous considérons les communautés émergées dans les graphes induits par les échanges entre entités du réseau d'interaction considéré.

Notre étude se focalise sur deux aspects : le premier aspect concerne l'étude et la classification des structures émergées. Le second aspect quant à lui traite de la robustesse de ces mêmes structures. Dans ce chapitre, nous décrivons notre modèle qui permet l'émergence des graphes sociaux que nous avons conçus. Nous étudions ces graphes sociaux selon certaines de leurs propriétés mathématiques et nous proposons une classification des différents types de ces graphes.

3.2 Plateforme d'émergence de structures sociales

Pour étudier l'émergence des structures sociales dans les réseaux de communication, nous considérons un graphe $G = (V, E)$, où V est l'ensemble des nœuds et E l'ensemble des arêtes (liens). Les nœuds peuvent partager des ressources dans le réseau. Un nœud offrant des ressources est dit "Fournisseur" ou "Serveur". La proportion de nœuds faisant office de fournisseur est notée par S et celle des ressources que contiennent ces nœuds est désignée par R . Les ressources sont regroupées par thématique. Par ailleurs, tous les nœuds, dès lors qu'ils interagissent dans le réseau, possèdent des centres d'intérêt (domaine d'intérêt). Les nœuds peuvent avoir un ou plusieurs intérêts et leur proportion est notée T . Tous ces paramètres varient d'une basse valeur à une forte valeur selon le nombre de nœuds dans le graphe $n = |V|$.

L'ensemble des nœuds fournisseurs est noté par V_s , $V_s \subseteq V$. Alors, S est défini par $S = \frac{|V_s|}{|V|}$.

Soit $v \in V_s$ un nœud fournisseur du graphe et soit $Res = \{r_1, r_2, \dots, r_k\}$ l'ensemble de toutes les ressources existantes dans le réseau. L'ensemble des ressources que le fournisseur v comporte est noté par $Res_v = \{r_i, \dots, r_j\} \subseteq Res$, avec $1 \leq i \leq j \leq k$. Le paramètre R

est donné par

$$R = \frac{\sum_{v \in V_P} \frac{|Res_v|}{|Res|}}{|V_s|}$$

De plus, soit $v \in V$ un nœud du graphe G . Le centre d'intérêt du nœud v est noté par un ensemble de toutes les ressources qui l'intéressent $I_v = \{r_i, \dots, r_j\} \subset Res$, avec $1 \leq i \leq j \leq k$. Le paramètre T est donné par

$$T = \frac{\sum_{v \in V} \frac{|I_v|}{|Res|}}{|V|}$$

L'émergence du graphe social induit par les interactions dans le réseau de communication est obtenue comme suit :

1. Tous les nœuds génèrent des requêtes selon leurs centres d'intérêts. Un nœud fournisseur qui satisfait la requête répond au nœud demandeur et un **lien d'interaction** est établi entre ces deux nœuds. En d'autres termes, le lien d'interaction entre les deux nœuds est ajouté si la requête du nœud demandeur est satisfaite par le nœud fournisseur. Par ailleurs, nous utiliserons le paradigme de l'attachement préférentiel tel que défini par Barabasi et Albert [10] pour obtenir des structures de graphe avec distribution de degrés qui suit une loi de puissance. Dans leur modèle, Barabasi et Albert utilisent le principe du "*rich get richer*" qui stipule que "*plus un nœud a un degré élevé, plus il a de la chance d'avoir de nouvelles connections*". Cette particularité est due à l'attachement préférentiel. Cela signifie, un nœud recherche d'abord dans les nœuds fournisseurs avec lesquels il a eu des échanges avant qu'il ne consulte d'autres fournisseurs. Ce type de mécanisme est détaillé dans le modèle PROSA [21]. Dans un tel modèle, les liens d'interaction obtenus sont considérés indépendamment de leur orientation.
2. Une fois que toutes les requêtes sont exécutées, un nouveau graphe d'interaction émerge. On note ce graphe $G' = (V', E')$ où $V' \subseteq V$ est l'ensemble des nœuds ayant interagi les uns avec les autres et E' l'ensemble des liens d'interaction (requêtes acceptées) entre les nœuds. Identiquement au modèle de PROSA [21], nous considérons G' comme un graphe non-orienté.
3. Nous assimilons les ressources à des documents. Par analogie avec la distribution du contenu dans [79], nous considérons que les fournisseurs contiennent plusieurs documents dans leurs centres d'intérêt et que les documents restants se trouvent dans d'autres centres d'intérêt.
4. Dans nos expérimentations, les différentes valeurs de n varient de 250 à 10000 nœuds. Les paramètres R et S varient entre 25%, 50% et 75%. Rappelons que pour les ressources, R représente la proportion de ressources que contient un nœud fournisseur.

Le paramètre T varie quant à lui entre 5%, 15% et 30% et désigne la proportion des différentes ressources qui intéressent le nœud demandeur. Dès lors que la détection des communautés dans les graphes est coûteuse en terme de calcul (complexité $O(n^3)$), le passage à l'échelle est limité à quelques milliers de nœuds [105]. Plusieurs algorithmes sont présentés dans la littérature, mais comme nous nous intéressons dans notre travail à l'étude de l'influence des paramètres R , S , et T , nous adoptons l'algorithme de bisection spectral proposé par Newman dans [90] pour la qualité de partitionnement du graphe en communautés.

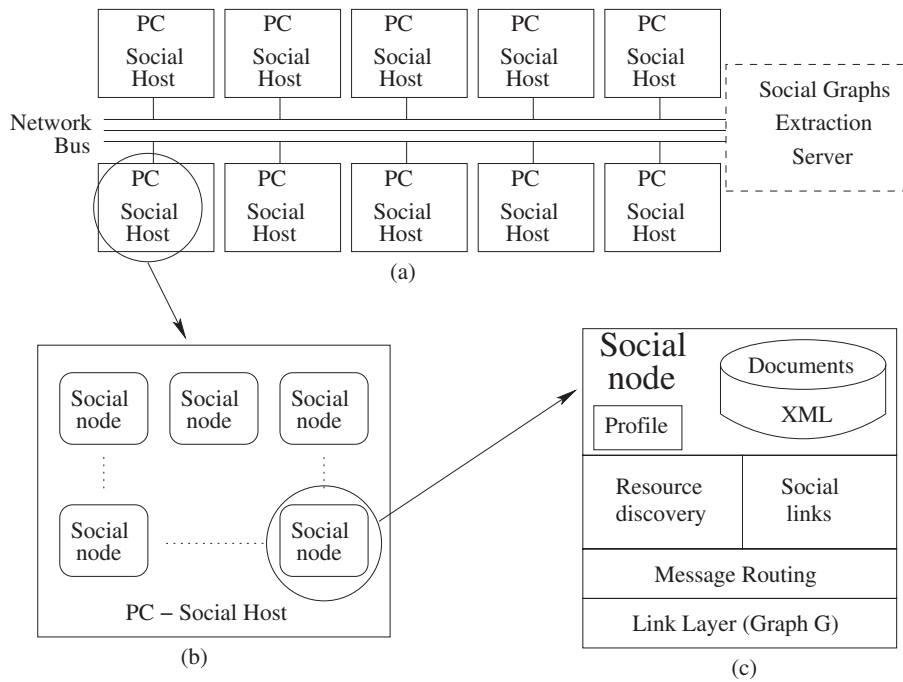


FIGURE 3.1 – Architecture de la plateforme

Comme cité précédemment, notre étude est scindée en 2 grandes parties. La première partie concerne l'étude de l'émergence des structures de communautés dans les réseaux d'interaction et la seconde partie sera dédiée à l'étude de la robustesse de ces mêmes structures. Notre objectif dans cette thèse est de déterminer les paramètres initiaux (la configuration du réseau) qui permettent l'émergence de communautés, et dans le cas d'émergence, déterminer les paramètres qui influent sur la robustesse des structures des communautés émergées.

Pour l'étude de la première partie de notre travail, nous avons défini une plateforme (framework) qui nous permet de récupérer le graphe social représentant les interactions entre les entités du réseau.

3.3 Description de la plateforme

La plateforme que nous avons conçue, définie et utilisée dans nos expérimentations est décrite dans la Figure 3.1. Un ensemble d'ordinateurs (PC) contiennent plusieurs nœuds sociaux. Ces nœuds ou *hosts* sont connectés à un serveur à travers un réseau local (Figure 3.1(a) et Figure 3.1(b)). Les serveurs tracent toutes les interactions et extraient le réseau social émergé (nous notons G' le graphe social ainsi émergé) une fois que toutes les requêtes sont traitées. Un nœud social est un objet composé d'un profil définissant son intérêt, en concordance avec le paramètre T et l'ensemble de ses ressources est défini en concordance avec les paramètres R et S .

Cette plateforme ainsi définie nous permet de traiter de l'émergence de communautés dans les réseaux d'interactions, comme l'indique la Figure 3.1. Cette plateforme comporte trois parties :

1. La première partie **Entrée du framework** qui n'est autre que le graphe représentant le réseau. Ce graphe est caractérisé par les trois paramètres R , S et T cités précédemment, la taille du graphe n et sa densité d (nombre de liens dans le graphe). Ces différents paramètres nous permettent de varier les types de réseaux en entrée et simuler différents types de réseaux.
2. La deuxième partie **Emergence du graphe social**. Une fois le réseau physique construit et les échanges entre les nœuds du réseau établis, une topologie virtuelle émerge, représentant les échanges entre les nœuds dans le réseau.
3. La troisième partie **Détection de communautés**. Une fois la topologie virtuelle obtenue (graphe social), nous appliquons un algorithme de détection de communautés pour déterminer les communautés d'intérêts dans le réseau social. Dans notre cas, nous utilisons l'algorithme spectral de Newman [90].

Dans ce qui suit, nous allons étudier les propriétés structurelles du graphe émergé G' .

3.4 Etude des structures émergées

3.4.1 Propriétés des structures émergées

Dans cette section, nous nous intéressons aux caractéristiques structurelles des graphes émergées des interactions dans le réseau. La figure 3.2 illustre un exemple du graphe émergé G' à partir du graphe initial G .

Dans cette étude, nous considérons trois métriques pour capturer les propriétés du réseau social. Ces métriques sont : le chemin moyen, la distribution de degrés et le coefficient

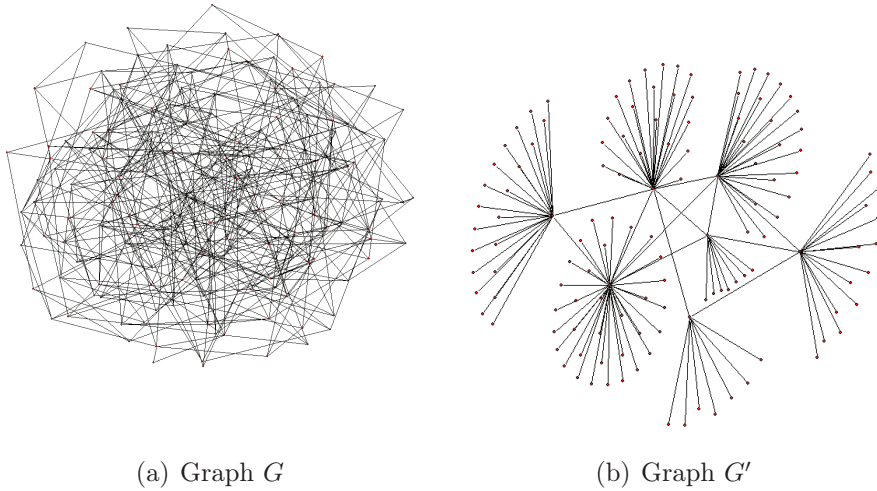


FIGURE 3.2 – Emergence du graphe d'interactions

de clustering. Les graphes sociaux sont souvent des graphes *petits mondes* ou *Scale-free*. Dans de tels graphes, la longueur moyenne d'un chemin entre deux nœuds doit être relativement petite par rapport à la taille du graphe social. Par ailleurs, la propriété *scale-free* induit une distribution de degrés qui suit une loi de puissance. De plus, les relations dans les réseaux sociaux sont souvent transitives. Par conséquent, la transitivité (coefficient de clustering) doit être relativement élevée.

3.4.1.1 Longueur moyenne des chemins (Average path length "APL") :

L'APL d'un graphe est défini comme étant la distance moyenne entre toutes les paires de nœuds dans le graphe. Notons que la distance entre deux nœuds est le plus court chemin entre eux. Par exemple, les réseaux *Scale free* ont un APL de l'ordre de $LnLn(n)$ (n représente le nombre de nœuds dans le réseau), alors que les réseaux *Small-world* ont un APL de l'ordre de $Ln(n)$ [5].

Dans notre étude, tous les graphes que nous avons obtenus des différentes configurations étudiées dans notre modèle, possèdent un APL compris entre $Log(n)$ et $LogLog(n)$. Cette importante caractéristique nous renseigne sur la nature de ces graphes émergés. En effet, cette première caractéristique confirme l'"*effet petit monde*", qui est une propriété "*Scale free*".

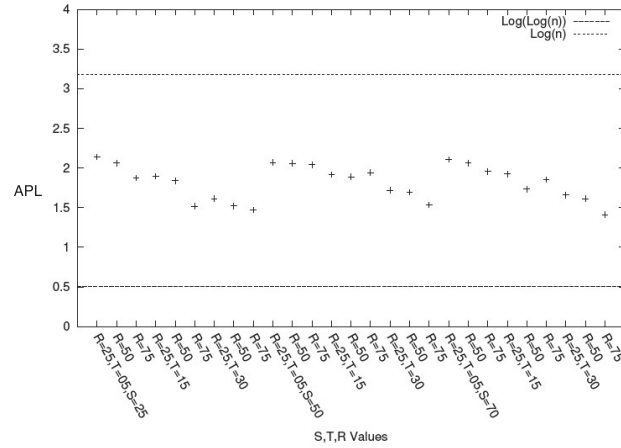


FIGURE 3.3 – APL moyen des graphes émergés de taille 1500 nœuds

		R faible	R moyen	R élevé
S faible	T faible	2.12	1.83	1.57
	T moyen	2.12	1.70	1.57
	T élevé	2.08	1.84	1.59
S moyen	T faible	2.14	1.94	1.69
	T moyen	2.10	2.01	1.75
	T élevé	2.08	1.97	1.41
S élevé	T faible	2.13	1.89	1.55
	T moyen	2.12	1.82	1.54
	T élevé	2.12	1.84	1.53

TABLE 3.1 – Valeurs de l'APL

3.4.1.2 Coefficient de Clustering ou Transitivité

En plus des petites valeurs de l'APL, les réseaux sociaux possèdent un coefficient de clustering relativement élevé [5]. En terme de topologie de graphes, le coefficient de clustering représente le nombre élevé de triangles dans le réseau, c'est à dire des ensembles de trois nœuds connectés entre eux. Le coefficient de clustering CC est élevé si deux nœuds possèdent un voisin ayant une forte probabilité d'être lui même connecté à ces deux nœuds. Le coefficient de clustering CC d'un nœud i ayant pour degré d_i est donné par :

$$C(i) = \frac{2 \times \text{nombre de liens entre les voisins de } i}{d_i(d_i - 1)}$$

Le coefficient pour le graphe tout entier n'est autre que la moyenne des différentes valeurs des coefficients de clustering de tous ses nœuds, soit : $CC = \frac{1}{n} \sum CC_i$.

La table 3.2 donne les valeurs de CC obtenus pour des graphes de 1500 nœuds.

		R low	R medium	R high
S low	T low	0.57	0.63	0.62
	T medium	0.61	0.63	0.64
	T High	0.62	0.63	0.63
S medium	T low	0.56	0.63	0.62
	T medium	0.61	0.62	0.64
	T High	0.62	0.63	0.64
S high	T low	0.59	0.63	0.63
	T medium	0.61	0.64	0.64
	T High	0.62	0.64	0.64

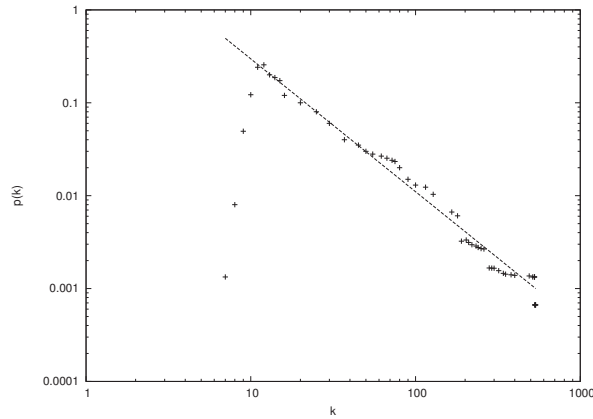
TABLE 3.2 – Valeurs moyennes du coefficient de clustering

Nous remarquons que les valeurs obtenues sont relativement élevées

3.4.1.3 Distribution de degrés

Le degré d'un nœud dans un réseau est le nombre de liens qui lui sont incidents. La probabilité $p(k)$ d'un degré k est la proportion de nœuds ayant pour degré k . La distribution de degrés peut aider à distinguer entre les différentes classes de réseau. Par exemple, une distribution de degrés de type *Gaussienne* est fréquemment présente dans les réseaux aléatoires, tandis qu'une distribution de type *Puissance* de type $p(k) \sim k^{-\gamma}$, indique que peu de nœuds ont des degrés élevés (hubs) et que la majorité des nœuds d'un tel réseau ont des degrés faibles. Ce second type de distribution de degrés caractérise les réseaux *Scale-free* [5].

Le tracé de la distribution de degrés de ces réseaux est souvent représenté dans une échelle logarithmique. Du moment que $p(k) \sim k^{-\gamma}$, alors $\log(p(k)) \sim -\gamma \log(k)$. Par conséquent, le tracé de la courbe ressemble à une droite. La figure 3.4 donne un exemple de la distribution de degrés pour une de nos configurations et la table 3.3 donne toutes les valeurs obtenues de γ pour un graphe de 1500 nœuds.

FIGURE 3.4 – Distribution de degrés d'un graphe émergé G' de 1500 nodes

Dans notre étude, les distributions de degrés obtenues des interactions entre les nœuds suivent une loi de puissance de type $p(k) \sim k^{-\gamma}$ et les valeurs de γ varient entre 1 et 3. Le tableau ci-après montre les valeurs de γ pour toutes les configurations (*i.e.* en concordance avec les proportions de nœuds fournisseurs S , de ressources R et des centres d'intérêt des nœuds T).

		R low	R medium	R high
S low	T low	2.85	2.35	2.02
	T medium	2.24	1.76	1.62
	T High	1.69	1.61	1.49
S medium	T low	2.33	1.81	1.63
	T medium	1.68	1.62	1.30
	T High	1.58	1.28	1.18
S high	T low	1.97	1.66	1.63
	T medium	1.69	1.44	1.23
	T High	1.40	1.12	1.01

TABLE 3.3 – Valeurs de γ pour un graphe de 1500 nœuds

3.4.1.4 Discussion

L'un des principaux objectifs de ce travail est l'étude des propriétés structurelles des structures de graphes émergées. En effet, la plupart des graphes obtenus possèdent des distributions de degrés suivant une loi de puissance. Par ailleurs, ces graphes ont des diamètres petits et des coefficients de clustering assez élevés. Ces trois caractéristiques nous

renseignent sur ces graphes qui sont de nature *Scale-free* (voir figure 3.5). Par conséquent, nous pouvons conclure que notre modèle génère des graphes sociaux induits des interactions entre les nœuds dans le réseau, ayant des propriétés proches des réseaux sociaux réels. La figure 3.5 montre un exemple de graphe avec $\gamma = 2.54$.



FIGURE 3.5 – Un graphe émergé avec $\gamma = 2.54$

Aspects des structures émergées La première partie de notre de notre travail consiste en l'étude et l'analyse des structures de graphes émergés des communications entre les nœuds dans le graphe physique. Pour ce faire, nous nous basons sur quelques propriétés de la théorie des graphes.

Les observations et les esquisses des différents résultats de nos expérimentations nous ont permis d'étudier les structures selon les trois propriétés suivantes :

1. Le **degré moyen des communautés** deg_moy . Il est donné

$$deg_moy(C) = \frac{\sum_i^n degre(i)}{n} \quad (3.1)$$

où, $degre(i)$ est le nombre de voisins du nœud i dans la communauté C .

2. Le **degré maximum** deg_max des communautés formées. Il représente la valeur maximale de tous les degrés des nœud d'une même communauté. Pour rappel, le degré d'un nœud i est le nombre d'arêtes qui lui sont incidentes.
3. Le **nombre de degré 1** nb_deg1 est le nombre de nœuds de degré 1, i.e. les nœuds feuilles dans le graphe. Du point de vue "communauté", il s'agit des nœuds qui ne sont connectés qu'à une seule communauté.
4. La **densité des communautés** d_comm . La densité moyenne des communautés d'un partitionnement s'obtient par le calcul de la densité de chaque communauté à part (considéré comme un graphe). Soit C une des communautés formées, et soient

n et m le nombre de nœuds dans C et le nombre d'arêtes dans C , respectivement. La densité de C est calculée comme suit :

$$d_comm(C) = \frac{m}{n \cdot (n - 1) / 2} \quad (3.2)$$

où, $n \cdot (n - 1) / 2$ est le nombre maximum d'arêtes que peut avoir un graphe de n nœuds (nombre d'arêtes d'un graphe complet de n nœuds).

5. Le "*Coefficient de Clustering (CC)*". Le CC est la proportion du nombre de triangles dans un graphe donné. Il est calculé comme la moyenne des CC de tous les nœuds du graphe et il est donné par l'équation :

$$CC(i) = \frac{2 * k}{degre(i) * (degre(i) - 1)} \quad (3.3)$$

où k représente le nombre d'arêtes entre les voisins du nœud i .

6. La *Distribution de degrés dis_deg*. Elle représente l'énumération du nombre de nœuds pour chaque degré. La distribution de degrés est représentée sur un graphique où l'axe des abscisses (x) contient les degrés et l'axe des ordonnées (y) contient le nombre de nœuds pour chaque degré.

3.4.2 Classification des structures de communautés émergées

L'étude structurelle des communautés émergées des communications dans le réseau, à travers les propriétés que nous avons décrites précédemment, nous permet de proposer une classification de ces communautés selon les structures de graphes connues dans la théorie des graphes tels que : l'étoile, la clique, l'arête, *etc*, et également selon les paramètres du réseau initial.

Une fois que les communautés sont formées, nous nous intéressons à la manière dont les nœuds sont interconnectés entre eux. En effet, plus les nœuds sont connectés plus la relation sociale entre eux est forte, et vice versa. En d'autres termes, nous observons que plus il y a des liens entre les nœuds, plus les communautés seront denses. Il revient par conséquent de mettre l'accent sur la corrélation entre les propriétés structurelles d'une communauté et les interactions induites entre les nœuds de la communauté.

Pour obtenir une classification des structures de communautés émergées aussi fidèle que possible, nous considérons un espace tridimensionnel dans lequel chaque dimension est représentée par l'un des critères suivant :

- La densité moyenne, " d_comm ".
- Le nombre de nœuds de degré 1, " nb_deg1 ".
- Le degré maximum de la communauté, " deg_max ".

Dans cette même idée de normalisation des deux autres propriétés nb_deg1 et deg_max , nous divisons le premier par n (taille de la communauté), sachant qu'un chemin entre 2 nœuds a tous les nœuds de degré 1. Pour le deuxième critère, nous le divisons par $n - 1$, le nombre maximum de voisins que peut avoir un nœud dans un graphe de taille n .

Nous remarquons des structures de communautés émergées qu'une relation "célébrité-fan" entre les nœuds apparaît (voir Figure 3.6).

- L'interaction célébrité-fan : combien de célébrités y-a-t-il dans la communauté que les fans connaissent et avec lesquelles ils interagissent ? ...
- L'interaction célébrité-célébrité : les célébrités connaissent-elles d'autres célébrités dans la communauté ? Combien ? ...
- L'interaction fan-fan : les fans connaissent-ils d'autres fans dans la communauté ? Combien ? ...

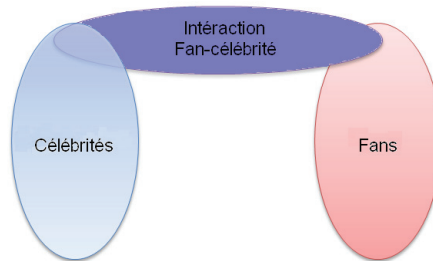


FIGURE 3.6 – Les interactions dans une communauté

Dans ce qui suit, nous discutons les coordonnées et les interactions induites de chaque classe.

- **Étoile** : Le nombre de nœuds de degré 1 de cette classe est élevé (≈ 1) et le degré maximum est aussi grand (≈ 1). Il apparaît clairement que le centre d'intérêt de l'étoile est la seule célébrité de la communauté. Tous les autres nœuds (feuilles) sont des fans.
- **Bistar** : Le nombre de nœuds de degré 1 du bistar est très grand (≈ 1) et le degré maximum est moyen ($\approx \frac{1}{2}$). Les deux centres du bistar sont les célébrités de la communauté. Tous les autres nœuds sont des fans. Dans cette classe, il n'existe que peu d'interactions entre les fans.
- **Clique** : Dans cette catégorie de graphe, le nombre de nœuds de degré 1 est nul ($= 0$) et le degré maximum est très grand ($= 1$). Chaque nœud connaît et interagit avec tous les autres nœuds de la communauté. La communauté est alors fortement connectée.

- **Sun** : Le graphe de type *Sun* est obtenu en ajoutant un nœud voisin à chaque nœud d'un graphe complet (clique). Le nombre de nœuds de degré 1 de cette classe est moyen ($\approx \frac{1}{2}$) et le degré maximum est aussi moyen ($\approx \frac{1}{2}$). Les feuilles de ce type de graphe sont les fans de la communauté et les nœuds du sous-graphe complet sont les célébrités. Chaque célébrité interagit avec presque toutes les autres célébrités, alors que les fans n'interagissent qu'avec une seule célébrité (ou au plus avec peu de célébrités). Dans cette classe, il n'y a presque pas d'interaction entre fans.
- **Caterpillar** : Les caterpillars sont des graphes ayant un nombre moyen de nœuds ayant le degré 1 (entre $\frac{1}{2}$ et des valeurs proches de 1) et le degré maximum est petit ($\approx \frac{\log(n)}{n-1}$). Dans ce type de graphes, il y a des célébrités qui n'interagissent pas beaucoup entre elles. Chaque célébrité a un groupe de fans avec lequel elle interagit. Un fan connaît uniquement une célébrité et n'interagit pas avec les autres fans.
- **Roue** : Dans cette classe de graphe, les nœuds de degré 1 sont quasiment inexistantes (≈ 0) et le degré maximum est très grand (≈ 1). Le centre de la *Roue* est l'unique célébrité de la communauté. Tous les autres nœuds sont des fans. Dans ce type de graphe, il existe une faible interaction entre les fans.
- **Fan** : Les graphes de ce genre regroupent les nœuds qui ont un degré moyen ($\approx \frac{1}{2}$) et le nombre de nœuds de degré 1 est pratiquement nul (≈ 0).
- **Arbre** : Les arbres sont des graphes possédant un nombre réduit ou moyen de nœuds de degré 1 ($\approx \frac{\log(n)}{n}$), cependant, dans certains cas particuliers, ce nombre peut-être élevé mais sans pour autant atteindre des valeurs proches de 1). Le nombre de nœuds de degré maximum est aussi petit ($\approx \frac{\log(n)}{n-1}$).

Nous pouvons constater, qu'une fois normalisé, deg_moy donne la même valeur que d_comm . En effet, considérons un graphe $G = (V, E)$, où $|V| = n$ et $|E| = m$. On a $d_comm = \frac{m}{\frac{n(n-1)}{2}} = \frac{2m}{n(n-1)}$. Aussi, on a $deg_moy = \frac{\sum_{v \in V} degree(v)}{n} = \frac{2m}{n}$. Sachant que le degré maximum d'un graphe à n nœuds est de $n - 1$, pour normaliser le deg_moy , on le divise par $n - 1$. Donc, on aura $deg_moy = \frac{2m}{n(n-1)}$, ce qui fait que l'utilisation de d_comm dans l'une des dimensions suffit sans le deg_moy .

3.4.2.1 Corrélation entre réseaux initiaux et structures de communautés

Selon la valeur de la densité, nous obtenons des structures de communautés différentes. En effet, pour une **faible densité**, les nœuds d'une même communauté n'ont pas tendance à se connecter entre eux et se contentent de se connecter presque à un même nœud. Cela se traduit par l'apparition de plusieurs nœuds de degré 1 et peu de nœuds de degré élevé. En théorie des graphes, les structures qui répondent à cette particularité sont des structures de type, chaîne ou cycle, chemin, arbre,...

Dans ce même esprit, nous pouvons observer que les types des structures de graphes de *densité moyenne* regroupent un grand nombre de nœuds de degré 1 et également un grand nombre de nœuds de degré élevé. En effet, un graphe ne peut avoir une densité moyenne et un grand nombre de nœuds de degré 1, ou très peu de nœuds de degré maximum. Donc, les combinaisons possibles sont celles s'apparentant à des structures de graphes de types éventail ou sun (soleil).

Pour ce qui est des structures de communautés ayant une densité moyenne forte, ceux-ci sont de type clique. La figure 3.7(a) résume ces structures :

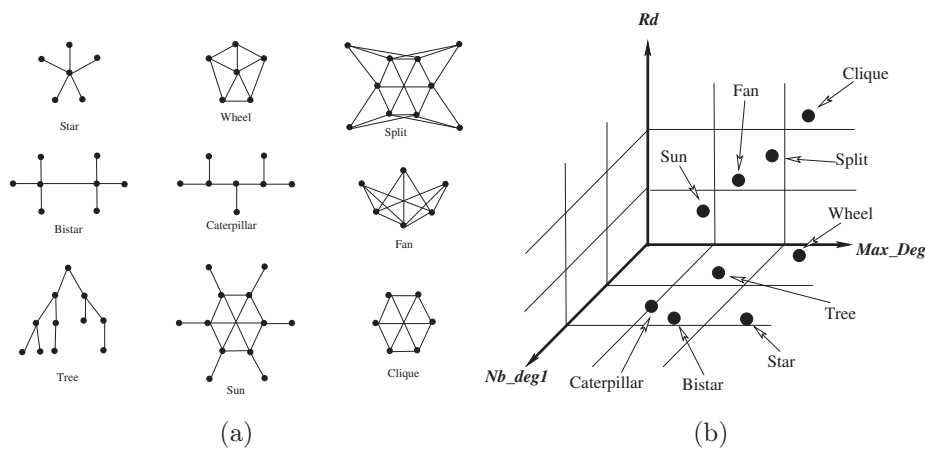


FIGURE 3.7 – (a) Les classes de graphes, (b) Classification structurelle

L'examen des différentes structures de communautés obtenues de nos expérimentations au travers des propriétés de la théorie des graphes, notamment les trois critères : densité moyenne (d_comm), nombre de nœuds de degré 1 (nb_deg1) et degré maximum (deg_max) nous ont permis de concevoir et de proposer une classification des structures de graphes émergées. La Figure 3.7(b) résume cette classification.

3.4.2.2 Impact des paramètres R , S et T sur la construction des communautés

Dans cette section, nous donnons un intérêt particulier à l'étude de l'influence des proportions de S , R et T sur les structures de communautés émergées. Le Tableau 3.4 donne la correspondance entre ces paramètres et les structures de communautés obtenues.

L'exploitation des différentes tendances contenues dans le Tableau 3.4 nous amène à définir un automate pour illustrer et récapituler l'influence des paramètres R , S et T sur les structures des communautés émergées. Il est à noter que les liens dans cet automate sont bidirectionnels. Tel qu'illustré par la Figure 3.8, si une configuration initiale donne des communautés de type *étoile* et si on diminue la proportion des centre d'intérêt des

			d_{comm}			nb_deg1			deg_max			Structure
			F	M	E	F	M	E	F	M	E	
S faible	R faible	T faible			X	X				X		Fan
		T moyen		X			X			X		Sun
		T fort	X				X				X	Wheel/Star
	R moyen	T faible			X		X			X		Fan
		T moyen	X			X			X			Tree
		T fort		X		X					X	Split
	R fort	T faible		X		X				X		Fan
		T moyen		X		X			X			Fan
		T fort	X					X		X		Bistar
S moyen	R faible	T faible			X	X				X		Split
		T moyen		X		X				X		Fan
		T fort	X				X				X	Wheel/Star
	R moyen	T faible			X			X	X			Sun
		T moyen		X		X				X		Fan
		T fort	X					X			X	Star
	R fort	T faible			X			X		X		Sun
		T moyen	X			X			X			Tree
		T fort	X			X					X	Wheel
S fort	R faible	T faible			X		X		X			Sun
		T moyen	X					X	X			Caterpillar
		T fort	X				X				X	Wheel/Star
	R moyen	T faible			X		X		X			Sun
		T moyen		X				X		X		Sun
		T fort	X				X				X	Wheel/Star
	R fort	T faible			X		X		X			Sun
		T moyen	X					X		X		Bistar
		T fort	X				X				X	Wheel/Star

TABLE 3.4 – Correspondance entre R , S , T et les structures de communautés. (**F** : Faible, **M** : Moyen, **E** : Elevé)

nœuds T , alors on obtiendra des communautés de type *soleil*. Par ailleurs, si nous considérons la diminution du nombre de fournisseurs dans le réseau, les communautés obtenues ressemblent beaucoup plus à des *caterpillar*. Enfin, en ajoutant plus de nœuds fournisseurs et en réduisant des ressources à une configuration initiale dont les communautés

émergées sont de type "fan graphs", on obtient des communautés de type "Soleil".

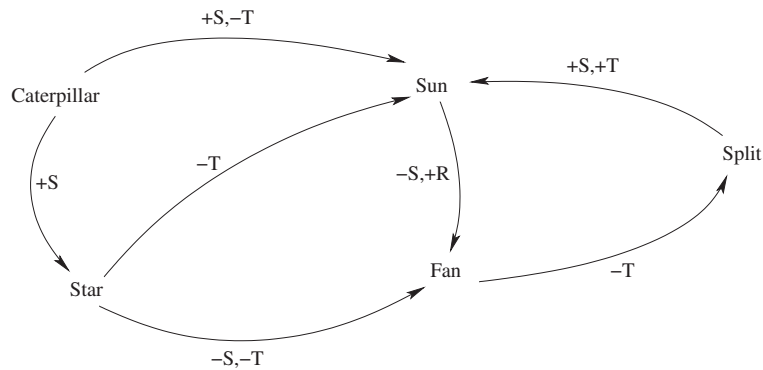


FIGURE 3.8 – Influence de R , S , T sur les structures de communautés

3.5 Conclusion

Dans ce chapitre, nous avons proposé un framework pour l'étude de l'émergence des structures de communautés dans les réseaux d'interaction dans lequel nous avons modélisé les réseaux initiaux par les graphes. Ce framework comporte deux parties : émergence de communautés et classification de celles-ci.

La partie émergence des communautés montre comment modéliser les réseaux d'interaction ainsi que comment extraire les structures sociales. La deuxième partie de notre framework s'intéresse, quand à elle, à la classification structurelle de ces communautés selon les trois propriétés décrites précédemment.

Notre framework a été implémenté et testé avec différentes configurations possibles. Nous avons fait varier les cinq paramètres initiaux (n , d , R , S , T) des graphes afin d'obtenir toutes les configurations possibles. Les résultats obtenus ont montré que seuls les trois paramètres (R , S et T) ont une influence sur les structures de communautés émergées et ce, indépendamment de la taille du graphe et de sa densité.

Cette classification structurelle tributaire des paramètres (R , S et T) ne se veut pas une classification pure et simple des communautés mais plutôt un pré-requis et une base de travail pour l'étude de la robustesse des réseaux d'interaction. C'est ainsi que toute la deuxième partie de notre travail est consacrée à l'étude de la robustesse des structures de communautés émergées.

Chapitre 4

Robustesse des communautés dans les réseaux de communication

Sommaire

4.1	Introduction	62
4.2	Objectifs de notre étude	62
4.3	Perturbations et Robustesse	63
4.3.1	Robustesse d'un système complexe	64
4.3.2	Mesures de la robustesse dans les réseaux complexes	65
4.3.3	Robustesse et communautés	65
4.3.4	Métrique pour la robustesse des structures des communautés	66
4.3.5	Positionnement et motivation de notre étude	66
4.3.6	Impact des paramètres R, S, T sur les structures de graphes	67
4.4	Etude de la robustesse des graphes émergés	67
4.4.1	Processus de perturbation du réseau	67
4.4.2	Évaluation de la robustesse	68
4.5	Analyse de la robustesse	71
4.5.1	Analyse de l'aspect global	72
4.5.2	Aspect local	73
4.6	Conclusions	80

4.1 Introduction

Les réseaux sociaux et l'étude des communautés d'intérêt qui émergent dans les réseaux d'interaction sont devenus un véritable challenge dans différents domaines de recherche. En effet, plusieurs recherches se sont intéressées à l'étude des types de relations entre les entités et leur impact sur le réseau considéré et sur la formation de ces communautés.

La détection de communautés dans un réseau est un moyen couramment utilisé pour simplifier les réseaux étant donné que le nombre de clusters est en général bien plus petit que le nombre de nœuds dans le réseau. Cependant, plusieurs interrogations concernant notamment les structures de ces communautés sont posées. Parmi celles-ci, nous pouvons citer celles afférentes à leur robustesse face à des perturbations ou bruits introduits sur ces structures (par exemple : disparition d'un lien entre deux nœuds ou crash d'un nœud important). Dans ce chapitre, nous nous intéressons à l'étude de la robustesse des structures de communautés ayant émergées des communications dans les réseaux d'interactions. Pour ce faire, nous avons défini un modèle ainsi qu'une plateforme d'évaluation de cette robustesse.

4.2 Objectifs de notre étude

La position de notre travail dans la problématique générale de l'étude de la robustesse des systèmes se positionne selon trois principaux objectifs :

- **Comprendre certains phénomènes**, notamment ceux relatifs à l'émergence de structures de communautés dans les réseaux d'interaction.
- **Prédire le comportement** des réseaux d'interactions. Quand on sait prédire un système, il est alors possible d'influer sur ce système. On peut prédire l'influence de chaque modification apportée et par conséquent choisir celle qui donne le comportement le plus intéressant. En d'autres termes, il nous sera plus facile de contrôler ce système.
- **Concevoir d'autres systèmes** à partir de notre compréhension, d'évolution souhaitée du système, et de la facilité de le contrôler.

Notre thèse s'inscrit dans ce contexte, c'est à dire dans la compréhension des réseaux sociaux d'interactions et notamment comprendre leur robustesse. Nous utilisons des modèles génériques afin d'identifier des concepts généraux potentiellement utiles dans de nombreux systèmes d'interactions.

4.3 Perturbations et Robustesse

Dans notre étude, nous nous intéressons plus précisément à la robustesse des communautés issues des réseaux d'interactions face à des perturbations.

Etudier la robustesse d'un système revient à étudier sa réaction face à des perturbations. En d'autres termes, s'il ne va pas changer de comportement ou sera profondément modifié. Nous essayons à travers notre modélisation de déterminer les configurations pour lesquelles ces systèmes sont sensibles aux perturbations, de définir parmi elles celles qui sont robustes et ayant le même comportement quelque soit la perturbation appliquée et celles qui sont sensibles à la moindre perturbation. Il y a lieu de préciser que dans notre étude, toutes les perturbations qu'on applique agissent sur la topologie du réseau d'interactions, c'est à dire nous nous intéressons aux paramètres qui font émerger ces topologies et non pas aux structures physiques du réseau considéré. Nous détaillerons notre méthode plus loin dans la partie définition de notre modèle.

Beaucoup de chercheurs ont traité l'aspect robustesse des réseaux. Nous pouvons citer le travail de Karrer *et al.* [59] qui ont proposé une méthode pour perturber les réseaux et mesurer le changement résultant dans les communautés. Ces changements sont utilisés pour comprendre la signification des communautés dans une variété de réseaux (réels et générés). Dans ce travail, les auteurs ont montré que la signification de la structure de communauté peut-être quantifiée en mesurant sa robustesse à des petites perturbations dans la structure du réseau. Dans [17], les auteurs ont exploré la robustesse par la mesure de la centralité aux différentes perturbations. Un autre travail intéressant est détaillé dans [55] où les auteurs ont étudié la réaction des réseaux complexes sujets à des attaques sur des nœuds ou des arêtes. Ils ont défini quatre stratégies d'attaque pour étudier la corrélation entre la centralité d'intermédierité (*Betweenness Centrality*) et le degré sur six graphes différents. Un autre travail intéressant a été publié dans [29]. Dans ce travail, les auteurs ont étudié la robustesse de 33 réseaux de métro à travers le monde. Assimilé à un graphe dont les nœuds représentent les stations et les liens entre eux représentent les chemins entre les stations, le réseau de métro est étudié de la même manière qu'un graphe en appliquant les propriétés théoriques pour étudier la complexité des réseaux de métro et déterminer quel est le réseau de métro le plus robuste. Cette robustesse est mesurée en utilisant le nombre de cycles présents dans le système.

Tous les travaux existants dans la littérature s'intéressent à l'étude la robustesse des graphes ou des réseaux du point de vue structurel, c'est à dire s'intéressent exclusivement à l'aspect structurel du graphe (nœuds et liens). Nous nous sommes intéressés ici plus

particulièrement aux structures sociales (communautés) qui émergent des interactions (échanges entre entités) dans les réseaux. Au chapitre précédent, nous avons présenté en détail comment émergent ces structures sociales à travers un modèle que nous avons défini. Nous avons également, dans ce même chapitre, décrit quelques propriétés intéressantes de ces structures de communautés telle que : la distribution des degrés des nœuds et leur interconnection entre eux [76]. Cette importante propriété nous renseigne sur la structure du réseau considéré. Par exemple, les réseaux avec des nœuds de degrés élevés qui tendent à se connecter avec d'autres nœuds à forts degrés sont plus exposés aux épidémies. En effet, un nœud infecté, avec un degré élevé, va rapidement infecter plusieurs autres nœuds à forts degrés également. L'inverse est vrai : un nœud à fort degré qui est connecté à des nœuds à faibles degrés ne va pas propager l'épidémie plus loin. Parallèlement, nous avons présenté dans le Chapitre 3 d'autres caractéristiques intéressantes (APL, coefficient de clustering, *etc*). Dans notre thèse, nous utilisons ces propriétés pour étudier et analyser la robustesse des structures sociales de communautés qui émergent des interactions dans les réseaux de communication.

4.3.1 Robustesse d'un système complexe

L'étude des propriétés des réseaux complexes est devenue très populaire. Par exemple, le réseau Internet est un cas d'étude intéressant. En effet, les applications du réseau Internet sont ubiquitaires et pervasives et sa structure est de type "Scale-free".

La robustesse d'un système est définie comme sa capacité à assurer, en cas d'une attaque ou panne, un fonctionnement minimum de ses services. Dans la théorie des graphes, un graphe est dit robuste si vouloir déconnecter une composante s'avère difficile. La robustesse d'un réseau dépend du type des attaques considérées.

Le terme "*Robustesse*" dans le contexte de réseaux sociaux a deux significations de connotations différentes. La robustesse d'un réseau concerne sa fiabilité [63] et son fonctionnement continu après une intervention. La robustesse d'un réseau est particulièrement pertinente dans les réseaux de communication à flux orienté. Une autre connotation du terme robustesse, que nous considérons dans notre étude, est la robustesse des mesures du réseau. Quand on associe le terme *robustesse* au terme *mesure*, le mot a une connotation beaucoup plus statistique. L'étude de la robustesse d'une mesure d'un réseau peut aussi faire référence à la "*sensibilité*" de cette même mesure.

Les réseaux sociaux présentent un niveau élevé de tolérance aux erreurs. Leur stabilité est attribuée à la façon dont les composants de ce réseau sont interconnectés ainsi qu'à la manière avec laquelle il fonctionne, c'est à dire la stabilité d'un système dépend de sa

nature (type). En effet, il est démontré que la vulnérabilité d'un système dépend de la classe de graphe dont il fait partie et, aussi, à la nature de l'attaque. Par exemple, les réseaux dits "scale-free" sont robustes aux attaques aléatoires, par contre, ils sont très vulnérables aux attaques ciblées (la sélection et la suppression d'un ensemble de nœuds jouent un rôle vital pour le maintien du réseau en mode connecté) [4,118]. La robustesse d'un réseau de type "scale-free" est enracinée dans la distribution non homogène de la connectivité, parce que la distribution en loi de puissance implique que la majorité des nœuds ne comptent que quelques liens, d'où, lors d'une attaque sur ce type de réseau, la probabilité qu'un nœud de faible degré soit la cible de l'attaque est élevée.

4.3.2 Mesures de la robustesse dans les réseaux complexes

Pour évaluer la résistance des réseaux contre les défaillances et les attaques, les mesures simples et précises doivent être sélectionnés. Toutefois, identifier l'ensemble des paramètres qui représente complètement la connectivité du réseau est une tâche complexe.

Dans [107], plusieurs métriques pour évaluer la robustesse des systèmes complexes ont été définies. Par exemple, la "*connectivité*" d'un graphe est une mesure importante de la robustesse du réseau [44]. La connectivité des nœuds (liens) détermine la robustesse du graphe face à des suppressions de nœuds (ou liens). Cependant, la connectivité d'un nœud ou d'un lien ne reflète que partiellement la capacité du graphe à conserver un certain degré de connexité après la suppression. D'autres mesures ont été introduites et étudiées incluant : "*Super connectivity*" [25], "*conditional connectivity*" [58], "*restricted connectivity*" [37], "*fault diameter*" [64], "*toughness*" [22], "*scattering number*" [57], "*tenacity*" [27], "*expansion parameter*" [6] and "*isoperimetric number*" [83].

4.3.3 Robustesse et communautés

Tel qu'il a été mentionné dans le chapitre précédent, la détection de structure de communautés a fait l'objet de plusieurs travaux dans la littérature. Ainsi, plusieurs méthodes sont définies [43], mais peu de travaux sont proposés pour l'étude de la robustesse de ces communautés. Dans ce qui suit, nous citons quelques travaux s'y rapportant.

Brian Karrer *et al.* [59] avaient proposé une méthode adaptée pour perturber les réseaux et quantifier les changements intervenus dans les structures de ces communautés. Dans ce travail, les auteurs étudient les communautés découvertes dans un graphe avant et après perturbations.

Dans [17], Borgatti *et al.* ont exploré la robustesse de la mesure de la centralité face à des erreurs. Les types d'erreurs sont : suppression de liens, suppression de nœuds, ajout

de liens et ajout de nœuds.

Dans [77], Clair P. Massen *et al.* ont étudié la thermodynamique des structures de communautés en introduisant des ensembles de partitions équilibrés. Les auteurs ont introduit une approche de partitionnement de réseaux en communautés qui ne détermine pas seulement la meilleure structure de la communauté, mais fournit également une gamme de techniques de caractérisation afin d'évaluer la façon dont cette structure est importante.

4.3.4 Métrique pour la robustesse des structures des communautés

Afin d'évaluer et quantifier les changements opérés dans les structures de communautés après une perturbation, plusieurs mesures ont été définies (cf. l'état de l'art de [114]). Dans ce papier, les auteurs ont présenté une étude organisée de 16 mesures externes de validation de partitionnement de graphes. Ils ont montré l'importance des mesures normalisées dans l'évaluation de la performance du partitionnement sur les distributions des classes de données déséquilibrées et résumé les principales propriétés de ces mesures externes.

Dans notre étude, nous considérons deux métriques : "Adjusted Rand Index (ARI)" [56] et "Normalized Mutual Information (NMI)" [24].

4.3.5 Positionnement et motivation de notre étude

Comme mentionné plus haut, la robustesse des réseaux est largement étudiée dans la littérature. Cependant, la plupart des méthodes proposées font référence à la perturbation structurelle de ces réseaux c'est à dire ajout ou suppression de nœuds ou de liens. Dans notre travail, nous nous intéressons à l'étude de la robustesse des communautés qui émergent des interactions entre entités dans les réseaux de communication, en ne prenant pas en considération leur aspect structurel, mais plutôt les paramètres ayant permis leur émergence. Cette façon d'aborder la robustesse des structures des graphes constitue une originalité du fait qu'aucune étude similaire n'a été réalisée jusqu'à ce jour. Par ailleurs, le modèle que nous avons défini n'est pas uniquement typique pour le cas des réseaux de communications qui partagent des ressources, mais peut-être généralisé pour d'autres types de réseaux tels que : les réseaux biologiques, les réseaux de transport, les réseaux économique ou autres (Cf. 4.1). Cette généralisation réside principalement dans l'identification des paramètres pertinents qui influent l'émergence de chaque type de réseaux sociaux.

Dans notre cas d'étude, qui concerne les réseaux d'interactions, nous considérons les paramètres R , S et T , définis dans le chapitre 3.

4.3.6 Impact des paramètres R , S , T sur les structures de graphes

Dans ce qui suit, nous allons nous intéresser à l'étude de l'influence du nombre de nœuds fournisseurs S , du nombre de ressources R et de la proportion du centre d'intérêt des nœuds T sur les structures de communautés émergées. Le tableau 3.4 (Cf. Chapitre 3) donne la correspondance entre les critères structurels.

De ce tableau, nous pouvons observer concernant la densité moyennes d_{comm} que le paramètre T a un impact principal. Plus T est restreint, plus les communautés sont denses et vice versa. Concernant le nombre de feuilles (nœuds de degré 1) nb_{deg1} , nous observons que pour des valeurs élevées (fortes) de S , le nombre de feuilles est également élevé ou moyen et pour des valeurs faibles ou moyennes de S , le nombre de feuilles est aussi faible ou moyen. Pour des valeurs moyennes de S , nous observons que c'est le paramètre R qui détermine l'influence. Plus il y a de ressources dans le réseau, plus le nombre de feuilles (nœuds de degré 1) est élevé. Pour le critère deg_{max} , nous pouvons observer qu'il prend des valeurs élevées quand le paramètre intérêt T est aussi élevé. De plus, dans le cas particulier où les trois paramètres S , R , et T ont des valeurs moyennes, nous obtenons des structures de communautés de type "fan". Ce type de graphe est composé de deux parties : la première partie est quasiment non connectée et la seconde est connectée mais avec une densité faible. Dans notre cas d'étude, la première partition est formée par les nœuds clients (qui ne sont pas des fournisseurs de ressources) et la seconde partie est, quant à elle, formée de nœuds fournisseurs. Un nœud client, dans ce cas, connecté à plus d'un fournisseur mais pas à d'autres clients, tandis qu'un nœud fournisseur peut être connecté à d'autres nœuds fournisseurs.

4.4 Etude de la robustesse des graphes émergés

4.4.1 Processus de perturbation du réseau

Tel que mentionné plus haut, notre étude de la robustesse se focalise sur les paramètres qui permettent l'émergence des structures de communautés dans les réseaux d'interaction. Pour ce faire, nous avons défini un modèle (cf. figure 4.1). Dans ce modèle, nous considérons deux configurations : *initiale* et *perturbée*. La différence entre ces deux configurations réside uniquement dans la distribution sémantique (des fournisseurs S , des ressources R et des centres d'intérêts des nœuds T (ressources désirées)). Pour chaque type de distribution, une stratégie de perturbation (ajout, suppression et déplacement) est définie.

L'intensité de ces perturbations varie entre 5%, 15% et 30%. Par exemple, la perturbation "suppression serveur" à 5% implique la suppression de 5% de la totalité des serveurs dans le réseau. Enfin, nous avons remarqué qu'une intensité de perturbation supérieure à 30% détruit complètement les structures de communautés, ce qui rend leur analyse difficile.

La première étape du processus "**génération des deux configurations**" (initiale et perturbée) implique que ces dernières possèdent le même nombre de nœuds et le même nombre de liens. Selon une combinaison des paramètres R , S et T des configurations initiales, nous procédons à des changements dans les valeurs initiales de ces paramètres pour obtenir des configurations perturbées.

La seconde phase de notre processus consiste à l'**étude de la robustesse des structures émergées** en comparant les structures émergées dans les configurations initiale et perturbée.

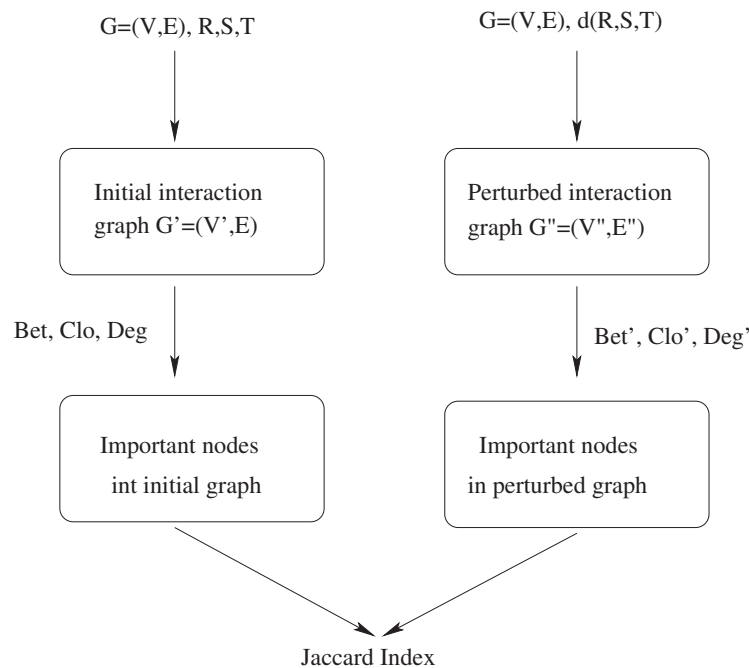


FIGURE 4.1 – Evaluation de la robustesse des communautés

4.4.2 Évaluation de la robustesse

Une mesure est robuste si un changement dans sa valeur d'entrée produit un léger changement dans sa valeur de sortie. Dans notre étude de la robustesse, nous considérons les mesures sur deux aspects différents relatives aux structures de communautés. Soient G'

et G'' deux graphes émergés. G' est le graphe initial et G'' est le graphe perturbé.

Le premier aspect est **Global** dans lequel on quantifie l'impact de l'introduction de la perturbation sur l'ensemble du graphe social émergé. Pour cet aspect, nous avons considéré trois métriques qui sont :

- ΔQ : La variation dans les valeurs de la modularité de G' et de G'' . La modularité que nous avons adoptée est celle proposée par Newman dans [90], qui détermine la qualité du partitionnement du graphe. Une bonne modularité indique un bon comportement des nœuds sur la formation des structures de communautés. Cette modularité est définie comme suit :

$$Q = \frac{1}{4m} \sum_{ij} \sum_r \left[A_{ij} \frac{k_i k_j}{2m} \right] S_{ir} S_{jr}$$

où, A est la matrice d'adjacence du graphe social, m le nombre de liens dans le graphe, k_i le degré du nœud i et S_{ir} une valeur binaire indiquant si i appartient ou non à la communauté r .

- $\Delta\gamma$: La variation de la valeur de γ entre G' et G'' . Cette mesure nous renseigne sur l'indice de scalabilité de la propriété scale-free du graphe. Avec cette valeur, on peut déterminer si la perturbation considérée a eu un impact sur la distribution de degrés du graphe.
- ΔCC : La variation de la valeur de la transitivité (Coefficient de clustering) CC entre G' et G'' . Cette mesure nous renseigne sur la connectivité des nœuds des communautés après une perturbation.

Toutes ces métriques sont calculées dans un sens relatif (la fraction de la différence entre les valeurs obtenues avant et après la perturbation sur la valeur obtenue avant la perturbation).

Le second aspect de notre étude est plutôt **Local** aux communautés. Il nous permet d'évaluer les changements opérés dans la façon dont les nœuds se sont regroupés dans ces communautés c'est à dire de déterminer la probabilité avec laquelle deux nœuds, qui étaient ensemble dans une même communauté, se trouvent aussi appartenir à une même communauté après perturbation. Pour pouvoir capturer cette information, nous utilisons les deux métriques *Adjuste Rand Index* [56] et *Normalized Mutual Information* [24] que nous allons détailler dans cette section.

4.4.2.1 Asect global :

Pour évaluer les changements intervenus sur les structures de communautés après chaque perturbation, nous avons utilisé trois mesures.

Variation dans la qualité du partitionnement : La fonction de qualité utilisée dans [90] mesure la qualité d'un bon partitionnement. Nous calculons cette fonction pour le même graphe avant et après perturbation :

$$\Delta Q = \frac{Q_2 - Q_1}{Q_1}$$

où, Q_1 et Q_2 représentent, respectivement, les valeurs de la qualité de partitionnement avant et après la perturbation.

Variation dans la valeur de γ : Tel que mentionné dans le chapitre précédent, les graphes obtenus à l'aide de notre modélisation, sont de type scale-free. Leur distribution de degrés suit une loi de puissance de type $p(k) \sim k^{-\gamma}$, où γ représente la pente de la courbe. Le calcul de la différence dans la valeur de γ , avant et après perturbation, nous renseigne sur la capacité pour un graphe à maintenir une apparence scale-free. Cette différence est calculée comme suit :

$$\Delta\gamma = \frac{\gamma_2 - \gamma_1}{\gamma_1}$$

où, γ_1 et γ_2 représentent, respectivement, la valeur de pente de la courbe avant et après perturbation.

Variation dans la valeur du coefficient de clustering : Comme détaillé dans le chapitre précédent, le coefficient de clustering est une mesure du degré auquel les nœuds d'un graphe ont tendance à se regrouper. La différence dans la valeur de cette mesure entre le graphe initial et le graphe perturbé nous renseigne sur les changements intervenus sur la structure du graphe considéré. Cette différence est calculée comme suit :

$$\Delta CC = \frac{CC_2 - CC_1}{CC_1}$$

où : CC_1 et CC_2 représentent, respectivement, les valeurs du coefficient de clustering avant et après perturbation.

4.4.2.2 Aspect local

Afin d'évaluer la qualité du partitionnement d'un graphe en communautés, avant et après introduction d'une perturbation, le *Adjust Rand Index* [56] est utilisé. Cette me-

sure est utilisée dès lors qu'on suppose que chaque individu (nœud) du réseau est affecté à une seule communauté. Soit $S = s_1, \dots, s_n$ un ensemble de n objets. On notera $U = u_1, \dots, u_R$ et $V = v_1, \dots, v_C$ deux classes de partitions différentes d'objets dans S tel que : $U_{i=1}^R u_i = S = U_{j=1}^C v_j$ et $u_i \cap u_{i'} = \emptyset = v_j \cap v_{j'}$ pour $1 \leq i \neq i' \leq R$ et $1 \leq j \neq j' \leq C$. Soit a le nombre de même paires d'objets placés dans les deux classes U et V , b le nombre de paires d'objets qui sont dans la classe U et n'appartiennent pas à la classe V , c le nombre de paires d'objets placés dans V et qui ne sont pas dans U et enfin, d le nombre de paires d'objets dans les différents classes dans les deux partitionnements. Les quantités a et b peuvent être interprétées comme étant les objets communs à U et V et les quantités c et d comme les objets de désagrément entre ces deux classes U et V .

On assume que U and V sont deux partitions choisies aléatoirement et que le nombre d'objets dans les deux classes est fixé. Alors, ARI est donné par :

$$ARI(U, V) = \frac{(a + b + c + d)(a + d) - ((a + b)(a + c) + (c + d)(b + d))}{(a + b + c + d)^2 - ((a + b)(a + c) + (c + d)(b + d))}$$

4.5 Analyse de la robustesse

Dans cette section, nous présentons une étude sur l'évolution des différentes caractéristiques du réseau selon les différentes perturbations. Nous considérons les perturbations relatives à deux paramètres : *Fournisseurs* et *Ressources*. Nous ne prenons pas en considération les perturbations relatives au centre d'intérêt des nœuds pour la simple raison que modifier l'ensemble de ressources valables sur le réseau ou modifier l'ensemble des requêtes (ressources désirées) donnent un comportement similaire. Nous considérons trois niveaux de perturbation : 5%, 15% et 30% pour chaque type de perturbation : *Ajout*, *Suppression* et *Déplacement* pour les deux paramètres *Fournisseurs* et *Ressources*.

4.5.1 Analyse de l'aspect global

Paramètre	Perturbation	5%	15%	30%
ΔCC	Ajout	3%	7%	11%
	Suppression	-4%	-10%	-22%
	Déplacement	1%	1%	2%
ΔAPL	Ajout	4%	16%	33%
	Suppression	-4%	-10%	-25%
	Déplacement	0%	-1%	2%
ΔQ	Ajout	36%	53%	91%
	Suppression	-4%	-16%	-80%
	Déplacement	0%	8%	5%
$\Delta \gamma$	Ajout	0%	4%	6%
	Suppression	-6%	-7%	-7%
	Déplacement	0%	-1%	2%

TABLE 4.1 – **Perturbations appliquées aux serveurs** : variations du *Coefficient de clustering* , *APL*, *Modularité*, et *Distribution de degrés*

Le tableau 4.1 montre l'évolution des quatre métriques selon les perturbations appliquées aux nœuds fournisseurs. Le coefficient de clustering est affecté par la suppression et l'ajout de serveurs. Ceci est dû à l'effet que chaque lien d'interaction a une extrémité connectée à un serveur. Par conséquent, modifier le nombre de serveurs, c'est modifier aussi le nombre de liens et le nombre de triangles dans le graphe. L'APL du graphe est également affecté par le nombre de serveurs. L'APL grandit proportionnellement avec le nombre de serveurs et vice versa. Du moment que les serveurs représentent des hubs dans le graphe social et si leur nombre est faible, alors l'APL du graphe est réduit. Nous avons observé également que la modularité (qualité du partitionnement du graphe en communautés) est améliorée quand l'APL est plus important. En effet, un graphe social ayant une forte concentration est facilement décomposable en communautés et le partitionnement obtenu est de bonne qualité. De plus, le paramètre γ est légèrement affecté par les perturbations sur les serveurs. En d'autres termes, les structures sociales émergées gardent leur aspect de la distribution de degrés de type puissance. D'une manière générale, les perturbations appliquées sur les serveurs n'affectent pas d'une manière significative les quatre paramètres discutés.

Paramètre	Perturbation	5%	15%	30%
ΔCC	Ajout	-5%	-2.11%	-15.47%
	Suppression	-6%	-9%	-59%
	Déplacement	0.00%	-3%	-3%
ΔAPL	Ajout	-1,36%	-2.09%	-8.26%
	Suppression	2.27%	2.81%	8.40%
	Déplacement	1.27%	3.31%	3.39%
ΔQ	Ajout	3.09%	9.77%	28.50%
	Suppression	-0.81%	-6.47%	-9.77%
	Déplacement	-1.08%	-2.65%	-9.60%
$\Delta \gamma$	Ajout	0.96%	2.25%	1.29%
	Suppression	2.05%	2.38%	3.36%
	Déplacement	0.28%	0.61%	2.83%

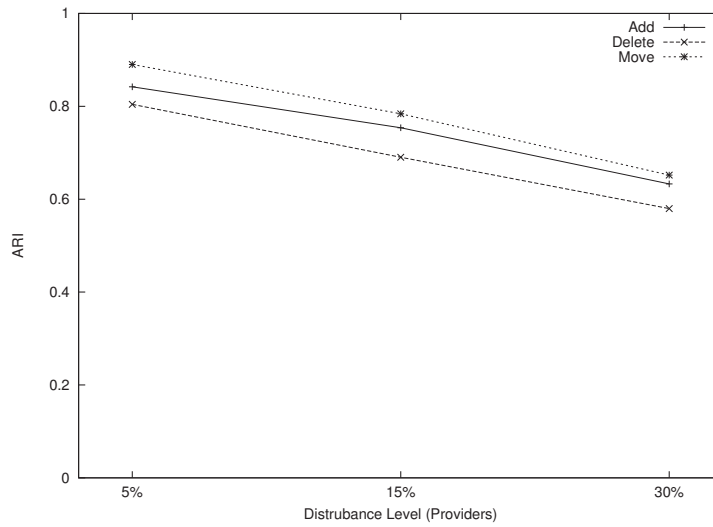
TABLE 4.2 – **Perturbations appliquées aux ressources** : variations du *Coefficient de clustering*, *APL*, *Modularité*, et *Distribution de degrés*

D'un autre côté, le Tableau 4.2 montre l'évolution des mesures globales selon les perturbations appliquées aux ressources. Le comportement est similaire aux perturbations appliquées aux serveurs avec une petite échelle d'impact. Cependant, nous pouvons voir dans la suite que les perturbations appliquées aux ressources ont un impact plus important sur l'aspect local des communautés.

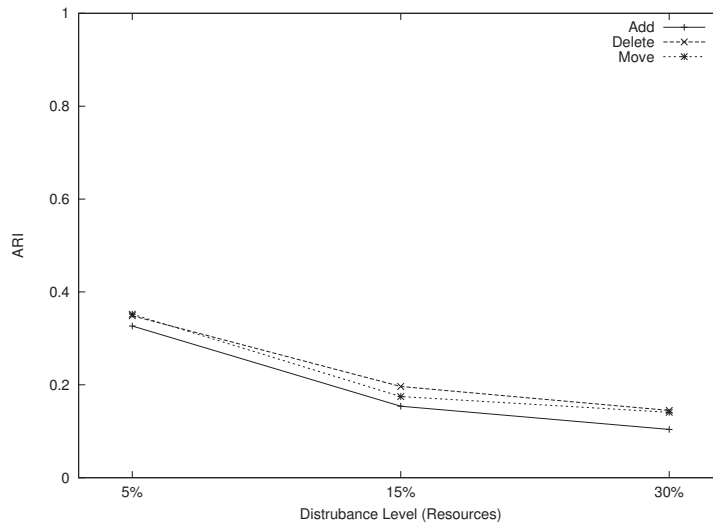
4.5.2 Aspect local

Dans cette section, nous nous focalisons sur l'étude de l'impact des différentes perturbations sur les structures de communautés émergées, de voir comment les perturbations peuvent affecter la façon avec laquelle les nœuds se regroupent et forment des communautés. Tel que mentionné précédemment, nous adoptons la mesure de l'index Rand ajusté (Adjust Rand Index) [56] pour quantifier cet impact. Plus cette mesure est forte, moins l'impact est important.

Nous considérons en premier les valeurs moyennes de "ARI" pour chaque niveau de perturbation (5%, 15% et 30%) pour les trois perturbations (ajout, suppression et déplacement) appliquées et sur les serveurs et sur les ressources. La figure 4.2 montre que l'impact des perturbations sur les serveurs évolue d'une manière linéaire tandis que l'impact des perturbations appliquées sur les ressources évoluent de façon plus rapide. Ceci illustre le fait que les interactions entre entités dans le réseau sont plus sensibles à la distribution de ressources qu'à la distribution des serveurs. Dès lors que les requêtes des nœuds (recherche



(a)

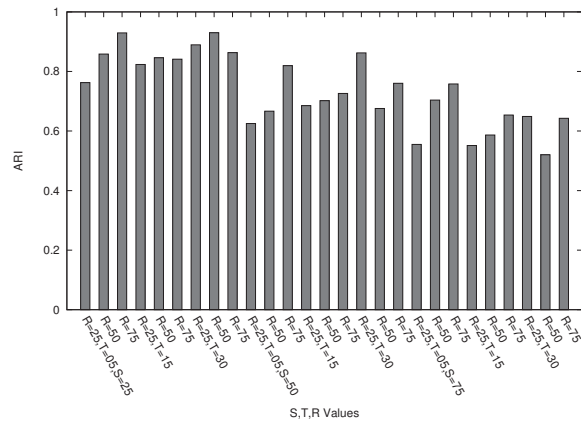


(b)

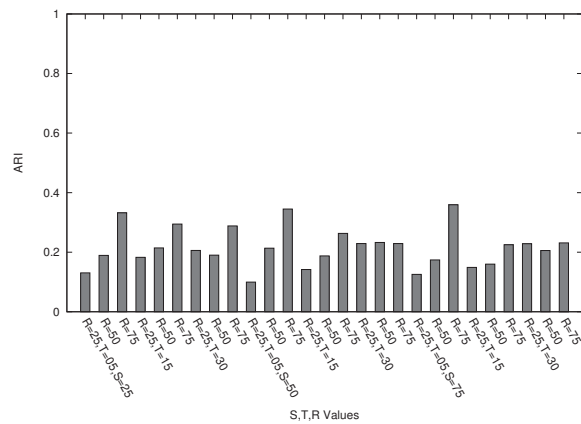
FIGURE 4.2 – Impacts moyens sur les serveurs et les ressources : selon l'intensité de la perturbation

des ressources désirées) sont générées à partir des centres d'intérêts des nœuds (ressources désirées) et si un serveur est déplacé ou si de nouveaux serveurs sont rajoutés au réseau, alors la probabilité de préserver les mêmes liens d'interaction est plus importante que si certains serveurs sont supprimés du réseau.

Par ailleurs, nous pouvons observer que l'ajout de nouvelles ressources dans le réseau est la perturbation qui produit le plus d'impact. En effet, ajouter des ressources rend ces dernières plus disponibles pour les nœuds et augmente la probabilité pour qu'un nœud soit



(a)



(b)

FIGURE 4.3 – Impacts moyens sur les serveurs (a) et sur les ressources (b) : selon les configurations

satisfait par un même serveur, ce qui change considérablement la façon dont les nœuds sont interconnectés dans le réseau. En outre, supprimer des ressources augmente la probabilité d'ajouter de nouveaux liens pour un nœud (un client peut-être connecté avec plusieurs serveurs). Ainsi, l'impact de la perturbation sera moins important que l'ajout de ressources. Déplacer des ressources a le même impact que la suppression de ressources sauf que les requêtes satisfaites (ressources satisfaites par des serveurs) ne vont pas changer dans le réseau perturbé.

La figure 4.3 montre la valeur moyenne des perturbations regroupées par type : perturbations sur les serveurs contre perturbations sur les ressources. Comme attendu, perturber la distribution des ressources a globalement plus d'impact que de perturber la distribution des serveurs. Dans la figure 4.4, nous donnons un classement des différentes perturbations.

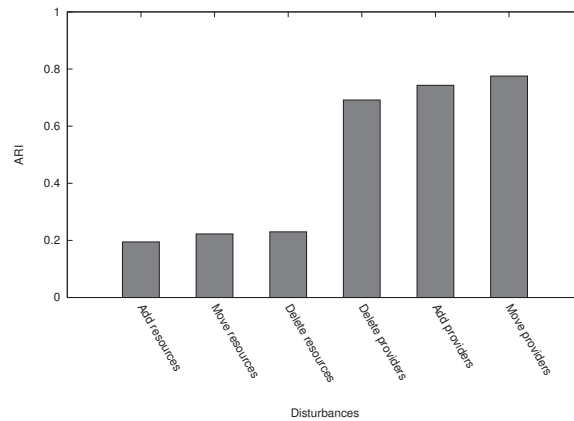


FIGURE 4.4 – Classement des impacts moyens pour toutes les perturbations

La perturbation ayant le plus d'impact sur les réseaux est "*Ajout de Ressources*", suivie de *suppression de ressources* et *déplacement de ressources*. Ajouter des ressources à des serveurs peut causer la perte de beaucoup de liens d'interaction qui étaient présents dans le graphe initial et ce, à cause du mécanisme de l'attachement préférentiel, tandis que la suppression ou le déplacement de ressources peut induire la création de nouveaux liens pour des ressources manquantes. En d'autres termes, pour les perturbations appliquées aux serveurs, force est de constater que la suppression de serveurs provoque beaucoup de changement dans les structures de communautés. En effet, supprimer des serveurs du réseaux incitent les nœuds clients à se regrouper autour d'un nombre restreint de serveurs, ce qui augmente la concentration dans les communautés, tandis que de nouveaux serveurs peut offrir la possibilité de réduire cette concentration mais tout en gardant les liens initiaux. Par conséquent, les changements doivent être moins important dans le second cas. De plus, déplacer des serveurs est la moins impactante des perturbations. Dès lors que la topologie sous-jacente peut influencer uniquement la performance de la recherche des ressources, les serveurs ayant des ressources exclusives vont maintenir leurs liens d'interactions même si leur emplacement dans le réseaux est modifié.

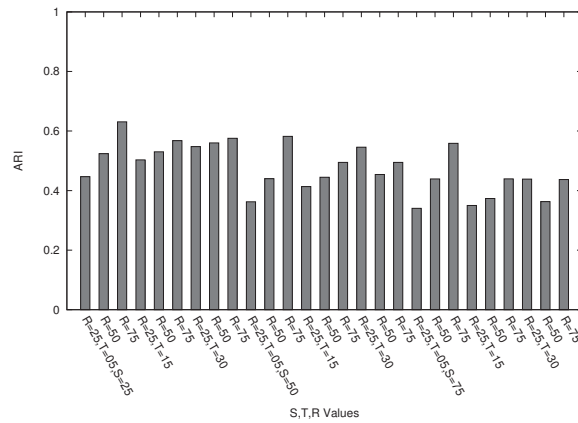
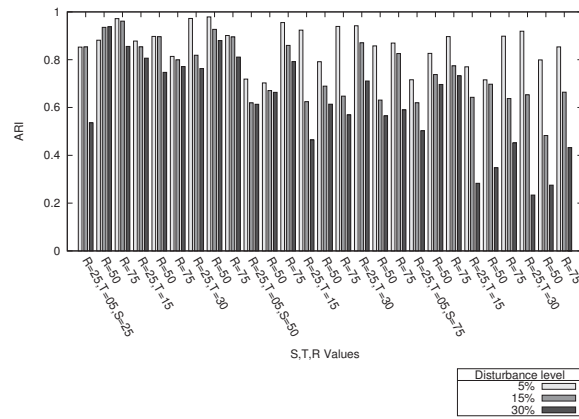
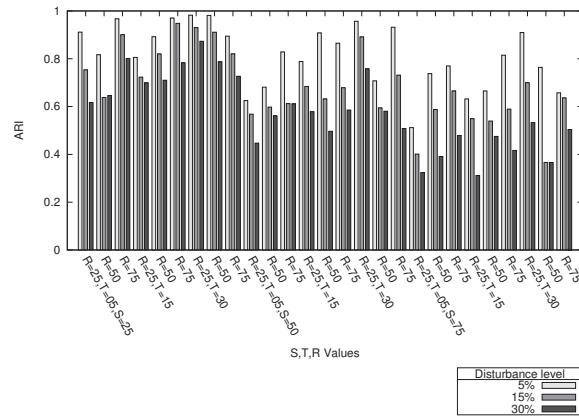


FIGURE 4.5 – Moyenne des impacts pour toutes les perturbations

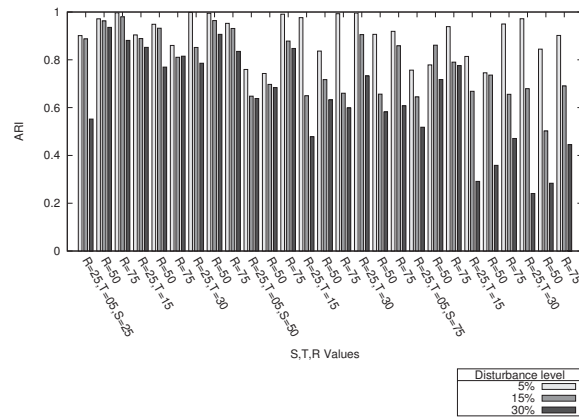
La figure 4.5 donne les moyennes pour toutes les perturbations (ajout, suppression et déplacement), avec toutes les intensités (5%, 15% et 30%) et ce, pour chaque configuration. Nous pouvons observer que les configurations avec peu de serveurs sont, globalement, plus robustes que les configurations avec beaucoup de serveurs. En outre, les configurations avec beaucoup de ressources sont plus robustes que celles possédant moins de ressources. Nous pouvons donc conclure que les configurations des réseaux les plus robustes sont celles ayant des faibles proportions de serveurs S combinées avec des proportions élevées de ressources R tel que le mettent en évidence les histogrammes 4.6 et 4.7



(a)

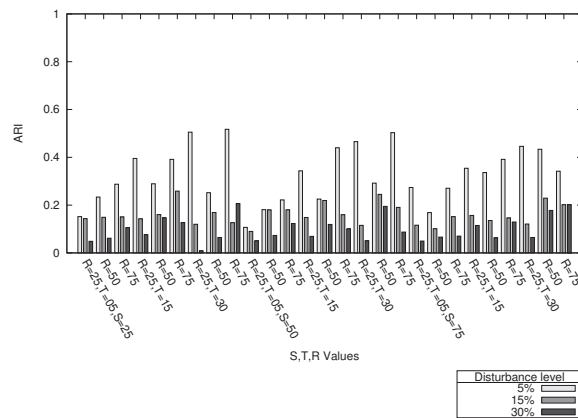


(b)

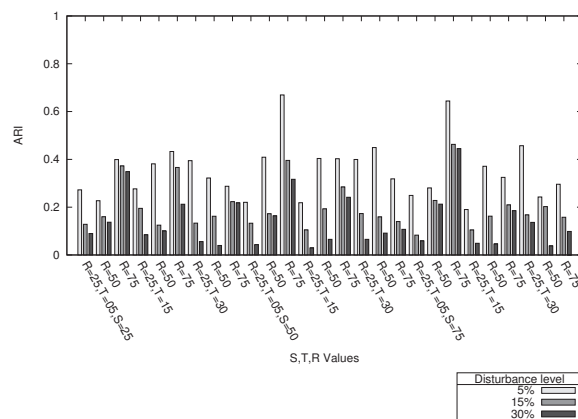


(c)

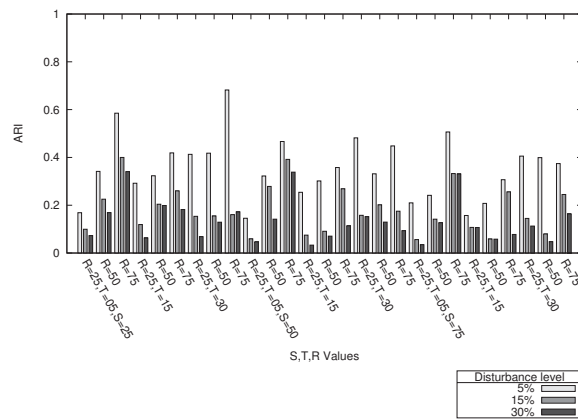
FIGURE 4.6 – Perturbations appliquées sur les serveurs



(a)



(b)



(c)

FIGURE 4.7 – Perturbations appliquées sur les ressources

4.6 Conclusions

A travers l'étude présentée dans ce chapitre, nous avons examiné l'effet de certaines perturbations sur les graphes sociaux émergés des communication dans les réseaux d'interaction. Nous avons défini un modèle qui nous a permis d'étudier de la robustesse de ces graphes sociaux émergés. Contrairement aux études classiques de la robustesse qui s'intéressent principalement aux aspects structurels (nœuds, liens), nous avons proposé une étude de la robustesse à travers des paramètres qui permettent l'émergence de ces structures sociales. A cet effet, nous avons étudié l'impact des perturbations appliquées sur les serveurs et les ressources dans un réseau d'interaction et les résultats que nous avons obtenus nous ont permis de mener une discussion sur l'importance de chaque configuration du réseau selon plusieurs métriques.

Notre modèle de l'étude de la robustesse peut-être généralisé à d'autres types de réseaux sociaux. La seule contrainte se présente dans l'identification des paramètres qui permettent l'émergence de ces réseaux sociaux. Par exemple, dans les réseaux de gènes, un des paramètres permettant cette émergence réside dans l'affinité entre ces gènes pour former une séquence. Par ailleurs, le cas d'étude que nous avons considéré peut aider les administrateurs des réseaux à mieux comprendre le comportement des réseaux qu'ils gèrent et par conséquent, leur permettre de mieux les administrer et prévenir les configurations spécifiques pour lesquelles le réseau est le moins robuste.

Chapitre 5

Analyse des positions des nœuds importants

Sommaire

5.1	Introduction	82
5.2	Détermination des nœuds importants	82
5.2.1	Centralité de degré (<i>Degree Centrality</i>)	83
5.2.2	Centralité de proximité (<i>Closeness Centrality</i>)	84
5.2.3	Centralité d'intermédiarité (<i>Betweenness Centrality</i>)	84
5.3	Evaluation de la robustesse des nœuds importants	85
5.4	Analyse de la robustesse	86
5.4.1	Effets de la perturbation sur la population des nœuds importants	86
5.4.2	Robustesse des différentes configurations (<i>selon les proportions</i> <i>R, S, T</i>)	90
5.4.3	Impact moyen sur les différentes perturbations	91
5.5	Conclusion	92

5.1 Introduction

L'analyse des réseaux sociaux découle de l'idée de l'existence d'une structure déterminée qui explique comment un individu connaît un autre individu directement ou indirectement. L'exemple parfait qui illustre cette idée est l'expérience de Milgram des *six degrés de séparation* [81]. Dans cette expérience, Milgram a montré que n'importe quelle paire d'individus sur terre sont en moyenne séparés par six intermédiaires.

L'analyse des réseaux sociaux examine la structure sociale des relations dans un groupe afin de découvrir les connections informelles entre les individus. Ces relations peuvent être de nature connaissance, de communication ou de prise de décision, *etc.* Présentement, nous retrouvons l'analyse des réseaux sociaux dans différents domaines. Celle-ci est devenue un outil important pour comprendre l'organisation qui relie les individus dans un réseau (par exemple déterminer l'individu le plus important dans un réseau).

Les réseaux sociaux sont des structures complexes qui capturent différentes relations existantes entre les individus formant ces réseaux. Ces relations peuvent être spécifiques selon la nature du réseau telles que : les relations d'amitié, l'intérêt commun, les échanges entre les individus textitetc.

Dans notre étude, nous nous focalisons sur les structures sociales émergées des communications dans les réseaux d'ordinateurs à partir du modèle proposé dans le chapitre 3.

Dans ce chapitre, nous étudions la position des nœuds importants dans les graphes sociaux émergés. A travers l'étude de ces positions, nous essayons d'évaluer la robustesse de ces structures. Cette façon de voir ne se veut pas une étude palliative à l'étude réalisée dans le chapitre précédent, mais plutôt une complémentarité à celle-ci.

5.2 Détermination des nœuds importants

Le principal concept des réseaux sociaux peut-être décrit comme un ensemble fini de nœuds connectés par des liens [46] [54] [111]. Usuellement, un nœud d'un réseau est dit "acteur", "individu", "entité" ou "entité sociale collective" [111], ou "consommateur" [115]. Dans le même sens, les arêtes entre nœuds sont aussi appelées "liens" ou "relations" [111]. Le nombre et le type des liens peuvent différer selon le type et la caractéristique de l'analyse des acteurs [54] [111].

L'analyse des réseaux sociaux [111] est un corpus de techniques mathématiques utilisées par des sociologues afin d'analyser des interactions sociales (analyse des relations sociales

entre entités du réseau). L'analyse des réseaux sociaux [111] offre certaines mesures pour déterminer l'importance des positions des nœuds dans le réseau social. Les plus utilisées parmi elles sont : *Centralité*, *Prestige*, *Accessibilité* et *Connectivité* [54,111]. Il existe plusieurs approches pour évaluer la centralité des individus [45] : la centralité de degré (*degree centrality*), la centralité de proximité (*closeness centrality*) et la centralité d'intermédiarité (*betweenness centrality*). La "centralité de degré" d'un acteur i prend en considération ses voisins directs [54]. La "centralité de proximité" montre combien i est proche d'autres acteurs dans le réseau [11]. Enfin, la "centralité d'intermédiarité" d'un acteur i spécifie sur combien de chemins différents celui-ci se retrouve dans le réseau social [45]. Un acteur i est plus important si plusieurs acteurs dans le réseau sont joints entre eux par l'intermédiaire de i [54]. Nous verrons plus loin les définitions détaillées de ces métriques.

Une autre propriété importante qui caractérise l'individu dans un réseau social est son "Prestige" [111] tel que le "degré prestige" et le "Rang prestige". Le "degré prestige" prend en compte le nombre d'utilisateurs v_i qui sont adjacents à i [111]. Le "rang prestige" est déterminé sur la base de l'état des utilisateurs dans le réseau et dépend, non seulement de la distance géodésique et du nombre de relations, mais aussi du statut des utilisateurs connectés à l'utilisateur i [60].

Tel que mentionné dans les chapitres précédents, nous considérons particulièrement les graphes sociaux issus des échanges entre entités dans les réseaux d'interactions. L'importance d'étudier ces types de réseaux sociaux est fondée sur l'importance théorique de ses individus ou ses membres en collectivité. En effet, certains nœuds sont plus importants que d'autres dans la définition de ce qui est appelé "identité sociale du réseau". Différentes mesures sont définies dans la littérature pour mettre en valeur les différences entre les nœuds importants et les nœuds non importants dans un réseau. Pour trouver ces nœuds importants dans le réseau social considéré, nous avons adopté trois métriques incluant la *centralité de degré*, la *centralité de proximité* et la *centralité d'intermédiarité*. Ces métriques sont les plus utilisées.

5.2.1 Centralité de degré (*Degree Centrality*)

Le degré d'un acteur dans un graphe est important. Il peut répondre à la question : quel est l'acteur le plus actif dans le réseau ?

Si $d(i)$ est le degré du nœud i , alors sa centralité de degré $C_D(i)$ n'est autre que son degré $d(i)$: $C_D(i) = d(i)$.

Cette mesure représente le nombre de liens qui sont incidents au nœud i . Nous notons que cette valeur dépend bien évidemment de la taille du réseau n . En effet, $n - 1$ est sa

valeur maximale. Consécutivement, une normalisation de cette valeur, $C_D(i) = \frac{d(i)}{n-1}$ est la proportion de nœuds qui sont adjacent à i .

Un nœud avec une centralité de degré élevée signifie que celui-ci est en contact avec beaucoup de nœuds dans le réseau. Ce nœud est connu par les autres nœuds comme étant un acteur important pour, par exemple, relayer l'information. Un acteur avec un faible degré est un acteur périphérique dans le réseau. Ce type de nœuds n'est pas très influent dans le réseau et sa disparition, par exemple, n'affecte pas vraiment le réseau. Un nœud isolé a une centralité de degré nulle.

5.2.2 Centralité de proximité (*Closeness Centrality*)

La deuxième mesure de la centralité d'un acteur est basée sur sa proximité ou sa distance par rapport aux autres nœuds. La centralité de proximité, pour un sommet, est l'inverse de la somme des distances entre le sommet considéré et tous les autres sommets du réseau. Dans le cas des réseaux sociaux, si on s'intéresse à maximiser cette somme, elle peut répondre à la question : quel est l'acteur le plus indépendant. Par contre, si on prend le minimum, elle peut répondre à la question : quel est l'acteur le plus efficace dans la transmission de l'information sur le réseau, le plus central est donc le moins loin des autres nœuds [100].

Soit $d(i, j)$ la distance entre les nœuds i et j . Le total des distances entre i et tous les autres nœuds est : $Cc(i) = \frac{1}{\sum_{j=1}^n d(i, j)}$. La valeur maximale de cette mesure, qui dépend de la taille du graphe n , est égale à $\frac{1}{n-1}$, ce qui signifie que le nœud i est adjacent à tous les autres nœuds. Si cette valeur est nulle, alors le nœud i est un nœud isolé. Dans [12], Beauchamp a suggéré la normalisation de cette quantité : $C'c(i) = (n-1)Cc(i)$

Cette normalisation entre 0 et 1 est considérée comme l'inverse de la distance moyenne entre le nœud i et tous les autres nœuds.

5.2.3 Centralité d'intermédiarité (*Betweenness Centrality*)

La centralité d'intermédiarité (*Betweenness Centrality*) est aussi une mesure essentielle pour l'analyse des réseaux sociaux. Shimbel [103] et Pitts [95] ont mis l'accent sur l'importance d'un acteur ainsi que de sa distance géodésique par rapport aux autres nœuds pour le calcul de centralité d'intermédiarité.

La centralité d'intermédiarité, pour un nœud, dépend directement du nombre des plus courts chemins qui passent par ce nœud. Elle répond à la question : quel est l'acteur qui

a le plus de contrôle sur l'échange d'information dans le réseau? le plus central est celui par lequel le plus d'acteurs doivent passer quand ils veulent atteindre les autres acteurs du réseau.

Soient $g_{jk}(i)$ le nombre de plus courts chemins entre les j et k passant par le nœud i et g_{jk} , le nombre total de plus courts chemins entre j et k , alors, la centralité d'intermédiarité du nœud i est donnée par la formule suivante : $C_B(i) = \sum_{j < k} \frac{g_{jk}(i)}{g_{jk}}$

Spécifiquement, il y a lieu de calculer d'abord le nombre de plus courts chemins liant l'acteur j et l'acteur k (Tous ces plus courts chemins doivent être de même longueur $d(n_j, n_k)$), et par la suite, déterminer combien, parmi ces plus courts chemins, passent par l'acteur i , pour tous les indices distincts i, j, k . Alors, la centralité d'intermédiarité pour i est donnée par :

$$C_B(i) = \sum_{j < k} \frac{g_{jk} n_i}{g_{jk}}$$

5.3 Evaluation de la robustesse des nœuds importants

Dans cette partie, nous considérons trois métriques : la centralité de degré (*degree centrality*), la centralité de proximité (*Closeness centrality*) et la centralité d'intermédiarité (*Betweenness centrality*), toutes définies plus haut (cf. figure 4.1).

Par nœud important, il est signifié tout nœud i ayant une valeur supérieure à la valeur moyenne de tous les nœuds. Nous nous intéressons à plusieurs proportions de $k\%$ de la population de nœuds importants, où k varie de 10 à 100. Le processus de sélection des nœuds importants est appliqué à la fois sur le graphe initial et sur le graphe perturbé et une comparaison utilisant l'*index de similarité de Jaccard* nous permet de déterminer les changements opérés au niveau de chaque configuration étudiée. Soit A et B deux ensembles. L'indice de Jaccard est défini comme suit :

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Dans notre contexte, les résultats des expérimentations montre que la plupart des nœuds ayant un degré élevé sont exclusivement des nœuds fournisseurs (serveurs). Les nœuds ayant une centralité de proximité élevée sont généralement des nœuds fournisseurs aussi, tandis que les nœuds ayant une centralité d'intermédiarité faible sont communément des nœuds clients. Ces observations sont importantes pour le reste de l'étude.

5.4 Analyse de la robustesse

Dans cette section, nous présentons l'étude de l'évolution des caractéristiques des différentes perturbations. Nous considérons les perturbations en tenant compte de deux paramètres : *Fournisseurs* et *Ressources*. Nous ne prenons pas en compte la perturbation appliquée sur les centres d'intérêt des nœuds pour la simple raison que modifier l'ensemble des ressources ou modifier l'ensemble des intérêts des nœuds (ressources désirées) nous donne des comportements similaires. Nous étudions ces changements selon deux niveaux : 5% et 15% pour les trois types de perturbations : *Ajout*, *Suppression* et *Déplacement* pour le *fournisseur* et les *ressources*.

Nous étudions en premier l'évolution de l'effet des perturbations sur la population des nœuds importants en concordance avec les proportions du paramètre k et ce, en regardant les trois centralités (degré, proximité et intermédiarité). Dans un second temps, nous nous intéressons à l'intensité de l'impact de la perturbation sur la population moyenne des nœuds importants. Enfin, nous nous focalisons sur la robustesse des différentes configurations face à des perturbations moyennes sur les fournisseurs et les ressources.

5.4.1 Effets de la perturbation sur la population des nœuds importants

Dans cette partie, les résultats présentés concerne le niveau de perturbation à 15%. Ceux avec 5% et 30% seront discutés plus loin.

5.4.1.1 Selon la centralité d'intermédiarité (*betweenness centrality*)

Les graphes de la Figure 5.1 montrent l'évolution des changements dans une population de nœuds importants avec une valeur élevée de centralité d'intermédiarité pour chaque type de configuration du réseau étudié. Nous pouvons observer sur ces mêmes graphiques que pour des proportions de k inférieures à 50%, la différence entre les valeurs obtenues change comme le montre l'indice de Jaccard, tandis que pour des valeurs élevées de k , des écarts sont visibles.

Pour toutes les perturbations, nous obtenons une valeur faible de la similarité de Jaccard pour des valeurs faibles de k (les premières proportions des nœuds importants). Par conséquent, le changement est perceptible dans les nœuds importants ayant une intermédiarité forte. Ceci illustre que la population des *nœuds d'articulation* (nœuds avec une centralité d'intermédiarité élevée) est très **sensible** à tous les types de perturbations. En effet, dès lors que les nœuds sont principalement des clients, n'importe quelle perturbation affecte la façon avec laquelle les clients sont connectés aux fournisseurs. Nous observons également

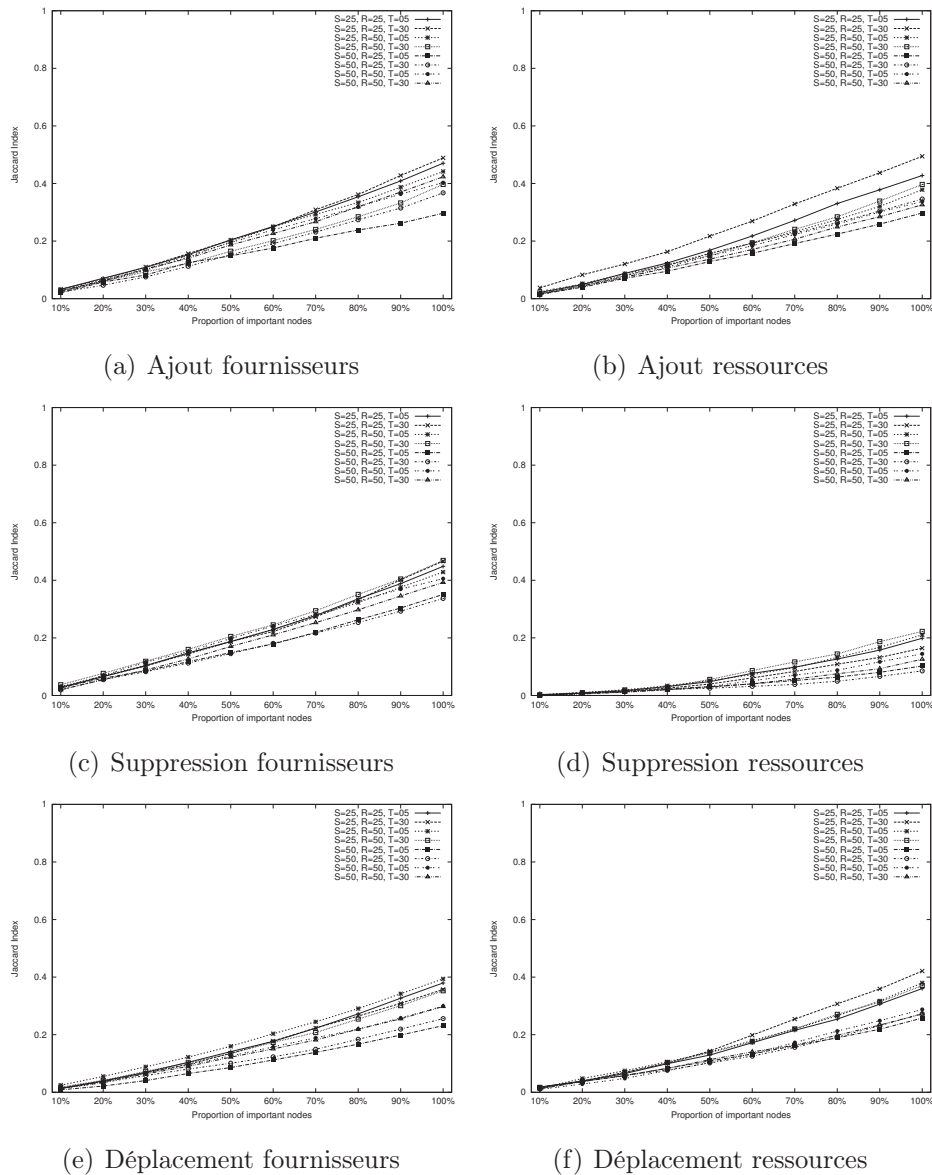


FIGURE 5.1 – Evolution des proportions des nœuds avec une centralité d'intermédiarité élevée

que pour des valeurs élevées de k , des valeurs similaires varient entre 0.3 et 0.5 pour toutes les perturbations exception faite pour la perturbation "suppression de ressources" pour laquelle la similarité est très faible (inférieure à 0.25). En effet, supprimer les ressources de certains fournisseurs incite les clients à changer leurs principaux nœuds fournisseurs, ce qui change considérablement la population de nœuds ayant une intermédiarité élevée.

Pour les perturbations "ajout de fournisseurs" et "ajout de ressources", la configuration

la plus robuste est celle qui contient une distribution de ressources faible (peu de fournisseurs, peu de ressources et une proportion des centres d'intérêt des nœuds élevée). Tandis que la configuration la moins robuste est telle que la proportion des ressources est élevée, celles des fournisseurs et des centres d'intérêt sont faibles. Ceci est dû au fait qu'un client ayant un large intérêt (ressources désirées par ce client) se connecte à plusieurs fournisseurs. D'où, dans un tel mécanisme d'attachement préférentiel, si nous ajoutons de nouveaux fournisseurs ou de nouvelles ressources dans le réseau considéré, ses nœuds renforcent leurs liens originaux et leurs positions intermédiaires, ce qui rend la configuration plus stable. D'un autre côté, si les intérêts des nœuds (ressources désirées) sont ciblés lorsque les fournisseurs sont nombreux dans le réseau mais ces fournisseurs ont uniquement peu de ressources, alors les connections changent souvent et par conséquent, la population des nœuds d'articulation change énormément. Pour la perturbation "supprimer fournisseurs", nous notons que la configuration la plus robuste est pour des faibles proportions des paramètres R, S, T et, la configuration la moins robuste est celle pour laquelle R est faible et S, T sont élevés. Lorsque les fournisseurs ne possèdent que peu de ressources et que les intérêts des nœuds sont restreints, la probabilité que la perturbation fasse changer de fournisseurs pour un client est faible. Contrairement à cela, si les intérêts des nœuds sont variés et qu'il existe plusieurs fournisseurs avec peu de ressources, la perturbation "suppression de ressources" induit beaucoup de changements dans le sens de la connection du client et du fournisseur. Pour ce type de perturbation, nous pouvons constater que la robustesse des différentes configurations dépend du ratio R/S . En effet, s'il existe beaucoup de ressources sur un nombre réduit de fournisseurs, alors la configuration du réseau est plus robuste et s'il y a peu de ressources sur beaucoup de serveurs, alors cette configuration est moins robuste. S'agissant des perturbations "Déplacement de fournisseurs" et "Déplacement de ressources", la configuration la moins robuste est celle qui contient beaucoup de fournisseurs et vice versa.

5.4.1.2 Selon la centralité de proximité (*closeness centrality*)

Les graphes de la Figure 5.2 montrent l'évolution des changements dans la population des nœuds avec une centralité de proximité élevée et ce, pour chaque type perturbation. Identiquement aux courbes obtenues pour la centralité d'intermédiarité, nous pouvons observer que pour des valeurs faibles de k , nous obtenons des valeurs faibles de l'indice de Jaccard. Toutefois, les écarts entre les différentes configurations sont perceptibles et, globalement, les valeurs de l'indice de Jaccard sont élevées par rapport à celles obtenues avec la centralité d'intermédiarité. Ceci montre que la population des nœuds fournisseurs (avec une forte centralité de proximité) est plus stable que la population des nœuds clients (nœuds avec une centralité d'intermédiarité élevée). Pour les proportions élevées de k , les

valeurs de la similarité de Jaccard se situent entre 0.4 et 0.9 pour toutes les perturbations, exception faite pour la perturbation "suppression ressources" pour laquelle la similarité est en dessous de 0.45.

Pour la perturbation "ajout de fournisseurs", les configurations les plus robustes sont celles où la distribution des ressources est forte (les proportions de R, S sont élevées). Quant aux configurations les moins robustes sont celles où les proportions de S, T sont basses. Quand il y a beaucoup de ressources dans le réseau, le client a tendance à voir toutes ses requêtes satisfaites par un même fournisseur même si de nouveaux fournisseurs viennent d'être ajoutés au réseau. Lorsque le nombre de fournisseurs dans le réseau n'est pas important et que les intérêts des nœuds clients sont restreints, alors la perturbation "ajout fournisseurs" va inciter les clients à se connecter à d'autres fournisseurs et de ce fait, ils changeront de fournisseurs principaux. Pour la perturbation "suppression de fournisseurs", la configuration la plus robuste est similaire à celle où la distribution R, S est forte. Par contre, la configuration la moins robuste est celle ayant peu de fournisseurs et des fortes proportions des paramètres R et T . En effet, "supprimer des fournisseurs" dans une telle configuration change considérablement les connections entre les clients et les fournisseurs. Pour la perturbation "déplacement de fournisseurs", la configuration la plus robuste est obtenue quand il y a peu de fournisseurs S et beaucoup de ressources désirées ou recherchées sur le réseau (T élevé). Les clients maintiennent les mêmes connections avec leur fournisseurs dans de telles configurations, même si certains fournisseurs sont déplacés. Pour ce qui est de la configuration la moins robuste, celle-ci est caractérisée par un nombre élevé de fournisseurs avec peu de ressources et que le nombre de ressources recherchées est important (T élevé). Dès lors que les clients ont le choix de se connecter à tel ou tel fournisseur, la topologie de la configuration considérée change en conséquence.

A ce stade, nous considérons les perturbations appliquées sur les ressources. Pour la perturbation "ajout de ressources", nous pouvons observer que le paramètre T (proportion des ressources recherchées dans le réseau) est le plus influent. La plupart des configurations sont robustes quand T est important et vice versa. Si les intérêts sont restreints, quand de nouvelles ressources sont ajoutées dans le réseau, il y a une forte probabilité que les clients changent de fournisseurs. Pour la perturbation "suppression de ressources", on peut observer que la perturbation a un fort impact pour toutes les configurations, dû essentiellement au mécanisme de l'attachement préférentiel. De plus, la perturbation "déplacement de ressources" impacte plus les configurations qui contiennent peu de fournisseurs (l'inverse est aussi vrai).

5.4.1.3 Selon la centralité de degrés *degree centrality*

De la Figure 5.3, nous pouvons observer que les courbes adoptent une forme linéaire et que la différence entre les configurations (combinaisons de R, S, T) est minimale : toutes les configurations ont pratiquement un comportement similaire face à des perturbations. En outre, les valeurs de l'index de Jaccard atteignent quasiment la valeur 1 pour $k = 100$ qui signifie que la population globale des nœuds ayant un degré élevé ne change pas ou peu après une perturbation.

Cette observation est due au fait que les fournisseurs sont des nœuds privilégiés pour avoir des degrés élevés. Par conséquent, et tel que mentionné plus haut, les nœuds ayant des degrés élevés sont uniquement des fournisseurs. Ainsi, notre algorithme d'interaction est conçu de sorte que, pour des valeurs faibles de k , nous obtenons des ensembles de fournisseurs qui ne se ressemblent pas. Cette particularité est propre à notre algorithme qui privilégie la recherche de ressources désirées auprès du serveur qui est le plus proche du nœud demandeur. Toutefois, si l'on considère la valeur de $k = 100$, nous obtenons certainement le même ensemble de nœuds ayant des degrés élevés (représentant les fournisseurs) et ce, même si la façon dont les clients sont connectés ces fournisseurs a changé.

5.4.2 Robustesse des différentes configurations (*selon les proportions R, S, T*)

Dans cette section, nous discutons en premier lieu de l'effet de l'intensité de la perturbation pour le comparer, par la suite, aux différentes configurations en concordance avec la valeur moyenne de la perturbation sur les fournisseurs et les ressources.

5.4.2.1 Les effets de l'intensité des perturbations

Les figures 5.4, 5.5 et 5.6 montrent la comparaison entre les valeurs moyennes de l'indice de Jaccard pour chaque configuration et toutes les configurations et ce, pour chacune des trois centralités (de proximité, de degré et d'intermédiarité).

Globalement, nous pouvons observer que certaines incohérences entre les intensités à 5% et 15% sont plus régulières pour la centralité d'intermédiarité que pour la centralité de proximité, tandis que la centralité de degrés est moins impactée par ces intensifications.

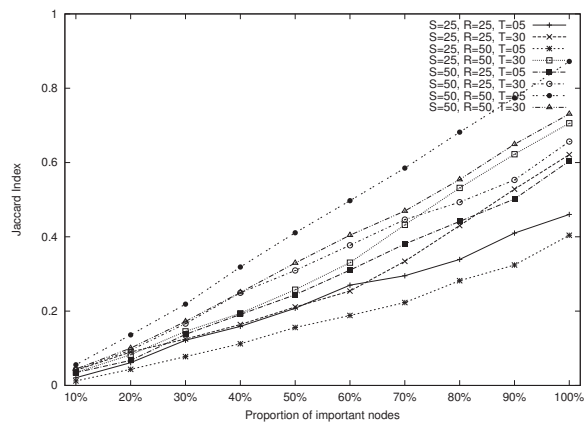
Par ailleurs, nous avons confirmé que la perturbation "suppression de ressources" est la plus impactante des perturbations et que la population des fournisseurs (ayant une centralité de proximité élevée) est plus stable que la population des clients (ayant une centralité d'intermédiarité élevée).

5.4.3 Impact moyen sur les différentes perturbations

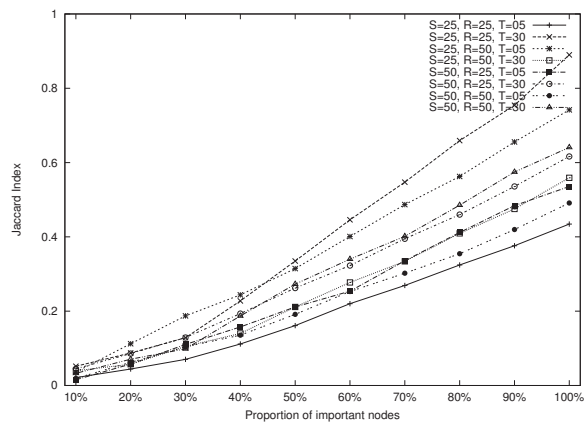
La figure 5.7 est une comparaison entre les impacts des perturbations appliquées sur les ressources ainsi que celles appliquées sur les fournisseurs pour les deux centralités intermédiarité et proximité. Nous pouvons remarquer que globalement, les perturbations appliquées sur les ressources ont plus d'impact que les perturbations appliquées sur les fournisseurs. De plus, la robustesse d'une configuration est nettement dépendante du ratio de ressources sur les fournisseurs R/S . Si ce ratio est petit, alors la configuration est fragile. Ceci explique le fait que s'il existe beaucoup de fournisseurs avec peu de ressources alors, les changements qui interviennent dans les liaisons entre les clients et les fournisseurs sont hautement altérés par les différentes perturbations.

5.5 Conclusion

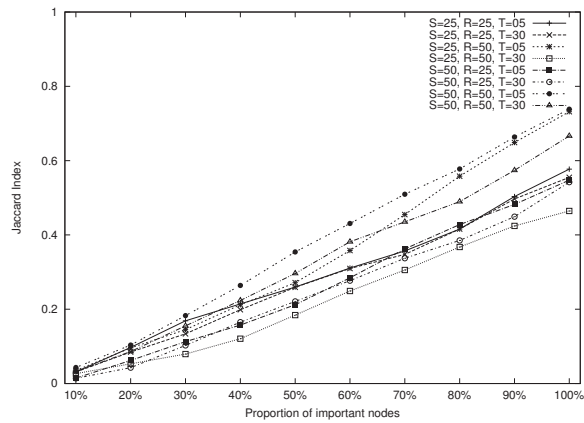
A travers l'étude des positions des nœuds importants, nous avons pu examiner les effets de certaines perturbations sur les structures de graphes qui émergent dans les réseaux de communication. Dans ce chapitre, nous nous sommes intéressés à étudier la robustesse des structures émergées des réseaux de communication par l'étude des nœuds importants selon trois centralités : la centralité de proximité, la centralité d'intermédiarité et la centralité de degrés. Les résultats que nous avons obtenus nous ont permis de mener des discussions notamment, sur l'impact de chaque perturbation sur les différentes configurations du graphe. Ces résultats nous ont permis aussi de déterminer, parmi toutes ces configurations, celles qui sont les plus robustes et les moins robustes en tenant compte à la fois d'une des trois centralités précitées.



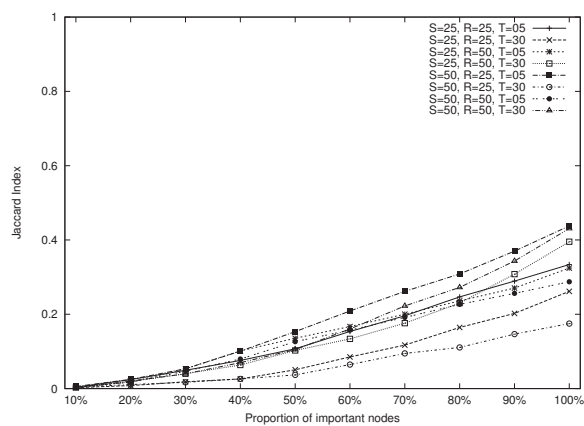
(a) Ajout fournisseurs



(b) Ajout ressources



(c) Supprimer fournisseurs



(d) Supprimer ressources

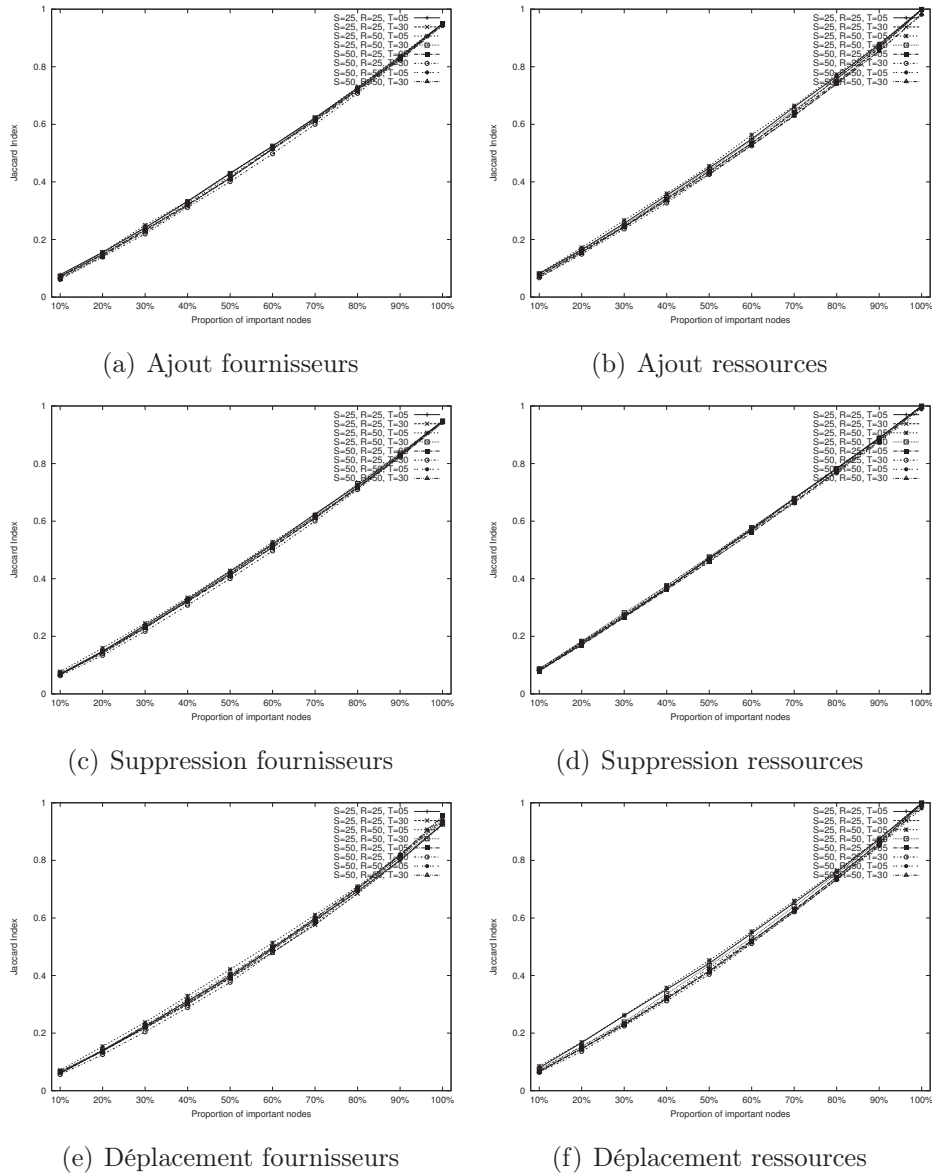


FIGURE 5.3 – Evolution des proportions des nœuds avec une centralité de degré élevée

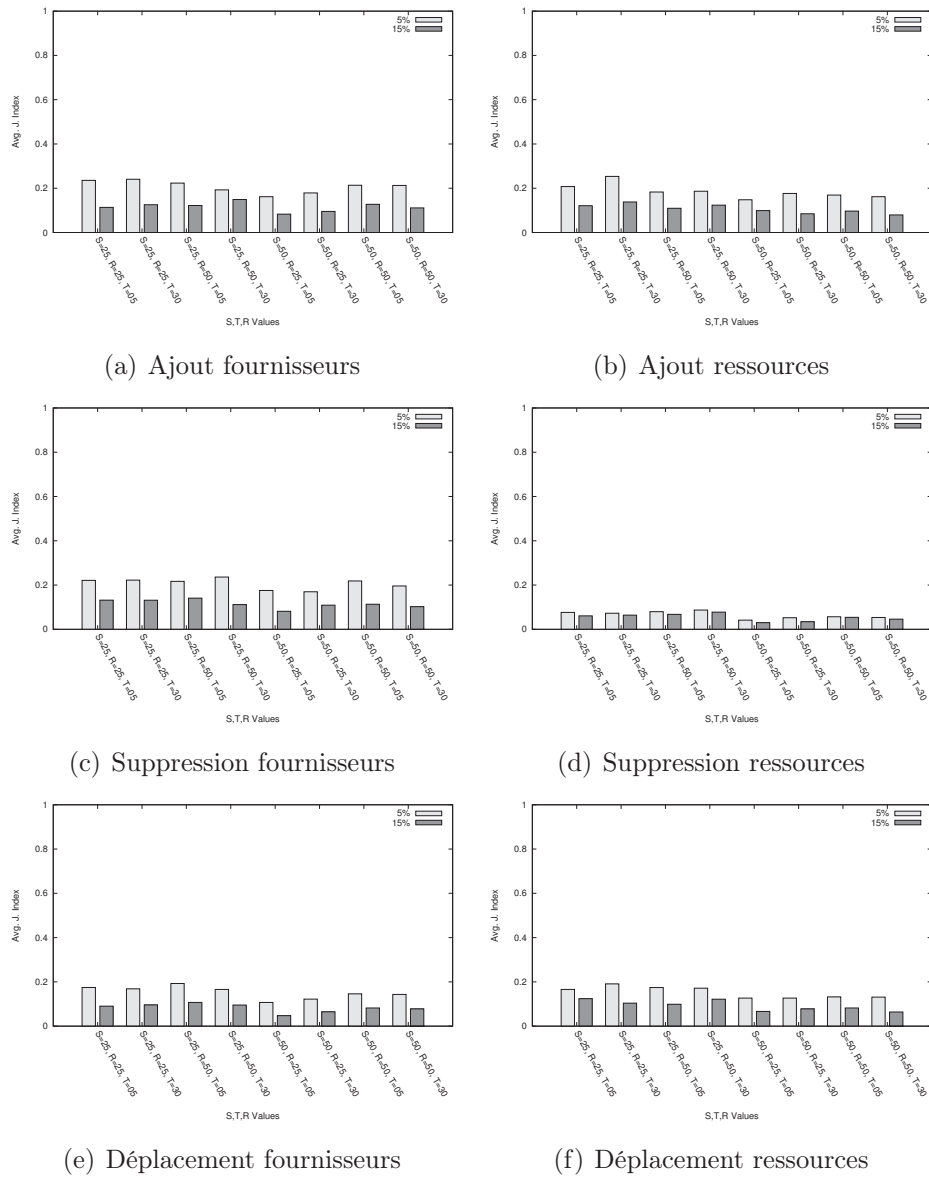


FIGURE 5.4 – Effets sur les nœuds à fort centralité de proximité

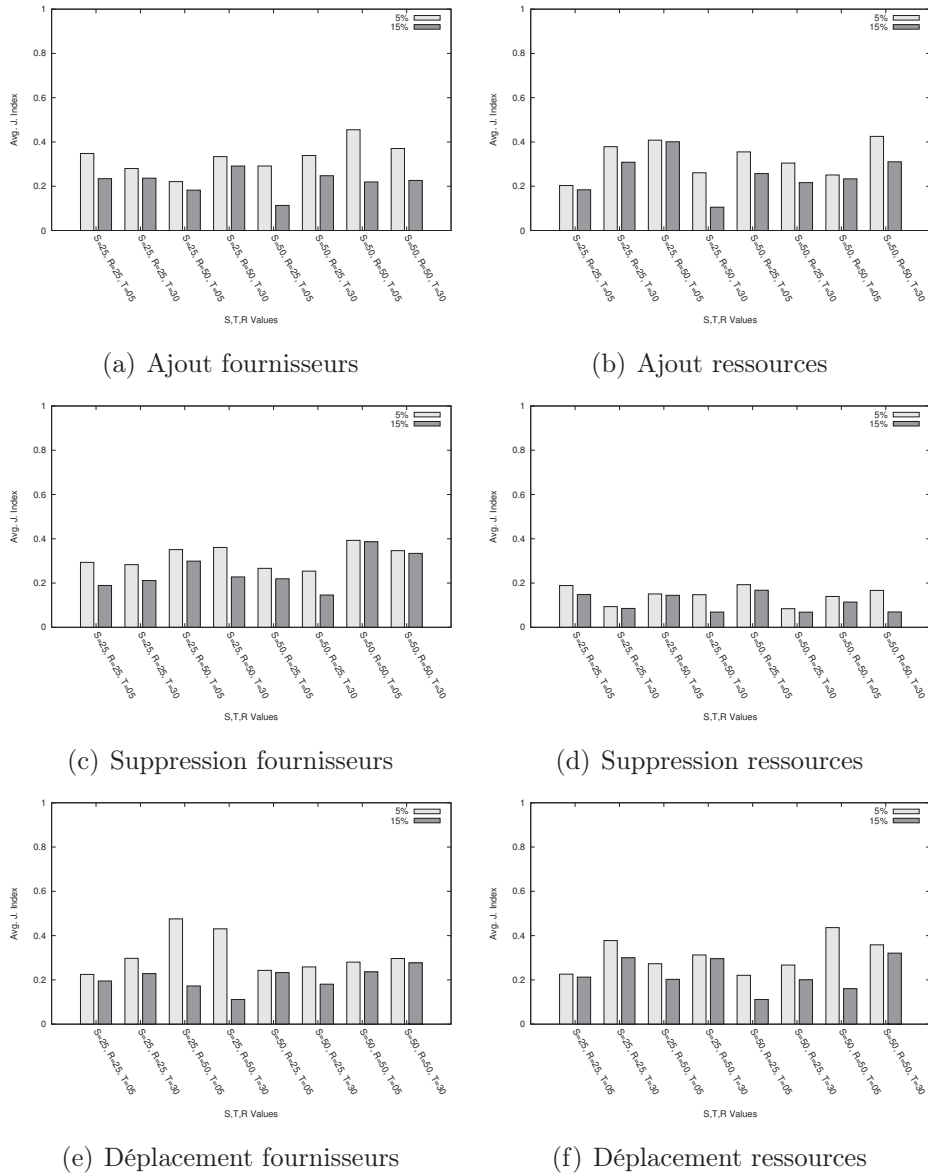


FIGURE 5.5 – Effets sur les nœuds à fort centralité de proximité

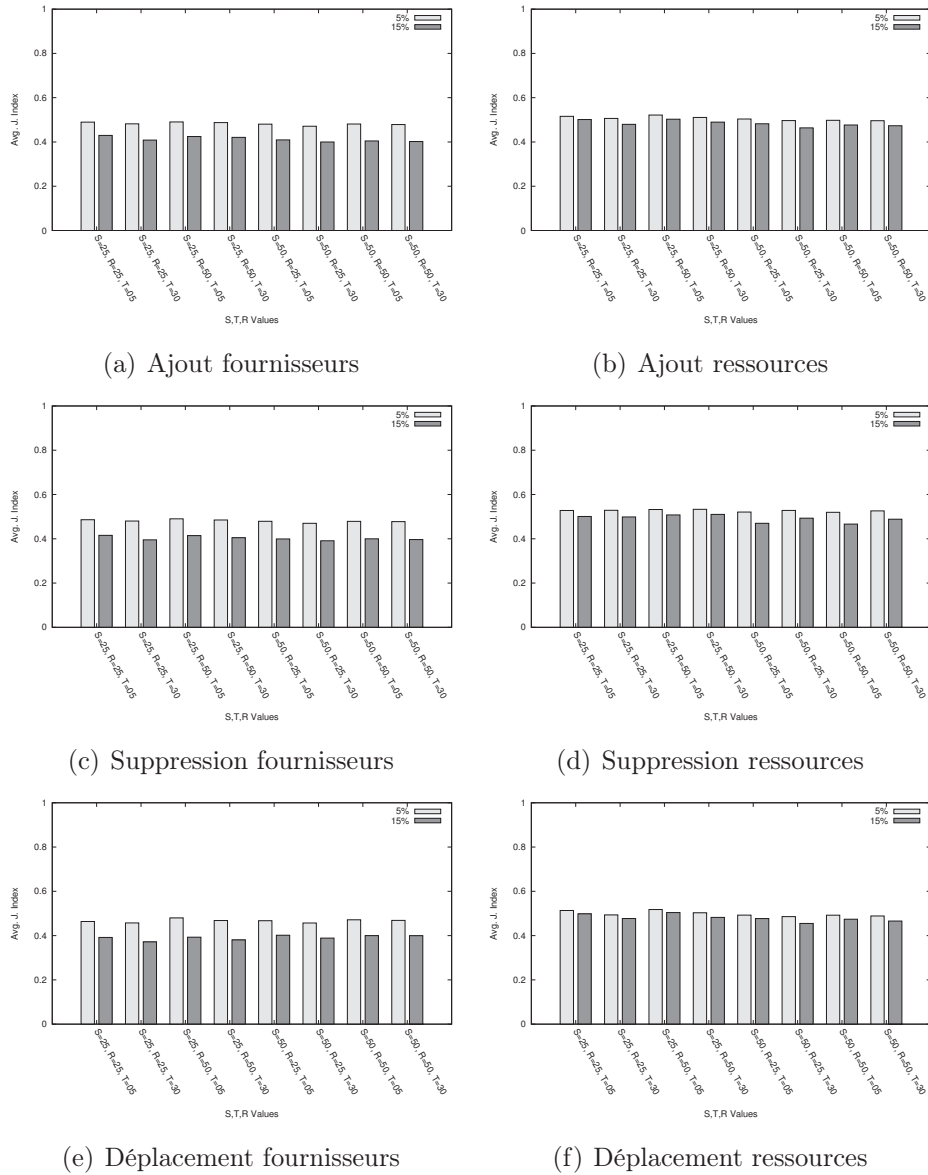
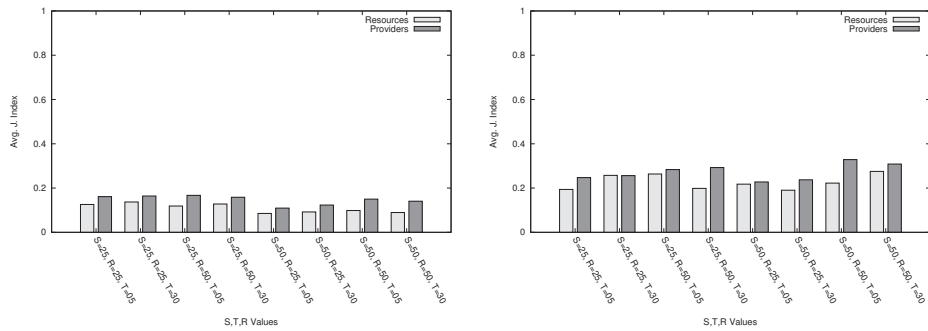


FIGURE 5.6 – Effets sur les nœuds à fort centralité de degrés



(a) Centralité d'intermédiation

(b) Centralité de proximité

FIGURE 5.7 – Impact moyen : Intermédiation et Proximité

Conclusion générale

Dans cette thèse, nous nous sommes intéressés à l'étude de la robustesse des structures sociales qui émergent dans les réseaux sociaux d'interaction. Notre travail est scindé en deux grand volets. Le premier volet traite de l'émergence de ces structures. Quand, au deuxième volet, celui-ci s'intéresse à l'étude de la robustesse de ces structures sociales.

Dans le premier volet, nous avons proposé un modèle (framework) pour l'émergence des communautés dans les réseaux. Ce framework modélise les réseaux par des graphes que nous avons caractérisés par un ensemble de paramètres initiaux du réseau (nombre de ressources dans le réseau, nombre de fournisseurs dans le réseau, *etc*). Les structures sociales (communautés) qui émergent ont été classées selon certaines propriétés mathématiques telles que la densité, le nombre de feuilles ou le degré max. Cette classification est nécessaire pour pouvoir étudier la robustesse de ces structures sociales. De plus, nous avons proposé une étude de la corrélation entre les paramètres du réseau initial et les types de structures de communautés susceptibles d'émerger. Cette étude des structures de communautés dans les réseaux est d'une grande importance pour comprendre le fonctionnement et le comportement des réseau considéré.

Le deuxième volet de notre thèse concerne l'étude de la robustesse de ces graphes sociaux. Pour réaliser notre étude, nous avons défini un autre modèle qui permet de comparer deux structures émergées (initiale et perturbée). Notre étude de la robustesse est constituée de trois parties : Une partie qui s'intéresse à la robustesse des structures émergées en tant qu'un seul graphe. La deuxième partie s'intéresse à la robustesse des structures émergées (communautés) prises séparément. Quant à la troisième partie, celle-ci s'intéresse à l'étude des nœuds importants dans le réseau. Ces trois études sont toutes complémentaires et nous permettent de mieux comprendre le fonctionnement des réseaux sociaux d'interaction et leurs organisations structurelles.

A travers notre étude, nous avons examiné l'effet de certaines perturbations sur les graphes sociaux dans les réseaux d'interaction. Contrairement aux études classiques de la robus-

tesse qui s'intéressent principalement aux aspects structurels (nœuds, liens), nous avons proposé une étude de la robustesse à travers des paramètres qui permettent l'émergence de ces graphes sociaux. Nous avons étudié l'impact des perturbations appliquées sur les serveurs et les ressources dans un réseau d'interaction et les résultats que nous avons obtenus nous ont permis de mener une discussion sur l'importance de chaque configuration du réseau selon plusieurs métriques.

Notre modèle pour l'étude de la robustesse n'est pas dédié à un type de réseau bien précis, au contraire, il peut-être généralisé à d'autres types de réseaux. La seule contrainte se présente dans l'identification des paramètres qui permettent l'émergence de ces réseaux sociaux. Par exemple, dans les réseaux de gènes, un des paramètres permettant cette émergence réside dans l'affinité entre ces gènes pour former une séquence. En effet, si on arrive à dénombrer les interactions possibles entre les gènes, on pourrait déterminer les paramètres permettant la formation de séquences et de manière similaire à notre étude, l'évaluation de la robustesse reviendrait alors à l'étude l'impact des perturbations sur les séquences de gènes. Par ailleurs, le cas d'étude que nous avons considéré peut aider les administrateurs des réseaux à mieux comprendre le comportement des réseaux qu'ils gèrent et par conséquent, leur permettre de mieux les administrer et prévenir les configurations spécifiques pour lesquelles le réseau est le moins robuste. En effet, de plus en plus, les applications actuelles intègrent les aspects sociaux dans leur fonctionnement et utilisent la notion de communautés afin de mieux adapter les traitements selon la communauté d'utilisateurs visés. Notre étude aidera non seulement l'administrateur à adapter les structures des communautés selon ses besoins spécifiques (besoin en diffusion, en filtrage d'interactions,...) mais aussi de placer les différentes ressources de manière à atteindre un niveau de robustesse permettant la pérennité de fonctionnement des applications.

Comme perspectives, nous envisageons d'appliquer notre modèle de l'étude la robustesse à d'autres types de réseaux d'interaction tels que les réseaux biologiques ou sur des réseaux génériques et réels avec des protocoles de communication différents de celui que nous avons défini. Cela permettra de valider nos modèles par une application et une étude de la robustesse des structures émergées.

Nous proposons également de définir de nouveaux paramètres spécifiques à chaque réseau qui permettent l'émergence des graphes sociaux d'élargir l'étude de la robustesse à d'autres propriétés théoriques et d'utiliser d'autres métriques permettant la quantification des changements survenus dans les structures de graphes après perturbation.

Nous souhaitons, à travers ces perspectives, définir un modèle plus complet et paramétrable à même d'être appliqué à n'importe quel réseau d'interaction.

Bibliographie

- [1] <<http://www.wormatlas.org/handbook/nshandbook.htm/nswiring.htm>>.
- [2] Lada A. Adamic. The small world web. pages 443–452. Springer, 1999.
- [3] William Aiello, Fan Chung, and Linyuan Lu. A random graph model for massive graphs. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, STOC '00, pages 171–180, New York, NY, USA, 2000.
- [4] R. Albert, H. Jeong, and A.L. Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794) :378–382, 2000.
- [5] Reka Albert, Hawoong Jeong, and Albert-Laszlo Barabasi. The diameter of the world wide web. *NATURE*, 401 :130, 1999.
- [6] N Alon. Eigen values and expanders. *Combinatorica*, 6 :83–96, 1986.
- [7] A. Arenas, J. Duch, A. Fernandez, and S. Gomez. Size reduction of complex networks preserving modularity. *New Journal of Physics*, 9, 2007.
- [8] A. Arenas, A. Fernandez, S. Fortunato, and S. Gomez. Motif-based communities in complex networks. *J. Phys. A : Math. Theor.*, 41 :224001, 2008.
- [9] A. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A : Statistical Mechanics and its Applications*, 311 :590–614, 2002.
- [10] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286 :509–512, 1999.
- [11] A. Bavelas. A mathematical model of group structure. *Human Organizations*, 7 :16–30, 1948.
- [12] Murray A. Beauchamp. An improved index of centrality. *Behavioral Science*, 10 :161–163, 1965.
- [13] Claude Berge. *Graphes et hypergraphes*. Dunod, 1970.
- [14] K. Bharat, A. Broder, M. Henzinger, P. Kumar, and S. Venkatasubramanian. The Connectivity Server : fast access to linkage information on the Web. *Computer Networks and ISDN Systems*, 30 :469–477, 1998.

- [15] M. Boguna, R. Pastor-Satorras, A. Diaz-Guilera, and A. Arenas. Models of social networks based on social distance attachment. *Physical Review E*, 70 :056122, 2004.
- [16] B. Bollobas. *Random Graphs*. Cambridge University Press, 2001.
- [17] S. Borgatti, K. Carley, and D. Krackhardt. On the robustness of centrality measures under conditions of imperfect data. *Social Networks*, 28, 2006.
- [18] A. Broder. Graph structure in the Web. *Computer Networks*, 33 :309–320, 2000.
- [19] A. Capocci, V. Servedio, F. Colaiori, and G. Caldarelli. Associative memory on a small-world neural network. *LNCS*, 3243 :181–188, 2004.
- [20] V. Carchiolo, M. Malgeri, G. Mangioni, and V. Nicosia. Emerging structures of p2p networks induced by social relationships. *Comput. Commun.*, 31(3) :620–628, 2008.
- [21] Vincenza Carchiolo, Michele Malgeri, Giuseppe Mangioni, and Vincenzo Nicosia. Prosa : P2p resource organisation by social acquaintances. *Agents and Peer-to-Peer Computing*, pages 135–142, 2008.
- [22] Vasek Chvatal. Tough graphs and hamiltonian circuits. *Discrete Mathematics*, 306(10-11) :910–917, 2006.
- [23] E. Cohen, A. Fiat, and H. Kaplan. Associative search in peer to peer networks : harnessing latent semantics. In *22nd IEEE INFOCOM*, volume 2, pages 1261–1271, 2003.
- [24] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience, 1991.
- [25] C. Suffel D. Bauer, F. Boesch and R. Tindell. Connectivity extremal problems and the design of reliable probabilistic networks. *The Theory and Application of Graphs*, ed. G. Chartrand, Y. Alavi, D. Goldsmith, L. Lesniak-Foster and D. Lick (Wiley), pages 45–54, 1981.
- [26] Leon Danon, Albert D. Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. In *J. Stat. Mech*, page 09008, 2005.
- [27] Dara and Moazzami. Tenacity of a graph with maximum connectivity. *Discrete Applied Mathematics*, 159(5) :367 – 380, 2011.
- [28] de Solla. Networks of Scientific Papers. *Science*, 149(3683), 1965.
- [29] Sybil Derrible and Christopher Kennedy. *Physica A : Statistical Mechanics and its Applications*, 389(17), 2010.
- [30] Pedro Domingos and Matthew Richardson. Mining the network value of customers. In *KDD*, pages 57–66, 2001.
- [31] L. Donetti and M. A. Munoz. Detecting network communities : a new systematic and powerful algorithm. *J. Stat. Mech. : Theor. Exp.*, 10 :10012, 2004.

-
- [32] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. In *Adv. Phys.*, 2002.
- [33] N. D. Doulamis, P. N. Karamolegkos., A. Doulamis, and L. Nikolakopoulos. Exploiting semantic proximities for content search over p2p networks. *Computer Communications*, 32(5) :814–827, 2009.
- [34] P. Erdős and A. Rényi. On random graphs, I. *Publicationes Mathematicae (Debrecen)*, 6 :290–297, 1959.
- [35] P. Erdos and A. Renyi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5 :17–61, 1960.
- [36] P. Erdos and A. Renyi. On the strength of connectedness of a random graph. *Acta Mathematica Scientia Hungary*, 12 :261–267, 1961.
- [37] Abdol-Hossein Esfahanian and S. Louis Hakimi. On computing a conditional edge-connectivity of a graph. *Inf. Process. Lett.*, 27(4) :195–199, 1988.
- [38] Alex Fabrikant, Elias Koutsoupias, and Christos H. Papadimitriou. Heuristically optimized trade-offs : A new paradigm for power laws in the internet. In *Proceedings of the 29th International Colloquium on Automata, Languages and Programming, ICALP '02*, pages 110–122, London, UK, UK, 2002. Springer-Verlag.
- [39] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the Internet topology. volume 29 of *SIGCOMM '99*, pages 251–262. ACM, 1999.
- [40] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Math Journal*, 23(98) :298–305, 1973.
- [41] S. Fortunato. Community detection in graphs. *Physics Reports*, 486 :75–174, 2010.
- [42] S. Fortunato and M. Barthelemy. Resolution limit in community detection. *PNAS*, 104 :36, 2007.
- [43] S. Fortunato and M. Barthelemy. Resolution limit in community detection. *PNAS*, 104 :36, 2007.
- [44] H. Frank and I. Frisch. Analysis and design of survivable networks. *Communication Technology, IEEE Transactions on*, 18(5) :501–519, 1970.
- [45] L. Freeman. Centrality in social networks : Conceptual clarification. *Social Networks*, 1 :215–239, 1979.
- [46] Laura Garton, Caroline Haythornthwaite, and Barry Wellman. Studying online social networks. *Journal of Computer Mediated Communication (JCMC)*, 3(1), 1997.

- [47] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12) :7821–7826, 2002.
- [48] P. Gleiser and L. Danon. Community structure in jazz. *Advances in Complex Systems*, 6 :565, 2003.
- [49] S. Gomez, P. Jensen, and A. Arenas. Analysis of community structure in networks of correlated data. *Physical Review E*, 80, 2009.
- [50] Ramesh Govindan and Hongsuda Tangmunarunkit. Heuristics for Internet map discovery. pages 1371–1380, 2000.
- [51] R. Guimera and L.A.N Amaral. Functional cartography of complex metabolic networks. *Nature*, 433 :895–900, 2005.
- [52] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in a network of human interactions. *Physical Review E*, 68(6) :065103, 2003.
- [53] R. Guimera, S. Mossa, A. Turtschi, and L. Amaral. The worldwide air transportation network : Anomalous centrality, community structure, and cities’ global roles. *Proc. Natl. Acad. Sci. U. S. A.*, 102 :7794–7799, 2005.
- [54] Robert A. Hanneman and Mark Riddle. *Introduction to social network methods*. University of California, Riverside, 2005.
- [55] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han. Attack vulnerability of complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 65, 2002.
- [56] Lawrence Hubert and Phipps Arabie. Comparing partitions. *J. of Classification*, 2(1) :193–218, 1985.
- [57] H. A. Jung. On a class of posets and the corresponding comparability graphs. *J. Comb. Theory, Ser. B*, 24(2) :125–133, 1978.
- [58] Xu Junming. On conditional edge-connectivity of graphs. *Acta Mathematicae Applicatae Sinica (English Series)*, 16 :414–419, 2000.
- [59] Brian Karrer, Elizaveta Levina, and M. E. J. Newman. Robustness of community structure in networks. *Phys. Rev. E*, 77, 2008.
- [60] E. Katz and F. Paul. *Personal Influence : The Part Played by People in the Flow of Mass Communications*. Transaction Publishers, 2005.
- [61] H. Kautz, B. Selman, and M. Shah. Referral web : combining social networks and collaborative filtering. *Commun. ACM*, 40(3) :63–65, 1997.
- [62] B. W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal*, 49(2) :291–308, 1970.

-
- [63] Minkyu Kim and Muriel Médard. Robustness in large-scale random networks. In *INFOCOM*, 2004.
- [64] M. S. Krishnamoorthy and b. Krishnamurthy. Fault diameter of interconnection networks. *Comput. Math. Appl.*, 13 :577–582, April 1987.
- [65] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the Web graph. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 57–65. IEEE Computer Society, 2000.
- [66] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, D. Sivakumar, Andrew S. Tomkins, and Eli Upfal. The web as a graph. 2000.
- [67] A. Lancichinetti and S. Fortunato. Community detection algorithms : A comparative analysis. *Physical Review E*, 80 :056117, 2009.
- [68] E. A. Leicht and M. E. J. Newman. Community structure in directed networks. *Physical Review Lettre*, 100 :118703, 2008.
- [69] Mark Levene and Alexandra Poulouvasilis. Web dynamics. *Software Focus*, 2(2) :60–67, 2001.
- [70] D. Lusseau, K. Schneider, O.J. Boisseau, P. Haase, E. Slooten, and S.M. Dawson. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. can geographic isolation explain this unique trait? *Behavioral Ecology and Sociobiology*, 54 :396, 2003.
- [71] S. Lemmouchi M. Haddad and H. Kheddouci. Robustness of emerged community in social network. *MEDES*, pages 477–479, 2009.
- [72] S. Lemmouchi M. Haddad and H. Kheddouci. Robustesse de communautés émergées des graphes sociaux issus des réseaux de communication. *EGC*, 2010.
- [73] S. Lemmouchi M. Haddad and H. Kheddouci. Robustness evaluation framework. 2010.
- [74] S. Lemmouchi M. Haddad and H. Kheddouci. Study of robustness of community emerged from exchanges in networks communication. *MEDES*, pages 189–196, 2011.
- [75] S. Lemmouchi M. Haddad and H. Kheddouci. Robustness study of emerged social communities from exchanges in networks. *Computer Communication Journal*, 2012.
- [76] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat. Systematic Topology Analysis and Generation Using Degree Correlations. *SIGCOMM 2006*, (4) :135–146, 2006.
- [77] C. P. Massen and J. P. K. Doye. Thermodynamics of community structure. *Preprint arXiv :cond-mat/0610077v1*, 2007.

- [78] P. D. McDaniel, S. Sen, O. Spatscheck, J. E. van der Merwe, W. Aiello, and C. R. Kalmanek. Enterprise security : A community of interest based approach. In *NDSS*, 2006.
- [79] Elke Michlmayr, Arno Pany, and Gerti Kappel. Using taxonomies for content-based routing with ants. *Comput. Networks*, 51(16) :4514–4528, 2007.
- [80] Peter Mika. Ontologies are us : A unified model of social networks and semantics. *J. Web Sem.*, 5(1) :5–15, 2007.
- [81] S. Milgram. The small world problem. *Psychology Today*, 2(1), 1967.
- [82] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs : Simple building blocks of complex networks. *Science*, 298(5594) :824–827, 2002.
- [83] Bojan Mohar. Isoperimetric numbers of graphs. *Journal of Combinatorial Theory, Series A*, 47(3) :274–291, 1989.
- [84] J.L. Moreno. Emotions mapped by new geography. *Physics Reports*, 1933.
- [85] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45 :167, 2003.
- [86] M. E. J. Newman. Detecting community structure in networks. *The European Physical Journal B*, 38(2) :321–330, 2004.
- [87] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69 :066133, 2004.
- [88] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *PHYS.REV.E*, 69, 2004.
- [89] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3) :036104, 2006.
- [90] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23) :8577–8582, June 2006.
- [91] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69 :026113, 2004.
- [92] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64, 2001.
- [93] G. Palla, I. Derenyi, L. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks. *Nature*, 435 :814–818, 2005.
- [94] J. Park and M E J Newman. A network-based ranking system for us college football. *Journal of Statistical Mechanics : Theory and Experiment*, 2005(10) :P10014, 2005.

-
- [95] F. R. Pitts. The medieval river trade network of russia revisited. *Social networks*, 1 :285–292, 1979.
- [96] P. Pons and M. Latapy. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10 :191, 2006.
- [97] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. *PNAS*, 101(9) :2658–2663, 2004.
- [98] S. Redner. How popular is your paper ? an empirical study of the citation distribution. *The European Physical Journal B*, 4 :131, 1998.
- [99] H. Rostami, J. Habibi, and E. Livani. Semantic partitioning of peer-to-peer search space. *Computer Communications*, 32(4) :619–633, 2009.
- [100] G. Sabidussi. The centrality index of a graph. *Psychometrika*, 31 :581–603, 1966.
- [101] J.W. Scannell, C. Blakemore, and M.P. Young. Analysis of connectivity in the catcerebral cortex. *Journal of Neuroscience*, 15 :1463, 1995.
- [102] John P. Scott. *Social Network Analysis : A Handbook*. SAGE Publications, 2000.
- [103] Alfonso Shimbel. Structural parameters of communication networks. *Bulletin of Mathematical Biology*, 15 :501–507, 1953.
- [104] H. Simon. On a class of skew distribution functions. *Biometrika*, 42, 1955.
- [105] Karsten Steinhaeuser and Nitesh V. Chawla. Community detection in a large real-world social network. In Huan Liu, John J. Salerno, and Michael J. Young, editors, *Social Computing, Behavioral Modeling, and Prediction*, pages 168–175. Springer US, 2008.
- [106] Steven H. Strogatz. Exploring complex networks. *Nature*, 410, 2001.
- [107] Ali Sydney, Caterina M. Scoglio, Phillip Schumm, and Robert E. Kooij. Elasticity : Topological characterization of robustness in complex networks. *CoRR*, abs/0811.4040, 2008.
- [108] C. Tantipathananandh, T. Berger-Wolf, and D. Kempe. A framework for community identification in dynamic social networks. In *13th ACM SIGKDD conference*, pages 717–726, 2007.
- [109] G. Tummarello and C. Morbidoni. The dbin platform : A complete environment for semantic web communities. *Web Semant.*, 6(4) :257–265, 2008.
- [110] Y. Upadrashta, J. Vassileva, and W. Grassmann. Social networks in peer-to-peer systems. In *IEEE HICSS conference*, page 200.3, 2005.
- [111] S. Wasserman and K Faust. *Social Network Analysis : Methods and Applications*. Cambridge University Press, New York., 1994.

- [112] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of small-world networks. 393 :440–442, 1998.
- [113] J. C. Willis and Yule G. U. some statistics of evolution and geographical distribution in plants and animals and their significance. *Nature*, 109, 1922.
- [114] Junjie Wu, Hui Xiong, and Jian Chen. Adapting the right measures for k-means clustering. In *ACM KDD*, pages 877–886, 2009.
- [115] S.J.H. Yang, J. Zhang, L. Lin, and J.J.P Tsai. Improving peer-to-peer search performance through intelligent social search. *Expert Systems with Application, (SCI)*, 36(7) :10312 – 10324, 2009.
- [116] W.W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33 :452–473, 1977.
- [117] B. Zhang, H. Li, Y. Liu, L. Ji, W. Xi, W. Fan, Z. Chen, and W.Y Ma. Improving web search results using affinity graph. In *28th ACM SIGIR conference*, pages 504–511, 2005.
- [118] Jichang Zhao and Ke Xu. Enhancing the robustness of scale-free networks. *CoRR*, abs/0905.2227, 2009.
- [119] H. Zhou and R. Lipowsky. Network brownian motion : A new method to measure vertex-vertex proximity and to identify communities and subcommunities. In *International conference on computational science*, pages 1062–1069, 2004.