

**UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS**

**DÉSINTÉGRATION D'UN RÉSEAU SOCIAL AU SEIN DE SES COMMUNAUTÉS**

**MÉMOIRE**

**PRÉSENTÉ**

**COMME EXIGENCE PARTIELLE**

**DE LA MAÎTRISE EN SCIENCES ET TECHNOLOGIES DE L'INFORMATION**

**PAR**

**BELLOUSSAIEF SAFAA**

**MAI 2020**

Ce mémoire a été évalué par un jury composé des personnes suivantes :

Dr. Michael Korwin-Pawlowski ..... Président du jury

Dr. Ana-Maria Cretu ..... Membre du jury

Dr. Rokia Missaoui ..... Directrice de recherche

Mémoire accepté le : 06-30-2020

# Table des matières

Liste des figures	iii
Liste des tableaux	iv
Remerciements	v
Résumé	vi
<b>1 Introduction</b>	<b>1</b>
<b>2 Notions de base</b>	<b>3</b>
2.1 Réseau social . . . . .	4
2.2 Outils d'analyse des réseaux sociaux . . . . .	6
2.2.1 Mesures de centralité de Freeman . . . . .	6
2.2.2 Notion de robustesse . . . . .	12
2.3 Types de réseaux . . . . .	16
<b>3 État de l'art</b>	<b>18</b>
3.1 Élimination des nœuds clés . . . . .	18
3.2 Élimination des liens clés . . . . .	19
3.3 Élimination des liens et nœuds clés . . . . .	20
3.4 Détection des communautés . . . . .	22
<b>4 Approche proposée</b>	<b>23</b>
4.1 Méthode $M_1$ : déstabilisation par élimination des nœuds clés . . . . .	24
4.2 Méthode $M_2$ : déstabilisation par élimination des liens clés . . . . .	25
4.3 Méthode $M_3$ : déstabilisation par élimination des liens clés et leurs nœuds adjacents	26
4.4 Variante $M_{11}$ de la méthode 1 . . . . .	27
4.5 Variantes de la méthode 2 : . . . . .	30
4.5.1 Variante 1 de la méthode 2 : $M_{21}$ . . . . .	30
4.5.2 Variante 2 de la méthode 2 : $M_{22}$ . . . . .	32
4.5.3 Variante 3 de la méthode 2 : $M_{23}$ . . . . .	32
4.6 Variantes de la méthode 3 : . . . . .	34
4.6.1 Variante 1 de la méthode 3 : $M_{31}$ . . . . .	34

4.6.2	Variante 2 de la méthode 3 : $M_{32}$ . . . . .	36
4.6.3	Variante 3 de la méthode 3 : $M_{33}$ . . . . .	36
<b>5</b>	<b>Analyse empirique</b>	<b>38</b>
5.1	Illustration . . . . .	39
5.1.1	Premier exemple . . . . .	39
5.1.2	Deuxième exemple . . . . .	45
5.2	Étude approfondie . . . . .	49
5.2.1	Complexité . . . . .	49
5.2.2	Tests et résultats . . . . .	51
<b>6</b>	<b>Conclusion et travaux futurs</b>	<b>63</b>
6.1	Conclusion . . . . .	63
6.2	Travaux futurs . . . . .	64

# Liste des figures

2.1	Réseau $G$ orienté et pondéré . . . . .	5
4.1	graphe $G$ après détection des communautés . . . . .	24
5.1	Graphe $G$ après détection des communautés . . . . .	41
5.2	Les graphes obtenus par la première méthode et sa variante $M_{11}$ . . . . .	44
5.3	Graphe $G$ après détection des communautés . . . . .	47
5.4	Statistiques générales sur toutes les exécutions . . . . .	53
5.5	Statistiques sur les graphes réels . . . . .	57
5.6	Statistiques sur les graphes synthétiques . . . . .	58
5.7	Statistiques sur les graphes synthétiques . . . . .	59
5.8	Statistiques sur les graphes synthétiques . . . . .	60
5.9	Statistiques sur les graphes synthétique de densité supérieure à 50% . . . . .	61
5.10	Statistiques sur les graphes synthétiques de densité inférieure à 50% . . . . .	62

# Liste des tableaux

2.1	Mesures de centralité des nœuds du graphe G . . . . .	10
2.2	Centralité d'intermediarité des liens du graphe G . . . . .	12
5.1	Méthodes proposées et leurs variantes . . . . .	39
5.2	Mesures de centralité des nœuds du graphe G . . . . .	40
5.3	Centralité d'intermédiarité des liens du graphe G . . . . .	41
5.4	Mesures de centralité des nœuds au sein des communautés . . . . .	42
5.5	Centralité d'intermédiarité des liens au sein des communautés . . . . .	43
5.6	Statistiques sur les variantes exécutées . . . . .	44
5.7	Mesures de centralité des nœuds du graphe G . . . . .	45
5.8	Centralité d'intermédiarité des liens du graphe G . . . . .	46
5.9	Mesures de centralité des nœuds au sein des communautés . . . . .	47
5.10	Centralité d'intermédiarité des liens au sein des communautés . . . . .	48
5.11	Statistiques sur les variantes exécutées . . . . .	48
5.12	Temps d'exécution en secondes . . . . .	56

# Remerciements

Je tiens, avant de présenter ce travail, à exprimer ma reconnaissance envers les personnes qui m'ont, de près ou de loin, apporté leur soutien. Qu'ils trouvent ici collectivement et individuellement l'expression de toute ma gratitude.

En premier lieu, je remercie tous les membres du jury, pour m'avoir fait l'honneur de participer à l'évolution de ce travail.

Je voudrais adresser toute ma reconnaissance à ma directrice de recherche, Dr. Missaoui pour sa patience, sa disponibilité et surtout ses judicieux conseils et recommandations tout au long de cette recherche.

Aussi, je ne pourrais laisser passer cette occasion sans saluer chaleureusement tous mes enseignants de l'UQO.

Enfin, je remercie infiniment ma famille qui ne cesse de demeurer avec amour un soutien moral indéfectible ainsi que tous mes amis pour leurs conseils et leur encouragement.

# Résumé

L'analyse des réseaux sociaux vise à identifier des structures, des groupes et des modèles cachés au sein d'un réseau social. Elle couvre plusieurs thématiques de recherche comme la détection de communautés, le démantèlement de réseaux de criminels et de trafiquants ou au contraire le renforcement de réseaux de professionnels ou d'amis, l'identification d'acteurs importants, et l'étude de la fragilité ou de la robustesse d'un réseau.

Le démantèlement d'un réseau en tant qu'approche visant à désintégrer une structure connectée tient à affaiblir la cohésion entre les membres de cette structure. Ceci est réalisé soit par l'élimination de liens clés entre les membres pour la formation de composants déconnectés ou encore par l'élimination des membres les plus influents. De ce fait, la déstabilisation vise, entre autres, la réduction de la propagation du flux d'informations ou la régression de l'influence.

Le présent mémoire de maîtrise propose une nouvelle approche de démantèlement d'un réseau social unimodal (un seul type de nœuds) orienté et pondéré. Cette approche se décline en trois méthodes et sept variantes distinctes. Elle se base dans un premier lieu sur la détection des communautés pour ensuite effectuer une détection et une suppression des liens et/ou nœuds clés au sein de chaque communauté. Cette approche descendante de démantèlement d'un réseau part de l'hypothèse qu'elle est plus efficace qu'une méthode de détection de nœuds et liens centraux au niveau global. Nous explorons de ce fait la possibilité de différencier les liens inter et intra-communautaires et donc avoir la possibilité de cibler les liens/nœuds importants au sein des communautés détectées.

Les aspects étudiés englobent l'identification des mesures à utiliser pour la détection des liens et/ou nœuds clés globalement ou au sein des communautés ainsi que l'estimation du niveau de démantèlement par le choix de mesures de robustesse du réseau. La validation de l'approche est faite sur deux réseaux réels et trois types de réseaux : petit-monde, aléatoire et invariant d'échelle en vue de mesurer la performance et la qualité de l'approche proposée.

# Abstract

Social network analysis is a field of research that aims at identifying groups, patterns and structures hidden within a social network. It covers several research areas such as community detection, destabilizing criminal or dealer networks or on the contrary strengthening networks of professionals or friends, identifying key actors, and the study of the fragility or the robustness of a network.

The dismantling of a network as an approach to disintegrate a connected structure is to weaken the cohesion between the members of that structure. This is achieved either by removing key links between members to form disconnected components or by removing the most influential members. Therefore, destabilization aims, among other things, to reduce the propagation of the flow of information or the regression of influence.

This master's thesis proposes a new approach to dismantling an oriented and weighted social network of one mode data (i.e., a single type of nodes). This approach comes in three methods and seven distinct variants. It is based firstly on the detection of communities and then it performs a detection and removal of links and / or key nodes within each community. This top-down approach to dismantling a network assumes that it is more efficient than a method of detecting core nodes and links at the global level. We therefore explore the possibility of differentiating between inter and intra-community links and therefore have the possibility of targeting important links / nodes within the detected communities detected.

The aspects studied include the identification of measures to be used for the detection of key links and / or nodes globally or within the communities as well as the estimation of the level of dismantling by the choice of network robustness measures. The validation of the approach is done on two real networks and three types of networks : small-world, random and free-scale ones in order to measure the performance and the quality of the proposed approach.

# Chapitre 1

## Introduction

Les réseaux sociaux constituent de nos jours l'environnement d'évolution et d'échange d'information entre les utilisateurs. Ils permettent également de maintenir les données organisées et accessibles au grand public. À ce niveau, les études se fondent sur la recherche des métriques qui caractérisent un réseau social et leur application à des fins d'identification des caractéristiques du réseau dans sa globalité ainsi que celles des nœuds et liens pris individuellement.

Parmi les champs d'études de l'analyse des réseaux sociaux, nous citons à juste titre la détection et l'étude de l'évolution des communautés, la prédiction de l'évolution d'un réseau, la mesure de sa fragilité ou de sa robustesse et la détection des acteurs et liens centraux véhiculant des flux d'information.

Il y a au moins deux raisons pour l'identification des acteurs (nœuds) influents dans un réseau social : soit dans le but de déstabiliser sa structure et affaiblir sa cohésion par l'élimination de ces nœuds ou à des fins d'une meilleure circulation du flux d'information. Pour arriver à analyser le fonctionnement de telles structures souvent complexes et fortement connectées, il est nécessaire d'avoir, en plus d'une connaissance de diverses méthodes proposées dans la littérature, une approche visant à minimiser le coût et augmenter la performance du modèle dégagé.

Le présent mémoire de recherche vise à explorer une nouvelle approche de démantèlement d'un réseau social afin de réduire la cohésion du réseau et la circulation de l'information entre ses membres. Notre approche est descendante puisqu'elle consiste en premier lieu à identifier puis isoler les communautés pour ensuite éliminer les liens et nœuds les plus critiques au sein de chacune des communautés en vue d'accroître leur démantèlement. Une telle approche pourrait être

---

appliquée à la déstabilisation d'un réseau de criminels, de trafiquants, ou même dans des domaines tels que la sociologie, l'économie, la santé, le marketing, etc.

Voici ci-après des questions que l'on s'est posé tout au début de cette recherche :

- Quelles sont les mesures à appliquer pour cibler les nœuds et/ou liens clés ?
- Quelles sont les mesures à concevoir, utiliser ou adapter pour déterminer le niveau de dés-intégration du réseau suite à un retrait d'un nœud et/ou lien ?
- Est-il vraiment plus fiable d'appliquer une approche descendante qui consiste à détecter les communautés avant d'appliquer l'approche de démantèlement au sein du réseau ?

Dans le cadre de ce mémoire, notre approche de démantèlement traite les réseaux à un mode de données ayant des liens orientés et pondérés. La validation de l'approche se fait sur trois types de réseaux synthétiques, à savoir le réseau petit-monde, le réseau aléatoire et le réseau invariant d'échelle ainsi que sur des réseaux réels en vue d'estimer la performance de l'approche en terme de précision du processus de démantèlement.

Ce document est organisé comme suit : nous dédions le chapitre 2 au rappel des notions de base sur l'analyse des réseaux sociaux et qui sont utilisées dans notre étude. Le chapitre 3 englobe un survol de l'état de l'art relatif au démantèlement des réseaux sociaux. Quant au chapitre 4, il présente la démarche à suivre pour atteindre les objectifs fixés et fournit une série de procédures utilisant des variantes d'élimination de nœuds et/ou de liens au niveau global ou par communauté du réseau. Le chapitre 5 est dédié à une analyse empirique des sept variantes de notre approche sur des réseaux réels ou synthétiques de trois types. Le chapitre 6 résume notre contribution et énumère les travaux futurs.

# Chapitre 2

## Notions de base

L'analyse des réseaux sociaux (ARS) traite de la complexité de composition de ces derniers ainsi que de leurs structures cachées. En effet, des techniques se sont basées d'une part, sur l'aspect conceptuel (théorie des graphes) et d'autre part, sur l'analyse et la fouille de données. Elles sont dédiées principalement au regroupement des nœuds pour la formation de communautés, à la prédiction des liens, à l'évolution du réseau et à l'identification des nœuds et des liens centraux.

L'ARS exploite la théorie des graphes et bien d'autres théories d'analyse de données pour la modélisation des réseaux et le traitement des données. Elle vise à cibler les liens et/ou les nœuds les plus centraux, à détecter les sous-graphes et à prédire les nouveaux liens qui peuvent être établis dans le futur sur la base de sa structure et ses changements dans le passé. Un réseau englobe des nœuds représentant des entités et des liens entre ces nœuds. Ces nœuds et liens peuvent comporter un ou plusieurs types. C'est à ce niveau qu'on peut classer les réseaux selon le mode de données en réseaux unimodaux (à un seul mode), bimodaux ou multimodaux. Quant au premier type, il ne comporte qu'un seul type de nœuds et de liens que ceux-ci entretiennent entre eux. Pour le réseau bimodal, il englobe deux types de nœuds comme les employés et leur lieu de travail. Comme indiqué auparavant, nous traitons des réseaux unimodaux ayant des liens orientés et pondérés.

Dans cette section, nous allons dans un premier temps définir les notions relatives aux réseaux sociaux ainsi que quelques formes de graphes. Ensuite, nous allons rappeler quelques mesures de centralité communément utilisées pour identifier l'importance des nœuds et des liens.

---

## 2.1 Réseau social

Un réseau social est un ensemble d'unités sociales (entités, nœuds ou acteurs) et des relations (liens ou arcs) que ces unités ont directement ou indirectement à travers des chemins relationnels de longueurs variables. On peut introduire à ce niveau la notion de réseau simple versus complexe. On peut différencier les deux types en terme de complexité, évolutivité ou dynamique. Un graphe est dit simple quand il remplit deux conditions, à savoir l'absence de boucle (lien d'un sommet vers lui-même) et l'absence de liens multiples entre deux nœuds.

Un réseau social est généralement représenté sous forme de graphe ou de matrice. Ces deux formats servent d'entrée à des méthodes d'analyse. Il existe différents types de graphes comme indiqué ci-après.

### Graphe non orienté et non valué

Soit un graphe  $G = (V, E)$  où  $V$  est un ensemble de sommets dans  $G$  et  $E$  est un ensemble de ses arêtes. Chaque arête liant une paire de nœuds indique la présence d'un lien entre les nœuds de cette paire sans en spécifier l'orientation. Un tel graphe est aussi appelé simple (ou à un seul mode de données). Dans ce cas, le graphe ne comprend ni boucle (lien d'un sommet vers lui-même) ni liens multiples. Pour des raisons de visualisation et de traitement de données, on recourt à la représentation de ce graphe sous forme de matrice ou de liste d'adjacence.

Pour un réseau unimodal, la matrice d'adjacence est carrée symétrique où ses composants prennent la valeur 0 ou 1 traduisant respectivement l'absence ou la présence de lien entre une paire de nœuds d'un seul type.

La liste d'adjacence, quant à elle, est une structure de données modélisant la liste de voisins de chaque sommet. C'est une matrice à deux colonnes où la première indique le sommet en question et la seconde présente l'ensemble de ses voisins.

### Graphe orienté et pondéré

Dans le cas où la circulation d'information varie par rapport au sens, les nœuds varient eux aussi entre transmetteurs et récepteurs. Les graphes orientés peuvent être représentés par une matrice asymétrique dans laquelle le sens d'un arc indique le sens de la communication. Dans le cas d'un graphe où on attribue une valeur de pondération positive ou négative à une arête existant entre deux nœuds, le graphe est dit valué ou pondéré. Ceci peut traduire la fréquence d'interaction (en matière

de temps ou de quantité de messages partagés) entre les nœuds ou l'intensité des relations sociales. Nous produisons en figure 2.1 un exemple de graphe orienté et pondéré qui va servir à illustrer notre approche.

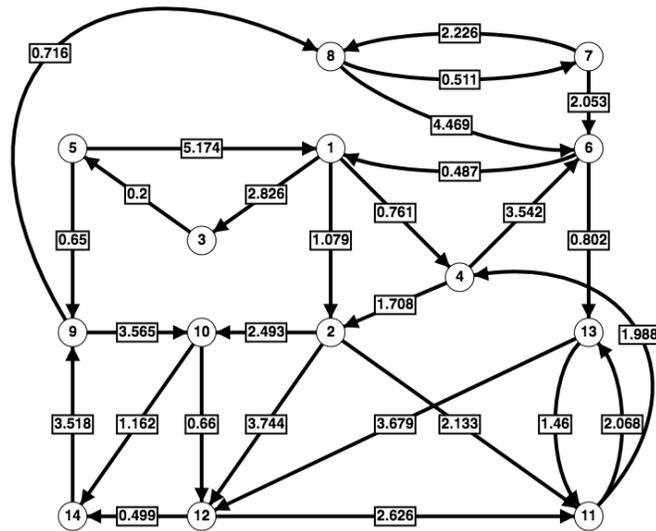


FIGURE 2.1 – Réseau  $G$  orienté et pondéré

Voici ci-dessous la matrice d'adjacence du graphe  $G$ .

$$\begin{pmatrix}
 0 & 1.079 & 2.286 & 0.760 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.493 & 2.132 & 3.743 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1.707 & 0 & 0 & 0 & 3.541 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 5.173 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.649 & 0 & 0 & 0 & 0 & 0 \\
 0.487 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.800 & 0 \\
 0 & 0 & 0 & 0 & 0 & 2.051 & 0 & 2.226 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 4.469 & 0.511 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.716 & 0 & 3.564 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.659 & 0 & 1.162 \\
 0 & 0 & 0 & 1.988 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.068 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.626 & 0 & 0 & 0.498 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.460 & 3.679 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3.517 & 0 & 0 & 0 & 0
 \end{pmatrix}$$

## Chemin et circuit

Un chemin  $\mu[x, y]$  dans un graphe orienté est une alternance de sommets et d'arcs consécutifs et sans redondance au niveau des arcs reliant la paire de nœuds  $(x, y)$ . Dans le cas d'un graphe orienté, *le circuit* représente une suite d'arcs consécutifs (chemins) dont les deux sommets extrêmes sont identiques. La notion correspondante dans les graphes non orientés est celle de *cycle*.

## Sous-graphe

Un graphe  $G' = (V', E')$  est dit sous-graphe de graphe  $G = (V, E)$  si et seulement si :

$$V' \subseteq V \quad \text{et} \quad E' \subseteq \{(x, y) \in E \mid x \in V' \wedge y \in V'\} \quad (2.1)$$

Il s'agit d'un sous-graphe couvrant lorsque  $V' = V$  et  $E' \subseteq E$ . Ainsi, tout graphe simple à  $n$  sommets est un sous-graphe couvrant du graphe complet  $K_n$  dans lequel il y a un lien entre toute paire de sommets. On pourra ainsi isoler et étudier des sous-réseaux en tenant compte de la position structurale des acteurs (mesures de centralité) et de la topologie du réseau.

## 2.2 Outils d'analyse des réseaux sociaux

### 2.2.1 Mesures de centralité de Freeman

#### Centralité de degré

La centralité de degré d'un nœud mesure son importance par le nombre de voisins auquel il est directement connecté. Elle est influencée par la nature du graphe (orienté ou non, pondéré ou non) et informe sur l'activité du nœud dans un réseau orienté en calculant le rapport entre le nombre de liens sortants du nœud et le degré maximal possible.

Un nœud ayant une centralité de degré élevée occupe une position avantageuse dans le réseau en terme de connectivité avec les autres nœuds, ce qui lui permet d'avoir un rôle clé dans la propagation de l'information [2].

Freeman [11] a introduit la centralité de degré d'un nœud  $i$  comme le nombre de sommets qui lui sont adjacents :

$$d_i = \sum_{j=1}^n A_{ij} \quad (2.2)$$

où  $n$  est le nombre total de nœuds,  $A_{ij}$  désigne le lien entre  $i$  et  $j$  qui vaut 0 ou 1.

En vue de tenir compte de la taille du réseau, cette mesure est normalisée selon l'équation ci-après :

$$D'_i = \frac{\sum_{j=1}^n A_{ij}}{n-1} \quad (2.3)$$

### Centralité d'intermédiarité

Cette centralité mesure le nombre de fois où un nœud sert d'intermédiaire dans les courts chemins (appelés distances géodésiques) liant deux autres nœuds. Il s'agit donc du degré de médiation d'un membre de réseau du point de vue du transfert d'information entre deux autres membres. Cette mesure paraît particulièrement appropriée dans l'étude de l'importance des nœuds agissant souvent comme des intermédiaires [11] et ayant un contrôle sur les données circulant entre les différents groupes de nœuds du réseau.

Ainsi, les sommets ayant une forte probabilité d'apparaître sur les chemins les plus courts entre deux nœuds ont une forte intermédiarité. Cette centralité est calculée pour un nœud  $i$  comme suit :

1. Déterminer pour chaque couple de nœuds  $(j, k)$  les plus courts chemins qui les relient.
2. Pour chaque paire  $(j, k)$ , on détermine la portion des plus courts chemins où le nœud  $i$  apparaît.
3. On effectue la sommation de ces portions pour toutes les paires  $(j, k)$  impliquant le nœud  $i$ .

Elle est introduite selon Freeman [11] comme suit :

$$B_i = \sum_j^n \sum_k^n b_{jk}(i) \quad (2.4)$$

Où  $n$  est le nombre total de nœuds du graphe, avec  $i \neq j \neq k$ ,  $b_{jk}(i)$  est une portion d'intermédiarité du nœud  $i$  associée aux plus courts chemins existant entre  $j$  et  $k$  avec  $b_{jk}(i) = 1$  si  $i$  intervient dans le chemin entre  $j$  et  $k$  et 0 sinon.

On introduit la centralité d'intermédiarité relative [11], variante entre 0 et 1 et prenant en compte la taille de réseau comme suit :

$$B'_i = \frac{2B_i}{n^2 - 3n + 2} \quad (2.5)$$

Où  $(n^2 - 3n + 2)/2$  représente la centralité maximale d'un nœud.

### Centralité de proximité

Elle est définie par le calcul de la distance qui sépare un nœud des autres nœuds du graphe. Dans un réseau social, cette mesure exprime la notion d'effort : un acteur ayant une grande valeur de centralité de proximité est capable de contacter la majorité des acteurs du réseau avec un minimum d'effort, c'est-à-dire avec un plus faible court chemin.

Cette centralité est donc calculée par la sommation des distances géodésiques entre un nœud  $i$  et tous les autres membres du réseau selon l'équation 2.6 :

$$P_i = \frac{1}{\sum_{j=1}^n d_{ij}} \quad (2.6)$$

Où  $d_{ij}$  représente la distance géodésique séparant  $i$  de  $j$ . À des fins de raffinement de cette mesure et dans le but de tenir compte de la taille du réseau, on introduit la centralité de proximité relative définie par l'équation 2.7 :

$$P'_i = \frac{n-1}{\sum_{j=1}^n d_{ij}} \quad (2.7)$$

## Autres mesures de centralité

### Centralité de vecteur propre

Cette mesure donne autant d'importance à l'acteur qu'il entretient des liens avec d'autres acteurs du réseau qui sont eux-mêmes importants. La qualification d'importance d'un acteur est indiquée par son poids. Cette centralité pour un nœud  $i$  est déterminée par la centralité des nœuds auxquels il est directement connecté. Elle est définie par Bonacich [4] comme une combinaison linéaire des

centralités des nœuds voisins immédiats du nœud  $i$  définie par l'équation 2.8 :

$$e_i = \frac{1}{\lambda} \sum_{j=1}^n A_{ij} e_j \quad (2.8)$$

où  $\lambda$  est la plus grande valeur du vecteur propre,  $A_{ij}$  est un élément de la matrice d'adjacence et  $e_j$  est la centralité de vecteur propre du nœud  $j$ .

Dans le cas où le réseau est orienté et face à la présence d'un nœud sans degré entrant (i.e., sans flèche vers ce nœud), la valeur de vecteur propre vaut 0 et se propage à tous les voisins directs et indirects. La centralité vaut donc 0 pour tous les nœuds du réseau. Cette lacune fait que cette mesure est inappropriée pour les réseaux orientés.

### Centralité de Katz

Proposée par Katz [17], cette centralité calcule l'influence relative d'un nœud au sein d'un réseau en mesurant non seulement le nombre de voisins immédiats mais aussi tous les autres nœuds du réseau connectés avec le nœud en question à travers ses voisins immédiats tout en pénalisant la contribution des nœuds éloignés. Elle est dédiée au graphe orienté et définie par :

$$k_i = \alpha \sum_{j=1}^n A_{ij} k_j + \beta \quad (2.9)$$

où  $\alpha$  est une constante appartenant à l'intervalle  $]0, 1/\lambda]$  et  $\lambda$  est la plus grande valeur de vecteur propre. En cas de présence des nœuds n'ayant pas de degré entrant, on assigne la valeur 1 à la constante  $\beta$  [21], ce qui palie à la lacune de la centralité de vecteur propre.

### Centralité de PageRank

Une variante de centralité de Katz, la centralité de *PageRank* surmonte une caractéristique indésirable de la première centralité, à savoir l'impossibilité de différencier la centralité d'un nœud important de ses voisins immédiats. Ceci est rectifié en partitionnant la centralité du nœud considéré important et en propageant une portion de sa centralité à ses voisins. Cette centralité pour le nœud  $i$  est définie [21] par l'équation 2.10 :

$$R_i = \alpha \sum_{j=1}^n A_{ij} \frac{R_j}{d_j^{out}} + \beta \quad (2.10)$$

Où  $R_j$  est la centralité *PageRank* du nœud  $j$  et  $d_j^{out}$  est le degré sortant (nombre de sommets recevant une flèche) de ce même nœud. Dans le cas où le nœud  $j$  n'a aucun lien sortant,  $d_j^{out}$  prend la valeur 1.

Le tableau 2.2 donne quelques mesures de centralité des nœuds du graphe  $G$  de la figure 2.1. On remarque que la centralité de degré n'est pas prise en compte car nous pensons que la centralité de proximité et celle d'intermédiarité sont bien plus importantes. En outre, le lecteur peut constater que les nœuds 6, 1 et 9 sont centraux dans le réseau puisque la moyenne des mesures est relativement la plus élevée. Ce tableau indique les centralités d'intermédiarité (*betweenness*), de *PageRank*, de proximité (*proximity*), et de la moyenne entre les trois premières mesures de centralité. Quant à la dernière mesure qui est *ProxiRank* que nous définissons comme la moyenne entre la centralité de proximité et de *PageRank*.

Id	Intermédiarité	PageRank	Proximité	Moyenne	ProxiRank
<b>6</b>	0.543	0.080	0.565	<b>0.396</b>	0.322
<b>1</b>	0.423	0.076	0.590	<b>0.363</b>	0.333
<b>9</b>	0.472	0.072	0.520	<b>0.355</b>	0.296
<b>12</b>	0.302	0.118	0.520	<b>0.313</b>	0.319
<b>8</b>	0.412	0.0268	0.481	<b>0.306</b>	0.254
<b>4</b>	0.269	0.076	0.541	<b>0.295</b>	0.309
11	0.247	0.132	0.481	0.287	0.306
7	0.412	0.013	0.419	0.281	0.216
13	0.225	0.110	0.500	0.278	0.305
2	0.214	0.047	0.565	0.275	0.306
10	0.230	0.074	0.520	0.274	0.297
5	0.285	0.0535	0.481	0.273	0.267
14	0.258	0.066	0.464	0.263	0.265
3	0.285	0.0503	0.419	0.251	0.234

TABLE 2.1 – Mesures de centralité des nœuds du graphe G

## Autres métriques

### Densité

Elle détermine la probabilité d'existence d'un lien au sein d'un réseau entre deux nœuds choisis aléatoirement et permet de définir la cohésion. Cette mesure peut être utilisée dans l'optique d'une

---

analyse socio-centrée. Une telle analyse consiste à mesurer la densité des liens autour d'un nœud donné et sert à dégager l'influence du nœud analysé sur la densité du sous-graphe auquel il appartient. Elle est définie par le rapport entre le nombre d'arêtes divisé par le nombre d'arêtes possibles donnée par  $n(n - 1)/2$  où  $n$  est le nombre total de nœuds.

### Degré moyen

C'est la moyenne de la centralité de degré définie comme suit :

$$\bar{d}(G) = \frac{\sum_{i \in n} d_i}{n} \quad (2.11)$$

Où  $n$  est le nombre de sommets de graphe  $G$  et  $d_i$  désigne le degré du nœud  $i$ . Le degré moyen est aussi calculé à travers la densité selon la formule ci-après :

$$\bar{d}(G) = Densite \times (n - 1) \quad (2.12)$$

### Centralité des liens

Diverses centralités mesurent l'importance des nœuds. Seule la centralité d'intermédiarité peut cibler les liens clés. Soit un lien  $l$  où sa centralité est calculée par le nombre des courts chemins passant par  $l$ . Cela se traduit par le nombre de fois où le lien  $l$  sert d'intermédiaire entre des nœuds. Le tableau ci-dessous illustre cette centralité calculée pour les liens du graphe  $G$  de la figure 2.1 :

Id	de	vers	intermédierité
<b>12</b>	6	1	<b>0.351</b>
<b>17</b>	8	7	<b>0.340</b>
<b>14</b>	7	6	<b>0.335</b>
<b>18</b>	9	8	<b>0.335</b>
<b>2</b>	1	3	<b>0.214</b>
.	.	.	.
15	7	8	0.005
6	2	12	0.000
10	5	1	0.000
16	8	6	0.000
21	10	14	0.000

TABLE 2.2 – Centralité d'intermediarité des liens du graphe G

De même, la détection des liens clés peut se faire à travers la mesure de l'impact de suppression d'un lien spécifique. En d'autres termes, on calcule l'efficacité du réseau avant et après le retrait du lien  $l$ . Cette variation est donnée par la formule suivante :

$$\Delta E = E(G) - E(G - l) \quad (2.13)$$

où  $E(G)$  et  $E(G - l)$  sont respectivement l'efficacité du réseau avant et après le retrait du lien  $l$ . Cette formule a été exploitée par plusieurs chercheurs [20, 25] dans la recherche de nœuds et de liens clés. Puisque l'efficacité du réseau est intimement liée à sa capacité de transmettre l'information, celle-ci peut être estimée par les formules de connectivité ou de compacité décrites ci-après.

### 2.2.2 Notion de robustesse

Le calcul de la robustesse d'un réseau est important pour mesurer sa performance suite à une attaque ou une panne. Un réseau est dit robuste si le flux d'information existant dans les circonstances normales trouve des chemins alternatifs en cas de panne ou d'attaque. L'estimation de la fragilité ou de la robustesse d'un réseau constitue l'un des champs d'étude des réseaux sociaux à des fins d'optimisation ou déstabilisation. Une présentation de ces mesures de robustesse est faite dans le but d'identifier celle qui permet de mieux différencier le niveau de robustesse des réseaux suite à un retrait de nœuds et/ou de liens. C'est à ce niveau que Ellens et Kooij [10] ont spécifié

diverses mesures de robustesse de réseau que nous citons ci-après pour ensuite les exploiter dans notre solution.

### Connectivité des nœuds

Dans le cas d'un graphe incomplet, la mesure  $k_v$  calcule le nombre maximal de nœuds à supprimer pour déconnecter le réseau. D'une manière duale, la connectivité de liens  $k_e$  calcule le nombre d'arêtes à supprimer pour démanteler un réseau.

Selon [9], on estime que  $k_v \leq k_e \leq \delta_{min}$  où  $\delta_{min}$  est le degré minimal des nœuds. Dans le cas d'un graphe complet  $K_n$ , on estime que  $k_v = n-1$ . Ceci vient du fait que plus la connectivité des arêtes ou des nœuds est importante, plus le réseau est robuste.

### Distance

Nous rappelons que la distance géodésique  $d_{ij}$  entre deux nœuds  $i$  et  $j$  est définie par la longueur du chemin le plus court entre  $i$  et  $j$ . Diverses mesures de robustesse sont basées sur la distance géodésique comme :

- **Le Diamètre**  $d$  qui représente la valeur maximale des distances géodésiques  $d_{ij}$ , soit  $d = \max(d_{ij})$ .
- **La distance moyenne** qui est la moyenne des distances géodésiques de toutes les paires de nœuds :

$$\bar{d} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n d_{ij} \quad (2.14)$$

Si  $\bar{d} = 1$ , cela signifie que tous les nœuds sont directement liés et donc le réseau est robuste.

- **L'efficacité** qui est utilisée pour les réseaux non connectés et donc à haut risque. Comme  $\bar{d}$ , cette mesure ne considère pas les chemins alternatifs et est définie comme suit :

$$\bar{d} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n \frac{1}{d_{ij}} \quad (2.15)$$

### Intermédiation

Elle indique le nombre de chemins les plus courts entre les paires de nœuds passant par un sommet ou arête  $x$ .

S'il existe plus d'un chemin plus court entre deux sommets, alors chacun des  $k$  chemins est compté

$1/k$  fois.

$$b_x = \sum_{i=1}^n \sum_{j=i+1}^n \frac{n_{ij}(x)}{n_{ij}} \quad (2.16)$$

où  $n_{ij}(x)$  est le nombre de chemins les plus courts entre  $i$  et  $j$  passant par  $x$ , et  $n_{ij}$  est le nombre total des chemins les plus courts entre les nœuds  $i$  et  $j$ . Basée seulement sur les chemins les plus courts, l'intermédiarité n'intègre pas les chemins alternatifs. Pour exprimer la robustesse du réseau dans sa globalité, des chercheurs ont introduit la moyenne d'intermédiarité de sommets et celles d'arêtes données respectivement comme suit :

$$\bar{b}_v = \frac{1}{2}(n-1)(\bar{d}+1) \quad (2.17)$$

$$\bar{b}_e = \frac{n(n-1)\bar{d}}{2m} \quad (2.18)$$

où  $n$  et  $m$  sont respectivement le nombre des sommets et d'arêtes du graphe  $G$ . Plus l'intermédiarité moyenne du réseau est petite, plus cette structure est robuste.

### Coefficient de regroupement (*clustering*)

Cette mesure donne la valeur moyenne des coefficients de regroupement locaux dans un réseau à  $n$  nœuds. Soit  $d_i$  le nombre des nœuds voisins de  $i$  et  $d_i(d_i-1)/2$  le nombre total des liens pouvant exister entre les voisins du nœud  $i$ . Soit  $c_i$  le coefficient de regroupement défini par le nombre  $e$  de liens des voisins de  $i$  divisé par le nombre possible de liens de ses voisins. Cette mesure réfère donc à la présence de relation triangulaire entre deux paires de nœuds voisins qui sont en rapport avec le nœud  $i$  en question. Elle est définie par :

$$\begin{aligned} C &= \frac{1}{n} \sum_{i \in V; d_i > 1} c_i \\ &= \frac{1}{n} \sum_{i \in V; d_i > 1} \frac{2}{d_i(d_i-1)} e_i \\ &= \frac{1}{n} \sum_{i \in V; d_i > 1} \frac{1}{d_i(d_i-1)} \sum_{j=1}^n \sum_{k=1}^n a_{ij} a_{jk} a_{ki} \\ &= \frac{1}{n} \sum_{i \in V; d_i > 1} \frac{1}{d_i(d_i-1)} (A^3)_{ii} \end{aligned} \quad (2.19)$$

où  $e_v$  est le nombre de liens parmi les voisins de  $v$ ,  $a_{ij}$  est l'élément de la ligne  $i$  et la colonne  $j$  de la matrice d'adjacence  $A$  ayant la valeur 1 en cas de présence de lien entre  $i$  et  $j$  et 0 sinon. Plus le coefficient de regroupement est important, plus le réseau est robuste.

## Performance totale

Lindelauf et al. [19] ont défini la performance totale du réseau comme étant le produit de sa performance moyenne  $I_g$  et de sa performance de secret (*secrecy efficiency*)  $S_g$ .

$$P_g = S_g \times I_g \quad (2.20)$$

dont les mesures  $I_g$  et  $S_g$  sont définies respectivement comme suit :

$$I_g = \frac{n \times (n - 1)}{T_g} \quad (2.21)$$

où  $T_g$  représente la distance totale du réseau définie comme la somme des distances géodésiques.

$$S_g = \sum_{i=1}^n \alpha_i \times \mu_i \quad (2.22)$$

Avec  $\alpha_i = \frac{1}{n}$  traduisant la probabilité qu'un nœud  $i$  soit découvert en tant que membre du réseau,  $\mu_i = 1 - \frac{d_i + 1}{n}$  où  $d_i$  est la centralité de degré du nœud  $i$  et  $\mu_i$  représente la fraction du réseau qui reste secrète après la détection du nœud  $i$ .

Adaptée seulement pour les réseaux connectés, la performance totale n'est pas fiable lorsque le réseau comporte plusieurs composants déconnectés. Certains courts chemins demeurent infinis entre des nœuds appartenant à des composants disjoints. Pour palier à ce problème entraînant la présence de valeurs infinies, la distance totale du réseau  $T_g$  est calculée par la distance géodesique maximale du réseau multipliée par  $(n - 1)$ .

## Taux de connectivité

Adaptée par Kong [18], le taux de connectivité (TC) traduit le rapport entre la connectivité du réseau suite au retrait d'un nombre de nœuds et la connectivité initiale. C'est de ce fait que  $TC^0 = TC_{max} = 1$  à l'itération 0 et que  $TC_{min} = 0$  qui est la valeur de connectivité du réseau déconnecté suite à la suppression de tous ses nœuds ou encore si la distance géodésique séparant les nœuds

restants tend vers  $\infty$ . Elle est définie à l'itération  $n_1$  comme suit :

$$TC = \frac{\sum_{i=1}^{n_1} C_i}{p * \sum_{i=1}^n C_i} \quad (2.23)$$

Ici,  $n_1$  est le nombre de nœuds restants dans le réseau après une suppression et  $p$  est le nombre de composants déconnectés à l'itération  $n_1$ . La compacité  $C_i$  d'un nœud  $i$  est la somme des compacités entre ce dernier et les autres nœuds du réseau initial. Elle correspond à la somme de l'inverse des distances géodésiques  $d_{ij}$  entre  $i$  et  $j$  donnée par l'équation suivante :

$$C_i = \sum_{i=1}^n \frac{1}{d_{ij}} \quad (2.24)$$

### La compacité

Elle renseigne sur la portion des nœuds qui peuvent se joindre via un chemin. C'est donc un outil de test de la cohésion du réseau.

$$C = \frac{\sum_{i=1}^n C_i}{n(n-1)} \quad (2.25)$$

La lacune de cette formule réside dans le fait qu'elle ne s'applique pas à tous les types de réseau. En effet, les réseaux non connectés présentent des nœuds appartenant à des composants différents et donc leur distance géodésique est indéfinie. Pour pallier à ce problème, on estime que cette distance est infinie et donc la compacité est donnée par  $\frac{1}{\infty}=0$ .

## 2.3 Types de réseaux

Nous validons notre approche sur des réseaux unimodaux orientés et pondérés en testant les méthodes proposées sur trois types de réseaux synthétiques ayant des caractéristiques distinctes, à savoir : le réseau aléatoire (*random network*), le réseau petit-monde (*small-world network*) et le réseau invariant d'échelle (*scale-free network*) ainsi que sur des réseaux réels.

Un réseau aléatoire est un réseau créé par un seul nœud isolé et des arêtes ajoutées entre des paires de nœuds sélectionnés uniformément de manière aléatoire. Le réseau aléatoire le plus courant est

---

$SN_{NP}$  dans lequel chaque lien possible entre les nœuds  $i$  et  $j$  est présent avec une probabilité indépendante  $p$ . Dans un tel modèle, tous les graphes avec  $n$  sommets et  $m$  arêtes ont la probabilité  $(1 - p)^{R-m}$  où  $R = \frac{1}{2}n(n - 1)$ . Vu comme une distribution de probabilité, le réseau aléatoire est composé d'une succession de marches aléatoires. Selon Hopkins et al. [16], il est caractérisé par une distance géodésique et un coefficient de regroupement moyennement faibles.

Dans les réseaux sociaux, les nœuds sont généralement connectés à d'autres nœuds via de très courts chemins. Ce phénomène est connu sous le nom de petit monde. Il présente les deux caractéristiques suivantes :

- la distance moyenne entre toute paire de sommets est faible.
- le niveau de regroupement local est élevé. Cela veut dire que les sommets sont généralement très connectés à leurs voisins immédiats. Le réseau est de ce fait vulnérable aux attaques par l'intermédiarité.

Finalement, le réseau invariant d'échelle est caractérisé par la croissance et l'attachement préférentiel. La croissance réfère au nombre de nœuds augmentant la taille du réseau dans le temps tandis que l'attachement préférentiel signifie que les nouveaux nœuds ont tendance à se connecter aux sommets ayant un grand degré. Selon Hidalgo et Barabasi [15], c'est un réseau dont les degrés suivent une loi de puissance :  $D_i \approx i^\gamma$  où  $\gamma$  est un paramètre situé en général entre 2 et 3. Ce type de réseau demeure robuste face aux attaques aléatoires. Cependant, il est vulnérable aux attaques par l'intermédiarité.

# Chapitre 3

## État de l'art

L'analyse des réseaux sociaux est conduite à des fins de découverte de relations et de structures cachées, d'identification d'acteurs clés, de prédiction d'évolution des réseaux et d'établissement de nouveaux liens [26].

Le démantèlement d'un réseau se fait à travers diverses approches qui visent toutes la régression du flux d'information, l'affaiblissement des nœuds et/ou liens clés dans le réseau et la dégradation de la robustesse du réseau par l'affaiblissement de son efficacité et de la cohésion au sein des ses communautés.

Nous allons classer ces approches en trois catégories : celles qui éliminent les nœuds importants, celles qui éliminent les liens clés et finalement celles qui visent à la fois les nœuds et leurs liens adjacents.

### 3.1 Élimination des nœuds clés

Des chercheurs ont mené des études sur le rôle des nœuds clés et l'impact de leur suppression sur le réseau. Ils se sont basés principalement sur des mesures de centralité pouvant identifier les membres importants d'un réseau.

En se servant de la centralité de degré et de vecteur propre, Memon et Larsen [20], dans leur approche de désintégration d'un réseau, ont étudié un réseau non orienté et non pondéré. Afin de détecter les éléments importants, ils ont eu recours à la hiérarchisation du graphe transformé en premier lieu en un réseau orienté. Les nœuds situés près de la racine de la hiérarchie ont plus tendance à gérer le flux d'information et donc se comporter comme des chefs de sous-groupes.

Une optimisation de l'algorithme de génération de l'arbre hiérarchique de Memon et Larsen a été menée par Chaurasia et Tiwari [6] qui ont fait appel à la centralité de Katz et *PageRank* dans la

construction de l'arbre pour un graphe non orienté et non pondéré. Les nœuds clés sont situés près de la racine de l'arbre lequel est donc généré sans se soucier de la direction des liens entre les nœuds.

Une autre approche basée sur l'indice de rôle ou de position (PRI : *Position Role Index*) a été proposée par Memon et Larsen [20]. Cet indice a pour rôle de détecter les meneurs dans un réseau en comparant l'efficacité de ce réseau avant et après la suppression d'un nœud. Plus cet écart est important, plus le nœud est une cible dans la déstabilisation du réseau. Sinon, il est considéré suiveur. On définit l'écart d'efficacité du réseau comme suit :

$$\Delta E = E(G) - E(G - n_i) \quad (3.1)$$

où  $n_i$  est le nœud du réseau  $G$  pour lequel on estime son importance.

Ces mêmes auteurs ont défini la centralité de dépendance pour mesurer le degré de dépendance d'un nœud par rapport à un autre nœud dans un graphe non orienté et non pondéré. Elle est définie comme suit :

$$CD_{ij} = \sum_{i \neq k, k \in G} \frac{d_{ij}}{N_k} + \Omega \quad (3.2)$$

Cette mesure indique la fréquence à laquelle le nœud  $i$  a besoin du nœud  $j$  pour communiquer avec d'autres nœuds du réseau.  $N_k$  est le nombre de chemins géodésiques liant  $i$  à  $k$  en passant par  $j$ ,  $d_{ij}$  est la distance géodésique entre la paire de nœuds  $i$  et  $j$  et la variable  $\Omega$  indique si un réseau est connecté ou non (1 ou 0).

Une centralité de dépendance faible indique ainsi la présence d'un nœud clé. Ceci est dû au fait que ce dernier, dans sa relation avec les autres nœuds du réseau, est quasiment indépendant de la présence d'intermédiaires dans sa communication avec le reste de réseau.

Berzinji et al. [2] ont fondé leur classification des nœuds d'un réseau sur trois principales mesures de centralité, à savoir la centralité de degré, la centralité d'intermédialité et la centralité de proximité. Les nœuds ayant des valeurs importantes de ces trois mesures sont considérés importants.

## 3.2 Élimination des liens clés

Han et al. [14] ont recours à l'analyse des réseaux sociaux en évaluant l'importance des liens selon leur efficacité. En effet, ils les catégorisent selon leur impact sur le réseau en :

- liens influents qui connectent les membres gérant les groupes et servant dans la propagation de l'information.
- liens à influence moyenne où les informations basiques sont transférées.
- liens à faible influence où des membres possèdent des communications occasionnelles avec d'autres membres.

Ils ont aussi discuté de l'aspect caché des liens qui peuvent être influents mais à cause de la faible fréquence de leur activation, ils paraissent faibles.

Wiil et al. [25], quant à eux, ont mis l'accent sur l'importance des liens pouvant renseigner sur les nœuds critiques du réseau. Ils ont introduit une valeur d'importance de lien définie par :

$$LI = p_l * w_l \quad (3.3)$$

où le poids  $w_l$  du lien  $l$  et  $p_l$  sa performance sont respectivement définis par :

$$w_l = \frac{\sum b_G}{\sum b_G - b_l} \quad (3.4)$$

avec  $\sum b_G$  représentant la centralité d'intermédierité totale des liens du graphe  $G$  et  $b_l$  décrivant la centralité d'intermédierité du lien  $l$ . Le poids d'un lien réfère donc à sa position dans le graphe.

$$p_l = \Delta P = P(G) - P(G - l) \quad (3.5)$$

où  $P(G)$  et  $P(G - l)$  sont respectivement la performance du réseau avant et après l'élimination du lien  $l$ .

### 3.3 Élimination des liens et nœuds clés

Dans un manuscrit, Sarr et al. [23] proposent une approche de démantèlement d'un réseau social en s'appuyant sur l'effet cascade. La méthode comporte deux phases : l'identification et l'élimination d'un nombre réduit de nœuds et arêtes cibles pour ensuite appliquer l'approche en cascade qui consiste à propager l'élimination à d'autres nœuds et liens jusqu'à l'atteinte d'un niveau prédéfini de désintégration.

L'optimisation consiste dans un premier lieu à choisir un nombre minimal de nœuds clés appelés déclencheurs. Cela a une influence sur les nœuds voisins des nœuds identifiés et peut entraîner leur suppression dans le but de réduire la cohésion par la formation de composants déconnectés.

Cibler les nœuds centraux qui sont les plus aptes à déclencher l'effet cascade présente donc la

---

première étape du processus. Cela permet la dislocation rapide du réseau à travers la détection des nœuds ayant une centralité de degré élevée (position sociale).

Kong [18] propose une variante de déstabilisation d'un réseau orienté et pondéré qui élimine les liens clés ainsi que les nœuds se trouvant aux extrémités. D'autres variantes d'une nouvelle approche de désintégration d'un réseau social sont appliquées soit par l'élimination des nœuds clés et les liens pendants, soit par la suppression des liens clés avec leurs nœuds adjacents. L'efficacité versus la dégradation d'un réseau est calculée moyennant une nouvelle formule d'efficacité, appelée le taux de connectivité, et présentée par la formule 2.23 dans le chapitre précédent. L'importance d'un nœud est calculée selon trois formules distinctes de centralité : la centralité de proximité, la mesure de *PageRank* et une moyenne de ces deux mesures *ProxiRank*. L'importance d'un lien est calculée à l'aide de sa centralité d'intermédierité.

Les résultats d'analyse de performance [18] indiquent que la méthode *ProxiRank* qui supprime les nœuds en commençant par ceux ayant une valeur élevée pour la moyenne des centralités de proximité et *PageRank* pourrait être avantageuse pour les réseaux petit-monde de petite taille. Par contre, lorsque le réseau comporte un grand nombre de nœuds et de liens, le choix dépend de l'objectif de la déstabilisation. Si l'on veut obtenir une déstabilisation optimale en retirant un nombre minimal de nœuds, cette méthode pourrait être également privilégiée. Toutefois, si le temps de calcul a une grande importance, la méthode lien-clé qui consiste à éliminer les liens importants incluant les nœuds adjacents sans aucune identification préalable des communautés est la plus fiable. Elle s'est avérée être la plus rapide méthode en cours de désintégration d'un réseau. Cependant, la faiblesse de cette méthode est le nombre élevé de sommets et d'arêtes à supprimer.

Une autre approche vise à collecter les nœuds potentiels dans le but de les retirer du réseau à des fins de raffinement. Tiwari et al. [24] ont fondé leur recherche sur les ensembles flous en attribuant une succession de rôles dans un mécanisme bien précis.

Une autre approche [1] vise la déstabilisation d'un réseau par la détection des communautés après avoir au préalable rajouté de nouveaux liens susceptibles d'apparaître dans le futur ou d'avoir été omis lors de la collecte ou l'analyse de données. Trois standards de prédiction de liens sont pris en compte, à savoir :

- le nombre de voisins communs pour chaque paire de nœuds.
- le nombre de voisins communs pondérés par les appels sortants
- le nombre de voisins communs pondérés par la hiérarchie.

---

## 3.4 Détection des communautés

Plusieurs algorithmes de détection de communautés ont été proposés dans la littérature. En particulier, il existe deux grandes catégories de procédures : celles dites hiérarchiques agglomératives et celles appelées hiérarchiques divisives. Dans la première catégorie, on part des nœuds et on constitue itérativement des groupes par fusion si la mesure de similarité choisie est forte. Dans le deuxième cas, on part du graphe entier et on procède à la formation de groupes de plus en plus fins en coupant les liens reliant des sommets peu similaires. L'exemple le plus connu d'approche divisive est l'algorithme de Girvan-Newman [7] alors que deux algorithmes communément utilisés dans la première catégorie (agglomérative) sont la procédure Walktrap [22] et la méthode Louvain [3]. Walktrap se base sur le principe de la marche aléatoire pour mesurer la similarité entre les nœuds. Plus précisément, la distance entre deux nœuds  $i$  et  $j$  est exprimée par la différence entre le comportement de deux marcheurs aléatoires partant respectivement des nœuds  $i$  et  $j$  et effectuant une marche de longueur fixe. Quant à l'approche Louvain, elle implante une méthode d'optimisation gloutonne locale de la modularité. À l'état initial, chaque nœud représente une communauté. Pour chaque nœud  $i$ , on évalue le gain de la modularité si on le déplace dans la communauté de ses voisins directs. Ainsi, on déplace  $i$  dans la communauté du voisin s'il maximise le gain de la modularité. Dans le cas où aucun gain n'est obtenu, le nœud reste dans sa communauté.

Une méthode divisive récente est celle de Bouguessa et al. [5] et comporte deux phases. La première étape initialise un nouveau partitionnement du réseau en regroupant les nœuds d'une même communauté à l'aide de la covariance des liens entre les nœuds et l'inertie inter-classes pour cibler les arcs à éliminer. Par la suite, les communautés sont identifiées par maximisation de la modularité.

# Chapitre 4

## Approche proposée

Comme indiqué dans les chapitres précédents, l'objectif de ce mémoire est d'étudier le démantèlement d'un réseau par élimination des liens et/ou des nœuds clés au sein de communautés déjà identifiées. La méthode proposée est applicable à des réseaux unimodaux orientés et pondérés. Notre approche est descendante puisqu'elle consiste en premier lieu à identifier puis isoler les communautés pour ensuite éliminer les liens et les nœuds les plus critiques au sein de chacune des communautés en vue de leur affaiblissement. Nous pensons qu'en procédant ainsi, la déstabilisation du réseau va s'effectuer d'une manière plus précise que si les suppressions sont faites au niveau global du réseau. Des tests seront alors menés pour confirmer ou infirmer cette hypothèse. Diverses questions se posent dans notre approche de désintégration d'un réseau social, à savoir :

- Existe-t-il un rapport entre le type de réseau (*aléatoire, petit-monde* ou *invariant d'échelle* ...) et l'une des variantes de notre approche de désintégration appliquée ? Autrement dit, est-ce qu'une variante de la méthode proposée fonctionne davantage pour un type de réseau plutôt qu'un autre ?
- Est-ce que l'identification en premier lieu des communautés facilite et améliore la qualité du démantèlement ?

Le processus de déstabilisation de notre approche comporte des variantes selon les deux dimensions suivantes : (i) présence versus absence d'identification de communautés, et (ii) suppression de nœuds, de liens ou des deux à la fois. Les étapes principales sont les suivantes :

1. Identifier ou non les communautés au sein du réseau.
2. Identifier les cibles à retirer du réseau. La cible peut être un nœud et/ou un lien.
3. Désintégrer le réseau par la suppression des cibles et calculer la nouvelle performance du réseau suite à son démantèlement.
4. Répéter le processus jusqu'à ce qu'un seuil prédéfini de démantèlement soit atteint.

Pour la détection de communautés, nous utilisons l'algorithme très connu appelé *Walktrap* [22] qui s'applique aux réseaux orientés et pondérés comme *Louvain* [7].

La figure 4.1 présente le réseau  $G$  après la détection de quatre communautés.

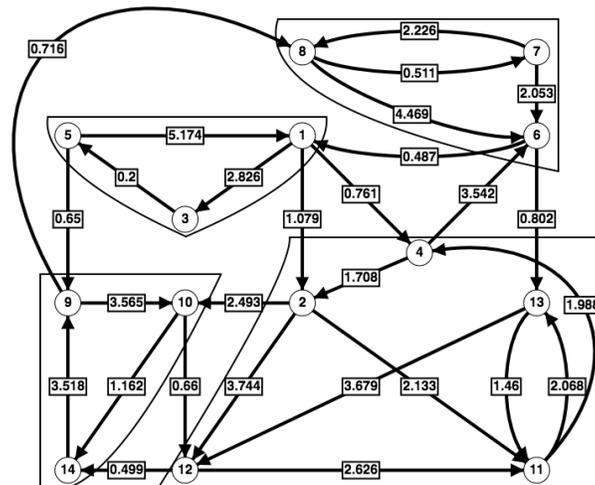


FIGURE 4.1 – graphe  $G$  après détection des communautés

Nous avons trois méthodes principales sans identification de communautés. La première méthode cible les nœuds clés avec les arcs qui leur lient, la seconde méthode élimine les liens clés seulement alors que la troisième méthode élimine non seulement les liens clés mais également les nœuds adjacents à ces liens.

Chacune de ces méthodes va ensuite être comparée à des variantes qui tiennent en compte des communautés au sein du graphe.

Nous commençons par présenter les trois méthodes principales notées  $M_1$ ,  $M_2$  et  $M_3$ .

## 4.1 Méthode $M_1$ : déstabilisation par élimination des nœuds clés

Dans la première méthode (cf. algorithme 1), on va calculer, sans identification de communautés, la mesure de centralité choisie pour identifier les nœuds les plus centraux ainsi que leurs liens directs. Le seuil  $s$  d'arrêt signifie que le réseau est démantelé jusqu'à ce que la performance finale du réseau atteigne  $s\%$  de la performance initiale.

L'algorithme 1 décrit les étapes de la première méthode qui ne considère pas la détection des communautés. Il consiste essentiellement à éliminer, à chaque itération, un nœud central ainsi que

**Algorithme 1** :  $M_1$  - Déstabilisation en supprimant les nœuds clés et leurs liens

---

```

1  Entrée  $G = (V, E)$  : graphe initial de  $n$  nœuds,  $s$  : seuil,  $Perf$  : formule de performance,
    $Ind$  : mesure de centralité
2  Sortie  $G' = (V', E')$  : graphe déstabilisé,  $P'$  : performance de  $G'$ , et  $i$  : nombre de nœuds
   supprimés
3   $G' \leftarrow G$ ;  $i \leftarrow 0$ ;
4   $P \leftarrow \text{Performance}(G, Perf)$ ;  $P' \leftarrow P$ ;
5   $Mat[n, 2] \leftarrow \text{Tri}(V, Ind)$ ;
6  tant que  $P' > s \times P$  faire
7  |    $i \leftarrow i + 1$ ;
8  |    $v \leftarrow Mat[i, 1]$ ;
9  |    $V' \leftarrow V' \setminus v$ ;
10 |    $E' \leftarrow E' \setminus Liens(v)$ ;
11 |    $P' \leftarrow \text{Performance}(G', Perf)$ ;
12 fin
13 Retourne  $(G', P', i)$ 
14

```

---

les liens qui lui sont associés. Ce processus se répète jusqu'à l'atteinte du seuil de désintégration. L'algorithme 1 donne les étapes de cette approche. La ligne 4 de l'algorithme calcule la performance du réseau initial selon la formule de performance  $Perf$  retenue et initialise  $P'$  à  $P$ . La ligne 5 place dans la matrice  $Mat$  de  $n$  lignes et deux colonnes (numéro du sommet et sa mesure de centralité) le résultat du tri de l'ensemble  $V$  des sommets de  $G$  par ordre décroissant de la mesure de centralité retenue. Les lignes 6 à 12 éliminent progressivement des nœuds jusqu'à ce que le seuil  $s$  soit atteint. On retourne finalement le graphe  $G'$  désintégré et sa performance  $P'$  ainsi que le nombre  $i$  de nœuds supprimés.

## 4.2 Méthode $M_2$ : déstabilisation par élimination des liens clés

La deuxième méthode vise à cibler les liens importants par le calcul de l'intermédiarité des liens. En premier lieu, nous calculons la liste des liens importants parmi les  $m$  arcs existants dans le réseau initial  $G$ . Le graphe  $G'$  présente le réseau après la désintégration et l'atteinte du seuil  $s$ . L'algorithme 2 résume les étapes de la méthode 2 sans considération de communautés. La différence entre cette procédure et celle de l'algorithme 1 réside dans le fait que la matrice  $Mat$  contient les  $m$  arêtes (et non pas les  $n$  sommets) du graphe et leur centralité d'intermédiarité et que la désintégration ne cible que les arêtes.

**Algorithme 2** :  $M_2$  - Déstabilisation en supprimant les liens clés

---

```

1  Entrée  $G = (V, E)$  : graphe Initial de  $m$  liens,  $s$  : seuil,  $Perf$  : formule de performance,
    $Ind$  : mesure de centralité
2  Sortie  $G' = (V', E')$  : graphe déstabilisé,  $P'$  : performance de  $G'$ , et  $i$  : nombre de liens
   supprimés
3   $G' \leftarrow G$ ;  $i \leftarrow 0$ ;
4   $P \leftarrow Performance(G, Perf)$ ;
5   $P' \leftarrow P$ ;
6   $Mat[m, 2] \leftarrow Tri(E, Ind)$ ;
7  tant que  $P' > s \times P$  faire
8  |    $i \leftarrow i + 1$ ;
9  |    $e \leftarrow Mat[i, 1]$ ;
10 |   $E' \leftarrow E' \setminus e$ ;
11 |   $P' \leftarrow Performance(G', Perf)$ ;
12 fin
13 Retourne  $(G', P', i)$ 
14

```

---

### 4.3 Méthode $M_3$ : déstabilisation par élimination des liens clés et leurs nœuds adjacents

L'approche consiste à désintégrer le réseau par élimination de liens importants incluant leurs nœuds adjacents. Cette élimination cible les liens dont leur centralité d'intermédiarité est la plus importante.

L'algorithme 3 illustre les étapes de la déstabilisation sans considération des communautés. Cette méthode est similaire à celle proposée par Kong [18] sauf que cette dernière est appliquée sans aucune identification préalable des communautés alors que dans notre approche, la suppression des liens importants et de leurs nœuds adjacents se fait selon deux situations : sans ou avec considération des communautés comme on le verra plus loin.

À ce stade, on cible uniquement les liens dans le réseau entier ainsi que leurs nœuds adjacents. Une élimination de tels liens dont la centralité d'intermédiarité est la plus élevée va contribuer à la désintégration du réseau par la formation des composants faiblement connectés.

**Algorithme 3 :**  $M_3$  - Déstabilisation en supprimant les liens clés et leurs nœuds adjacents

---

```

1  Entrée  $G = (V, E)$  : graphe Initial de  $n$  nœuds,  $s$  : seuil,  $Perf$  : formule de performance,
    $Ind$  : Mesure de centralité
2  Sortie  $G' = (V', E')$  : graphe déstabilisé,  $P'$  : performance de  $G'$ , et  $i$  : le nombre de
   nœuds supprimés
3   $G' \leftarrow G$ ;  $i \leftarrow 0$ ;
4   $P \leftarrow Performance(G, Perf)$ ;  $P' \leftarrow P$ ;
5   $Mat[m, 2] \leftarrow Tri(V, Ind)$ ;
6  tant que  $P' > s \times P$  faire
7  |    $i \leftarrow i + 1$ ;
8  |    $e \leftarrow Mat[i, 1]$ ;
9  |    $VA \leftarrow Adjacents(e)$ ;
10 |    $V' \leftarrow V' \setminus VA$ ;
11 |    $E' \leftarrow E' \setminus (e \cup Liens(VA))$ ;
12 |    $P' \leftarrow Performance(G', Perf)$ ;
13 fin
14 Retourne  $(G', P', i)$ 
15

```

---

## 4.4 Variante $M_{11}$ de la méthode 1

Cette variante de la méthode 1 suit la même démarche d'identification et de suppression des nœuds clés et de leurs liens sauf qu'elle s'applique à chacune des communautés préalablement détectées. Comme on peut s'en douter, la centralité d'un nœud au sein de sa communauté peut être différente de celle de ce même nœud dans le réseau pris dans son intégralité. Ceci est dû au fait qu'un nœud central dans une petite communauté peut être moins important au niveau global du réseau. Pour pallier à ce problème, les nouvelles centralités sont calculées en multipliant les anciennes mesures de centralité par  $\frac{n_1}{n}$  où  $n_1$  et  $n$  sont respectivement le nombre de nœuds de la communauté à laquelle appartient le nœud  $i$  et  $n$  le nombre total de nœuds du graphe  $G$ .

L'algorithme 4 résume les étapes de la variante  $M_{11}$  qui compare la performance de la première méthode  $M_1$  en cas de prise en compte ou non de l'existence de communautés. Il prend comme entrée le graphe  $G$  à déstabiliser, le seuil de déstabilisation  $s$ , la formule de performance  $Perf$  et un indicateur pour la mesure de centralité  $Ind$ . Il retourne les graphes  $G'$  et  $G''$  qui présentent respectivement les graphes désintégrés sans et avec considération des communautés et leurs performances ainsi qu'une variable booléenne  $Statut$  indiquant si l'utilisation des communautés a mieux déstabilisé le réseau ou non pour un même nombre de nœuds retirés.

Au départ, on utilise l'algorithme 1 pour la désintégration du réseau dans sa globalité ce qui permet

---

d'obtenir le nombre  $nb$  de nœuds supprimés ainsi que la nouvelle performance  $P'$  (ligne 3). Ces deux éléments sont ensuite exploités dans le reste de l'algorithme. La ligne 4 effectue l'initialisation de variables et la détection de communautés est faite à la ligne 5. Quant aux lignes 6 à 9, elles permettent la collecte des nœuds et de leur centralité au sein des communautés. La ligne 10 sert à stocker dans la nouvelle matrice  $Mat_1$  le résultat du tri des nœuds contenus dans la première colonne de  $Mat$  selon un ordre décroissant de la centralité placée dans la seconde colonne de cette même matrice. On supprime ensuite  $nb$  nœuds du graphe  $G''$  (lignes 11 à 15) ainsi que leurs liens pendants et on calcule sa performance  $P''$  (ligne 16).

Une comparaison entre les deux performances termine le processus. Si la performance  $P''$  relative à la déstabilisation au sein des sous-groupes est inférieure à  $P'$  relative au réseau désintégré sans considération des communautés, on affirme que la déstabilisation ciblant les communautés est plus performante avec le même nombre de nœuds éliminés.

---

**Algorithme 4 :** Variante  $M_{11}$  : Déstabilisation en supprimant les nœuds clés au sein des communautés

---

```

1  Entrée  $G = (V, E)$  : graphe Initial de  $n$  nœuds,  $s$  : seuil,  $Perf$  : formule de performance,
    $Ind$  : Mesure de centralité
2  Sortie  $G' = (V', E')$  : graphe déstabilisé,  $P'$  : performance de  $G'$ ,  $G'' = (V'', E'')$  :
   graphe déstabilisé,  $P''$  : performance de  $G''$ , et  $statut$  : Valeur indiquant si la détection de
   communautés a amélioré la déstabilisation ou non
3   $(G', P', nb) \leftarrow M_1(G, s, Perf, Ind)$ 
4   $G'' \leftarrow G$ ;  $Mat[n, 2] \leftarrow \emptyset$ ;
5   $Com \leftarrow Communautés(G)$ ;
6  pour  $c \in Com$  faire
7  |    $V_c \leftarrow Noeuds(c)$ ;
8  |    $Mat[n, 2] \leftarrow Mat[n, 2] \cup Tri(V_c, Ind)$ ;
9  fin
10  $Mat[n, 2] \leftarrow Tri(Mat[n, 2])$ ;
11 pour  $i$  de 1 à  $nb$  faire
12 |    $v \leftarrow Mat[i, 1]$ ;
13 |    $V'' \leftarrow V'' \setminus v$ ;
14 |    $E'' \leftarrow E'' \setminus Liens(v)$ ;
15 fin
16  $P'' \leftarrow Performance(G'', Perf)$ ;
17 si  $P'' < P'$  alors
18 |    $statut \leftarrow Vrai$ ;
19 sinon
20 |    $statut \leftarrow Faux$ ;
21 fin
22 Retourne  $(G', P', G'', P'', statut)$ 

```

---

---

## 4.5 Variantes de la méthode 2 :

Pour la méthode 2 éliminant uniquement les liens clés, on a trois variantes qui considèrent la détection des communautés au sein du graphe comme phase préliminaire de désintégration.

### 4.5.1 Variante 1 de la méthode 2 : $M_{21}$

La première variante  $M_{21}$  explicitée dans l'algorithme 5 suit les étapes de détérioration du réseau en ciblant uniquement les liens entre les communautés. Cette procédure utilise l'algorithme 2 pour éliminer les liens clés au niveau du réseau dans sa globalité et conserve l'information sur le nombre  $nb$  de liens supprimés et la nouvelle performance  $P'$  du réseau désintégré sans considération des communautés (Ligne 3). Ensuite, on calcule la même liste des liens les plus centraux (ligne 4) mais n'extrait que les liens intercommunautaires (ligne 5). Ensuite, on procède à supprimer un par un les liens intercommunautaires jusqu'à l'atteinte des  $nb$  liens supprimés ou jusqu'à l'épuisement des liens intercommunautaires (lignes 7 à 14). Enfin, la performance  $P''$  calculée (ligne 15) est comparée à  $P'$  (ligne 16). La valeur de la variable booléenne *statut* indique si le recours à la détection des communautés a amélioré ou non la désintégration du réseau.

---

**Algorithme 5** : Variante  $M_{21}$  - Déstabilisation en supprimant les liens clés intercommunautaires

---

```

1  Entrée  $G = (V, E)$  : graphe Initial de  $m$  liens,  $s$  : seuil,  $Perf$  : formule de performance,
    $Ind$  : Mesure de centralité
2  Sortie  $G' = (V', E')$  : graphe déstabilisé,  $P'$  : performance de  $G'$ ,  $G'' = (V'', E'')$  :
   graphe déstabilisé,  $P''$  : performance de  $G''$ , et  $statut$  : Valeur indiquant si la détection
   des communautés a amélioré la déstabilisation
3   $(G', P', nb) \leftarrow M_2(G, s, Perf, Ind); G'' \leftarrow G$  ;
4   $Mat[m, 2] \leftarrow Tri(E, Ind)$  ;
5   $Inter \leftarrow InterCom(G)$ ;
6   $i \leftarrow 0; j \leftarrow 0; p \leftarrow |Inter|$ ;
7  tant que  $i < MIN(nb, p)$  et  $j < m$  faire
8  |    $j \leftarrow j + 1$ ;
9  |    $e \leftarrow Mat[j, 1]$ ;
10 |   si  $e \in Inter$  alors
11 |   |    $E'' \leftarrow E'' \setminus e$ ;
12 |   |    $i \leftarrow i + 1$ 
13 |   fin
14 fin
15  $P'' \leftarrow Performance(G'', Perf)$  ;
16 si  $P'' < P'$  alors
17 |    $statut \leftarrow Vrai$ ;
18 sinon
19 |    $statut \leftarrow Faux$ ;
20 fin
21 Retourne  $(G', P', G'', P'', statut)$ 

```

---

### 4.5.2 Variante 2 de la méthode 2 : $M_{22}$

La deuxième variante de la même méthode 2 est  $M_{22}$  laquelle consiste à cibler les liens intra-communautaires. Il n'est pas nécessaire de montrer l'algorithme de cette variante. La seule différence avec l'algorithme 5 est le fait d'extraire de la liste déjà calculée des liens centraux seulement les liens *Intra* (intra-communautés) au lieu des liens *Inter*. Cela veut dire qu'on va extraire de la liste triée des liens (selon leur centralité d'intermédiation) ceux qui sont à la fois intra-communautaires (au sein des communautés) et centraux au niveau global.

### 4.5.3 Variante 3 de la méthode 2 : $M_{23}$

Quant à la troisième variante  $M_{23}$  de la méthode 2 décrite par l'algorithme 6, elle ne différencie pas entre les liens intra-communautaires ou inter-communautaires. Par contre, elle définit un coefficient d'appartenance  $a(i, j)$  pour tous les liens du réseau après la détection des communautés et multiplie ce coefficient par l'indicateur de mesure de centralité *Ind* qui est dans ce cas la centralité d'intermédiation des liens. Ce coefficient d'appartenance  $a(i, j)$  du lien  $l$  allant de  $i$  vers  $j$  est calculé comme suit :

$$a(i, j) = \frac{nb(C_i) + nb(C_j)}{n} \quad (4.1)$$

où  $nb(C_i)$  est le nombre de nœuds de la communauté à laquelle  $i$  appartient,  $nb(C_j)$  est le nombre de nœuds de la communauté à laquelle  $j$  appartient et  $n$  est le nombre total de nœuds.

La principale différence avec l'algorithme 5 est que l'indicateur de mesure de centralité *Ind'* dans l'algorithme 6 est la ré-évaluation de la centralité *Ind* en considérant les communautés comme indiqué à la ligne 5. De plus, dans la boucle *tant que* (lignes 7 à 11), il n'y a plus de vérification du type (inter-communautaire ou intra-communautaire) de lien.

---

**Algorithme 6 :** Variante  $M_{23}$  : Déstabilisation en supprimant les liens clés avec considération des communautés

---

```

1  Entrée  $G = (V, E)$  : graphe Initial de  $m$  liens,  $s$  : seuil,  $Perf$  : formule de performance,
    $Ind$  : Mesure de centralité
2  Sortie  $G' = (V', E')$  : graphe déstabilisé,  $P'$  : performance de  $G'$ ,  $G'' = (V'', E'')$  :
   graphe déstabilisé,  $P''$  : performance de  $G''$ , et  $statut$  : Valeur indiquant si la détection
   des communautés a amélioré la déstabilisation
3   $(G', P', nb) \leftarrow M_2(G, s, Perf, Ind)$   $G'' \leftarrow G$ ;
4   $CommununautesLiens \leftarrow Com(G)$ ;
5   $Mat[m, 2] \leftarrow Tri(E, Ind')$  ;
6   $i \leftarrow 0$ ;
7  tant que  $i < nb$  faire
8  |    $i \leftarrow i + 1$ ;
9  |    $e \leftarrow Mat[i, 1]$ ;
10 |    $E'' \leftarrow E'' \setminus e$ ;
11 fin
12  $P'' \leftarrow Performance(G'', Perf)$  ;
13 si  $P'' < P'$  alors
14 |    $statut \leftarrow Vrai$ ;
15 sinon
16 |    $statut \leftarrow Faux$ ;
17 fin
18 Retourne  $(G', P', G'', P'', statut)$ 

```

---

---

## 4.6 Variantes de la méthode 3 :

### 4.6.1 Variante 1 de la méthode 3 : $M_{31}$

Comme indiquée précédemment, la méthode 3 cible les liens centraux et leurs nœuds adjacents. Quant à ses variantes, la première, notée  $M_{31}$ , cible à la fois les liens intercommunautaires (ponts entre les diverses communautés) et leurs nœuds adjacents afin d'assurer l'isolement des sous-groupes du réseau et donc aboutir à la désintégration de ce dernier.

La variante  $M_{31}$  est décrite par l'algorithme 7, et la seule différence qu'elle apporte par rapport à l'algorithme 5 est l'ajout d'une vérification qu'aucun des deux nœuds adjacents au lien à supprimer n'a pas été déjà éliminé. Cette vérification se fait de la ligne 10 à la ligne 18.

---

**Algorithme 7 :** Variante  $M_{31}$  : Déstabilisation en supprimant les nœuds adjacents des liens clés intercommunautaires

---

```

1  Entrée  $G = (V, E)$  : graphe Initial de  $n$  nœuds,  $s$  : seuil,  $Perf$  : formule de performance,
    $Ind$  : mesure de centralité
2  Sortie  $G' = (V', E')$  : graphe déstabilisé,  $P'$  : performance de  $G'$ ,  $G'' = (V'', E'')$  :
   graphe déstabilisé,  $P''$  : performance de  $G''$ , et  $statut$  : Valeur indiquant si la détection
   des communautés a amélioré la déstabilisation
3   $(G', P', nb) \leftarrow M_3(G, s, Perf, Ind)$   $G'' \leftarrow G$ ;  $NœudsRetirés \leftarrow \emptyset$ ;
4   $Mat[n, 2] \leftarrow Tri(V, Ind)$ ;
5   $Inter \leftarrow InterCom(G)$ ;
6   $i \leftarrow 0$ ;  $j \leftarrow 0$ ;  $p \leftarrow |Inter|$ ;
7  tant que  $i < MIN(nb, p)$  et  $j < m$  faire
8  |    $j \leftarrow j + 1$ ;
9  |    $e \leftarrow Mat[j, 1]$ ;
10 |   si  $e \in Inter$  alors
11 |   |    $VA \leftarrow adjacents(e)$ ;
12 |   |   si  $VA[1] \notin NœudsRetirés$  ET  $VA[2] \notin NœudsRetirés$  alors
13 |   |   |    $i \leftarrow i + 1$ ;
14 |   |   |    $NœudsRetirés \leftarrow NœudsRetirés \cup VA$ ;
15 |   |   |    $V'' \leftarrow V'' \setminus VA$ ;
16 |   |   |    $E'' \leftarrow E'' \setminus (e \cup Liens(VA))$ ;
17 |   |   fin
18 |   fin
19 fin
20  $P'' \leftarrow Performance(G'', Perf)$ ;
21 si  $P'' < P'$  alors
22 |    $statut \leftarrow Vrai$ ;
23 sinon
24 |    $statut \leftarrow Faux$ ;
25 fin
26 Retourne  $(G', P', G'', P'', statut)$ 

```

---

---

### 4.6.2 Variante 2 de la méthode 3 : $M_{32}$

La deuxième variante nommée  $M_{32}$  de la méthode 3 se base sur l'identification des liens intracommunautaires les plus importants et leurs nœuds adjacents. Aussi, il n'est pas nécessaire de montrer l'algorithme de cette variante. La seule différence avec l'algorithme 7 est qu'on va utiliser la liste *Intra* au lieu de *Inter* des liens ciblés.

### 4.6.3 Variante 3 de la méthode 3 : $M_{33}$

Quant à la troisième variante  $M_{33}$  de la méthode 3, elle ne différencie pas entre les liens intracommunautaires ou inter-communautaires. Elle tient compte des nouvelles mesures de centralité des liens avec considération des communautés en calculant le coefficient d'appartenance comme expliqué dans la section 4.5.

L'algorithme 8 qui décrit la troisième variante  $M_{33}$  de la méthode 3 comporte deux différences par rapport à l'algorithme 7 : on utilise l'indicateur  $Ind'$  au lieu de  $Ind$  à la ligne 5 qui calcule la centralité d'intermédiarité des liens pondérée en se référant au coefficient d'appartenance et en tenant compte des communautés détectées. Ensuite, on supprime dans la boucle *tant que* (Ligne 7 à 17) les paires de nœuds adjacents aux liens détectés.

---

**Algorithme 8 :** Variante  $M_{33}$  :Déstabilisation en supprimant les liens clés avec considération des communautés et leurs nœuds adjacents

---

```

1  Entrée  $G = (V, E)$  : graphe Initial de  $m$  liens,  $s$  : seuil,  $Perf$  : formule de performance,
    $Ind$  : Mesure de centralité
2  Sortie  $G' = (V', E')$  : graphe déstabilisé,  $P'$  : performance de  $G'$ ,  $G'' = (V'', E'')$  :
   graphe déstabilisé,  $P''$  : performance de  $G''$ , et  $statut$  : Valeur indiquant si la détection
   des communautés a amélioré la déstabilisation
3   $(G', P', nb) \leftarrow M_3(G, s, Perf, Ind)$   $G'' \leftarrow G$ ;  $NœudsRetirés \leftarrow \emptyset$ ;
4   $CommununautesLiens \leftarrow Com(G)$ ;
5   $Mat[m, 2] \leftarrow Tri(E, Ind')$ ;
6   $i \leftarrow 0$ ;  $j \leftarrow 0$ ;
7  tant que  $i < nb$  ET  $j < m$  faire
8  |    $j \leftarrow j + 1$ ;
9  |    $e \leftarrow Mat[j, 1]$ ;
10 |    $VA \leftarrow adjacents(e)$ ;
11 |   si  $VA[1] \notin NœudsRetirés$  ET  $VA[2] \notin NœudsRetirés$  alors
12 | |    $i \leftarrow i + 1$ ;
13 | |    $NœudsRetirés \leftarrow NœudsRetirés \cup VA$ ;
14 | |    $V'' \leftarrow V'' \setminus VA$ ;
15 | |    $E'' \leftarrow E'' \setminus (e \cup Liens(VA))$ ;
16 |   fin
17 fin
18  $P'' \leftarrow Performance(G'', Perf)$ ;
19 si  $P'' < P'$  alors
20 |    $statut \leftarrow Vrai$ ;
21 sinon
22 |    $statut \leftarrow Faux$ ;
23 fin
24 Retourne  $(G', P', G'', P'', statut)$ 

```

---

# Chapitre 5

## Analyse empirique

Dans le but de valider notre approche, nous avons codé les méthodes proposées dans le chapitre précédent en langage Python. La génération synthétique des réseaux et la détection de communautés constituent la première phase qui précède le démantèlement. Les trois types de réseaux testés sont : *aléatoires*, *invariants d'échelle* et *petit monde*. Quant au démantèlement, il a été réalisé sur plusieurs itérations successives suivant la démarche décrite précédemment. En se basant sur les résultats d'expérimentations préliminaires sur des réseaux de grandes tailles, on peut tirer quelques conclusions permettant le raffinement de l'approche proposée laquelle vise à mesurer le potentiel de la détection des communautés dans le démantèlement du réseau. Pour se faire, on s'est basé sur la mesure de robustesse (appelée *Perf* dans les algorithmes) et calculée selon la performance totale, le taux de connectivité ou la compacité du réseau.

Afin de comparer les méthodes proposées et leurs variantes en terme de qualité de désintégration, nous avons mené des tests sur un PC fonctionnant sous Windows 10 et doté d'un processeur intel Core i7 à 3.0 GHz et d'une mémoire RAM de 8 giga octets.

Le tableau 5.1 résume les méthodes proposées, leurs variantes associées, l’algorithme de la méthode de base, l’algorithme de la variante ainsi que la cible détectée. Une cible peut être un nœud et/ou un lien.

Méthode de base	Variante	Algorithme	Cible
1	$M_{11}$	4	Nœuds
2	$M_{21}$	5	Liens inter
2	$M_{22}$	-	Liens intra
2	$M_{23}$	6	Liens*
3	$M_{31}$	7	Liens inter + nœuds adjacents
3	$M_{32}$	-	Liens intra + nœuds adjacents
3	$M_{33}$	8	Liens* + nœuds adjacents

TABLE 5.1 – Méthodes proposées et leurs variantes

Dans le tableau, la cible *Liens\** réfère aux liens dont la centralité d’intermédiarité est recalculée selon le coefficient d’appartenance en tenant compte des communautés détectées.

## 5.1 Illustration

Pour étudier le fonctionnement des variantes des trois méthodes proposées, et avant de faire des exécutions répétées sur de grands graphes pour tirer des statistiques, nous avons procédé à une étude du comportement des algorithmes sur des graphes de petite taille afin de pouvoir les visualiser via des graphiques ou des tableaux.

### 5.1.1 Premier exemple

Le premier exemple à étudier est un graphe orienté et pondéré de 14 nœuds, 28 liens et de densité égale à 15.38%. La performance initiale du graphe est de 0.286. Le tableau 5.2 montre les mesures de centralité des nœuds à savoir la centralité d’intermédiarité, la centralité *PageRank*, la centralité de proximité, une moyenne des trois centralités citées ainsi qu’une centralité *ProxiRank* qui est une moyenne de la centralité de proximité et de *PageRank*.

Id	intermédierité	<i>PageRank</i>	Proximité	Moyenne	<i>ProxiRank</i>
<b>6</b>	0.543	0.080	0.565	<b>0.396</b>	0.322
<b>1</b>	0.423	0.076	0.590	<b>0.363</b>	0.333
<b>9</b>	0.472	0.072	0.520	<b>0.355</b>	0.296
<b>12</b>	0.302	0.118	0.520	<b>0.313</b>	0.319
<b>8</b>	0.412	0.026	0.481	<b>0.306</b>	0.254
<b>4</b>	0.269	0.076	0.541	<b>0.295</b>	0.309
11	0.247	0.132	0.481	0.287	0.306
7	0.412	0.013	0.419	0.281	0.216
13	0.225	0.110	0.500	0.278	0.305
2	0.214	0.047	0.565	0.275	0.306
10	0.230	0.074	0.520	0.274	0.297
5	0.285	0.053	0.481	0.273	0.267
14	0.258	0.066	0.464	0.263	0.265
3	0.285	0.050	0.419	0.251	0.234

TABLE 5.2 – Mesures de centralité des nœuds du graphe G

Des nœuds comme 6, 1, 9 et 12 semblent les plus centraux car ils présentent une moyenne de centralité importante par rapport au reste des nœuds du graphe. Une telle moyenne reflète le rôle que peut jouer un nœud central en tant qu'intermédiaire dans la circulation de l'information (centralité d'intermédierité), sa capacité de contacter la majorité des nœuds avec un minimum d'effort (centralité de proximité) et sa centralité de *PageRank*.

Quant au tableau 5.3, il illustre la centralité d'intermédierité des liens du graphe. Cette centralité sert à calculer le nombre de fois qu'un lien  $l$  sert de pont entre deux nœuds dans le plus court chemin qui les relie. Les liens dont les identifiants sont 12, 17, 14 et 18 sont les plus centraux.



d'appartenance aux communautés  $a(i, j)$  comme expliqué au chapitre précédent. Ces deux mesures sont affichées respectivement dans les tableaux 5.4 et 5.5. On remarque que les nœuds 1, 3, 5, 11 et 9 présentent une moyenne de centralité importante et sont donc centraux au sein de leur communauté alors que les nœuds les plus importants dans le réseau global sont les nœuds 6, 1, 9 et 12.

Id	intermédiarité	PageRank	Proximité	Moyenne	ProxiRank
<b>1</b>	0.071	0.333	1.000	<b>0.555</b>	0.666
<b>3</b>	0.071	0.333	1.000	<b>0.555</b>	0.666
2	0.035	0.162	0.800	0.354	0.481
4	0.035	0.156	0.666	0.307	0.411
<b>5</b>	0.035	0.333	1.000	<b>0.500</b>	0.666
<b>11</b>	0.017	0.303	1.000	<b>0.451</b>	0.651
12	0.000	0.216	0.800	0.338	0.508
13	0.000	0.161	0.666	0.276	0.414
6	0.000	0.50	1.000	0.499	0.748
7	0.000	0.215	1.000	0.405	0.607
8	0.000	0.286	1.000	0.428	0.643
<b>9</b>	0.000	0.333	1.000	<b>0.444</b>	0.666
<b>10</b>	0.000	0.333	1.000	<b>0.444</b>	0.666
<b>14</b>	0.000	0.333	1.000	<b>0.444</b>	0.666

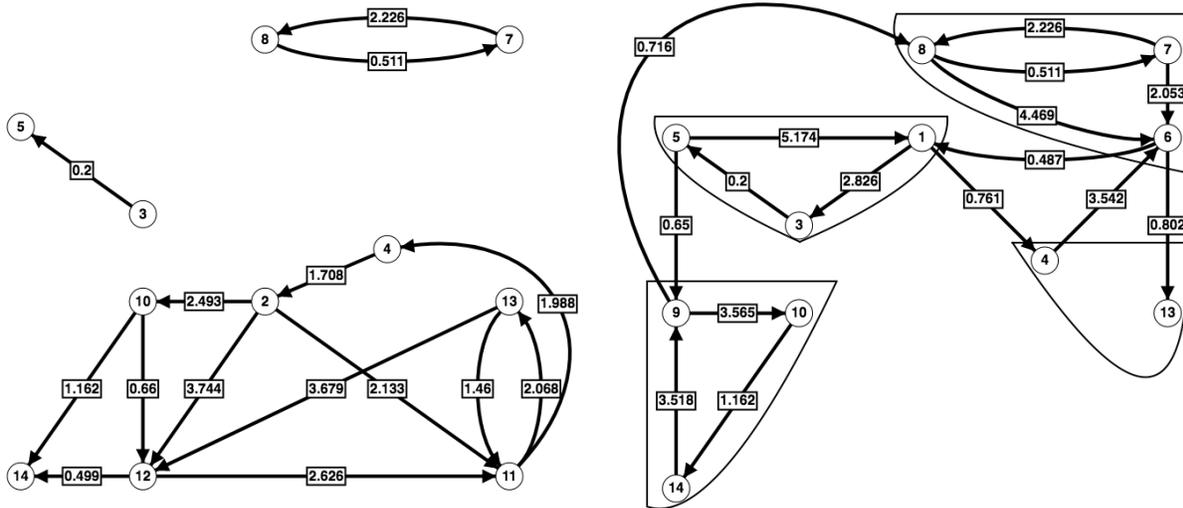
TABLE 5.4 – Mesures de centralité des nœuds au sein des communautés

Les liens ayant une intermédiarité pondérée élevée sont ceux d'identifiants 12, 17, 14 et 18. Cette intermédiarité pondérée dépend de l'appartenance du lien à sa communauté. On remarque que des liens comme celui dont l'identifiant est 2 a une intermédiarité globale importante comparée à celle pondérée (0.214 contre 0.09). Ce lien du nœud 1 vers le nœud 3 est intracommunautaire et présente peu d'importance lorsque les communautés sont prises en compte.

Id	de	vers	intermédiation	intermédiation pondérée
<b>12</b>	6	1	0.351	<b>0.150</b>
<b>17</b>	8	7	0.340	<b>0.145</b>
<b>14</b>	7	6	0.335	<b>0.143</b>
<b>18</b>	9	8	0.335	<b>0.143</b>
<b>25</b>	12	14	0.186	<b>0.133</b>
<b>22</b>	11	4	0.142	<b>0.102</b>
9	4	6	0.137	0.0981
2	1	3	0.214	0.091
7	3	5	0.214	0.091
11	5	9	0.214	0.091
.	.	.	.	.
16	8	6	0.000	0.000
21	10	14	0.000	0.000

TABLE 5.5 – Centralité d'intermédiation des liens au sein des communautés

La figure 5.2 présente le graphe  $G'$  suite au démantèlement du réseau en ciblant les nœuds clés dans le réseau entier. Cette approche désintègre le réseau en supprimant les nœuds 6, 1 et 9. Quant au graphe  $G''$ , il présente le réseau après la désintégration en tenant compte des communautés et l'élimination du même nombre de nœuds que le graphe  $G'$ . Après la détection des communautés, la suppression s'est faite sur les nœuds 11, 2 et 12 contrairement à la désintégration du réseau entier où les nœuds 6, 1 et 9 sont supprimés.



(a) Graphe  $G'$  après suppression des nœuds dans le réseau entier (b) Graphe  $G''$  après suppression des nœuds avec considération des communautés

FIGURE 5.2 – Les graphes obtenus par la première méthode et sa variante  $M_{11}$

Nous mesurons les performances du premier graphe pour chaque méthode après la fin de l'étape 1 (sans détection de communautés) et à la fin de l'étape 2 (avec détection de communautés) dans le tableau 5.6.

Variante	Cible	Supprimés	Sans détection ( $G'$ )		Avec détection ( $G''$ )	
			Perf	Temps (s)	Perf	Temps (s)
$M_{11}$	Nœuds	2	<b>0.04</b>	0.196	<b>0.241</b>	0.066
$M_{21}$	Liens inter	6	0.064	1.184	0.06	0.086
$M_{22}$	Liens intra	6	0.064	1.184	0.07	0.069
$M_{23}$	Liens*	6	0.064	1.184	0.064	0.078
$M_{31}$	Nœuds des liens inter	5	0.021	0.126	0.017	0.055
$M_{32}$	Nœuds des liens intra	5	0.021	0.126	0.021	0.063
$M_{33}$	Nœuds des liens*	5	0.021	0.126	0.021	0.055

TABLE 5.6 – Statistiques sur les variantes exécutées

Ces résultats donnent une idée sur les méthodes appliquées à notre petit exemple. En effet, la détection des nœuds clés dans les communautés et la suppression de leurs liens pendant dégrade mieux le réseau (c.-à-d. aboutit à une performance plus faible) sans détection de communautés comme le montre la variante  $M_{11}$  du tableau 5.6. Par contre, les six autres variantes aboutissent à des résultats

similaires aussi bien en présence qu'en absence de communautés. Des tests supplémentaires sur des réseaux plus grands vont confirmer ou infirmer l'hypothèse de départ, à savoir que l'approche descendante visant à d'abord identifier les communautés améliore le niveau de désintégration du réseau.

### 5.1.2 Deuxième exemple

Le deuxième exemple est un réseau synthétique aléatoire de 15 nœuds avec une probabilité d'existence de liens égale à 0.5. Après la génération du graphe, nous avons obtenu un graphe de 93 liens de densité égale à 44.29%. La performance initiale du graphe est de 0.94. Les tableaux 5.7 et 5.8 montrent respectivement les mesures de centralité des nœuds et des liens du graphe.

Id	intermédiarité	PageRank	Proximité	Moyenne	ProxiRank
<b>6</b>	0.447	0.0811	0.875	<b>0.467</b>	0.478
<b>9</b>	0.395	0.094	0.823	<b>0.437</b>	0.459
<b>11</b>	0.319	0.072	0.823	<b>0.405</b>	0.448
<b>3</b>	0.309	0.079	0.933	<b>0.440</b>	0.506
<b>13</b>	0.271	0.056	0.777	<b>0.368</b>	0.417
2	0.242	0.083	0.823	0.383	0.453
12	0.223	0.048	0.777	0.349	0.413
10	0.185	0.056	0.823	0.355	0.440
1	0.180	0.083	0.777	0.347	0.430
8	0.161	0.0365	0.777	0.325	0.407
15	0.157	0.029	0.700	0.295	0.364
4	0.147	0.060	0.736	0.315	0.398
5	0.138	0.109	0.736	0.328	0.423
14	0.138	0.055	0.777	0.323	0.416
7	0.133	0.050	0.666	0.283	0.358

TABLE 5.7 – Mesures de centralité des nœuds du graphe G

Les nœuds ayant une moyenne de centralité élevée sont ceux jouant un rôle critique dans le réseau. Ces nœuds sont 6, 9, 11, 3 et 13.

Quant aux liens d'identifiant 51, 9, 34, 19 et 38, ils jouent le rôle de pont de communications entre les divers composants du réseau.

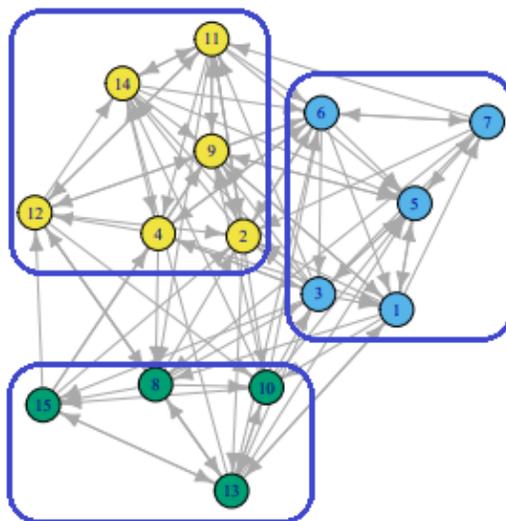
Id	de	vers	intermédialité
<b>51</b>	9	6	<b>0.300</b>
<b>9</b>	2	9	<b>0.176</b>
<b>34</b>	6	3	<b>0.152</b>
<b>19</b>	3	11	<b>0.147</b>
<b>38</b>	6	12	<b>0.133</b>
74	12	13	0.133
60	11	2	0.104
56	10	9	0.090
6	1	10	0.080
61	11	4	0.076
.	.	.	.
87	14	9	0.000
91	15	9	0.000

TABLE 5.8 – Centralité d’intermédialité des liens du graphe G

Comme pour le premier graphe, nous affichons : (i) le graphe après la détection de ses trois communautés dans la figure 5.3, (ii) la centralité des nœuds au sein de leur communautés dans le tableau 5.9 et (iii) la centralité des liens multipliée par le coefficient d’appartenance à des communautés dans le tableau 5.10.

Comme présenté précédemment, les mesures de centralité donnent une idée sur l’importance des nœuds dans le réseau. En tenant compte de la détection des communautés, ces centralités dépendent de l’appartenance à une communauté particulière comme expliqué dans le chapitre 4. Les nœuds ayant une moyenne de centralité importante sont les nœuds 1, 3, 5, 10 et 13, contrairement aux nœuds centraux au niveau globale qui sont les nœuds 6, 9, 11, 3 et 13.

Les liens les plus intermédiaires en considérant les communautés sont ceux ayant les identifiants 51, 9, 74 et 34.

FIGURE 5.3 – Graphe  $G$  après détection des communautés

Id	intermédiarité	PageRank	Proximité	Moyenne	ProxiRank
<b>1</b>	0.066	0.193	1.000	<b>0.464</b>	0.596
<b>3</b>	0.066	0.209	1.000	<b>0.469</b>	0.604
<b>5</b>	0.066	0.327	1.000	<b>0.509</b>	0.663
2	0.026	0.140	0.833	0.346	0.486
4	0.026	0.094	0.833	0.331	0.464
11	0.026	0.243	1.000	0.436	0.621
6	0.016	0.077	1.000	0.375	0.538
7	0.016	0.192	1.000	0.414	0.596
9	0.013	0.217	1.000	0.417	0.608
12	0.0133	0.150	1.000	0.394	0.575
8	0.000	0.0542	1.000	0.351	0.527
<b>10</b>	0.000	0.325	1.000	<b>0.441</b>	0.662
<b>13</b>	0.000	0.389	1.000	<b>0.463</b>	0.694
15	0.000	0.230	1.000	0.410	0.615
14	0.000	0.153	1.000	0.384	0.576

TABLE 5.9 – Mesures de centralité des nœuds au sein des communautés

Id	de	vers	intermédierité	intermédierité pondérée
<b>51</b>	9	6	0.300	<b>0.240</b>
<b>9</b>	2	9	0.176	<b>0.140</b>
<b>74</b>	12	13	0.133	<b>0.106</b>
<b>34</b>	6	3	0.152	<b>0.101</b>
19	3	11	0.147	0.098
38	6	12	0.133	0.088
60	11	2	0.104	0.083
61	11	4	0.076	0.060
6	1	10	0.080	0.053
56	10	9	0.0904	0.048
.	.	.	.	.
87	14	9	0.000	0.000
91	15	9	0.000	0.000

TABLE 5.10 – Centralité d’intermédierité des liens au sein des communautés

### Résultats préliminaires

Les résultats des variantes sur ce deuxième graphe illustratif sont affichés dans le tableau 5.11.

Variante	Cible	Supprimés	Sans détection ( $G'$ )		Avec détection ( $G''$ )	
			Perf	Temps (s)	Perf	Temps (s)
$M_{11}$	Nœuds	5	<b>0.298</b>	0.4	<b>0.69</b>	0.110
$M_{21}$	Liens inter	15	0.431	1.705	0.428	0.159
$M_{22}$	Liens intra	15	<b>0.431</b>	1.705	<b>0.554</b>	0.083
$M_{23}$	Liens*	15	0.431	1.705	0.423	0.147
$M_{31}$	Nœuds des liens inter	3	0.289	0.242	0.289	0.159
$M_{32}$	Nœuds des liens intra	3	<b>0.289</b>	0.242	<b>0.722</b>	0.09
$M_{33}$	Nœuds des liens*	3	0.289	0.242	0.289	0.059

TABLE 5.11 – Statistiques sur les variantes exécutées

L’application des sept différentes variantes au deuxième réseau de type aléatoire a abouti à une désintégration plus intéressante sans détection préalable de communautés pour les variantes  $M_{11}$ ,

---

$M_{22}$  et  $M_{32}$  tel qu'illustré par le tableau 5.11. Par contre, l'application des autres variantes au sein de communautés pré-identifiées a conduit à un même niveau de dégradation du réseau.

## 5.2 Étude approfondie

Nous avons testé les algorithmes sur onze graphes orientés et valués distincts :

- Un graphe réel de 217 nœuds et de 2672 liens fourni par [12] qui est un réseau orienté et pondéré de personnes vivant dans une résidence située dans le campus universitaire national Australien. Les liens entre les personnes réfèrent aux amitiés entretenues. Ce graphe est de densité égale à 5.7%.
- Un graphe réel de 2539 nœuds et de 12969 liens fourni par [13] qui a été créé à partir d'une enquête réalisée en 1994/1995 auprès d'élèves. Chaque élève a été invité à énumérer ses cinq meilleurs amis (hommes et femmes). Un nœud représente un élève et un lien  $l_{ij}$  signifie que l'élève  $i$  a choisi l'élève  $j$  comme ami. Des poids de lien plus élevés indiquent plus d'interactions et un poids de lien nul montre qu'il n'y a aucune activité commune du tout. Ce graphe est de densité de 0.2%.
- Trois réseaux synthétiques aléatoires de 100, 1000 et 3000 nœuds, chacun crée avec une densité de lien égale à 20%, 50% et 80%.
- Trois réseaux synthétiques invariants d'échelle de 100, 1000 et 3000 nœuds afin de mesurer les performances des sept variantes à une plus grande échelle.
- Trois réseaux synthétiques petit-monde de 100, 1000 et 3000 nœuds, tous créés avec  $k = 5$ . Le but de ces réseaux est de vérifier si certaines variantes sont plus performantes pour un type de réseau que pour un autre.

Chacun de ces graphes a été testé trois fois en utilisant les variantes présentées au chapitre 4 et en utilisant les mesures de robustesse (performance) suivantes : la performance globale, la compacité et le taux de connectivité.

### 5.2.1 Complexité

En plus des mesures de performance, la mesure des complexité temporelle (temps d'exécution) de chacune des méthodes est un point important pour décider quelle méthode est la plus efficace à choisir. On calcule donc la fonction grand O de chacune des méthodes proposées dans le chapitre précédent, et on commence par les trois principaux algorithmes (sans détection préalable de communautés) :

**Algorithme 1 :** Il réalise d'abord le tri des noeuds en  $O(n \log n)$ . Ensuite, il effectue au maximum  $n$  (nombre de noeuds) itérations de la boucle *Tant Que*, on a donc une complexité linéaire  $O(n)$ . Dans chacune de ses itérations, on a une suppression d'un noeud qui a une complexité constante  $O(1)$  mais on a aussi un calcul de la performance du graphe qui est une opération de complexité quadratique  $O(n^2)$ .

La complexité de chacune des itérations est donc égale à  $O(n^2 + 1) = O(n^2)$ . La complexité totale de l'algorithme 1 est alors de  $O(n \times n^2) = O(n^3)$ .

**Algorithme 2 :** Quant à l'algorithme 2, le tri coûte  $O(m \log m)$  et le nombre maximal d'itérations de la boucle *Tant Que* est de complexité linéaire par rapport à  $m$  qui représente le nombre de liens du réseau (puisque on parcourt les arcs et non pas les noeuds), mais la complexité de chacune des itérations ne change pas et vaut  $O(n^2)$ . La complexité totale de l'algorithme 2 est de  $O(m \times n^2)$ .

**Algorithme 3 :** Le tri coûte  $O(m \log m)$  et le nombre d'itérations dans l'algorithme 3 est au maximum égal à  $n$ , on a donc une complexité linéaire  $O(n)$ . Pour chacune des itérations, la complexité vaut  $O(n^2)$ , ce qui nous amène à une complexité globale de l'algorithme  $O(n \times n^2) = O(n^3)$  comme pour l'algorithme 1.

Généralement, on a un nombre de liens  $m$  supérieur au nombre de noeuds  $n$ , on a donc  $O(m \times n^2) > O(n^3)$ . Les algorithmes 1 et 3 prennent donc moins de temps que l'algorithme 2, sauf lorsque les réseaux ont une densité très faible.

**Les variantes :** On passe ensuite au calcul de la complexité des variantes des principales méthodes (avec détection de communautés).

Quant à la première variante  $M_{11}$  : La variante du premier algorithme effectue séquentiellement :

1. Une détection de communautés qui a une complexité de  $O(n \times m)$  pour l'utilisation de *Walktrap*
2. Une boucle *Pour* sur toutes les communautés laquelle est une opération linéaire sachant que le nombre des communautés est largement inférieur au nombre des noeuds. La complexité est alors  $O(n)$ .
3. Une boucle *Pour* sur les noeuds supprimés qui est une opération de complexité  $O(n)$ . On note que dans les itérations de cette boucle, on a seulement des opérations de complexité constante  $O(1)$  qui ne change pas la complexité totale de la boucle.

4. Un calcul de la performance du graphe qui est une opération quadratique  $O(n^2)$ .

La complexité globale de  $M_{11}$  est  $O(nm + n + n + n^2)$ , en supposant que  $m$  qui est le nombre total des liens est toujours supérieur à  $n$  représentant le nombre de noeuds. On a donc une complexité de  $O(n \times m)$ .

En ce qui concerne les trois variantes de la deuxième méthode, à savoir  $M_{21}$ ,  $M_{22}$  et  $M_{23}$ , on a les mêmes étapes 1 et 4 que la variante  $M_{11}$  précédemment énumérées, on n'a plus l'étape 2, et l'étape 3 s'effectue sur les liens supprimés et non pas sur les noeuds. La complexité de l'étape 3 est donc  $O(m)$  au lieu de  $O(n)$ . La complexité globale est donc égale à  $O(nm + m + n^2) = O(nm)$ .

Pour les variantes  $M_{31}$ ,  $M_{32}$  et  $M_{33}$  de la méthode 3, on a exactement les étapes 1, 3 et 4 de la variante  $M_{11}$ . La complexité globale est égale à  $O(nm + n + n^2)$  et donc d'ordre  $O(nm)$ .

On remarque que sur toutes les variantes, c'est le coût de la détection des communautés qui hausse la complexité des méthodes.

Aussi, on remarque qu'on a toujours une complexité faible pour les variantes par rapport aux algorithmes principaux seulement parce que ces derniers se chargent de définir le nombre de suppressions à effectuer par ces variantes ultérieurement. Si les variantes vont être plus tard utilisées de manière indépendante, on devra effectuer une mesure de performance du réseau après chaque suppression et on ajoute le coût quadratique de cette opération à l'étape 3 dans toutes les variantes. Le coût de cette étape sera donc égal à  $O(n^3)$  pour les variantes  $M_{11}$ ,  $M_{31}$ ,  $M_{32}$  et  $M_{33}$  et égale à  $O(m * n^2)$  pour les variantes  $M_{21}$ ,  $M_{22}$  et  $M_{23}$ .

Le coût global des variantes sera donc aussi égale à  $O(n^3)$  pour les variantes  $M_{11}$ ,  $M_{31}$ ,  $M_{32}$  et  $M_{33}$ , et égale à  $O(m * n^2)$  pour les variantes  $M_{21}$ ,  $M_{22}$  et  $M_{23}$ . Dans ce cas, le coût de chaque variante serait toujours égale à sa méthode principale.

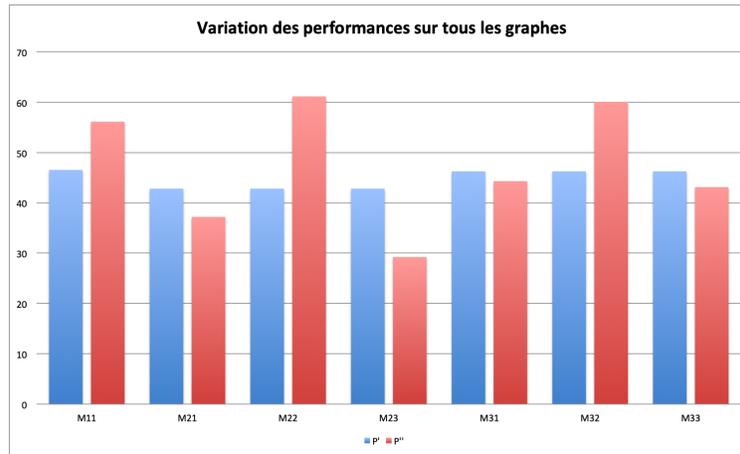
On note donc au final que la mesure de performance donne le coût le plus important à toutes les méthodes proposées mais qu'en pratique, il sera plus faible sachant que les réseaux ont généralement une faible densité et donc un nombre de liens davantage de l'ordre de  $O(n)$  que de  $O(n^2)$ .

### 5.2.2 Tests et résultats

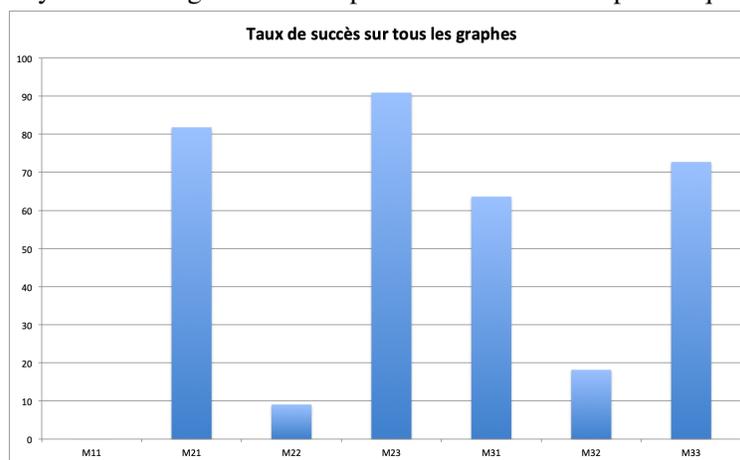
Pour optimiser l'exécution des algorithmes, nous avons créé deux copies de chaque graphe. L'un en utilisant la librairie *Networkx* et l'autre en utilisant la librairie *Python-iGraph*. Ainsi, nous effectuons chaque calcul en utilisant la librairie qui optimise le plus ce calcul. Nous effectuons chaque suppression dans les deux copies pour garder la conformité des deux graphes. Pour comparer l'impact de la détection de communautés dans la désintégration du réseau, on procède, pour chaque test, comme suit :

1. On désintègre le réseau avec chacune des trois méthodes principales sans considérer les communautés et on garde la performance obtenue ainsi que le nombre de suppressions effectuées.
2. Pour chacune des sept variantes proposées au chapitre 4, on identifie les communautés et on procède au retrait du même nombre d'éléments obtenu à la première phase (sans considération des communautés) comme suit : méthode 1 pour la variante  $M_{11}$  ; méthode 2 pour les variantes  $M_{21}$ ,  $M_{22}$  et  $M_{23}$  et méthode 3 pour les variantes  $M_{31}$ ,  $M_{32}$  et  $M_{33}$  et on garde les performances obtenues.
3. On compare la dégradation de la performance obtenue avec détection de communautés ( $P''$  obtenue à l'étape 2) par rapport à celle obtenue sans considération des communautés ( $P'$  obtenue à l'étape 1) de chaque méthode. Si  $P'' < P'$ , on conclut que la détection de communautés est un succès puisqu'elle a assuré une meilleure désintégration du réseau. Sinon, il s'agit d'un échec.

Pour chaque variante et pour chaque graphe, on a effectué trois exécutions. La figure 5.4a montre la dégradation de la performance  $P'$  (barre en bleu) sans considération des communautés et la nouvelle (dégradation de) performance  $P''$  (barre rouge) en considérant les communautés. L'axe des  $X$  dans les histogrammes présente les sept variantes des méthodes proposées. L'axe des  $Y$  présente la proportion de la nouvelle performance par rapport à celle du graphe initial avant désintégration. La figure 5.4b indique le taux de succès de la prise en compte des communautés pour les diverses variantes du démantèlement. Ce taux est calculé en divisant le nombre de succès obtenus ( $P'' < P'$ ) sur le nombre de tests réalisés pour un même graphe.



(a) La moyenne des dégradations de performances obtenue par chaque variante



(b) Le taux de succès de chaque variante

FIGURE 5.4 – Statistiques générales sur toutes les exécutions

On remarque que les variantes  $M_{21}$  et  $M_{23}$  sont les deux variantes qui offrent une dégradation de performance favorisant l'utilisation de la détection des communautés comme étape préliminaire de désintégration du réseau. De même, les variantes  $M_{31}$ ,  $M_{33}$  offrent aussi un taux de succès acceptable mais minime. Ces résultats ne peuvent être définitifs puisque les graphes que nous avons utilisés diffèrent beaucoup l'un de l'autre en terme de type et de densité.

Aussi, nous effectuons des tests additionnels pour chaque type de graphe utilisé, à savoir les réseaux *réels*, *aléatoires*, *invariants d'échelle* et *petit-monde*. Les résultats sont affichés dans les figures 5.5, 5.6, 5.7 et 5.8.

Pour les deux réseaux réels, le premier (résidents d'un campus australien) qui a une densité 5.7% semble vulnérable aux attaques par la centralité d'intermédiarité. En effet, les résultats des va-

---

riantes  $M_{21}$  et  $M_{31}$  sont les plus intéressants en terme de dégradation du réseau. Ces deux variantes suppriment les liens intercommunautaires, ce qui favorise la formation de composants isolés. Par contre, le deuxième réseau réel des élèves, ayant une densité faible de 0.2% est vulnérable aux attaques qui visent les liens sans distinction. Les variantes  $M_{23}$  et  $M_{33}$  qui recalculent la centralité d'intermédiarité des liens en tenant compte du coefficient d'appartenance aux communautés ciblent en premier lieu les liens intracommunautaires les plus centraux des communautés pour ensuite cibler les ponts entre les diverses communautés afin d'isoler ces dernières. De ce fait, la dégradation du réseau se fait à deux niveaux simultanément, à savoir l'affaiblissement des sous-groupes et la suppression des liens entre eux.

Le réseau aléatoire (cf. chapitre 2) est un réseau créé par un seul nœud isolé et des arêtes ajoutées entre des paires de nœuds sélectionnés de manière uniformément aléatoire. Les tests menés (figure 5.6) sur ce type de graphe ayant une distance géodésique et un coefficient de regroupement moyennement faibles montrent que le réseau est mieux désintégré sans la détection préalable des communautés. La variante  $M_{11}$  est de ce fait moins performante. Ceci vient du fait que la répartition des liens est quasiment uniforme entre les nœuds. Les variantes ciblant les liens intra ( $M_{22}$  et  $M_{32}$ ) semblent les moins pertinentes car en déconnectant les membres d'une même communauté tout en laissant des liens entre ces communautés, le flux d'information reste transmissible. Par contre, les variantes qui présentent la suppression des liens intercommunautaires ( $M_{21}$ ) et la suppression des liens intercommunautaires et leurs nœuds adjacents ( $M_{31}$ ) améliorent la désintégration du réseau. Ceci peut être lié à la densité élevée des réseaux aléatoires qui sont vulnérables aux attaques ciblant les ponts de communication entre les communautés. Une telle attaque engendre la formation des sous-groupes isolés dont leurs nœuds sont fortement connectés. Les variantes  $M_{23}$  et  $M_{33}$  qui ciblent les liens d'intermédiarité pondérée donnent le même niveau de démantèlement que sans détection de communautés.

Comme présenté précédemment, le réseau invariant d'échelle est vulnérable aux attaques des liens (surtout les ponts) et des nœuds clés. Cependant, ces caractéristiques ne se vérifient pas lors de la considération des communautés pour les variantes  $M_{11}$  et  $M_{21}$ . Par contre, les variantes  $M_{23}$  et  $M_{33}$  qui éliminent les liens dont l'intermédiarité pondérée est la plus importante sans distinction entre liens intercommunautaires et intracommunautaires offrent une meilleure dégradation.

Quant au réseau petit monde dont la distance moyenne entre toute paire de sommets est faible et le niveau de regroupement local est élevé, les expérimentations montrent que ce réseau est vul-

---

néralable aux attaques surtout des ponts (liens centraux) mais pas des sommets qui sont généralement très connectés. Ceci est montré par les résultats de la désintégration du réseau pour les variantes  $M_{23}$  et  $M_{33}$  qui éliminent les liens dont l'intermédiarité pondérée est la plus importante sans distinction entre liens intercommunautaires et intracommunautaires. Cette intermédiarité varie selon l'importance de la communauté à laquelle un lien appartient.

Les résultats obtenus suite à la classification des graphes réels versus des graphes synthétiques n'offrent pas de nouvelles précisions quant aux meilleures variantes de désintégration, mais ils éliminent définitivement les variantes  $M_{11}$ ,  $M_{22}$  et  $M_{32}$  qui n'ont jamais amélioré la désintégration grâce à la détection des communautés en ciblant les liens intracommunautaires.

Pour tester l'impact de la densité des réseaux sur les variantes, nous avons considéré des graphes ayant une densité supérieure à 50% et ceux avec une densité inférieure à 50%. Les statistiques obtenues sont affichées dans les figures 5.9 et 5.10.

On remarque que les variantes  $M_{21}$  et  $M_{31}$  favorisent la dégradation des réseaux pour les graphes de densité supérieure à 50%. Ces deux variantes ciblent les liens intercommunautaires avec ou sans leurs nœuds adjacents. Ainsi, pour déstabiliser les réseaux ayant une densité importante, il semble avantageux de cibler les ponts entre communautés pour les isoler. Un sous-groupe isolé est incapable de diffuser l'information dans le réseau au complet même si ses composants sont fortement connectés. Par contre, en présence d'une faible densité, les réseaux semblent vulnérables aux attaques par l'intermédiarité des liens mais sans distinction entre liens intercommunautaires et intracommunautaires. Les variantes  $M_{23}$  et  $M_{33}$  sont probablement mieux appropriées pour les réseaux de faible densité.

Le tableau 5.12 présente le temps d'exécution pour les différentes variantes proposées avec et sans détection des communautés. Les colonnes du tableau représentent les différents réseaux alors que les lignes expriment les variantes testées de  $M_{11}$  à  $M_{33}$ .  $G'$  et  $G''$  présentent respectivement le réseau après démantèlement sans et avec détection préalable des communautés.

Les quatre réseaux qu'on a choisis de tester comme exemple pour le calcul de temps d'exécution sont : le graphe réel de 217 nœuds et 2672 liens (densité de 5.7%), un graphe aléatoire de 1000 nœuds et 246239 liens (densité presque égale à 25%), un graphe invariant d'échelle de 1000 nœuds et 1655 liens et enfin un graphe petit-monde de 1000 nœuds et 4000 liens et  $k=5$ .

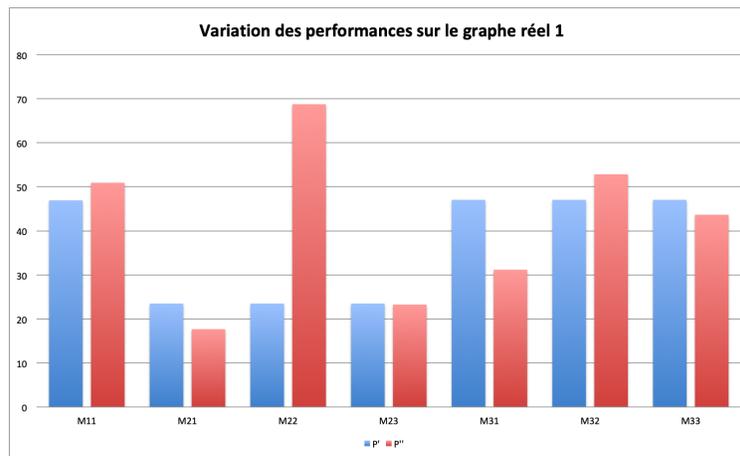
On remarque que la première colonne de chaque réseau ( $G'$  : sans considération de communautés) montre que les temps d'exécution de la méthode 1 et la méthode 3 ont le même ordre de grandeur

alors que la méthode 2 prend considérablement plus de temps. Cela confirme notre étude de complexité. La différence est quand même moins considérable pour le graphe invariant d'échelle car cela est probablement dû au fait que le nombre de liens pour ce réseau est assez proche du nombre de nœuds (1655 et 1000 respectivement).

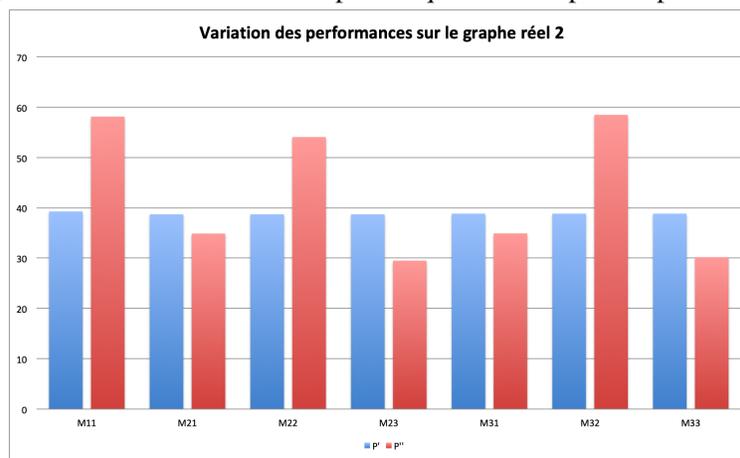
Quand à la deuxième colonne de chaque réseau ( $G''$  : avec considération de communautés), on remarque que toutes les méthodes ont le même ordre de grandeur. C'est ce qu'on avait prévu dans l'étude de complexité vu que la deuxième étape avait une complexité  $O(n * m)$  pour toutes les variantes.

Variante	Réel		Aléatoire		Invariant d'échelle		Petit-Monde	
	$G'$	$G''$	$G'$	$G''$	$G'$	$G''$	$G'$	$G''$
$M_{11}$	36'694.565	3'431.04	1'098.99	444.525	1'078.926	162.909	1'016.799	167.914
$M_{21}$	359'477.705	3'856.721	8'468.48	439.282	1'161.56	165.306	3'600.375	170.738
$M_{22}$	359'477.705	3'846.785	8'468.48	439.487	1'161.56	165.356	3'600.375	172.706
$M_{23}$	359'477.705	3'829.785	8'468.48	503.71	1'161.56	165.385	3'600.375	172.711
$M_{31}$	37'849.893	3'543.091	1'133.129	400.219	1'063.8	154.96	1'002.771	165.131
$M_{32}$	37'849.893	3'542.076	1'133.129	396.45	1'063.8	99.947	1'002.771	166.462
$M_{33}$	37'849.893	3'549.834	1'133.129	458.691	1'063.8	94.323	1'002.771	163.961

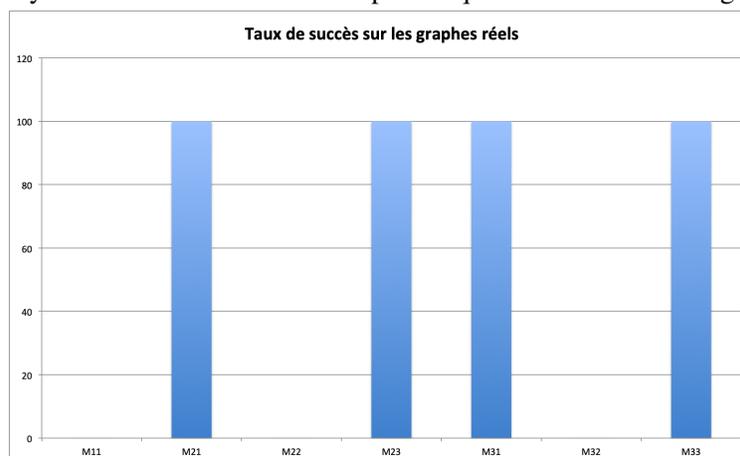
TABLE 5.12 – Temps d'exécution en secondes



(a) La moyenne des variations obtenue par chaque variante pour le premier graphe réel

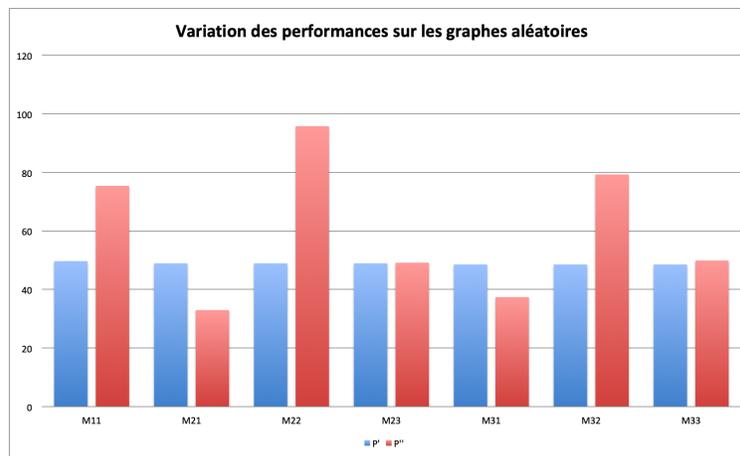


(b) La moyenne des variations obtenue par chaque variante le second graphe réel

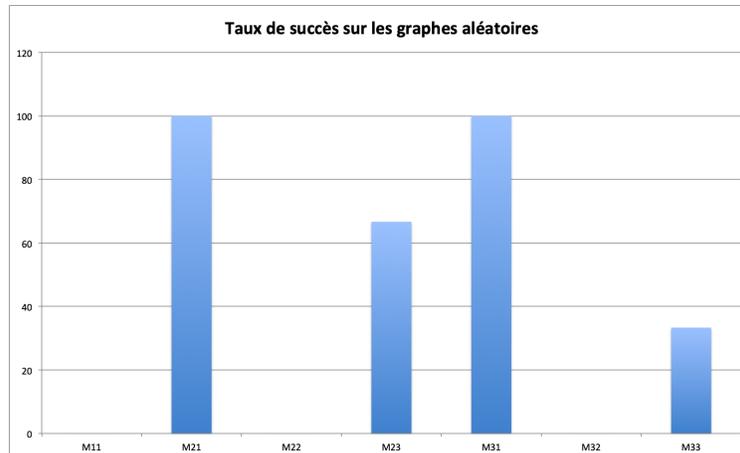


(c) Le taux de succès de chaque variante

FIGURE 5.5 – Statistiques sur les graphes réels

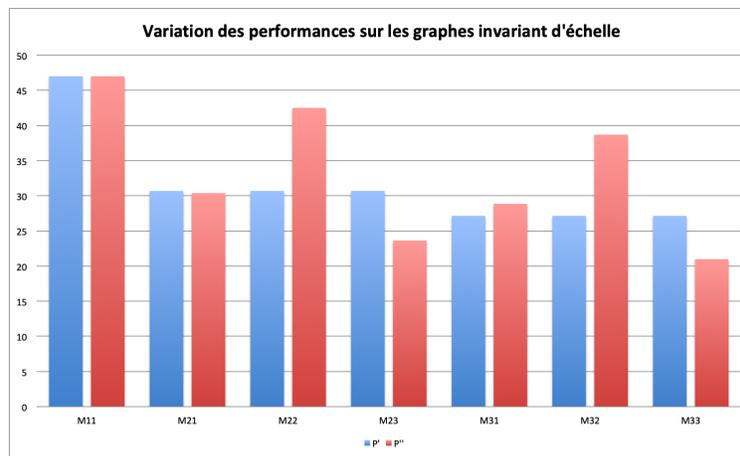


(a) La moyenne des variations obtenue par chaque variante

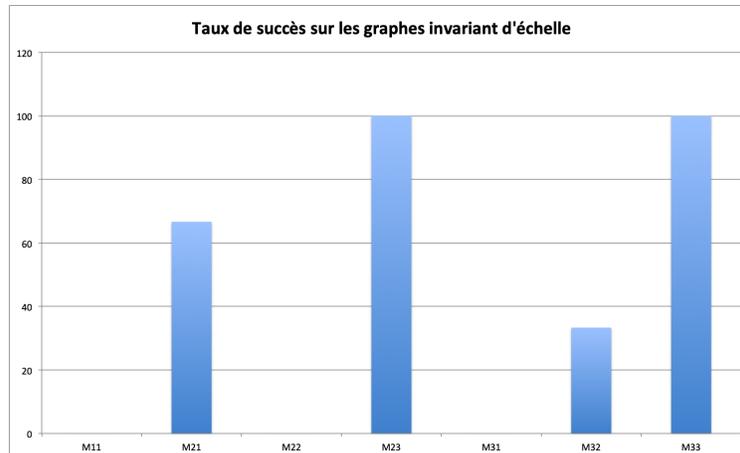


(b) Le taux de succès de chaque variante

FIGURE 5.6 – Statistiques sur les graphes synthétiques

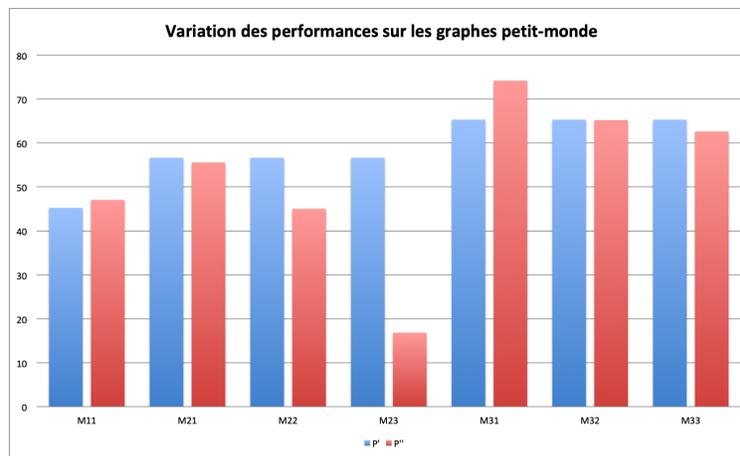


(a) La moyenne des variations obtenue par chaque variante

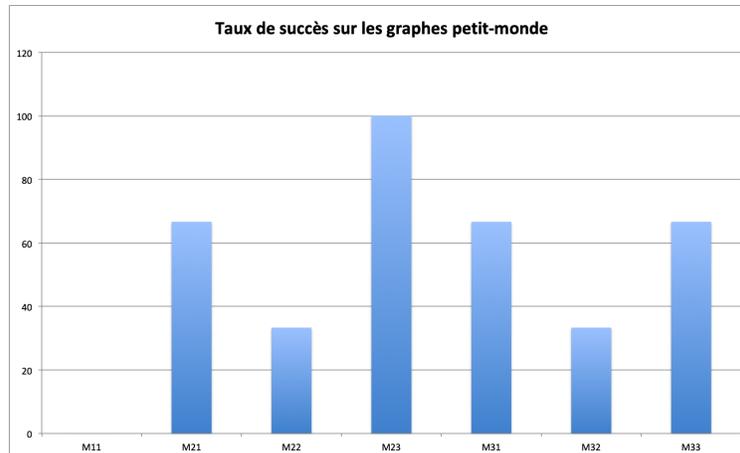


(b) Le taux de succès de chaque variante

FIGURE 5.7 – Statistiques sur les graphes synthétiques

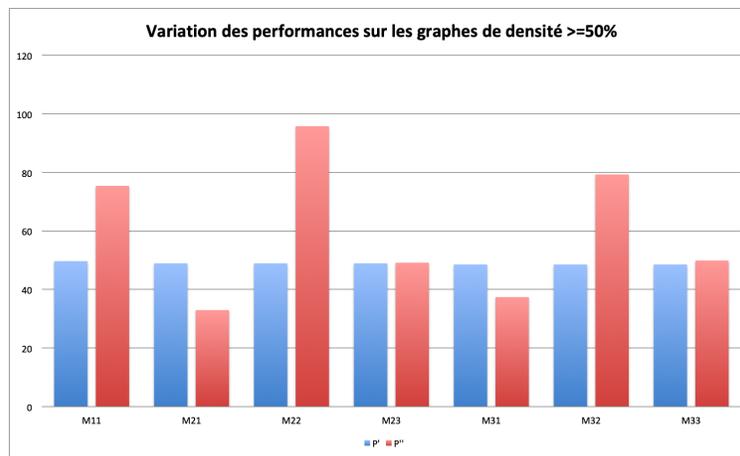


(a) La moyenne des variations obtenue par chaque variante

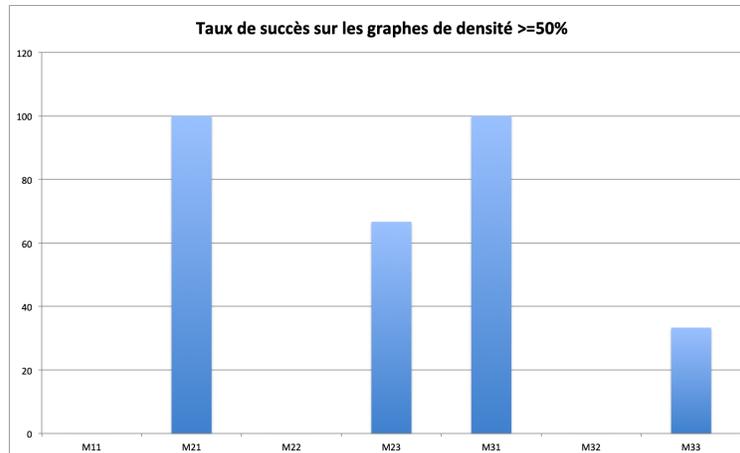


(b) Le taux de succès de chaque variante

FIGURE 5.8 – Statistiques sur les graphes synthétiques

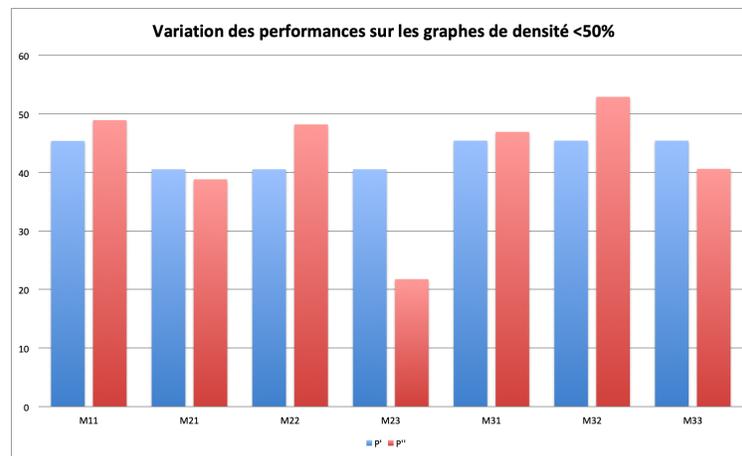


(a) La moyenne des variations obtenue par chaque variante

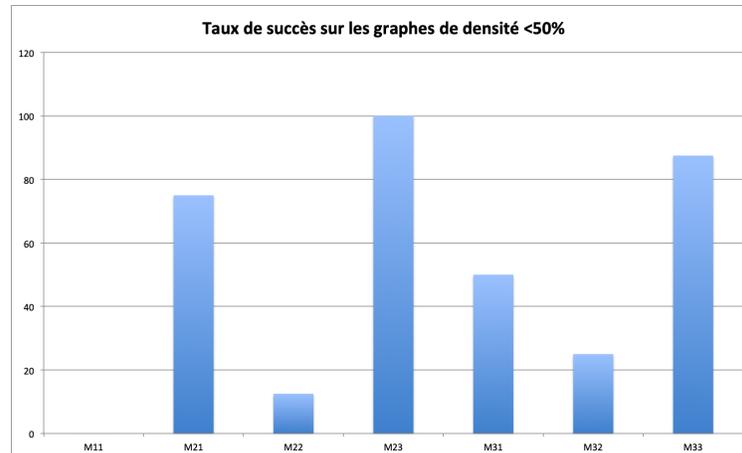


(b) Le taux de succès de chaque variante

FIGURE 5.9 – Statistiques sur les graphes synthétique de densité supérieure à 50%



(a) La moyenne des variations obtenue par chaque variante



(b) Le taux de succès de chaque variante

FIGURE 5.10 – Statistiques sur les graphes synthétiques de densité inférieure à 50%

# Chapitre 6

## Conclusion et travaux futurs

### 6.1 Conclusion

Dans le cadre de ce mémoire, nous avons présenté une approche descendante de démantèlement d'un réseau social. Cette approche identifie les communautés pour ensuite cibler et éliminer les noeuds et/ou liens importants dans les sous-groupes dans la perspective d'obtenir une désintégration plus prononcée que si les suppressions s'effectuaient au niveau global du réseau. Pour une identification adéquate des communautés, nous avons mené des comparaisons avec les trois algorithmes : *Walktrap*, *Louvain* et *Label Propagation*. Le choix de ces algorithmes est dicté d'une part, par le fait qu'ils traitent le type de réseau choisi, à savoir unimodal orienté et pondéré, et d'autre part, ils sont également très connus et considérés comme les plus performants et les plus précis.

Nous avons proposé trois méthodes de démantèlement d'un réseau social en s'attaquant aux noeuds et/ou liens importants avec ou sans considération des communautés existantes. Cela a débouché vers la proposition de diverses variantes. Pour identifier les éléments clés du réseau, nous avons tenu compte de mesures de centralité qui nous paraissent les plus aptes à repérer les noeuds et/ou liens centraux. Quant à l'estimation de la robustesse du réseau avant et après démantèlement, nous avons identifié les mesures les plus crédibles à savoir la performance totale, la compacité et le taux de connectivité.

Nous avons mené des tests de dégradation de la robustesse (ou performance) sur deux réseaux réels et trois types de réseaux synthétiques, à savoir le réseau petit-monde, le réseau invariant d'échelle et le réseau aléatoire avec des densités variables. Les résultats obtenus indiquent que la détection de communautés peut effectivement améliorer la désintégration d'un réseau en supprimant les liens

---

inter-communautaires ou encore en ciblant les nœuds adjacents à des liens inter-communautaires. On remarque qu'avec les variantes  $M_{22}$  et  $M_{31}$  qui concernent respectivement la suppression des liens intracommunautaires ( $M_{22}$ ) et des liens intercommunautaires avec leurs nœuds adjacents ( $M_{31}$ ), la détection de communautés améliore la désintégration du graphe puisque les graphes les plus démantelés en terme de réduction de la performance globale sont ceux obtenus grâce à ces variantes. Elles ont en commun la suppression des liens intercommunautaires. Par contre, nous notons que les variantes ciblant les liens et/ou nœuds au sein des communautés n'offrent aucune amélioration à la désintégration. Dans ces cas, la détection de communautés ne contribue nullement à la désintégration et aboutit même à des graphes plus performants que ceux obtenus sans la détection des communautés.

Quant aux variantes  $M_{23}$  et  $M_{33}$  qui ne différencient pas entre liens intercommunautaires et intracommunautaires mais offrent plutôt un nouvelle centralité pondérée par un coefficient d'appartenance à une communauté, elles donnent une amélioration comparée à la désintégration dans le réseau entier.

Quant aux différents types du réseaux testés, les expérimentations montrent que pour le réseau aléatoire, les variantes  $M_{21}$  et  $M_{31}$  qui ciblent les liens intercommunautaires semblent les plus prometteuses. Ceci est dû à la vulnérabilité de ce réseau aux attaques de ponts. Pour le réseau invariant d'échelle et celui du petit-monde, les variantes  $M_{23}$  et  $M_{33}$  qui ciblent toutes les deux les liens avec une centralité d'intermédierité pondérée (par l'appartenance d'un sommet à une communauté) semblent les plus aptes à désintégrer ces réseaux après prise en compte de communautés.

## 6.2 Travaux futurs

Nous comptons explorer les pistes de recherche suivantes :

- initier le processus de démantèlement par une restauration préalable du réseau suivie de la détection de ses communautés. Cette étape aura pour but d'avoir au préalable rajouté de nouveaux liens susceptibles d'apparaître dans le futur ou d'avoir été omis ou caché lors de la collecte ou l'analyse de données. De ce fait, la détection des communautés qui suit la restauration peut générer des sous-groupes stables dont les nœuds sont fortement connectés.
- découvrir, adapter ou utiliser d'autres mesures de robustesse d'un réseau orienté et pondéré
- proposer une formule du calcul du seuil de dégradation de la performance du réseau (au lieu de demander à l'utilisateur une valeur subjective) via l'inertie interclasse ou à travers des stratégies d'attaques optimales basées sur la recherche Tabu [8].

- 
- mener des tests empiriques plus approfondis sur de plus grands réseaux réels et synthétiques dans le but d’obtenir des résultats plus concluants sur le potentiel des sept variantes des méthodes proposées dans la désintégration d’un réseau avec ou sans identification préalable de communautés.

# Bibliographie

- [1] Bahulkar, A., Szymanski, B. K., Baycik, N. O. et Sharkey, T. C. [2018], Community detection with edge augmentation in criminal networks, *in* U. Brandes, C. Reddy et A. Tagarelli, eds, 'IEEE/ACM 2018 International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018, Barcelona, Spain, August 28-31, 2018', IEEE Computer Society, pp. 1168–1175.
- [2] Berzinji, A., Kaati, L. et Rezine, A. [2012], Detecting key players in terrorist networks, pp. 297–302.
- [3] Blondel, V. D., Guillaume, J.-L., Lambiotte, R. et Lefebvre, E. [2008], 'Fast unfolding of communities in large networks', *Journal of statistical mechanics : theory and experiment* **2008**(10), P10008.
- [4] Bonacich, P. [1987], 'Power and centrality : A family of measures', *American Journal of Sociology* **92**(5), 1170–1182.
- [5] Bouguessa, M., Missaoui, R. et Talbi, M. [2014], A novel approach for detecting community structure in networks, *in* '26th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2014, Limassol, Cyprus, November 10-12, 2014', pp. 469–477.
- [6] Chaurasia, N. et Tiwari, A. [2013], 'Efficient algorithm for destabilization of terrorist networks', *International Journal Information Technology and Computer Science (IJITCS)* **5**, 21–30.
- [7] Clauset, A., Newman, M. et Moore, C. [2005], 'Finding community structure in very large networks', *Physical review. E, Statistical, nonlinear, and soft matter physics* **70**, 066111.
- [8] Deng, Y., Wu, J. et jin Tan, Y. [2016], 'Optimal attack strategy of complex networks based on tabu search', *Physica A : Statistical Mechanics and its Applications* **442**(C), 74–81.
- [9] Diestel, R. [2010], *Graph Theory*, Vol. 173 of *Graduate Texts in Mathematics*, Springer, Heidelberg ; New York.
- [10] Ellens, W. et Kooij, R. [2013], 'Graph measures and network robustness'.
- [11] Freeman, L. C. [1978], 'Centrality in social networks conceptual clarification', *Social Networks* p. 215.

- [12] Freeman, L. C., Webster, C. M. et Kirke, D. M. [1998], ‘Exploring social structure using dynamic three-dimensional color images.’  
**URL:** [http://konect.uni-koblenz.de/networks/moreno\\_o\\_z](http://konect.uni-koblenz.de/networks/moreno_o_z),
- [13] Freeman, S. C. et Freeman, L. C. [Residence hall network dataset – KONECT, April 2017], ‘The networkers network : A study of the impact of a new communications medium on sociometric structure.’  
**URL:** [http://konect.uni-koblenz.de/networks/moreno\\_h\\_ealth](http://konect.uni-koblenz.de/networks/moreno_h_ealth).
- [14] Han, J., Kamber, M. et Pei, J. [2011], *Data Mining : Concepts and Techniques, 3rd edition*, Morgan Kaufmann.  
**URL:** <http://hanj.cs.illinois.edu/bk3/>
- [15] Hidalgo, C. A. et Barabasi, A. [2008], ‘Scale-free networks’, *Scholarpedia* **3**(1), 1716.
- [16] Hopkins, A., Advisor, Fine, D. S. et Hopkins [2010], Graph theory, social networks and counter terrorism, Technical report, University of Massachusetts Dartmouth.
- [17] Katz, L. [1953], ‘A new status index derived from sociometric analysis’, *Psychometrika* **18**(1), 39–43.
- [18] Kong, T. [2016], Déstabilisation d’un réseau sociale par élimination de nœuds et de liens clés, Master’s thesis, UQO. mémoire de maîtrise.
- [19] Lindelauf, R., Borm, P. et Hamers, H. [2008], ‘The influence of secrecy on the communication structure of covert networks’, *Social Networks* **31**.
- [20] Memon, N. et Larsen, H. [2006], Practical algorithms for destabilizing terrorist networks, pp. 389–400.
- [21] Newman, M. [2010], *Networks : An Introduction*, Oxford University Press, Inc., New York, NY, USA.
- [22] Pons, P. et Latapy, M. [2006], ‘Computing communities in large networks using random walks.’, *J. Graph Algorithms Appl.* **10**, 191–218.
- [23] Sarr, I., Missaoui, R. et Lagarriague, A. [2015], ‘Déstabilisation parcimonieuse d’un réseau social par effet cascade’, *manuscrit* .
- [24] Tiwari, S., Shrivastava, S. et Gupta, M. [2015], ‘A new fuzzy based approach for destabilization of terrorist network’, *International Journal of Computer Applications* **129**, 21–27.
- [25] Wiil, U., Gniadek, J. et Memon, N. [2010], Measuring link importance in terrorist networks, pp. 225–232.
- [26] Xu, J. et Chen, H.-c. [2005], ‘Criminal network analysis and visualization’, *Commun. ACM* **48**, 100–107.