UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

DÉSTABILISATION D'UN RÉSEAU SOCIAL PAR ÉLIMINATION DE NŒUDS ET DE LIENS CLÉS

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN SCIENCES ET TECHNOLOGIES DE L'INFORMATION

PAR

TITHARY KONG

OCTOBRE 2016

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 : 27 octobre 2016

Remerciements

Cette recherche fut intéressante et très enrichissante. Cependant, elle requit beaucoup de travail et de temps, ce qui pourrait facilement décourager une étudiante à temps partiel telle que moi. Heureusement, ma directrice de mémoire, le Dr. Missaoui a su m'encourager et me motiver tout au long de ce mémoire. Je tiens à la remercier pour son dévouement, ses conseils et recommandations. Grâce à ce mémoire, j'ai acquis de nouvelles connaissances qui contribueront à diversifier mes compétences en informatique.

D'autre part, une grande partie de cette recherche a été réalisée grâce au langage R. De ce fait, j'aimerais remercier toute l'équipe qui a implémenté cet outil de développement et de l'avoir mis à la disposition du grand public. La qualité et la disponibilité de la documentation m'ont permis d'apprendre et de me servir de cet outil pour mener diverses expérimentations.

Table des matières

Liste des figures

Liste des tableaux

Résumé

L'analyse de réseaux sociaux se base sur plusieurs théories dont la théorie des graphes pour étudier diverses facettes de ces réseaux telles que la détection de communautés, l'identification d'acteurs influents, la prédiction de l'évolution des structures, l'estimation de la fragilité ou de la robustesse d'un réseau ainsi que sa déstabilisation volontaire.

La déstabilisation d'un réseau social consiste à démanteler une structure interconnectée en vue d'affaiblir sa cohésion par la formation de plusieurs composantes déconnectées et/ou l'élimination d'acteurs clés, réduisant ainsi la propagation de l'information ou de l'influence au sein du réseau. Ce phénomène se produit dans des situations réelles comme le démantèlement de réseaux de criminels ou l'affaiblissement de réseaux rivaux.

Le présent mémoire de maîtrise vise à explorer et implémenter quelques variantes d'une nouvelle méthode de déstabilisation d'un réseau social orienté et pondéré. Celle-ci identifie un certain nombre de nœuds et d'arêtes à éliminer dans le but de réduire la cohésion du réseau et la circulation de l'information. Les aspects étudiés couvrent l'identification de la mesure de centralité des nœuds et des liens à adopter et la proposition d'une nouvelle formule d'estimation de l'ampleur de la désintégration du réseau. La validation de la solution est effectuée sur des réseaux réels ainsi que des réseaux générés synthétiquement selon deux types connus: petit-monde (*small-world*) et aléatoire.

Abstract

Social network analysis is based on several theories, including the graph theory, to study various aspects of these networks such as community detection, key node identification, prediction of network evolution, network fragility, robustness and destabilization.

Social network destabilization aims at reducing its cohesiveness structure by producing disconnected components and/or removing key players. This phenomenon happens in real situations such as criminal network dismantling or competitor networks weakening. The expected outcome is the reduction of the information and the influence inside the network.

This Master Degree thesis intends to investigate and implement a few variants of a new network destabilization strategy. The solution identifies nodes and links to remove in order to significantly reduce the network cohesiveness and decrease the flow of information. The aspects to be studied include the computation of nodes and links importance, taking into consideration directed and weighted links in the dismantling process and designing a new formula to compute the degree of the network destabilization. The strategy will be validated on real networks as well as generated networks based on two models: small-world and random networks.

Chapitre 1

Introduction

Nous vivons dans un monde où la technologie ne cesse de progresser. Ces progrès technologiques ont changé notre façon de vivre et de travailler. Prenons l'internet par exemple, cette technologie a sans conteste révolutionné notre façon de communiquer. Grâce à l'internet, on peut communiquer de façon électronique permettant ainsi d'acheminer un message à son destinataire presqu'instantanément. Cette communication électronique a apporté des bienfaits considérables. En effet, on peut maintenant communiquer plus facilement avec les personnes éloignées géographiquement. De plus, tel que mentionné dans le rapport 2013 de France Digital [30], l'internet a mis en place des réseaux sociaux comme *Facebook*, *Twitter*, *LinkedIn* dont le nombre de connexions ne cesse de grimper. Ces réseaux facilitent la communication et permettent aux membres éloignés tels des parents et amis d'entretenir divers types de liens. Le grand public voit donc de grands avantages à adhérer à ces réseaux.

Selon Durland et Fredericks [12], la popularité des réseaux sociaux sur internet ainsi que la disponibilité de leurs données constituent l'un des facteurs contribuant à l'augmentation de l'intérêt porté à l'analyse des réseaux sociaux dans les dernières décennies. Les auteurs de [31] mentionnent que cette analyse trouve application dans plusieurs domaines incluant la sociologie, l'anthropologie, la politique, l'économie, la criminologie, le marketing, etc.

L'analyse de réseaux sociaux comporte plusieurs thèmes tels que l'importance des nœuds et des liens, la détection et l'évolution de communautés ainsi que la prédiction de l'évolution des réseaux. D'une vue globale [12], cette analyse est utilisée pour, entre

autres, identifier les meneurs dans une organisation, mesurer la collaboration dans les équipes, identifier la structure cachée du terrorisme, planifier un réseau de transport et étudier la propagation de maladies et de virus dans le domaine de la santé publique.

Dans la présente recherche, nous concentrons l'effort sur l'identification de meneurs dans une organisation. Nous étudions les techniques d'analyse de réseaux sociaux et de graphes afin de développer et tester une approche permettant de cibler les nœuds et les liens clés dont l'élimination mène à une déstabilisation rapide du réseau.

1.1 Champs d'application

La déstabilisation d'un réseau est applicable dans de multitudes de domaines. Nous ne citons que quelques-uns dans cette section.

Un des principaux champs d'application de la déstabilisation d'un réseau social est le combat contre les réseaux de terroristes ainsi que d'autres réseaux de crimes organisés. De nos jours, nous savons que les forces de l'ordre travaillent sans relâche pour démanteler les réseaux de trafiquants de drogues, de proxénètes ou d'exploitation juvénile, etc.

La déstabilisation de réseaux sociaux peut également être utilisée dans le domaine de la santé afin de réduire les risques de propagation de maladies. Par exemple, en identifiant les individus à haut risque et en les isolant, on diminue ainsi la probabilité de transmission de maladies. D'ailleurs, les auteurs de [23] ont étudié le contrôle de maladies chez les animaux de ferme en se servant l'analyse de réseaux sociaux. Ces derniers affirment que les fermes qui sont fortement connectées à d'autres fermes sont plus susceptibles de propager les maladies. En retirant ou en isolant ces dernières du réseau de fermes, on réduit la probabilité d'épidémie dans la population animale.

Bien qu'elle offre diverses utilités légales et légitimes, la déstabilisation de réseaux sociaux peut cependant être exploitée par des individus ou organisations malhonnêtes ayant pour objectif d'affaiblir les organisations rivales afin de prendre le monopole sur un champ d'activité quelconque. C'est un phénomène qui se produit dans des

situations réelles comme la guerre entre des clans de motards ou de la mafia.

Dans les exemples précédents, nous avons vu que la déstabilisation est appliquée dans le but d'affaiblir un réseau. Mais il faut noter qu'elle peut aussi être utilisée pour étudier la robustesse des réseaux tels que les réseaux de télécommunication ou d'ordinateurs. Par exemple, elle peut être utilisée pour simuler un bris par le retrait d'un nœud afin de mesurer et analyser le fonctionnement du réseau suite à une défectuosité éventuelle. Dans [32], les auteurs ont simulé le retrait des nœuds importants pour étudier la robustesse et l'optimisation d'un réseau. Dans [11], une étude similaire est réalisée par le retrait des liens.

Dans le présent mémoire, nous étudions la déstabilisation dans le contexte de démantèlement de réseaux. Des techniques d'analyse de réseau social (ARS) seront expérimentées pour bien identifier les meneurs ou les acteurs influents dans le but de les retirer du réseau.

1.2 Problématique

On a reconnu l'importance de l'analyse de réseaux sociaux dans le combat contre le terrorisme, bien avant l'attaque du 11 septembre 2001 [29]. John Arquilla et David Ronfeldt [2] ont d'ailleurs publié un livre avant cette attaque. Le livre intitulé *Networks and netwars : The future of terror, crime, and militancy* présentait les concepts de réseaux dans une organisation criminelle moderne. La théorie de ce livre est que dans le monde d'aujourd'hui, la guerre se déroule sur l'internet et est livrée par des groupes de terroristes, de criminels ou d'extrémistes qui forment entre eux des réseaux bien connectés mais malheureusement partiellement visibles aux autorités.

Après l'attaque du 11 septembre 2001, les organismes gouvernementaux ont déployé d'importants efforts dans le domaine de l'analyse de réseaux sociaux pour combattre le terrorisme [29]. Depuis lors, plusieurs études ont été réalisées dans le but de trouver des méthodes efficaces pour la déstabilisation de ce type de réseau.

Bien que plusieurs méthodes aient été développées pour démanteler un réseau social, certaines méthodes n'ont pas encore été explorées. De plus, la plupart des études

réalisées considèrent des réseaux à un seul mode sans orientation ni pondération des liens. Dans le cadre de ce mémoire, nous nous proposons de mettre au point une approche de déstabilisation qui considère les réseaux à un mode de données ayant des liens orientés et pondérés. En effet, ces deux caractéristiques apportent des précisions considérables dans la sélection des nœuds et des liens clés.

Par ailleurs, plusieurs chercheurs reconnaissent que l'analyse des réseaux de criminels ne peut pas se limiter à l'application des mesures typiques de centralité pour identifier et éliminer les nœuds et les liens les plus critiques. C'est la raison pour laquelle nous allons développer, adapter et expérimenter plusieurs mesures identifiées dans le chapitre 2 dans le processus de déstabilisation.

1.3 Contribution

Dans ce mémoire, nous proposons d'expérimenter trois méthodes de déstabilisation d'un réseau social. En premier lieu, nous déstabilisons un réseau par élimination de nœuds clés identifiés par des mesures de centralité que nous verrons en détail dans le prochain chapitre. Dans la deuxième méthode, nous tiendrons compte des communautés dans le processus de déstabilisation du même réseau. Cette méthode identifiera les communautés cliques du réseau puis sélectionne les nœuds appartenant à plusieurs cliques comme cibles à retirer du réseau. Finalement, pour la troisième méthode, le réseau est déstabilisé par élimination de liens clés incluant leurs nœuds adjacents.

Pour valider ces trois méthodes, nous utiliserons les réseaux générés synthétiquement ayant les liens orientés et pondérés. Puisqu'il existe différents modèles de réseaux, nous utilisons les réseaux du petit-monde et les réseaux aléatoires. Nous comparons l'efficacité de chaque méthode sur les réseaux testés afin de déterminer quelle méthode est mieux adaptée pour chacun des deux modèles de réseaux expérimentés. Afin de mesurer l'impact de la déstabilisation, nous proposons une nouvelle formule qui permet d'estimer l'efficacité du réseau et donc sa dégradation suite à l'élimination de nœuds et de liens clés.

1.4 Contenu

Nous débutons ce travail dans le chapitre 2 par un court rappel sur la théorie des graphes et les mesures de centralité. Nous présentons surtout les notions qui sont utilisées dans la présente recherche. Le chapitre 3 expose les méthodes de déstabilisation et de calcul d'efficacité d'un réseau déjà proposées dans la littérature. Dans le chapitre 4, nous présentons notre approche de déstabilisation avec les trois variantes. Par la suite, nous discutons des résultats obtenus à la suite d'une analyse empirique de chacune des méthodes étudiées et nous concluons dans le chapitre 5.

Chapitre 2

Graphes et réseaux

Les réseaux sont omniprésents dans notre monde. À titre d'exemple, on peut nommer les réseaux de neurones, de circulation sanguine, de rivières, d'amitié, de professionnels, de transport, etc. Chaque réseau accomplit une fonction bien spécifique. Par exemple, l'objectif du réseau de transport est d'assurer la liaison entre chaque site du réseau, alors que le réseau d'amitié permet à ses membres de garder contact et de partager de multiples informations.

Un réseau est un ensemble de points reliés entre eux par des lignes. Les points sont appelés nœuds ou sommets et les lignes sont appelées des liens ou des arêtes [7, 31]. Dans un réseau social, les nœuds représentent les individus à l'intérieur du réseau et les liens représentent les relations que ceux-ci entretiennent entre eux. La signification des relations dépend du réseau en question. Par exemple, dans un réseau d'amis, une relation signifie "ami de" alors que dans une organisation, une relation pourrait signifier "se rapporte à" [7].

Comme un réseau social peut comporter un ou plusieurs types de nœuds et de liens, nous allons rappeler brièvement deux types de réseau : ceux à un seul mode de données et ceux à deux modes de données. Un réseau à un mode de données compte un seul type de nœuds et un seul type de liens. Dans un réseau à deux modes de données, il existe deux types de nœuds [16] avec des liens entre eux comme par exemple les individus et les événements auxquels ils assistent. Dans le cadre de ce mémoire, nous analyserons uniquement des réseaux sociaux à un seul mode de données avec

des arcs orientés et pondérés.

2.1 Représentation d'un réseau

L'analyse de réseaux sociaux utilise deux outils mathématiques pour représenter les données d'un réseau : les graphes et les matrices [31].

2.1.1 Graphe non orienté et non valué

Un graphe G est défini comme un ensemble de nœuds et de liens de la forme $G = (V, E)$ où V est l'ensemble de sommets (*vertices*) et E est l'ensemble d'arêtes (*edges*). Le graphe d'un réseau social est composé d'acteurs (nœuds ou individus) et de relations (liens, arêtes ou arcs). Celui-ci permet de visualiser quel acteur est lié à quel autre acteur. Un réseau non orienté est utilisé lorsque la direction des liens n'a pas de signification particulière ou lorsque toutes les relations sont réciproques [7]. La figure 2.2 en est un exemple de réseau non orienté. Cependant, lorsque le réseau comporte un trop grand nombre d'acteurs et de relations, comme on peut le constater dans la figure 2.1, il devient difficile, voire impossible de visualiser le graphe. Il faut alors utiliser un autre mode de représentation du réseau : une matrice ou une liste d'adjacence.

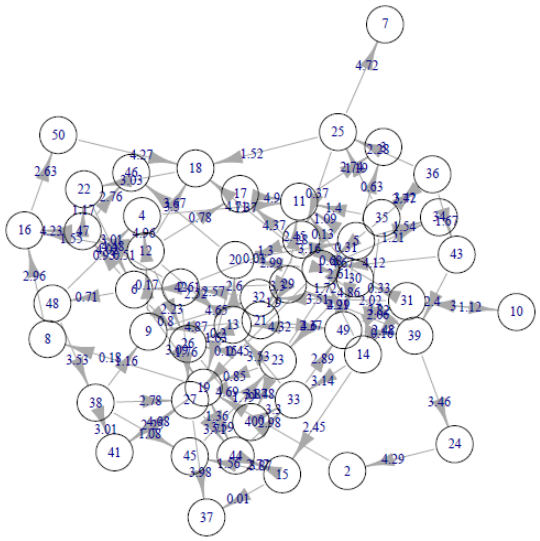


FIGURE 2.1 – Graphe d'un réseau social

La représentation matricielle comporte des avantages sur la représentation graphique d'un réseau. À part la lisibilité mentionnée précédemment, il y a le traitement par ordinateur. Plusieurs analyses de réseau nécessitent des calculs mathématiques et algorithmes complexes exigeant le recours aux programmes informatiques. Pour un ordinateur, la matrice est le mode de représentation le plus simple. De ce fait, les données de réseaux sont généralement enregistrées dans une matrice.

La matrice représentant un réseau unimodal est nommée matrice d'adjacence. C'est un tableau dont le nombre de lignes et de colonnes correspond au nombre d'acteurs. Quel que soit le nombre d'individus dans le réseau, la matrice d'adjacence est impérativement une matrice carrée où chacun de ses éléments représente la relation entre les acteurs. Une valeur 1 dans la cellule signifie qu'il existe une relation directe entre l'acteur se trouvant sur la ligne et celui qui se trouve sur la colonne. Lorsqu'il n'y a pas de relation directe, la cellule prend la valeur 0. Dans un réseau non orienté, la matrice d'adjacence est toujours symétrique puisque les relations entre les nœuds sont valides dans les deux directions.

La figure 2.2 montre la représentation graphique d'un réseau comportant huit acteurs alors que le tableau 2.1 donne la représentation matricielle de ce même réseau qui est non orienté et non valué.

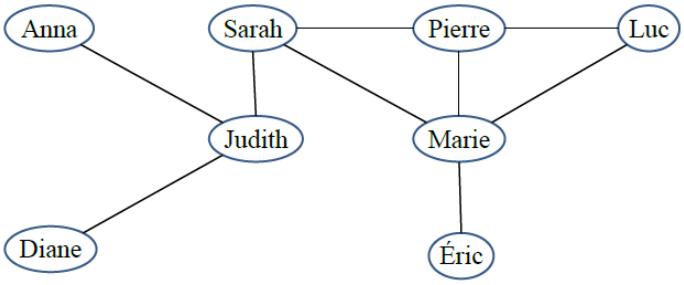


FIGURE 2.2 – R1 : Réseau non orienté à un seul mode de données

	Marie	Pierre	Luc	Sarah	Éric	Judith	Diane	Anna
Marie	0	1	1	1	1	0	0	0
Pierre	1	0	1	1	0	0	0	0
Luc	1	1	0	0	0	0	0	0
Sarah	1	1	0	0	0	1	0	0
Éric	1	0	0	0	0	0	0	0
Judith	0	0	0	1	0	0	1	1
Diane	0	0	0	0	0	1	0	0
Anna	0	0	0	0	0	1	0	0

TABLE 2.1 – R1 : Matrice d'adjacence d'un réseau non orienté et non valué

Une autre façon de représenter un réseau est la liste de liens ou d'adjacence [27]. C'est une matrice à deux colonnes où la première colonne indique le nœud de départ et la deuxième indique le nœud où le lien aboutit. La matrice comporte autant de lignes qu'il y a de liens dans le réseau. Pour un graphe non orienté, les éléments des deux colonnes sont commutatifs. Le tableau 2.2 montre la liste de liens du réseau R1.

Nœud 1	Nœud 2
Marie	Pierre
Marie	Luc
Marie	Sarah
Marie	Eric
Pierre	Luc
Pierre	Sarah
Sarah	Judith
Judith	Diane
Judith	Anna

TABLE 2.2 – R1 : Liste de liens

2.1.2 Graphe orienté et valué

Pour certains réseaux, il y a un sens spécifique de la circulation d'information. Dans ce cas, les liens ont des flèches indiquant la direction de l'information. La figure 2.3 montre un exemple d'un graphe orienté indiquant quels acteurs sont des transmetteurs d'information et lesquels sont des destinataires. Notons également que ce graphe n'est pas valué car il n'y a pas de pondération sur les liens.

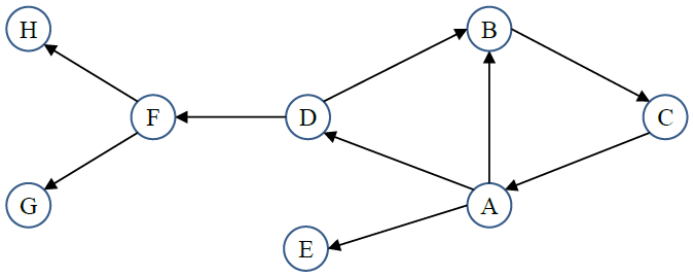


FIGURE 2.3 – R2 : Réseau orienté

La matrice d'adjacence d'un tel réseau a une signification du sens de circulation d'information. Par convention, le transmetteur se trouve sur une ligne de la matrice alors que le destinataire se trouve dans une colonne. Par exemple, si l'acteur Bob envoie l'information à l'acteur Alice, alors l'élément (Bob, Alice) de la matrice prend la valeur 1 alors qu'une valeur 0 signifie que Bob ne transmet pas d'information directement à Alice.

Le tableau 2.3 montre un exemple de matrice pour un graphe orienté. Dans cette matrice, la cellule (F, H) contient la valeur 1 alors que la cellule (H, F) a une valeur de 0. Cela signifie que F transmet l'information à H mais l'inverse n'est pas vrai. On voit donc que contrairement au graphe non orienté, la matrice d'adjacence d'un graphe orienté n'est pas symétrique. Notons toutefois que la symétrie va exister si toutes les relations du réseau sont réciproques. Dans ce cas, il est préférable d'utiliser un réseau non orienté.

Dans la section précédente, nous avons vu que la matrice d'adjacence d'un graphe non valué (tableau 2.2) contient seulement des valeurs binaires, soit 1 ou 0. Mais il y a des situations où les liens du graphe sont pondérés afin de signifier leur importance relative. Dans un tel graphe, le lien n'indique pas seulement la relation directe entre les acteurs, il donne aussi la possibilité de quantifier chaque relation du réseau. On a alors un graphe valué (ou graphe pondéré) où la pondération des liens exprime l'intensité de la relation [20] entre deux acteurs. La matrice d'adjacence contient alors des nombres réels positifs ou négatifs. La signification de la valeur des liens dépend du réseau étudié. Par exemple, dans un réseau de transport, la valeur du lien entre les sites A et B peut indiquer la distance ou le coût de transport entre ces deux sites alors que dans un réseau d'amitié, la valeur du lien pourrait signifier le niveau d'appréciation entre deux acteurs.

L'orientation et la pondération des liens sont des caractéristiques indépendantes l'une de l'autre. Il existe effectivement des réseaux orientés sans être valués et vice-versa. Mais il existe également des réseaux orientés ayant des liens pondérés permettant de mesurer leur importance relative. La figure 2.4 et le tableu 2.3 présentent respectivement le graphe et la matrice d'adjacence d'un réseau orienté et valué.

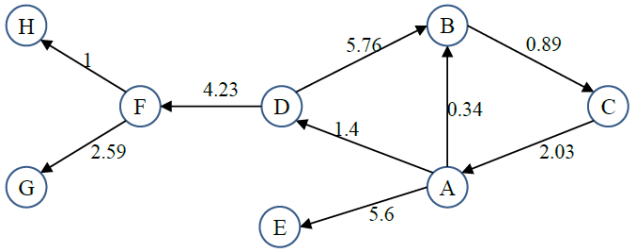


FIGURE 2.4 – R3 : Réseau orienté et valué

	A	B	C	D	E	F	G	H
A	0	0.34	0	1.40	5.60	0	0	0
B	0	0	0.89	0	0	0	0	0
C	2.03	0	0	0	0	0	0	0
D	0	5.76	0	0	0	4.23	0	0
E	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	2.59	1
G	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0

TABLE 2.3 – R3 : Matrice d'adjacence du réseau

2.2 Les mesures d'un réseau

La théorie des graphes et l'analyse de réseau social offrent de nombreuses mesures permettant d'analyser les réseaux, incluant les mesures globales du réseau et les mesures individuelles des nœuds et des liens. Dans le cadre de ce mémoire, nous explorons seulement quelques-unes de ces mesures.

2.2.1 Mesures globales d'un réseau

Les mesures globales d'un réseau sont utilisées pour analyser le réseau dans son ensemble. Elles sont également indispensables lorsqu'il y a un besoin de comparer différents réseaux. Les mesures les plus simples sont sans doute le nombre de nœuds et le nombre de liens. Ces mesures sont cependant très primaires car elles fournissent des informations très minimales. Afin d'obtenir des données significatives sur un réseau, les mesures de cohésion sont de mise.

La cohésion est un indice de connectivité du réseau. Elle donne le degré de fluidité d'information. Plusieurs mesures peuvent être utilisées pour déterminer la cohésion d'un réseau. Nous présentons ici quelques-unes de ces mesures.

Densité

La densité est la probabilité qu'un lien existe entre une paire de nœuds choisis au hasard. Elle donne un indice sur la concentration de la population du réseau. Cette mesure est exprimée par le nombre de liens divisé par le nombre maximal de liens possibles lequel est donné par $n(n-1)/2$ où n est le nombre de nœuds. Cette mesure est appropriée pour comparer deux réseaux ayant approximativement la même taille. Par contre, lorsque la taille des réseaux à comparer est trop différente, cette mesure n'est pas recommandée car un petit réseau aura probablement une meilleure densité qu'un grand réseau. Par exemple, dans un réseau de dix individus, on peut s'attendre à ce que chaque individu soit connecté avec plusieurs autres. Cependant, si le réseau comporte cent individus, il est peu probable que chacun soit connecté avec un grand nombre d'entre eux. Dans ce cas, la comparaison de densités donne des résultats biaisés [7].

Degré moyen

Le degré moyen est la moyenne de la centralité de degré des nœuds du réseau (voir section 2.2.2 pour la centralité de degré). Cette mesure est sans doute mieux adaptée que la densité pour comparer deux réseaux de tailles très différentes. Mentionnons qu'il existe une relation entre ces deux mesures qui est donnée par :

$Degré\ moyen = Densité * (n-1)$ où n est le nombre de nœuds.

Composants

Un composant est un ensemble de nœuds dans lequel chacun peut rejoindre les autres par un chemin. Ce sont des sous-groupes qui sont isolés du reste du réseau. Plus le réseau compte de composants, moins bonne est la circulation d'information. Bien qu'il soit possible de l'utiliser pour comparer les réseaux, cette mesure comporte une lacune car il est très probable d'obtenir le même nombre de composants pour des réseaux ayant une cohésion complètement différente. Ainsi donc, cette mesure ne

suffit pas à elle seule mais peut être intégrée dans une formule d'estimation de la cohésion.

Connectivité (*Reachability*)

La connectivité est la proportion des paires de nœuds pouvant se rejoindre par un chemin. Autrement dit, les paires de nœuds se trouvant dans le même composant. Cette mesure est donnée par

$$\frac{\sum_{i=1}^{n}\sum_{j\neq i}^{n} r_{ij}}{n(n-1)} \tag{2.1}$$

où $r_{ij} = 1$ si i et j sont joignables. La connectivité peut être utilisée pour évaluer le changement dans le réseau. Par exemple, pour calculer la cohésion après le retrait d'un nœud. Plus la valeur de la connectivité est élevée, meilleure est la cohésion du réseau et il y a donc une meilleure circulation de l'information. La faiblesse de cette mesure est qu'elle ne considère pas la distance entre les paires de nœuds. Chaque paire aura la même note pourvu que les deux nœuds puissent se rejoindre. Pour cette raison, il peut être préférable d'utiliser sa variante qui est la compacité (*Compactness*) [7] qui pénalise les paires de nœuds par la distance géodésique les séparant l'un de l'autre. La compacité du réseau est calculée par l'équation 2.2

$$\frac{\sum_{i=1}^{n} C_i}{n(n-1)} \tag{2.2}$$

où $C_i = \sum_{j\neq i}^{n} \frac{1}{d_{ij}}$ et d_{ij} est la distance géodésique entre i et j. Le désavantage de la compacité est qu'elle ne peut être calculée lorsque le graphe n'est pas connecté. En effet, pour deux nœuds qui ne sont pas dans le même composant, leur distance géodésique n'est pas définie puisque ces derniers ne peuvent se rejoindre. Cependant, on peut tout de même adapter la formule pour résoudre ce problème. Le raisonnement est que si les deux nœuds ne peuvent se rejoindre, c'est qu'il n'existe pas de chemin entre eux. On déduit alors que la distance géodésique qui les sépare est infiniment grande, et la compacité entre les deux nœuds en question est alors donnée par $\frac{1}{\infty} = 0$. En adoptant ce raisonnement, on pourra assigner la valeur 0 à la compacité de deux nœuds se trouvant dans deux composants différents.

2.2.2 Mesures de centralité

L'efficacité d'un réseau social dépend grandement de la cohésion entre les acteurs. Une des activités de la déstabilisation d'un réseau consiste alors à l'explorer dans le but de déterminer puis d'anéantir sa cohésion. Afin d'atteindre cet objectif, il convient d'étudier comment chaque acteur contribue à cette cohésion par le biais des relations qu'il entretient avec les autres membres du réseau. Tel que mentionné précédemment, la théorie des graphes offre plusieurs mesures permettant d'étudier les acteurs dans le réseau. Nous analysons donc les mesures de centralité et de connectivité individuelles des nœuds. Celles-ci servent à identifier les acteurs et les relations apportant une contribution importante à la cohésion du réseau.

La centralité d'un acteur est un indice qui permet d'estimer sa position relative à l'intérieur du réseau. Plus un acteur est central, plus il a une position favorable [31]. Il pourrait alors exercer beaucoup d'influence et apporter une contribution importante à l'efficacité du réseau. Dans cette section, nous présentons les mesures de centralité et de connectivité qui seront exploitées dans cette présente recherche.

Centralité de Freeman

En 1979, Linton C. Freeman [15] a établi trois mesures de centralité : la centralité de degré (*degree centrality*), la centralité d'intermédiarité (*betweeness centrality*) et la centralité de proximité (*closeness centrality*). Bien que d'autres mesures aient été exploitées depuis, les centralités de Freeman demeurent fondamentales pour l'analyse de réseaux sociaux.

Centralité de degré

Linton Freeman a établi la centralité de degré pour un graphe non orienté où la centralité d'un nœud repose sur le nombre de ses voisins immédiats. Plus un nœud a des voisins, plus il est central. Cette centralité pour le nœud i est définie par l'équation 2.3 et représente le nombre de voisins (nœuds adjacents) de i.

$$d_i = \sum_{j=1}^{n} A_{ij} \tag{2.3}$$

où n est le nombre total de nœuds, A_{ij} est un élément de la matrice prenant la valeur 1 si le nœud i a une relation directe avec le nœud j, sinon, c'est la valeur 0.

L'ampleur de d_i dépend de la taille du réseau. Plus le réseau est large, plus il y a une grande probabilité de trouver des nœuds ayant un degré élevé. Dans certaines situations, il peut être plus approprié d'utiliser une centralité de degré qui est indépendante de la taille du réseau. Dans ce cas, la centralité de degré est normalisée selon l'équation suivante :

$$d'_i = \frac{\sum_{j=1}^{n} A_{ij}}{n-1} \tag{2.4}$$

Le tableau 2.4 donne la liste des centralités relatives de degré du réseau non orienté R1 présenté dans la figure 2.2. On peut constater qu'il est facile de calculer la centralité de degré. Par contre, celle-ci est assez minimaliste car elle tient compte uniquement du nombre de voisins, ce qui suggère que les nœuds ayant le même degré exercent le même niveau d'influence dans le réseau. Dans la réalité, deux nœuds ayant le même nombre de voisins n'ont pas nécessairement le même pouvoir d'influence.

Dans un réseau orienté, chaque nœud peut avoir un degré entrant (*in-degree*) et un degré sortant (*out-degree*). Le degré sortant se trouve dans le vecteur ligne alors que le degré entrant se trouve dans le vecteur colonne de la matrice d'adjacence. Dans certains cas, ces degrés peuvent tous les deux être utilisés [27].

Centralité de proximité

La centralité de proximité donne l'importance à un nœud qui est globalement plus proche de tous les autres nœuds. Cela signifie qu'un acteur est central s'il peut contacter les autres sans dépendre d'acteurs intermédiaires. Cette centralité peut être définie comme étant la distance qu'un acteur doit parcourir pour joindre les autres acteurs du réseau [17]. Plus cette distance est petite, plus l'acteur est central. Cette description repose sur la notion de distance géodésique entre deux nœuds i et j. Cette dernière représente le plus court chemin entre ces deux nœuds. Pour calculer la centralité de proximité d'un nœud i, il faut d'abord trouver la plus courte distance entre i et chaque autre nœud du graphe séparément. La somme de ces distances est le

cumul des distances géodésiques entre i et les autres nœuds. La centralité de proximité de i correspond alors à l'inverse de cette somme et est obtenue par l'equation 2.5

$$c_i = \frac{1}{\sum_{j=1}^{n} d_{ij}} \tag{2.5}$$

où d_{ij} représente le plus court chemin, *i.e.* le nombre minimum d'arêtes reliant les nœuds i et j. Encore une fois, cette mesure dépend du nombre de nœuds dans le réseau. Afin d'obtenir une centralité de proximité relative d'un nœud du graphe, l'équation 2.5 est normalisée au niveau du réseau par l'équation 2.6. La centralité de proximité relative des nœuds du réseau non orienté R1 de la figure 2.2 est présentée dans le tableau 2.4.

$$c'_i = \frac{n-1}{\sum_{j=1}^{n} d_{ij}} \tag{2.6}$$

Lorsque le réseau n'est pas valué, cette centralité compte tout simplement le nombre minimum d'arêtes entre deux nœuds. Dans un réseau valué, il est important de bien définir la pondération des liens car le plus court chemin pourrait être erroné. Prenons par exemple le réseau R4 de la figure 2.5, avec les liens pondérés. Dans ce réseau, il existe deux chemins possibles entre les nœuds A et D, soient A-D et A-C-D. Lorsque l'on veut obtenir le plus court chemin entre ces derniers, on obtiendrait A-C-D puisqu'il a une distance totale de 3 qui est effectivement plus courte que A-D laquelle a une distance de 5. Dans le cas où la pondération signifie le coût ou le temps de parcours, la centralité de proximité donne une valeur qui est en accord avec la structure du réseau. Cependant, si la valeur des liens représente le nombre de fois que deux amis se rencontrent, on aura alors une proximité erronée car le plus court chemin devrait plutôt être A-D.

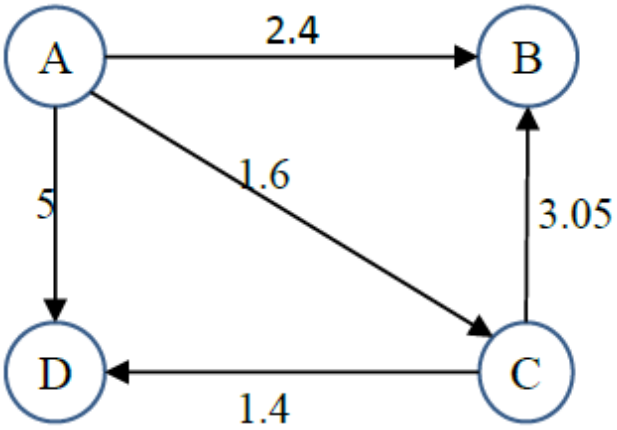


FIGURE 2.5 – R4 : Réseau orienté et valué

Centralité d'intermédiarité

La centralité d'intermédiarité indique la fréquence qu'un nœud se trouve sur un chemin liant deux autres nœuds. Dans un réseau social, plus un acteur sert d'intermédiaire, plus il a le potentiel pour contrôler la communication. À titre d'exemple, prenons *Marie* dans le réseau R1 de la figure 2.2. On voit clairement qu'elle sert d'intermédiaire entre les acteurs *Pierre* et *Éric*, *Luc* et *Éric*, *Sarah* et *Éric*, etc. Dans cet exemple, *Marie* est intermédiaire dans des voies de communication différentes contrairement à *Luc* qui ne sert d'intermédiaire à aucun acteur du réseau. On peut alors en déduire que l'acteur *Marie* est potentiellement le plus central car la circulation d'information à l'intérieur du réseau dépend de ce dernier. Le calcul de la centralité d'intermédiarité du nœud i se fait selon les étapes suivantes :

1. Établir les intermédiarités partielles du nœud i entre chaque paire de nœuds j et k.
2. Calculer la somme des intermédiarités partielles pour obtenir la centralité d'intermédiarité de i.

En résumé, la centralité d'intermédiarité du nœud i est établie par

$$b_i = \sum_{j}^{n} \sum_{k}^{n} b_{jk}(i) \tag{2.7}$$

où $i \neq j \neq k$, n est le nombre total de nœuds dans le graphe et $b_{jk}(i)$ est une intermédiarité partielle du nœud i. Si i se trouve entre les nœuds j et k, alors $b_{jk}(i)$ est

égale à 1, sinon, $b_{jk}(i)$ vaut 0.

De façon similaire aux centralités décrites précédemment, la centralité d'intermédiarité est aussi dépendante de la taille du réseau. La centralité d'intermédiarité relative d'un nœud i est alors exprimée par l'équation 2.8. À titre d'exemple, la centralité d'intermédiarité des nœuds du réseau non orienté R1 de la figure 2.2 est calculée dans le langage R et présentée dans le tableau 2.4.

$$b'_i = \frac{2b_i}{n^2 - 3n + 2} \tag{2.8}$$

Nœud	**Centralité**		
	Degré	**Proximité**	**Intermédiarité**
Marie	0.57	0.5833	0.38
Pierre	0.43	0.5385	0.1
Luc	0.29	0.4118	0
Sarah	0.43	0.6364	0.57
Eric	0.14	0.3889	0
Judith	0.43	0.5385	0.52
Diane	0.14	0.3684	0
Anna	0.14	0.3684	0

TABLE 2.4 – Centralités des nœuds du réseau R1 (figure 2.2)

Étant donné le caractère minimaliste de la centralité de degré, les centralités d'intermédiarité et de proximité sont certainement plus appropriées dans la majorité des cas. En revanche, ces deux mesures sont plus coûteuses en terme de temps de calcul car il faut parcourir tout le graphe pour chaque paire de nœuds. Selon Freeman, la centralité des nœuds d'un graphe peut être déterminée soit en utilisant la centralité de degré, la centralité d'intermédiarité ou la centralité de proximité. Toutefois, de nouvelles centralités ont été proposées. Certaines d'entre elles sont dérivées des mesures de centralité de Freeman.

Centralité de vecteur propre

La centralité de vecteur propre (*eigenvector centrality*) est définie par Bonacich [5]. C'est une extension de la centralité de degré. Selon Bonacich, un nœud est important non seulement parce qu'il a beaucoup de voisins, mais aussi parce qu'il a des voisins qui sont eux-mêmes importants. Donc, un nœud ayant des voisins qui sont eux-mêmes très connectés, obtiendrait un score plus élevé que celui ayant le même nombre de voisins moins connectés. La centralité de vecteur propre d'un nœud i est définie par :

$$e_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} e_j \tag{2.9}$$

qui indique que la centralité de vecteur propre du nœud i est proportionnelle à la somme des centralités de ses voisins immédiats. La constante λ 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.

Bien qu'elle représente une amélioration de la centralité de degré, la centralité de vecteur propre n'est pas bien adaptée pour les réseaux orientés. Le problème existe dans le cas où le réseau contient des nœuds sans degré entrant. Prenons par exemple, le réseau orienté R3 de la figure 2.4 où tous les nœuds ont un degré entrant. On voit que la centralité de vecteur propre peut être calculée. Cependant, si on modifie légèrement le réseau R3 (figure 2.4) de sorte qu'il comporte au moins un nœud (le nœud A) sans degré entrant, on obtient alors le réseau R3' (figure 2.6) ayant une centralité de 0 pour tous les noœuds. Ceci est dû au fait que la centralité de vecteur propre de A vaut 0. Cette valeur est alors propagée aux nœuds adjacents à A, qui à leur tour, la propagent à leurs voisins. Conséquemment, tous les nœuds du réseau ont une centralité de vecteur propre de 0. Le tableau 2.5 illustre cette lacune en comparant la centralité de vecteur propre de deux réseaux similaires où tous les nœuds du réseau R3 ont des voisins entrants alors que le réseau R3' contient un nœud sans degré entrant.

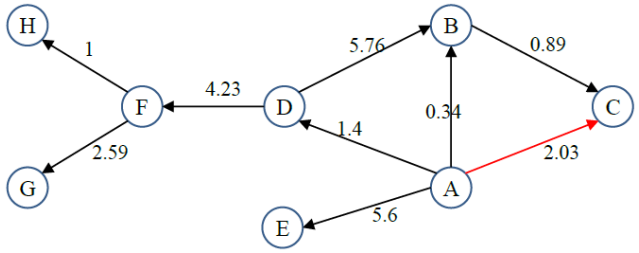


FIGURE 2.6 – R3' : Réseau orienté avec un nœud sans degré entrant

Nœud	Centralité de vecteur propre	
	Réseau R3	**Réseau R3'**
A	0.67	0
B	1	0
C	0.82	0
D	0.55	0
E	0.55	0
F	0.45	0
G	0.37	0
H	0.37	0

TABLE 2.5 – Comparaison de centralités de vecteur propre

Centralité de Katz

Comme décrit plus haut, la centralité de vecteur propre se calcule moins bien pour un graphe orienté. La centralité de Katz (*Katz centrality*) établie par Leo Katz [19] est certainement mieux adaptée pour ce type de graphes. Elle compense la centralité de vecteur propre à l'aide d'une constante β accordée à chaque nœud du graphe [27]. La centralité de Katz pour le nœud i est définie par :

$$k_i = \alpha \sum_{j=1}^{n} A_{ij} k_j + \beta \tag{2.10}$$

où la constante α doit être prise dans l'intervalle $]0, 1/\lambda]$ et λ est la plus grande valeur de vecteur propre. En général, on assigne la valeur 1 à la constante β, laquelle permet d'attribuer une valeur non nulle aux nœuds n'ayant pas de degré entrant.

Centralité de PageRank

L'élément indésirable de la centralité Katz est qu'un nœud ayant une grande centralité donne également une grande centralité à tous les nœuds vers lesquels celui-ci pointe [27]. La centralité de PageRank (*PageRank centrality*) qui est une variante de la centralité de Katz vient remédier à ce problème. Afin d'éviter qu'un nœud ayant une centralité élevée partage toute sa centralité avec les nœuds vers lesquels il pointe, seulement une proportion de sa centralité sera transférée à ses voisins destinataires. Cette centralité pour le nœud i est définie par la relation :

$$p_i = \alpha \sum_{j=1}^{n} A_{ij} \frac{p_j}{d_j^{out}} + \beta \tag{2.11}$$

où p_j est la centralité *PageRank* du nœud j et d_j^{out} est le degré sortant de ce dernier. Donc, même si le nœud j a une centralité élevée, il transférera seulement une fraction de sa centralité à ses voisins. Pour éviter le problème de division par zéro, d_j^{out} prend la valeur 1 lorsque le nœud j n'a aucun lien sortant.

Selon Wikipédia [35], l'algorithme de PageRank est utilisé dans le moteur de recherche de Google afin de mesurer la popularité des sites web. Les sites recevant un nombre élevé de liens provenant d'autres sites sont importants car ils ont une grande valeur de centralité PageRank. Cela signifie que cette centralité favorise les nœuds ayant un degré entrant élevé.

Centralité des liens

La majorité des centralités présentées précédemment s'appliquent uniquement à des nœuds. Parmi elles, seule la centralité d'intermédiarité de Freeman permet de détecter un lien central. La centralité d'un lien l correspond au nombre de courts chemins passant par l. En d'autres mots, c'est le nombre de fois qu'un lien l sert d'intermédiaire ou de pont entre deux nœuds. Ce nombre est également calculé à l'aide de l'équation 2.7.

Une autre façon d'identifier les liens clés est de calculer leur importance qui est donnée par la variation de l'efficacité. Cette variation est calculée selon l'équation 2.12.

$$\Delta E = E(G) - E(G - i) \tag{2.12}$$

où $E(G)$ et $E(G-i)$ sont respectivement l'efficacité du réseau avant et après le retrait du lien ou du nœud i. Cette technique a été utilisée par plusieurs analystes [24, 34] 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 obtenue en calculant la valeur de la connectivité ou la compacité du réseau décrites dans la section 2.2.1.

Le tableau 2.6 donne les mesures de centralité des liens du réseau R3 (figure 2.4). Dans ce tableau, l'importance des liens a été calculée à l'aide de l'équation 2.2 pour mesurer le niveau de fluidité de l'information. Comme on peut le constater, il n'existe pas de corrélation évidente entre la centralité d'intermédiarité et l'importance des liens. Par conséquent, le résultat de la déstabilisation sera différent pour chacune de ces mesures comme nous allons l'illustrer à travers nos tests empiriques.

Lien	**Centralité d'intermédiarité**	**Importance du lien**
C – A	0.71	0.05
D – F	0.57	0.07
A – D	0.57	0.04
B – C	0.52	0.06
F – H	0.24	0.06
F – G	0.24	0.05
A – E	0.19	0.05
D – B	0.19	0.01
A – B	0.14	0.06

TABLE 2.6 – Centralité vs importance des liens

2.3 Équivalence

L'analyse de réseaux sociaux ne se limite pas à la centralité des nœuds et des liens. Les auteurs de [7, 17] ont présenté la théorie de l'équivalence qui pourrait être appropriée dans la désintégration de réseaux. Bien qu'elle ne fasse pas l'objet de notre recherche, nous décrivons tout de même brièvement cette théorie afin d'avoir une compréhension globale sur le sujet. L'équivalence est une mesure permettant d'identifier les acteurs qui sont équivalents dans un réseau.

Équivalence structurelle

L'équivalence structurelle existe lorsque deux acteurs ou plus ont le même nombre de liens et sont connectés exactement aux mêmes voisins. Ces derniers ont exactement les mêmes voisins sortants et entrants et sont réciproquement substituables. La figure 2.7 montre que les acteurs E et F sont structurellement équivalents. Ils ont tous deux un seul lien et sont tous deux connectés au même acteur B. Également, on peut voir que H et I qui sont tous deux connecté à D sont structurellement équivalents.

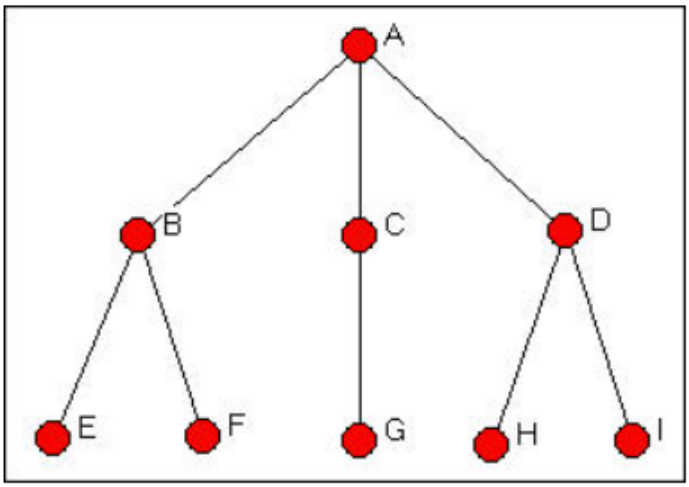


FIGURE 2.7 – Équivalence des acteurs [17]

Équivalence automorphique et régulière

Il existe d'autres formes d'équivalence dont les conditions sont moins strictes que l'équivalence structurelle. Certains acteurs peuvent être d'une certaine façon considé-

rés équivalents sans être connectés exactement aux mêmes voisins. Dans la figure 2.7, les acteurs B et D sont automorphiquement équivalents car ils se situent tous les deux au même niveau hiérarchique. De plus, ils se rapportent tous les deux au même patron et chacun a deux subordonnés.

L'équivalence régulière regroupe les acteurs dont les liens ont les mêmes profils, c'est-à-dire, ils ont le même rôle social. Ainsi deux mères de famille sont régulièrement équivalentes par leur rôle, elles sont toutes des mères. Si on regarde encore la figure 2.7, on voit que les acteurs B, C et D sont régulièrement équivalents. Ils sont tous les trois situés au milieu dans l'organisation. Chacun a un patron et chacun a des employés.

2.4 Les communautés

Les communautés sont des sous-groupes d'individus d'un réseau. Il existe plusieurs types de communautés, y compris les cliques et les factions [7].

Une clique est un sous-ensemble d'acteurs dans lequel chacun est lié à tous les autres. Formellement, c'est un sous-graphe complet. En général, une clique doit comporter un minimum de trois acteurs. De plus, il n'existe aucune restriction quant à l'appartenance aux cliques. Certains acteurs peuvent participer à plus d'une clique alors que d'autres n'appartiennent à aucune clique.

Quant aux factions, ce sont des sous-groupes cohésifs à l'intérieur d'un réseau. Dans ce type de regroupement, un acteur n'appartient qu'à une seule et unique faction. Précisons également que les acteurs font nécessairement partie d'une des factions. Il n'existe donc pas d'acteurs sans faction comme dans le cas des cliques.

2.5 Modèles de réseaux

Puisque nous validons notre stratégie avec différents modèles de réseaux, nous décrivons brièvement dans la présente section les trois modèles de réseaux existants : le réseau aléatoire (*random network*), le réseau petit-monde (*small-world network*) et le réseau invariant d'échelle (*scale-free network*).

Dans un réseau aléatoire, les liens sont placés au hasard. On pourrait affirmer que le graphe résultant est "uniforme" car la grande majorité des nœuds ont approximativement le même nombre de liens [13]. De ce fait, il serait difficile d'identifier un meneur étant donné que la plupart des nœuds sont similaires.

Les réseaux *petit-monde* de Watts et Strogatz [33] comportent de nombreuses paires de nœuds dont la distance géodésique (nombre de pas) est faible, *i.e.* une distance de six pas ou moins. Grâce à cette caractéristique, la diffusion d'information est plus facile que dans les autres types de réseaux. Une deuxième propriété est que dans les réseaux petit-monde, le coefficient de regroupement [1] (*clustering coefficient*) est élevé. Cela signifie que les membres ont une grande tendance à former des communautés.

Quant aux réseaux invariants d'échelle de Barabasi [3], ils comportent des nœuds dont la probabilité de connexion (degré) suit une loi de puissance, $D_i \approx i^{\gamma}$ où γ est un paramètre situé en général entre 2 et 3. Ce type de réseau compte un petit nombre de super nœuds (*hubs*) qui possèdent un nombre élevé de voisins alors que la majorité des nœuds du réseau ont un petit nombre de voisins. Un tel réseau est résistant aux attaques aléatoires. Il est en revanche très vulnérable aux attaques ciblées.

1. Le coefficient d'un graphe est une mesure de connectivité des nœuds. Il permet de mesurer le degré de connexion du voisinage des divers nœuds.

En résumé, Adelaide Hopkins [18] a classifié ces trois modèles de réseaux selon les caractéristiques propres à chacun. Le tableau 2.7 donne les caractéristiques de chaque modèle.

Modèle de réseau	**Caractéristiques**
Aléatoire	- Distance géodésique moyenne faible, - Coefficient de regroupement faible
Petit-monde	- Distance géodésique moyenne ou faible, - Coefficient de regroupement élevé, - Vulnérable aux attaques par l'intermédiarité (*bridge attacks*)
Invariant d'échelle	- Distribution de degré suit la loi de puissance, - Robuste contre les bris et les attaques aléatoires, - Vulnérable aux attaques ciblant les super nœuds, - Vulnérable aux attaques par l'intermédiarité

TABLE 2.7 – Modèles de réseau

Chapitre 3

État de l'art

L'étude menée par Chatterjee [9] a révélé que la direction d'un réseau de criminels est généralement composée d'un président, de vice-présidents et de chefs de cellules qui collaborent ensemble pour accomplir des activités criminelles du réseau. De ce fait, ces personnes forment le cerveau qui assure le fonctionnement des opérations au sein du réseau. L'objectif de la déstabilisation d'un réseau social est de provoquer son disfonctionnement ou d'inhiber ses activités.

Carley et al. [8], mentionnent qu'il existe trois principaux indices de déstabilisation d'un réseau. Premièrement, le réseau est déstabilisé lorsque la circulation d'information est dramatiquement réduite. Le second indice de déstabilisation est noté lorsque les preneurs de décision sont incapables d'atteindre un consensus ou de l'atteindre dans un temps raisonnable. Finalement, si le réseau perd de l'efficacité dans l'exécution de ses activités opérationnelles, on peut effectivement conclure qu'il est déstabilisé.

Pour mesurer cet indice de déstabilisation, la réduction d'efficacité du réseau est calculée à la suite du retrait d'un nœud ou d'un lien clé. Cette réduction calculée selon l'équation 2.12 (chapitre 2) est utilisée par plusieurs chercheurs [10, 24, 34] pour mesurer l'effet provoqué par le retrait d'un ou plusieurs nœuds. Plus la réduction de l'efficacité (ΔE) est élevée, plus le réseau est déstabilisé. Certains chercheurs considèrent cette mesure comme l'indice d'importance ou de performance pouvant identifier les nœuds et les liens clés.

D'après les connaissances tirées des paragraphes précédents, une des activités cruciales dans la déstabilisation d'un réseau consiste alors à identifier les acteurs et les relations clés dont le retrait contribue significativement à déstabiliser le réseau. Nous présentons dans cette section les travaux réalisés dans le but d'identifier les acteurs ou les liens clés à retirer d'un réseau de criminels.

3.1 Élimination de nœuds clés

Memon et Larsen [24] ont proposé plusieurs méthodes pour identifier les nœuds clés dont le retrait conduit à une déstabilisation du réseau. Tout d'abord, ils ont développé deux algorithmes pour la construction d'un arbre hiérarchique afin de découvrir les acteurs importants dans un réseau non orienté et non pondéré. Plus un acteur est près de la racine de l'arbre, plus il est important. Pour ce faire, ils ont proposé la démarche suivante :

1. Transformer un graphe non orienté en un graphe orienté en se servant des centralités de degré et de vecteur propre.
2. Construire un arbre hiérarchique à partir du graphe orienté.

Durant la première étape, on compare la centralité de degré ou de vecteur propre pour chaque paire de nœuds. La direction du lien part du nœud qui a la plus grande centralité vers celui ayant une centralité moins élevée. Lorsqu'il y a égalité de centralité, le lien est alors ignoré. À la fin de cette étape, on obtient un graphe avec des arcs (liens) orientés. Ce graphe est alors utilisé à la seconde étape où la relation parent-enfant est déterminée par la direction des arcs. Ainsi, le nœud ayant le lien sortant est considéré comme parent de celui qui reçoit le lien. À la fin du processus, on obtient un arbre hiérarchique permettant d'identifier les acteurs clés, c'est-à-dire ceux se trouvant près de la racine de l'arbre hiérarchique.

Une autre méthode proposée par Memon et Larsen est l'indice de rôle ou de position (*PRI : Position Role Index*). Selon eux, un réseau de criminels est composé de meneurs et de suiveurs. La méthode *PRI* permet de détecter les meneurs en cal-

culant l'écart d'efficacité du réseau (ΔE) engendré par le retrait de chaque nœud. Plus le retrait d'un nœud engendre un grand écart, plus la déstabilisation du réseau est significative. Par conséquent, le nœud en question est une cible pour le retrait. L'importance d'un nœud est alors donnée par la variation ΔE calculée selon l'équation 2.12. Les nœuds ayant une valeur ΔE élevée sont critiques pour le réseau et sont considérés comme des meneurs. À l'inverse, les nœuds dont le retrait occasionne un faible écart sont considérés comme des suiveurs. En général, ils sont moins connectés, ce qui explique qu'ils ont moins d'impact sur l'efficacité du réseau.

La centralité de dépendance (*Dependency Centrality*) est une autre mesure proposée par ces deux auteurs. Elle détermine la fréquence qu'un nœud dépende d'un autre nœud du réseau. Cette approche utilise un graphe sans orientation ni pondération qui détermine la centralité de dépendance des nœuds selon la relation suivante :

$$DC_{ij} = \sum_{i \neq k, k \epsilon G} \frac{d_{ij}}{N_k} + \Omega \tag{3.1}$$

où DC_{ij} est la centralité de dépendance du nœud i par rapport au nœud j. Cette mesure indique la fréquence à laquelle i a besoin du nœud j pour communiquer avec d'autres nœuds du réseau. N_k est le nombre de chemins géodésiques de i vers k passant par j et d_{ij} est la distance géodésique entre i et j. Ω prend la valeur 1 lorsque le graphe est connecté et 0 autrement.

Ainsi, les nœuds ayant une centralité de dépendance faible par rapport aux autres nœuds du graphe sont des meneurs potentiels. En général, ils ont un grand nombre de liens directs avec les autres nœuds, ils sont donc moins dépendants des autres pour communiquer avec le reste du réseau.

Berzinji et al. [4] ont proposé une méthode qui utilise une combinaison de trois centralités pour détecter les acteurs principaux dans un réseau de terroristes. La stratégie consiste à calculer les centralités suivantes pour chaque acteur du réseau :

1. La centralité de degré
2. La centralité d'intermédiarité
3. La centralité de proximité.

Une fois ces centralités calculées pour chaque acteur, on peut identifier les nœuds qui sont les plus importants. En effet, les acteurs centraux sont ceux ayant des valeurs élevées pour la plupart des trois centralités calculées. Toutefois, les auteurs n'ont pas été explicites sur la manière de tenir compte et d'intégrer ces centralités.

Selon Charausia et Tiwari [10], l'arbre hiérarchique de Memon et Larsen peut être construit autrement. Les auteurs ont alors proposé un nouvel algorithme utilisant les centralités de Katz et de *PageRank* pour construire l'arbre hiérarchique. Similaire à la méthode discutée précédemment, cet algorithme utilise un réseau non orienté et non pondéré. Dans cet algorithme, on compare la centralité de *PageRank* ou de Katz pour chaque paire de nœuds. Le nœud possédant la plus grande centralité devient le parent de l'autre. Lorsqu'il y a égalité de centralité entre deux nœuds quelconques, le lien est alors ignoré. Contrairement à la méthode précédente, la relation parent-enfant est obtenue directement en comparant la centralité de deux nœuds. L'arbre hiérarchique est construit en une seule étape.

Sarr et al. [21] ont développé une méthode de désintégration d'un réseau social en utilisant le modèle de propagation d'information. Cette méthode consiste en premier lieu à trouver le déclencheur à retirer du réseau. Par la suite, le modèle de propagation d'information est appliqué afin d'identifier les nouveaux nœuds ciblés pour l'élimination. Le processus d'élimination en cascade est ainsi répété jusqu'à ce que le niveau de démantèlement atteigne un seuil établi par l'utilisateur. Dans le but de démanteler un réseau à moindre coût, ces auteurs utilisent les techniques d'analyse de réseau social permettant de trouver un déclencheur efficace. Pour atteindre cet objectif, deux stratégies distinctes ont été utilisées pour trouver le déclencheur qui va initialement provoquer une déstabilisation optimale. La première stratégie consiste à identifier le déclencheur comme étant le nœud ayant la centralité de degré la plus élevée. La deuxième stratégie consiste à identifier les communautés à l'intérieur du réseau. Ensuite, un nœud est choisi dans chaque communauté pour former le déclencheur.

Allsup et al. [1] ont expérimenté trois stratégies d'interruption de réseaux d'exploitation sexuelle des enfants. Ces derniers ont testé la déstabilisation de réseaux contenant du matériel de pornographie juvénile par élimination de nœuds clés. Les

trois mesures utilisées pour cibler les nœuds à retirer sont le degré entrant, le degré sortant et la centralité d'intermédiarité. Afin de quantifier l'impact causé par le retrait des nœuds, plusieurs mesures globales ont été utilisées telles que la densité, le nombre de liens, la cohésion basée sur la connectivité des liens ainsi que la distance moyenne des courts chemins dans le réseau. Ces stratégies ont été validées sur deux réseaux distincts dont le premier est le réseau des sites web contenant du matériel d'exploitation sexuelle des enfants. Le second réseau contient des individus qui sont des propriétaires des sites web du premier réseau. Le résultat a démontré que l'attaque par la centralité d'intermédiarité est plus efficace sur le réseau de sites web alors que le réseau de propriétaires est plus vulnérable aux attaques par le degré entrant et le degré sortant.

3.2 Élimination de liens clés

Jusqu'ici, la plupart des analystes accordent un grand intérêt à l'importance des nœuds. Ceci est dû au fait que lorsqu'un nœud est éliminé, les liens adjacents à ce nœud sont éliminés automatiquement. Cependant, Wiil et al. [34] soutiennent que l'on doit également considérer l'importance des liens. En fournissant l'effort de trouver les liens importants, on découvre certainement les nœuds importants. En effet, un nœud qui est lié à un autre nœud via un lien important, devient lui-même important. Ils se sont donc intéressés à la déstabilisation de réseaux sociaux par élimination de liens au lieu des nœuds. À cette fin, ils ont développé une équation pour calculer l'importance d'un lien qui est donné par le produit de sa performance et de son poids :

$$LI = p_k * w_k \tag{3.2}$$

où p_k est la performance du lien k et w_k est son poids. La performance d'un lien correspond à l'écart de performance du réseau suite au retrait de ce lien. Cette performance est définie par :

$$p_k = \Delta P = P(G) - P(G - k) \tag{3.3}$$

où P(G) et $P(G-k)$ sont respectivement la performance du réseau avant et après la suppression du lien. Le poids du lien k est calculé par l'équation 3.4

$$w_k = \frac{\sum b_G}{\sum b_G - b_k} \tag{3.4}$$

où $\sum b_G$ représente la somme de centralité d'intermédiarité des liens du graphe G et b_k est la centralité d'intermédiarité du lien k. Plus le poids d'un lien est grand, plus ce lien est important.

Tel que mentionné précédemment, la plupart des approches présentées ci-haut considèrent des réseaux non orientés et non pondérés. Cela nous motive à explorer des stratégies similaires mais en utilisant des réseaux orientés ayant des liens pondérés. De plus, puisque chaque réseau possède des caractéristiques qui lui sont spécifiques, il n'existe donc pas de méthodes universelles pouvant démanteler efficacement tous les types de réseaux. Pour cette raison, il convient d'étudier de nouvelles stratégies en dépit de l'existence de méthodes performantes.

Chapitre 4

Démarche proposée

Tel que spécifié dans les chapitres précédents, l'objectif de ce mémoire est d'étudier la déstabilisation de réseaux sociaux à un mode de données avec des liens orientés et pondérés dans le contexte de démantèlement de réseaux illégaux. Ce phénomène de déstabilisation se produit dans plusieurs situations réelles, soit pour des raisons légitimes comme dans des réseaux de criminels, trafiquants ou fraudeurs, soit pour des raisons illégitimes telle que la désagrégation d'un groupe influent de politiciens ou professionnels par malhonnêteté ou malveillance. Les questions que nous nous posons sont les suivantes :

- Quelles sont les mesures à appliquer pour identifier les nœuds et les liens les plus importants d'un réseau de criminels sachant que les dirigeants ont peu de liens sortants et que les subalternes de bas niveau peuvent avoir plusieurs liens entrants ? Doit-on viser le sommet hiérarchique ou plutôt la base d'un réseau ?
- Faut-il adapter ces mesures au type de réseau sachant qu'il existe trois grandes catégories de structures avec leurs propres caractéristiques [18] : les réseaux aléatoires, invariants d'échelle (*scale-free*) et du petit-monde (*small-world*) ?
- Quelles mesures faut-il utiliser, adapter ou concevoir pour quantifier le degré de démantèlement d'un réseau suite à l'élimination de nœuds et liens clés ?
- Quel critère d'arrêt du processus de démantèlement faut-il utiliser sans obligatoirement introduire un seuil fourni par l'utilisateur ?

Les premiers éléments de réflexion sur le sujet au sein de l'équipe LARIM ainsi que la consultation de documents internes [21, 14] nous amènent à explorer et valider deux approches distinctes (avec possiblement quelques variantes) pour la déstabili-

sation d'un réseau de criminels : l'élimination de nœuds centraux et l'élimination de liens importants incluant les nœuds adjacents. Ces deux stratégies de déstabilisation seront exécutées sur les modèles de réseaux petit-monde et aléatoires. Tel que mentionné précédemment, ces réseaux auront des liens orientés et pondérés et seront générés synthétiquement dans le langage R [28].

Puisque la signification des liens diffère d'un réseau à l'autre, nous définissons les liens en tant que relations de transmission d'information. Un lien de l'acteur A vers l'acteur B signifie que A envoie l'information à B. Quant à la pondération d'un lien, elle représente la périodicité de la communication. Par exemple, une pondération de 3 signifie qu'en moyenne, l'information est transmise tous les trois jours (ou heures, semaines, etc.). Cette définition de pondération a été choisie pour fin de compatibilité avec les calculs de proximité. En effet, tel que mentionné dans la sous-section 2.2.2, la centralité de proximité favorise les liens ayant de faibles pondérations qui correspondent à de fréquentes interactions entre les membres du réseau.

Notons que la déstabilisation d'un réseau social réel est un processus laborieux. La collecte d'informations sur les membres du réseau est fastidieuse et exige beaucoup de temps sans nécessairement aboutir à des réseaux reflétant la réalité. Aussi, nous tenons à préciser que la collecte de données est exclue de ce mémoire. Les réseaux servant de validation sont générés synthétiquement. Pour utiliser les données de réseaux réels, il suffit de bâtir la matrice d'adjacence pour chaque réseau testé au lieu de la générer. Le processus de déstabilisation se déroule alors dans l'ordre suivant :

1. Identifier les cibles à retirer du réseau. La cible peut être un nœud et/ou un lien.
2. Supprimer les cibles.
3. Mesurer l'impact du retrait des cibles sur le réseau.
4. Répéter le processus jusqu'à ce qu'une condition d'arrêt prédéfinie soit atteinte.

4.1 Étape 1 : Identifier les cibles

La déstabilisation d'un réseau social débute par l'identification de la ou des cibles à retirer du réseau. Pour ce faire, nous calculons l'importance de ces derniers à l'aide

de différentes mesures. Notre première stratégie de déstabilisation consiste à éliminer les nœuds centraux. Dans cette section, nous décrivons deux variantes pour identifier les nœuds clés à supprimer du réseau. Nous nous servons du réseau R5 pour effectuer les calculs permettant d'identifier les nœuds clés. Comme on peut le constater dans la figure 4.1 et le tableau 4.1, c'est un réseau orienté et pondéré comportant 15 nœuds et 30 liens. Ce réseau est de type petit-monde généré dans le langage R.

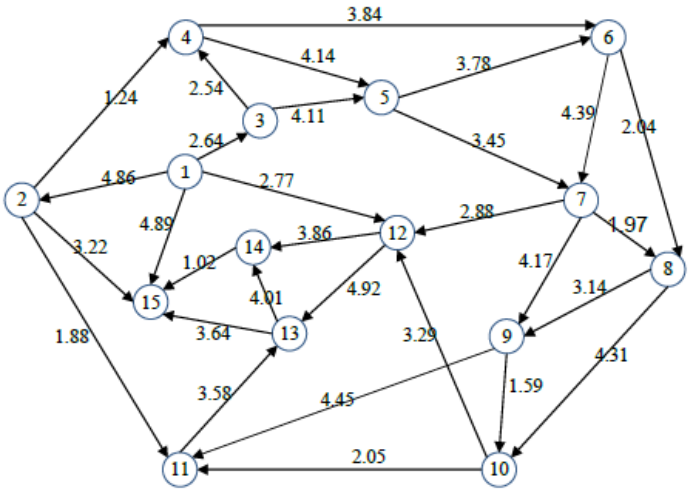


FIGURE 4.1 – R5 : Réseau orienté et valué

Lien	Pondération
1 - 2	4.86
1 - 3	2.64
1 - 12	2.77
1 - 15	4.89
2 - 4	1.24
2 - 11	1.88
2 - 15	3.22
3 - 4	2.54
3 - 5	4.11
4 - 5	4.14
4 - 6	3.84
5 - 6	3.78
5 - 7	3.45
6 - 7	4.39
6 - 8	2.04

Lien	Pondération
7 - 8	1.97
7 - 9	4.17
7 - 12	2.88
8 - 9	3.14
8 - 10	4.31
9 - 10	1.59
9 - 11	4.45
10 - 11	2.05
10 - 12	3.29
11 - 13	3.58
12 - 13	4.92
12 - 14	3.86
13 - 14	4.01
13 - 15	3.64
14 - 15	1.02

TABLE 4.1 – R5 : Liste de liens

Nous avons vu dans le chapitre 2 que l'analyse de réseaux sociaux offre plusieurs mesures de centralité permettant d'identifier les nœuds importants dans un réseau. Pour la première stratégie, les deux variantes suivantes sont testées séparément sur le même réseau :

1. Identifier les nœuds clés en se servant du nombre de cliques auxquelles ils appartiennent.
2. Identifier les nœuds clés à l'aide d'une combinaison de mesures de centralité.

4.1.1 Identification par le nombre de cliques

La première variante consiste à déterminer dans un premier temps les communautés cliques du réseau. Par la suite, les nœuds appartenant à un nombre élevé de cliques tel que décrit dans le chapitre 2 sont ciblés pour la suppression. Cette variante que nous appelons la méthode *Clique* est similaire à la méthode utilisée dans [21] laquelle on choisit un nœud aléatoirement à l'intérieur des communautés alors que nous, nous proposons d'utiliser et tester la déstabilisation en éliminant les nœuds clés identifiés par le nombre de participations aux communautés cliques. Puisqu'un acteur peut faire partie de plus d'une clique, il serait intéressant de découvrir lesquels appartiennent au plus grand nombre de cliques et les retirer du réseau. En effet, les acteurs impliqués dans plusieurs cliques sont plus susceptibles de propager de l'information et d'exercer de l'influence sur le réseau. Il va de soi que la méthode ne s'applique pas en l'absence de cliques dans un réseau.

Le tableau 4.2 classe les nœuds du réseau R5 de la figure 4.1 par ordre d'importance selon le nombre de cliques. Il montre que les nœuds *5, 6, 7, 8* et *9* participent à un plus grand nombre de cliques que les autres. Ils seront donc les premiers à être retirés du réseau.

Tel que décrit dans la section 2.4, les cliques sont des communautés dans lesquelles chaque acteur est connecté à tous les autres dans cette structure. Pour éviter les cas triviaux, nous travaillons uniquement avec les cliques ayant au moins trois acteurs.

Un autre point important concernant les communautés cliques est que dans le langage R, ces dernières sont identifiées en ignorant l'orientation des liens. Ceci est dû

Nœud	Nombre de cliques
5	3
6	3
7	3
8	3
9	3
4	2
10	2
13	2
14	2
15	2
1	1
2	1
3	1
11	1
12	1

TABLE 4.2 – R5 : Nombre de cliques

au fait que les cliques existent seulement si la cohésion est maximale, ce qui ne peut être obtenu lorsque le réseau est orienté [7]. Par conséquent, bien que nous effectuions nos tests sur des réseaux orientés, ces derniers sont traités comme des réseaux non orientés dans la détermination des communautés cliques. Nous tenons tout de même à expérimenter les cliques car cela apporte une diversité à nos stratégies de désintégration. En effet, nous pouvons ainsi comparer nos méthodes de déstabilisation avec et sans considération de l'orientation des liens.

4.1.2 Identification par une combinaison de centralités

La seconde variante consiste tout simplement à identifier les nœuds clés d'une manière globale sans aucune référence aux communautés. L'identification se fait sur la base d'une ou plusieurs mesures de centralité comme par exemple *PageRank*. Cette variante est similaire à la technique proposée par Charausia et Tiwari [10] uniquement dans le cas où la centralité de *PageRank* est retenue alors que nous envisageons de tester une combinaison de mesures de centralité des nœuds.

Pour cette approche, nous calculons la centralité de *PageRank* ainsi que la centralité de proximité des nœuds. Puisque le langage R offre plusieurs paramètres dans

le calcul de ces centralités, précisons que ces dernières sont calculées comme suit :

Centralité de PageRank : page_rank(g,directed = TRUE, weights = poids)

Centralité de proximité : centr_clo(g, mode = 'in', normalized = TRUE)

où g est le graphe ;
$directed = TRUE$ indique que le réseau est orienté ;
$weights = poids$ pour considérer la pondération des liens ;
$mode =' in'$ pour considérer les liens entrants ;
$normalized = TRUE$ pour normaliser la centralité au niveau du réseau.

Une fois les valeurs de centralité obtenues, la moyenne est calculée pour chaque nœud afin d'identifier ceux ayant une moyenne élevée. Le tableau 4.3 donne la centralité de proximité et de *PageRank* des nœuds ainsi que la moyenne de leurs centralités respectives. Notons que ce tableau classe les nœuds en ordre décroissante de leur centralité. Bien que l'ordre d'importance des nœuds soit similaire pour les deux centralités, il existe tout de même une légère différence entre ces dernières. Nous constatons ainsi que les cinq nœuds clés sont *15, 14, 13, 12* et *11* pour les trois mesures. Par contre, on voit que le nœud *7* se trouve au 9^e rang selon la centralité de proximité alors qu'il est en 6^e position selon la centralité de PageRank. Afin de normaliser la position des nœuds, nous optons pour la moyenne de ces deux centralités. Dorénavant, nous nous référerons à cette variante sous le nom de *ProxiRank*. Nous avons quand même retenu les deux mesures de centralité (*Proximité* et *PageRank*) parce qu'elles nous paraissent également appropriées pour identifier les nœuds à éliminer.

4.1.3 Identification de liens importants

La deuxième approche consiste à démanteler le même réseau par élimination de liens importants incluant les nœuds adjacents sans aucune identification préalable des communautés. Nous référons à cette méthode sous le nom de *Lien-clé*. Tel que mentionné dans la section 2.2.2, ces liens peuvent être déterminés soit par leur centralité d'intermédiarité, soit par l'écart d'efficacité causé par leur retrait du réseau (voir l'équation 2.12). Dans ce mémoire, l'identification de liens importants se fait sur la base de la centralité d'intermédiarité dans le but de vérifier la théorie de Hopkins [18] mentionnant que le réseau petit-monde est vulnérable aux attaques fondées sur cette

Nœud	Centralité de proximité
15	0.359
14	0.2917
13	0.2373
12	0.1728
11	0.1687
10	0.14
9	0.1273
8	0.1157
7	**0.1061**
6	0.0972
5	0.0897
4	0.0828
2	0.0714
3	0.0714
1	0.0667

Nœud	Centralité de PageRank
15	0.1754
13	0.1267
14	0.1033
11	0.0853
12	0.072
7	**0.0695**
9	0.0641
10	0.0575
6	0.0533
8	0.0472
5	0.0459
4	0.0316
2	0.0254
3	0.0229
1	0.0199

Nœud	Centralité moyenne
15	0.2672
14	0.1975
13	0.182
11	0.127
12	0.1224
10	0.0988
9	0.0957
7	**0.0878**
8	0.0814
6	0.0752
5	0.0678
4	0.0572
2	0.0484
3	0.0472
1	0.0433

TABLE 4.3 – R5 : Centralité de proximité et PageRank

centralité.

Selon Wiil et al. [34], les nœuds qui communiquent via un lien important deviennent eux-mêmes importants. En suivant ce raisonnement, nous supprimons non seulement le lien important mais aussi ses deux nœuds adjacents. De cette façon, la déstabilisation progresse plus rapidement que si seulement le lien important est supprimé. Le tableau 4.4 donne la centralité d'intermédiarité normalisée des liens du réseau R5 de la figure 4.1. Ce tableau est trié en ordre décroissant de la centralité d'intermédiarité des liens. Tout indique que les nœuds *5,7* et *12* seraient les premiers à retirer du réseau avec éventuellement d'autres nœuds selon le seuil établi par l'utilisateur.

Lien	Centralité d'intermédiarité
5-7	0.1209
7-12	0.1099
6-8	0.1044
12-14	0.1044
4-6	0.0989
8-10	0.0824
10-11	0.0659
3-4	0.0604
9-10	0.0549
14-15	0.0549
4-5	0.0495
10-12	0.0495
11-13	0.0495
1-3	0.044
2-4	0.044

Lien	Centralité d'intermédiarité
3-5	0.044
7-9	0.033
8-9	0.033
12-13	0.033
2-11	0.022
6-7	0.022
1-12	0.0165
13-14	0.0165
1-2	0.011
7-8	0.011
13-15	0.011
1-15	0.0055
2-15	0.0055
5-6	0.0055
9-11	0

TABLE 4.4 – R5 : Centralité d'intermédiarité des liens

4.2 Étape 2 : La déstabilisation

Après avoir calculé la centralité des nœuds et des liens, le processus de déstabilisation peut débuter. La façon la plus simple est de retirer les nœuds ou les liens un par un en commençant par celui ayant la centralité la plus élevée. C'est donc un long processus pour un réseau ayant beaucoup de nœuds et de liens. Dans la section 4.2.1, nous présentons deux options permettant d'accélérer la simulation de la déstabilisation.

4.2.1 Déterminer les nœuds cibles

Dans le but d'accélérer le processus de déstabilisation, nous choisissons d'éliminer les groupes de nœuds ou de liens au lieu de les supprimer un par un. Une des alternatives de groupement de nœuds est l'utilisation de la moyenne et l'écart-type des valeurs de centralité. Les nœuds cibles sont ceux dont la centralité se situe dans l'intervalle $[S, C_{max}]$ où C_{max} est la valeur maximale de la centralité et $S = C_{moy} + (k * E_{typ})$. C_{moy} représente la centralité moyenne des nœuds ou des liens, E_{typ} est l'écart-type et k est un nombre entier positif qui doit être choisi de sorte que S s'approche le plus

possible de la valeur de C_{max} sans toutefois la dépasser.

Une autre option de groupement est de choisir les nœuds sur la base du rang centile de leur centralité. Par exemple, supprimer tous les nœuds dont la centralité se situe au 90^e centile de la valeur maximale. Les tableaux 4.5 et 4.6 donnent la liste des nœuds se trouvant au 90^e et qui seraient les cibles à supprimer dans la première itération pour chacune des trois méthodes proposées.

Méthode Clique	
Nœud	**Nb. cliques**
5	3
6	3
7	3
8	3
9	3
4	2
10	2
13	2
14	2
15	2
1	1
2	1
3	1
11	1
12	1
90^e centile : 3	
Cibles : 5, 6, 7, 8, 9	

Méthode ProxiRank	
Nœud	**Centralité moyenne**
15	0.2672
14	0.1975
13	0.182
11	0.127
12	0.1224
10	0.0988
9	0.0957
7	0.0878
8	0.0814
6	0.0752
5	0.0678
4	0.0572
2	0.0484
3	0.0472
1	0.0433
90^e centile : 0.1913	
Cibles : 14, 15	

TABLE 4.5 – Rang centile des centralités des nœuds

Méthode Lien-clé			
Lien	**Centralité**	**Lien**	**Centralité**
5 - 7	0.1209	3 - 5	0.044
7 - 12	0.1099	7 - 9	0.033
6 - 8	0.1044	8 - 9	0.033
12 - 14	0.1044	12 - 13	0.033
4 - 6	0.0989	2 - 11	0.022
8 - 10	0.0824	6 - 7	0.022
10 - 11	0.0659	1 - 12	0.0165
3 - 4	0.0604	13 - 14	0.0165
9 - 10	0.0549	1 - 2	0.011
14 - 15	0.0549	7 - 8	0.011
4 - 5	0.0495	13 - 15	0.011
10 - 12	0.0495	1 - 15	0.0055
11 - 13	0.0495	2 - 15	0.0055
1 - 3	0.044	5 - 6	0.0055
2 - 4	0.044	9 - 11	0
90^e centile : 0.1044 Cibles : 5, 6, 7, 8, 12, 14			

TABLE 4.6 – Rang centile des centralités des liens

Dans le cadre de ce mémoire, le groupement sur la base du rang centile a été retenu pour sa simplicité. Notons également que lors des expérimentations, nous avons constaté que pour les réseaux de grande taille mais de densité faible, la méthode *Lien-clé* supprime tous les nœuds du réseau en une seule itération lorsque le 90^e centile est appliqué pour déterminer les liens cibles. Un tel résultat ne nous permet pas de valider la performance de la méthode puisqu'il est impossible de suivre la progression de la déstabilisation. Afin d'observer une déstabilisation progressive, nous avons choisi le 96^e centile pour les petits réseaux et le 99.9^e centile pour les grands réseaux.

4.2.2 Déterminer la condition d'arrêt

Dans le processus de déstabilisation, nous surveillons la variation de quatre mesures globales du réseau : la densité, le degré moyen, le nombre de composants ainsi que la variation de l'efficacité. Nous fixons alors la condition d'arrêt sur la base du degré de dégradation de l'efficacité. Rappelons que la variation de l'efficacité est donnée par $\Delta E = E(G) - E(G - i)$ où $E(G)$ est l'efficacité du réseau initial et $E(G - i)$ est

l'efficacité du réseau suite au retrait du nœud i. Plus un retrait provoque un grand écart d'efficacité, plus le réseau est déstabilisé. Le processus se termine lorsque ΔE atteint une proportion définie par l'utilisateur.

4.3 Étape 3 : Mesurer le résultat de déstabilisation

Tel que décrit dans la sous-section 2.2.1, l'analyse de réseaux sociaux offre plusieurs mesures globales permettant de caractériser les réseaux. Dans cette section, nous analysons l'effet de la déstabilisation en surveillant la densité, le degré moyen, le nombre de composants déconnectés ainsi que la variation de l'efficacité du réseau obtenue suite à la suppression de nœuds. Ces mesures sont prises sur le réseau initial et sur le réseau résultant. Les trois premières mesures sont simples à calculer. Pour l'efficacité en revanche, il existe plus d'une façon de la calculer. Nous explorons au moins trois formules afin de choisir celle qui convient le mieux.

4.3.1 Efficacité du réseau

Une des formules les plus simples est sans doute la formule de la compacité soit l'équation 2.2 présentée dans la sous-section 2.2.1. Nous la reprenons ci-dessous pour faciliter la consultation.

$$\frac{\sum_{i \neq j} \frac{1}{d_{ij}}}{n(n-1)}$$

Celle-ci permet de mesurer l'indice de connectivité du réseau. Plus cet indice est élevé, plus la cohésion du réseau est élevée et plus la circulation d'information est fluide. Puisque l'efficacité d'un réseau est étroitement liée à sa capacité de diffuser l'information, la compacité peut effectivement servir de mesure d'efficacité.

Cependant, nous avons constaté que cette mesure comporte une lacune dont il faut tenir compte. L'expérimentation a révélé que dans certains cas, l'efficacité du réseau obtenue par le calcul de compacité augmente après une suppression de nœuds alors qu'on s'attend à la voir diminuer. Dans une telle situation, il devient impossible de conclure si le retrait des nœuds cibles a provoqué la détérioration attendue

du réseau. Le tableau 4.7 montre un exemple de cette lacune. Ce tableau contient le résultat de déstabilisation par la méthode *ProxiRank* qui supprime les nœuds en commençant par ceux ayant une valeur élevée pour la moyenne des centralités de proximité et PageRank. Afin d'observer la progression du résultat, les nœuds ont été supprimés un par un. On peut constater qu'après le retrait du nœud *13*, l'efficacité du réseau a augmenté au lieu de diminuer par rapport à l'efficacité du réseau initial. Cette augmentation a engendré un ΔE négatif ce qui va à l'encontre de l'objectif de la déstabilisation. Pour un deuxième exemple, regardons l'itération 9. Après le retrait du nœud *8*, on voit qu'il ne reste plus que six nœuds. Pourtant, l'efficacité a une plus grande valeur qu'à l'itération 0 où le réseau comportait encore tous les 15 nœuds. Par conséquent, l'écart d'efficacité est également négatif pour cette itération. D'ailleurs, nous pouvons observer que contre toute attente, cet écart est négatif dans presque toutes les itérations au lieu d'augmenter graduellement. Cette lacune nous amène alors à douter de la pertinence de cette mesure et donc à l'écarter.

Itération	**Nœud retiré**	**Efficacité**	ΔE	**Nœuds restants**
0		0.0932	0	15
1	15	0.0919	0.0013	14
2	14	0.0962	-0.0043	13
3	**13**	**0.1012**	**-0.005**	**12**
4	11	0.1025	-0.0013	11
5	12	0.104	-0.0015	10
6	10	0.1073	-0.0033	9
7	9	0.1157	-0.0084	8
8	7	0.1183	-0.0026	7
9	**8**	**0.1262**	**-0.0079**	**6**
10	6	0.1398	-0.0136	5

TABLE 4.7 – Efficacité basée sur la compacité

La deuxième mesure d'efficacité expérimentée dans ce mémoire est la *performance totale* définie par Lindelauf et al. [22]. Ces auteurs ont défini la performance totale du réseau comme étant le produit de sa performance moyenne et de sa performance en termes de secret. Cette performance totale du réseau est calculée selon l'équation 4.1.

$$P_g = S_g * I_g \tag{4.1}$$

où S_g est la performance en terme de secret du réseau. Celle-ci est définie à l'aide de deux paramètres qui sont la probabilité d'exposition et la probabilité de détection des individus. Cette performance S_g est calculée selon l'équation 4.2. I_g représente la performance moyenne du réseau et est calculée par l'équation 4.3 :

$$S_g = \sum_{i=1}^{n} \alpha_i * \mu_i \tag{4.2}$$

$$I_g = \frac{n * (n-1)}{T_g} \tag{4.3}$$

où $\alpha_i = \frac{1}{n}$, $\mu_i = 1 - \frac{d_i+1}{n}$ et d_i est la centralité de degré du nœud i. Cette performance est fondée sur le principe que chaque nœud i a une probabilité α_i d'être découvert en tant que membre du réseau. Lorsque i est découvert, celui-ci exposera une fraction du réseau représentée par $1 - \mu_i$ où μ_i est la fraction du réseau qui demeure secrète après la détection de i. T_g représente la distance totale du réseau qui est donnée par la somme des plus courts chemins.

La lacune de la formule de *performance totale* est qu'elle fonctionne seulement avec les réseaux connectés. Lorsque le réseau comporte plusieurs composants, certains des courts chemins sont indéfinis puisque les nœuds sont incapables de se rejoindre. Pour pallier à ce problème, nous calculons la distance totale, T_g en remplaçant les valeurs infinies par la plus grande valeur des plus courtes distances augmentée de la valeur 1 de sorte qu'elles soient supérieures aux autres valeurs du réseau.

Nous avons expérimenté la formule de la *performance totale* afin de calculer l'efficacité du réseau après le retrait des nœuds. Le tableau 4.8 montre le résultat de déstabilisation dont l'efficacité est calculée par l'équation 4.1.

Itération	**Nœud retiré**	**Efficacité**	ΔE	**Nœuds restants**
0		**0.0768**	**0**	**15**
1	15	0.0767	0.0001	14
2	14	0.0768	0	13
3	13	0.077	-0.0002	12
4	11	0.0771	-0.0003	11
5	12	0.0761	0.0007	10
6	10	0.0804	-0.0036	9
7	9	0.0953	-0.0185	8
8	7	0.0937	-0.0169	7
9	8	0.1051	-0.0283	6
10	**6**	**0.1212**	**-0.0444**	**5**

TABLE 4.8 – Efficacité calculée par la *performance totale*

Similaire à la formule de la *compacité*, l'efficacité par la *performance totale* du réseau augmente alors que la déstabilisation progresse. En effet, nous observons qu'après l'itération 10, il ne reste plus que cinq nœuds. Pourtant, l'efficacité est de 0.1212 qui est plus élevée que celle du réseau initial à 0.0768 et un écart d'efficacité (ΔE) négatif. De plus, on peut constater que l'évolution du ΔE ne suit aucune logique de la déstabilisation. Ce résultat d'expérimentation nous a alors amené à conclure que cette formule ne nous permet pas de mesurer adéquatement l'impact de la désintégration du réseau.

Dans la recherche de la formule d'efficacité appropriée, nous proposons une nouvelle formule appelée *Taux de connectivité* (TC) du réseau à l'itération m qui est définie par l'équation 4.4.

$$TC = \frac{\sum_{i=1}^{m} C_i}{p * \sum_{i=1}^{n} C_i} \tag{4.4}$$

où n est le nombre de nœuds du réseau initial, m est le nombre de nœuds restants (non encore éliminés) à la suite d'une suppression, p est le nombre de composants déconnectés, C_i est la compacité du nœud i donnée par la somme des compacités

entre ce dernier et les autres nœuds du réseau initial. Formellement, $C_i = \sum_{j=1}^{n} \frac{1}{d_{ij}}$ et d_{ij} est la plus courte distance entre i et j. Le tableau 4.9 présente le résultat de calcul de l'efficacité selon l'équation 4.4.

Itération	**Nœud retiré**	**Efficacité (TC)**	ΔE	**Nœuds restants**
0		1	0	15
1	15	0.8544	0.1456	14
2	14	0.7668	0.2332	13
3	13	0.6829	0.3171	12
4	11	0.5763	0.4237	11
5	12	0.4782	0.5218	10
6	10	0.3949	0.6051	9
7	9	0.3311	0.6689	8
8	7	0.2804	0.7196	7
9	8	0.1934	0.8066	6
10	6	0.1429	0.8571	5

TABLE 4.9 – Efficacité calculée par le taux de connectivité

Contrairement aux deux premières formules d'efficacité expérimentées précédemment, on peut constater que l'efficacité calculée selon le *Taux de connectivité* diminue graduellement à mesure que la déstabilisation progresse et tel qu'anticipé, la valeur de ΔE suit une progression ascendante. Afin de nous assurer qu'il existe toujours une diminution progressive de la valeur de TC, nous établissons trois lemmes.

Lemme 1 : La valeur maximale de TC est 1 pour tout réseau R.
Démonstration :
Soit R un réseau comportant n nœuds et supposons que

$$Cm = \sum_{i=1}^{m} C_i$$

$$Cn = \sum_{i=1}^{n} C_i$$

On a alors $TC = \frac{Cm}{p*Cn}$ et $TC_{max} = TC^0$ où TC^0 est le *Taux de connectivité* du réseau à l'itération 0, i.e., avant que le processus de déstabilisation ne débute. Puisque le

nombre de nœuds restants est le même que le nombre de nœuds du réseau, on a donc $m = n$, ce qui implique que $Cm = Cn$. D'autre part, étant donné qu'aucun nœud n'a encore été retiré du réseau, il est trivial que le nombre de composants déconnectés est égal à 1. TC^0 est alors donné par $TC^0 = \frac{Cn}{Cn} = 1$.

Lemme 2 : La valeur minimale de TC est toujours égale à 0 peu importe le réseau
Démonstration :
Soit R un réseau comportant n nœuds et supposons que

$$Cm = \sum_{i=1}^{m} C_i$$

$$Cn = \sum_{i=1}^{n} C_i$$

On a alors $TC = \frac{Cm}{p*Cn}$. Donc $TC_{min} = \frac{Cm}{p*Cn} = 0$ si et seulement si $Cm = 0$ et que $p * Cn > 0$

Pour avoir $Cm = 0$, l'une des conditions suivantes doit être satisfaite :

- Tous les nœuds sont supprimés du réseau
- Chaque nœud non supprimé forme un composant isolé

Lorsque tous les nœuds sont supprimés, il est trivial que $Cm = 0$. Regardons le cas où il reste des nœuds isolés. Dans ce cas, on sait que le plus court chemin entre ces derniers n'existe pas. La distance géodésique étant indéfinie, la compacité d'un nœud i est alors donnée par $C_i = \frac{1}{\infty} = 0$, donc, $Cm = 0$.

Vérifions maintenant que la relation $p * Cn > 0$ est toujours vraie. On sait que $Cn > 0$ tout au long du processus de déstabilisation car c'est la somme de compacité des nœuds du réseau à l'état initial. Quant à p, rappelons que c'est le nombre de composants déconnectés, ce qui implique qu'il sera toujours supérieur à 0. Il est donc trivial que $p * Cn > 0$ dans toutes les étapes de la déstabilisation, et donc 0 est la plus petite valeur possible pour TC_{min}.

Lemme 3 : La valeur de TC diminue en fonction du nombre de nœuds restants
Démonstration :
Soit R un réseau comportant n nœuds et supposons que

$$Cm = \sum_{i=1}^{m} C_i$$

$$Cn = \sum_{i=1}^{n} C_i$$

On a alors $TC = \frac{Cm}{p*Cn}$. De plus, on sait que initialement $m = n$ et que $TC^0 = 1$. Supposons qu'à l'itération k, x nœuds sont supprimés du réseau. Donc, m à l'itération k est donné par $m^{(k)} = m - x$. Ce qui implique que $m^{(k)} < m$ et donc $Cm^{(k)} < Cm$.

Supposons maintenant qu'à l'itération $k + 1$, x autres nœuds sont à leur tour retirés du réseau. Le nombre de nœuds non supprimés m à l'itération $k + 1$ est alors donné par $m^{(k+1)} = m^{(k)} - x$. Ce qui implique que $m^{(k+1)} < m^{(k)} < m$ et que $Cm^{(k+1)} < Cm^{(k)} < Cm$.

On constate donc que la valeur de Cm qui est le numérateur diminue en fonction du nombre de nœuds supprimés alors que Cn qui fait partie du dénominateur reste le même. Puisque p varie seulement à la hausse, nous pouvons confirmer que le dénominateur de la formule *Taux de connectivité* sera toujours supérieur à la valeur de Cm tout au long du processus de déstabilisation. Ce qui démontre que $\frac{Cm^{(k+1)}}{p*Cn} < \frac{Cm^{(k)}}{p*Cn} < \frac{Cm}{p*Cn}$ et donc $TC^{(k+1)} < TC^{(k)} < TC^0$. Nous pouvons alors affirmer avec certitude que la valeur de TC à l'itération k sera impérativement inférieure à celle de l'itération précédente. La relation $TC^{(k+1)} < TC^{(k)}$ est donc toujours vraie.

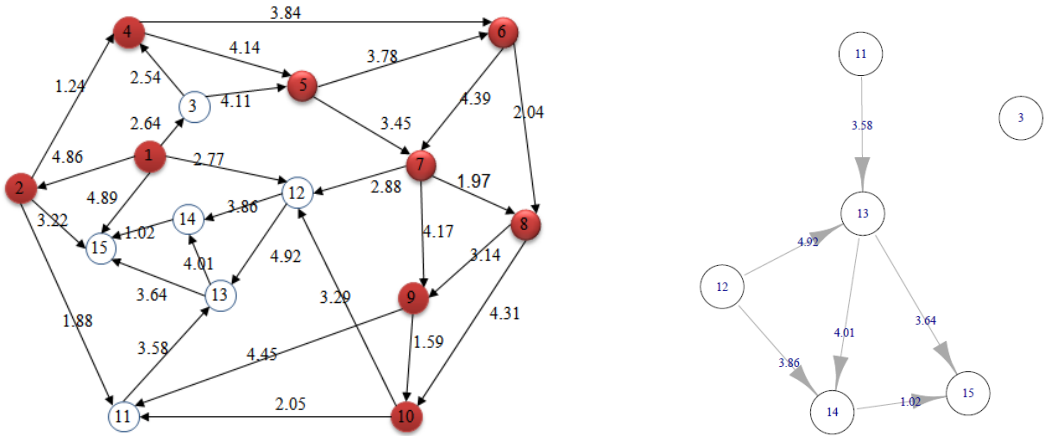
En conclusion, l'efficacité selon le *Taux de connectivité* sera toujours dans l'intervalle [0,1]. De plus, cette formule nous garantit que l'efficacité et l'écart d'efficacité évoluent en fonction de la taille du réseau résultant, ce qui nous permet de quantifier adéquatement l'impact de la déstabilisation. C'est la raison pour laquelle elle a été retenue pour ce mémoire.

4.3.2 Autres mesures globales du réseau

Tel que mentionné précédemment, l'efficacité n'est pas la seule mesure utilisée pour valider la déstabilisation. Dans le présent mémoire, nous expérimentons quatre mesures globales :la densité, le degré moyen, le nombre de composants isolés et la variation de l'efficacité du réseau. Nous rappelons que $\Delta E = TC(G) - TC(G - i)$ puisque nous avons retenu TC comme mesure. Le tableau 4.10 montre un exemple de statistiques de déstabilisation par la méthode *Clique* effectuée sur le réseau R5 de la figure 4.1.

Itération	**Nœuds retirés**	**Densité**	**Degré moyen**	**Nombre de composants**	**Nb. retirés**	**Efficacité (TC)**	ΔE
0		0.1429	0.2857	1	0	1	0
1	5, 6, 7, 8, 9	0.1778	0.3556	1	5	0.7062	0.2938
2	10	0.1944	0.3889	1	6	0.6229	0.3771
3	1, 2	0.1667	0.3333	2	8	0.3062	0.6938
4	4	0.2	0.4	2	9	0.2706	0.7294

TABLE 4.10 – Statistiques de déstabilisation du réseau R5 par la méthode Clique



(a) R5 : Réseau initial.

(b) Réseau R5 déstabilisé.

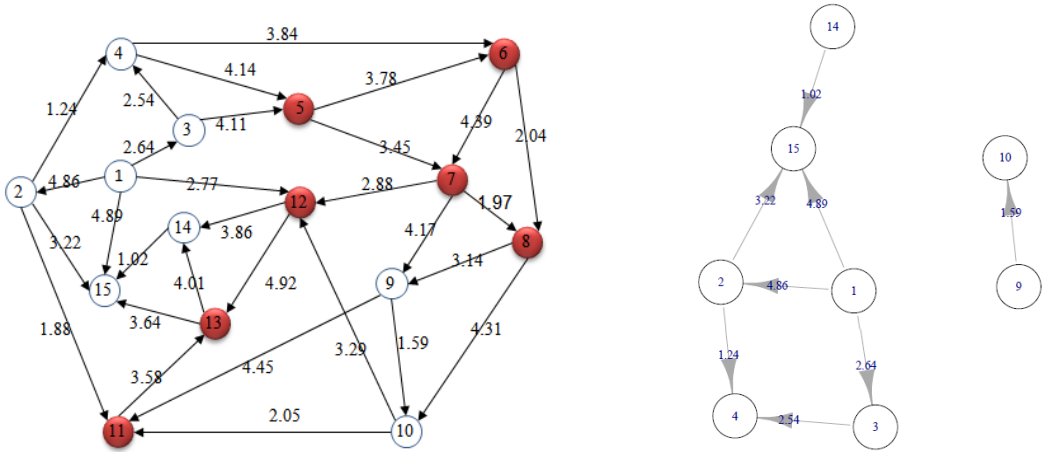
FIGURE 4.2 – Réseau R5 déstabilisé par la méthode Clique.

Le tableau 4.10 confirme qu'initialement, le réseau a une efficacité égale à 1. À l'itération 1, 5 nœuds sont retirés du réseau et l'efficacité a chuté à 0.7062 avec un ΔE de 0.2938. À la fin du processus, il reste une efficacité d'environ 27% et un ΔE de 73% approximativement, ce qui respecte la condition d'arrêt pour laquelle le seuil d'efficacité est fixé à une réduction de 70%. Dans la figure 4.2, on peut voir le graphe initial où les nœuds cibles sont identifiés en rouge. À la fin du processus de déstabilisation, il ne reste plus que six nœuds dont un est complétement isolé.

Regardons un deuxième exemple présenté dans le tableau 4.11 et la figure 4.3 où la déstabilisation est réalisée avec la méthode *Lien-clé*.

Itération	**Nœuds retirés**	**Densité**	**Degré moyen**	**Nombre de composants**	**Nb. retirés**	**Efficacité (TC)**	ΔE
0		0.1429	0.2857	1	0	1	0
1	5, 7, 12	0.1364	0.2727	1	3	0.8093	0.1907
2	6, 8, 11, 13	0.1429	0.2857	2	7	0.2406	0.7594

TABLE 4.11 – Statistiques de déstabilisation du réseau R5 par la méthode Lien-clé



(a) R5 : Réseau initial. (b) Réseau R5 déstabilisé.

FIGURE 4.3 – Réseau R5 déstabilisé par la méthode Lien-clé.

Comme on peut le constater, sept nœuds sont retirés du réseau et l'efficacité finale est d'environ 24% avec un ΔE de 76% approximativement. Bien entendu, il est encore trop tôt pour tirer des conclusions. Mais si on compare rapidement les deux méthodes utilisées dans les deux exemples précédents, la méthode *Lien-clé* semble plus performante car elle supprime moins de nœuds et inflige un plus grand dommage au réseau. En effet, la méthode *Clique* supprime neuf nœuds et pourtant, elle cause moins de dommage car l'efficacité finale est de 27% dont 3% de plus que la méthode *Lien-clé* qui supprime seulement sept nœuds et réduit l'efficacité du réseau de 76%. Nous en discutons davantage dans les prochaines sous-sections lors des expérimentations plus rigoureuses sur un volume plus important de données.

4.4 Analyse empirique

Afin de valider notre démarche, nous nous servons d'outils d'analyse de réseaux sociaux tels que *Ucinet* [6] et du langage R [28] très utilisés dans ce domaine de recherche. Les réseaux sont générés pour ensuite être déstabilisés selon les méthodes décrites dans le présent chapitre. La déstabilisation est effectuée en plusieurs itérations successives et chacune est exécutée en trois étapes telles que décrites dans les sections précédentes. Pour chaque itération, tous les nœuds dont la centralité se situe à valeur du centile prédéfini sont supprimés. Après la suppression, nous mesurons le résultat à l'aide de la densité, le degré moyen, le nombre de composants ainsi que l'efficacité obtenue par l'équation 4.4 qui est fondée sur le taux de connectivité des nœuds. Le processus de déstabilisation est répété jusqu'à ce que l'efficacité du réseau soit réduite d'un pourcentage fixé par l'utilisateur. Dans le cadre de cette recherche, nous avons choisi une réduction de 70%.

La validation est effectuée sur deux modèles de réseaux : *petit-monde* et *aléatoire* qui sont générés synthétiquement. Nous avons écarté les réseaux *invariants d'échelle* car ce type de réseau est très vulnérable aux attaques ciblées. Les expérimentations ont démontré que ces derniers sont démantelés après seulement deux ou trois itérations.

4.4.1 Données

Ayant pour objectif d'obtenir un résultat suffisamment concluant, la simulation est appliquée sur plusieurs séries de réseaux de différentes tailles et différentes valeurs de densité. En premier lieu, deux séries de réseaux de petite taille sont générés pour chacun des deux modèles : *petit-monde* et *aléatoire*. Ces réseaux comportent respectivement 100 et 200 nœuds. Par la suite, trois séries de réseaux de plus grande taille sont également expérimentés. Ces derniers comportent respectivement 1000, 2000 et 3000 nœuds. Tous les réseaux sont orientés et pondérés. Afin de varier la composition des réseaux, nous avons généré plusieurs séries de différentes densités pour les réseaux de 200 et de 1000 nœuds. Pour chaque série, cinq réseaux sont générés pour un total de quatre-vingt réseaux [1] pour les deux modèles combinés. Les tableaux 4.12 et 4.13 présentent les caractéristiques des réseaux testés. Chaque réseau de ces tableaux est déstabilisé par les trois méthodes décrites dans la section 4.1 : *ProxiRank*, *Clique* et *Lien-clé*.

Réseaux petit-monde		
Nombre de nœuds	**Densité**	**Nombre de réseaux testés**
100	0.05	5
200	0.02, 0.04, 0.05, 0.06, 0.08	5
1000	0.05, 0.1, 0.2, 0.3	20
2000	0.05	5
3000	0.05	5

TABLE 4.12 – Séries de réseaux petit-monde expérimentés

1. (2 * (5 + 5 + 20 + 5 +5)) = 80

Réseaux aléatoires		
Nombre de nœuds	**Densité**	**Nombre de réseaux testés**
100	0.05	5
200	0.02, 0.04, 0.05, 0.06, 0.08	5
1000	0.05, 0.1, 0.2, 0.3	20
2000	0.05	5
3000	0.05	5

TABLE 4.13 – Séries de réseaux aléatoires expérimentés

Ayant pour objectif de diversifier les données d'essai, nous avons également utilisé deux réseaux réels tirés de la base de données des réseaux de scientifiques rendus disponibles par M.E.J Newman. Le premier est un réseau de collaboration sur la théorie de haute énergie (*High-energy theory collaborations : hep-th*) fourni par [26]. Le deuxième est le réseau de collaboration de scientifiques (*Coauthorships in network science : netscience*) fourni par [25]. Mentionnons toutefois que nous n'avons pas utilisé ces réseaux tels quels. Ces derniers étant non connectés, nous avons extrait le plus grand composant de chacun de ces réseaux pour construire des réseaux de validation. Le tableau 4.14 présente les caractéristiques de ces derniers. Notons également que seulement les méthodes *ProxiRank* et *Lien-clé* sont testées avec ces réseaux.

Composant de	**Nombre de nœuds**	**Nombre de liens**	**Densité**
hep-th	5835	13815	0.0008
netscience	379	914	0.0127

TABLE 4.14 – Caractéristiques des réseaux réels testés

4.4.2 Statistiques de déstabilisation

Afin de faciliter la comparaison entre chacune des méthodes expérimentées, nous enregistrons les statistiques de déstabilisation du réseau. À la fin du processus, nous analysons le degré de démantèlement à l'aide de plusieurs mesures globales incluant la densité, le degré moyen, le nombre de composants ainsi que l'écart de l'efficacité

(ΔE) laquelle mesure la dégradation du réseau. Le tableau 4.15 présente un exemple de statistiques de déstabilisation d'un réseau.

It.	**Nœuds retirés**	**Densité**	**Degré moyen**	**Nombre de composants**	**Nb. retirés**	**Efficacité (TC)**	ΔE
0		0.0487	0.0974	1	0	1	0
1	16, 36, 46, 14, 32, 78	0.0466	0.0931	1	6	0.9326	0.0674
2	87, 40, 70, 74, 86	0.0472	0.0945	1	11	0.8814	0.1186
3	82, 39, 67, 81	0.0476	0.0952	1	15	0.8404	0.1596
4	38, 65, 77, 18, 20, 45, 47	0.048	0.0959	1	22	0.7756	0.2244
5	71, 2, 11, 21, 31, 35, 37, 41, 42, 59, 60, 64, 68	0.0486	0.0971	1	35	0.6389	0.3611
6	27, 33, 29, 34	0.047	0.094	1	39	0.6	0.4
7	23, 25, 30, 43	**0.0467**	**0.0934**	1	**43**	0.5635	0.4365
8	19, 26, 15, 55	0.0483	0.0965	1	47	0.5205	0.4795
9	13, 51, 3, 7, 8, 17, 22, 24, 44, 50	**0.0482**	**0.0963**	2	**57**	0.2123	0.7877

TABLE 4.15 – Exemple de statistiques de déstabilisation d'un réseau

Ce tableau nous permet d'observer que les trois premières mesures, dont la densité, le degré moyen et le nombre de composants ne sont pas significatives dans la quantification du niveau de désintégration du réseau. Par exemple, on voit que la densité globale du réseau ne varie pas en fonction du nombre de nœuds supprimés. En effet, nous constatons qu'après l'itération 9, cinquante-sept nœuds sont éliminés. Pourtant, la densité est supérieure à celle après l'itération 7 où seulement quarante-trois nœuds sont retirés. Selon cette mesure, le réseau n'est pas détérioré. Au contraire, il est même en meilleure condition après le retrait de ses membres. Nous avons également observé un résultat similaire pour le degré moyen. Quant au nombre de composants, il n'a pas beaucoup d'intérêt lorsqu'il est utilisé seul. Toutefois, nous l'avons intégré dans le calcul d'efficacité du réseau c'est-à-dire l'équation 4.4 que nous reprenons ici-bas :

$$TC = \frac{\sum_{i=1}^{m} C_i}{p* \sum_{i=1}^{n} C_i}$$

Dans cette équation, p est le nombre de composants. Comme on peut le voir, plus il y a de composants déconnectés, plus l'efficacité du réseau diminue, ce qui reflète bien le degré de fluidité de l'information au sein du réseau.

Puisqu'il est impossible de quantifier le niveau de démantèlement à l'aide de ces trois mesures, nous comparons les méthodes *ProxiRank*, *Clique* et *Lien-clé* en nous servant uniquement de la mesure d'efficacité et du temps de traitement pris par le CPU. En outre, nous calculons le pourcentage de nœuds supprimés par chaque méthode pour atteindre le niveau de désintégration (ΔE) préétabli.

4.5 Résultats partiels

Les expérimentations ont été exécutées sur cinq séries totalisant 40 réseaux aléatoires et 40 réseaux petit-monde. Cependant, nous présentons dans cette section seulement les résultats pour deux séries de réseaux : la série des 200 nœuds avec densité variable et la série des mille nœuds avec une densité de 20%. De plus, mentionnons que les expérimentations sont menées sur un ordinateur ayant les caractéristiques suivantes :

Système d'opération :	Windows 10 Home
Type de système :	Intel(R) Core(TM) i5-4440 CPU @ 3.10 GHz
Processeur :	64 bits
Mémoire (RAM) :	8.00 GB

4.5.1 Résultats partiels pour de petits réseaux

Nous présentons dans cette sous-section les statistiques pour cinq petits réseaux. Le tableau 4.16 présente le sommaire des résultats obtenus pour les réseaux *petit-*

monde ayant 200 nœuds à densité variable.

Réseaux petit-monde à 200 nœuds avec densité variable					
Méthode	**Densité**	**Nb. retirés**	**ΔE finale**	**% retirés**	**Temps (sec)**
ProxiRank	0.02	91	0.7040	45	0
	0.04	89	0.7040	44	0.02
	0.05	97	0.7095	48	0.02
	0.06	88	0.7044	44	0
	0.08	97	0.7084	48	0
Moyenne		91	0.7055	45	0.008
Lien-clé	0.02	153	0.7500	71	0
	0.04	140	0.7096	69	0
	0.05	154	0.8724	62	0
	0.06	144	0.7082	71	0
	0.08	152	0.7693	69	0
Moyenne		148	0.7601	68	0
Clique	0.02	149	0.5510	95	0.19
	0.04	156	0.6054	90	1.91
	0.05	154	0.6269	86	1.15
	0.06	129	0.7006	64	12.39
	0.08	154	0.6104	88	21.96
Moyenne		147	0.6210	84	3.91

TABLE 4.16 – Statistiques des réseaux petit-monde à 200 nœuds

Ce tableau montre que le processus de déstabilisation ne s'arrête pas exactement au seuil établi pour l'écart d'efficacité. Par exemple, dans le cas du réseau avec une densité de 0.05, la méthode *Lien-clé* s'arrête avec un ΔE de 87%. Dans d'autres cas, elle s'arrête lorsque tous les nœuds du réseau sont supprimés. Lorsque cette situation se produit, nous prenons le résultat de l'itération précédente comme dans le cas du réseau avec une densité de 0.08 où on voit que la méthode *Clique* supprime 154 nœuds et réduit l'efficacité de 61%. On peut alors constater qu'il est difficile de comparer la performance des méthodes avec ces résultats bruts. Nous avons alors normalisé le nombre de nœuds retirés pour le ΔE de 70% car c'est la condition d'arrêt préétablie. Cette normalisation nous permet de voir le pourcentage de nœuds supprimés (dans

la colonne *% retirés*) pour réduire l'efficacité de 70%.

Pour ce groupe de réseaux, la méthode *ProxiRank* supprime en moyenne 91 nœuds sur 200 pour approximativement réduire l'efficacité de 71% alors que la méthode *Lien-clé* a éliminé 148 nœuds pour une réduction de 76%. La méthode *Clique*, pour sa part, a supprimé en moyenne 147 nœuds et a réduit l'efficacité du réseau de 62%. Après normalisation, les trois méthodes ont éliminé respectivement 45, 68 et 84% des nœuds. Comme nous pouvons le constater, la méthode *Clique* supprime plus de nœuds et affiche un temps d'exécution nettement plus élevé que les deux autres méthodes.

Les figures 4.4 et 4.5 donnent la représentation graphique du tableau 4.16. La figure 4.4 montre le pourcentage moyen de nœuds supprimés alors que la figure 4.5 donne le temps moyen d'exécution des trois méthodes.

Pour les réseaux *aléatoires*, les statistiques de déstabilisation sont présentées dans le tableau 4.17 pour une série de cinq réseaux ayant 200 nœuds à densité variable. Similairement aux réseaux *petit-monde*, le nombre de nœuds retirés est normalisé en pourcentage pour une réduction de l'efficacité (ΔE) de 70%. Ces valeurs sont également présentées dans les figures 4.4 et 4.5. Celles-ci permettent une meilleure visualisation du pourcentage de nœuds supprimés et le temps d'exécution pour chaque méthode. En terme de précision, nous constatons que pour ce groupe de réseaux, les trois méthodes ont supprimé approximativement le même nombre de nœuds. Par contre, la méthode *Lien-clé* semble plus avantageuse en terme de temps. Cependant, l'écart est tout de même négligeable entre *ProxiRank* et *Lien-clé*. Bien que ces résultats semblent suivre une tendance, il est encore trop tôt pour tirer des conclusions.

Pour ce qui est de la méthode *Clique*, elle a donné une fois de plus une bien piètre performance avec un temps d'exécution largement supérieur aux deux autres méthodes. Notons toutefois que l'écart est moins grand que dans le cas des réseaux *petit-monde*. Selon les résultats obtenus pour les deux types de réseau, cette méthode prend davantage plus de temps lorsque le réseau comporte un grand nombre de liens, surtout pour les réseaux *petit-monde*. Tout laisse croire que cette méthode n'est pas bien adaptée au démantèlement de réseaux.

Réseaux aléatoires à 200 nœuds avec densité variable					
Méthode	**Densité**	**Nb. retirés**	**ΔE finale**	**% retirés**	**Temps (sec)**
ProxiRank	0.02	143	0.7077	71	0
	0.04	143	0.7041	71	0.01
	0.05	138	0.7071	68	0
	0.06	143	0.7011	71	0.05
	0.08	138	0.703	69	0.01
Moyenne		142	0.7332	70	0.014
Lien-clé	0.02	144	0.7101	71	0
	0.04	144	0.7074	71	0
	0.05	129	0.8335	54	0
	0.06	145	0.7169	71	0
	0.08	150	0.7635	69	0
Moyenne		141	0.7517	67	0
Clique	0.02	122	0.6120	70	0.01
	0.04	118	0.5900	70	0.16
	0.05	129	0.6422	70	0.08
	0.06	141	0.7057	70	0.42
	0.08	123	0.6146	70	0.26
Moyenne		128	0.6417	70	0.1675

TABLE 4.17 – Statistiques des réseaux aléatoires à 200 nœuds

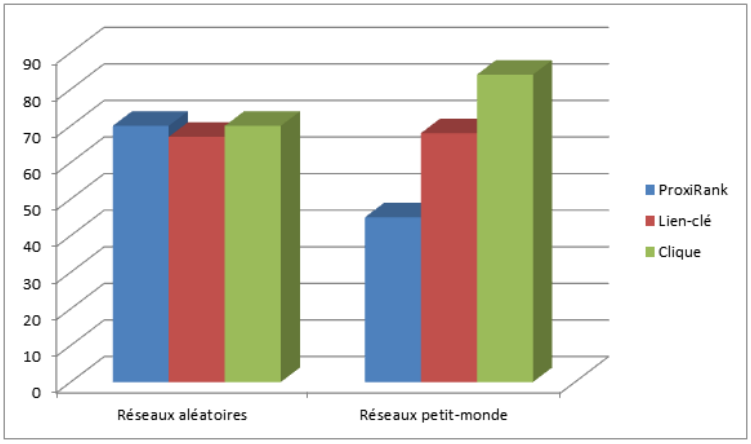


FIGURE 4.4 – Pourcentage moyen de nœuds supprimés pour les réseaux de 200 nœuds avec densité variable

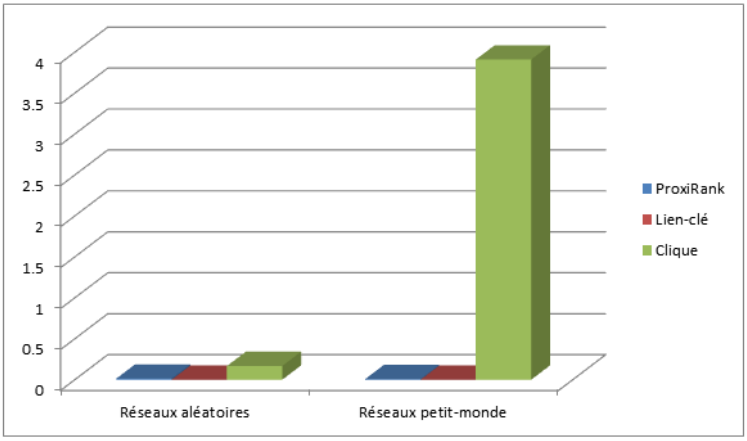


FIGURE 4.5 – Temps moyen d'exécution pour les réseaux de 200 nœuds avec densité variable

4.5.2 Résultats partiels des grands réseaux

Tel que nous l'avons observé dans la sous-section 4.5.1, la méthode *Clique* s'est révélée très coûteuse en terme de temps. En effet, le délai d'exécution est très élevé comparé à celui des deux autres méthodes. De ce fait, nous avons décidé de ne pas la considérer pour les grands réseaux. La comparaison se fait alors entre les méthodes *ProxiRank* et *Lien-clé* seulement. Les tableaux 4.18 et 4.19 présentent les résultats de déstabilisation effectuée sur cinq réseaux *aléatoires* et cinq réseaux *petit-monde* comportant mille nœuds et une densité de 20%.

Réseaux petit-monde à 1000 nœuds avec une densité de 20%					
Méthode		**Nb. retirés**	**ΔE finale**	**% retirés**	**Temps (sec)**
ProxiRank	1	462	0.7008	46	23.07
	2	461	0.7002	46	10.77
	3	461	0.701	46	12.08
	4	461	0.7003	46	10.31
	5	462	0.7007	46	9.72
Moyenne		461	0.70	46	13.19
Lien-clé	1	965	0.9583	70	0.06
	2	967	0.9655	70	0
	3	956	0.9602	70	0.07
	4	959	0.9545	70	0.02
	5	961	0.9636	70	0.06
Moyenne		962	0.96	70	0.042

TABLE 4.18 – Statistiques des réseaux petit-monde à 1000 nœuds

Regardons d'abord le tableau 4.18. Celui-ci donne les résultats obtenus pour les réseaux *petit-monde*. Nous constatons que ces résultats sont similaires à ceux obtenus pour les petits réseaux de 200 nœuds présentés dans la sous-section précédente. Nous pouvons effectivement confirmer que le temps d'exécution est plus élevé pour la méthode *ProxiRank* que pour *Lien-clé*, soit en moyenne 13.19 secondes par rapport à 0.042 seconde. En revanche, il est très évident que *ProxiRank* est plus précise car elle a supprimé en moyenne 46% des nœuds alors *Lien-clé* doit en supprimer 70% pour

atteindre le même niveau de déstabilisation.

Pour ce qui est des réseaux *aléatoires*, les résultats sont présentés dans le tableau 4.19. Les deux méthodes suppriment en moyenne le même nombre de nœuds, soit 70%. Toutefois, la méthode *Lien-clé* est nettement plus rapide avec un temps moyen d'exécution de 0.108 seconde alors que *ProxiRank* affiche un délai de 3.134 secondes, un écart très significatif. Mais ce ne sont que des résultats partiels pour le moment. Dans la prochaine section, nous présentons des résultats généralisés pour toutes les séries de réseaux testés dans le cadre de cette recherche.

Réseaux aléatoires à 1000 nœuds avec une densité de 20%					
Méthode		**Nb. retirés**	**ΔE finale**	**% retirés**	**Temps (sec)**
ProxiRank	1	698	0.7004	70	5.05
	2	698	0.7005	70	2.76
	3	699	0.7015	70	2.87
	4	702	0.7044	70	2.38
	5	700	0.7016	70	2.61
Moyenne		699	0.70	70	3.134
Lien-clé	1	853	0.8544	70	0.09
	2	854	0.8554	70	0.17
	3	1000	1	70	0.1
	4	843	0.8444	70	0.07
	5	1000	1	70	0.11
Moyenne		910	0.91	70	0.108

TABLE 4.19 – Statistiques des réseaux aléatoires à 1000 nœuds

4.5.3 Résultats partiels des réseaux réels

L'extraction de composants provenant des réseaux *hep-th* et *netscience* de Newman nous a permis de construire deux réseaux réels servant de données d'essai. Le tableau 4.20 montre les statistiques de déstabilisation de ces réseaux.

Composants des réseaux de Newman					
Méthode	**Réseau**	**Nb. retirés**	**ΔE finale**	**% retirés**	**Temps (sec)**
ProxiRank	hep-th	6	0.7504	0.10	3.3
	netscience	1	0.8755	0.21	0
	Moyenne	4	0.8130	0.15	1.65
Lien-clé	hep-th	4	0.8335	0.06	1.21
	netscience	4	0.7536	0.98	0
	Moyenne	4	0.7936	0.52	0.605

TABLE 4.20 – Statistiques des réseaux réels

Ce tableau confirme qu'en moyenne, la méthode *ProxiRank* supprime moins de nœuds que *Lien-clé*. Elle est de ce fait plus précise. Une fois de plus, nous constatons que la méthode *Lien-clé* est beaucoup plus rapide.

4.6 Résultats généralisés

Nous avons précédemment exposé les résultats partiels obtenus spécifiquement pour les réseaux de petite taille comportant deux cents nœuds et pour les grands réseaux de mille nœuds. Nous présentons sous peu les résultats généralisés incluant tous les réseaux dont la taille et la densité sont décrites dans les tableaux 4.12 et 4.13. Rappelons qu'au total, quatre-vingt réseaux ont été expérimentés pour les deux modèles de réseaux. Dans les tableaux 4.21 et 4.24, nous présentons le pourcentage moyen des nœuds supprimés. Tout d'abord, le pourcentage moyen est calculé pour les séries comprenant chacune cinq réseaux. Par la suite, la moyenne globale des nœuds supprimés est calculée au niveau des méthodes afin de comparer leur performance générale.

Moyenne globale pour les réseaux petit-monde				
Méthode	Nb. nœuds	Densité	% nœuds retirés	Temps (sec)
ProxiRank	100	0.05	50	0.000
	200	0.02, 0.04, 0.05, 0.06, 0.08	45	0.008
	1000	0.05, 0.1, 0.2, 0.3	47	12.404
	2000	0.05	47	27.408
	3000	0.05	47	10.262
		Moyenne	47	
Lien-clé	100	0.05	49	0.002
	200	0.02, 0.04, 0.05, 0.06, 0.08	68	0.000
	1000	0.05, 0.1, 0.2, 0.3	71	0.047
	2000	0.05	69	0.104
	3000	0.05	70	1.104
		Moyenne	66	
Clique	100	0.05	63	0.008
	200	0.02, 0.04, 0.05, 0.06, 0.08	84	3.910
	1000	0.05, 0.1, 0.2, 0.3	non testé	non testé
	2000	0.05	non testé	non testé
	3000	0.05	non testé	non testé
		Moyenne	73	

TABLE 4.21 – Statistiques globales pour les réseaux petit-monde

En premier lieu, observons le tableau 4.21 qui donne le résultat moyen calculé pour chaque série de réseaux *petit-monde*. Par exemple, cinq réseaux de deux mille nœuds ont été testés. Pour ces derniers, le pourcentage moyen des nœuds supprimés est de 47% et un temps d'exécution moyen de 27.40 secondes. Par la suite, les valeurs moyennes des séries sont utilisées pour calculer la moyenne globale des nœuds supprimés au niveau des méthodes de déstabilisation.

Mentionnons que le temps moyen d'exécution a été calculé uniquement au niveau des séries et non au niveau des méthodes car le délai d'exécution varie en fonction de la taille du réseau, de la densité et du rang centile choisi pour le regroupement des nœuds cibles. Le délai augmente s'il y a plus de nœuds ou si le réseau est plus dense. Ce fait peut être observé dans le tableau 4.21 pour les réseaux de mille et de deux

mille nœuds pour la méthode *ProxiRank*. Toutefois, si l'on choisit de réduire le rang centile pour grouper les cibles à éliminer, le temps d'exécution sera également réduit car il y aura plus de nœuds supprimés dans chaque itération. Par exemple, le réseau de deux mille nœuds prend 27.40 secondes alors que le réseau de trois mille nœuds prend seulement 10.26 secondes. Ceci est dû au fait que le 99.9^e est utilisé lors de la désintégration des réseaux de deux mille nœuds et le 96^e centile est utilisé dans la déstabilisation des réseaux de trois mille nœuds.

Bien que le temps moyen d'exécution n'est pas calculé au niveau des méthodes, nous pouvons tout de même observer qu'en général, la méthode *Lien-clé* est plus rapide que les deux autres méthodes. Par exemple, il a fallu en moyenne 10.26 secondes pour désintégrer le réseau de trois mille nœuds par la méthode *ProxiRank* alors qu'il n'en faut que 1.10 secondes pour la méthode *Lien-clé*. L'un des facteurs contribuant à la lenteur de *ProxiRank* est certainement le calcul de la centralité des nœuds. Il faut se rappeler que cette méthode est fondée sur la moyenne de deux centralités que sont celles de *PageRank* et de proximité. Ce qui signifie qu'à chaque itération, cette méthode doit recalculer deux centralités alors que la méthode *Lien-clé* recalcule seulement une centralité. Donc, plus le réseau est grand, plus l'écart de temps d'exécution est grand entre ces deux méthodes.

Un autre facteur expliquant l'écart de temps entre *ProxiRank* et *Lien-clé* est le fait que les liens ont des valeurs centralité d'intermédiarité très rapprochées. Afin de nous en convaincre, regardons les tableaux 4.22 et 4.23. Ces deux tableaux présentent les statistiques de déstabilisation d'un même réseau aléatoire par les deux méthodes. Le tableau 4.22 montre que la méthode *ProxiRank* a déstabilisé le réseau en 308 itérations. Pendant ce temps, le tableau 4.23 montre que la méthode *Lien-clé* a déstabilisé le même réseau en seulement trois itérations. Sachant que le 99.9^e centile est utilisé pour regrouper les nœuds cibles, nous pouvons présumer que beaucoup de liens ont la même valeur de centralité ou des valeurs quasi égales. D'ailleurs, nous remarquons que cette méthode a supprimé 932 nœuds seulement après la première itération. Ceci explique donc son temps d'exécution très avantageux par rapport à la méthode *ProxiRank*.

It.	**Nœuds retirés**	**Densité**	**Degré moyen**	**Nombre de composants**	**Nb. retirés**	**Efficacité (TC)**	ΔE
0		0.05	0.1	1	0	1	0
1	1142, ...	0.05	0.0999	1	5	0.9973	0.0027
2	429, ...	0.0499	0.0999	1	9	0.9952	0.0048
3	74, 1133, 1558	0.0499	0.0998	1	12	0.9937	0.0063
4	914, 1820, 1843	0.0499	0.0998	1	15	0.9921	0.0079
...	...	...	...	...	...	...	...
306	887, 203	0.041	0.0821	1	1394	0.3008	0.6992
307	1357	0.041	0.0821	1	1395	0.3003	0.6997
308	989, 1478, 1871	0.041	0.082	1	1398	0.2989	0.7011

TABLE 4.22 – Statistiques de déstabilisation par la méthode ProxiRank

It.	**Nœuds retirés**	**Densité**	**Degré moyen**	**Nombre de composants**	**Nb. retirés**	**Efficacité (TC)**	ΔE
0		0.05	0.1	1	0	1	0
1	1210, 226, ...	0.0495	0.0991	1	932	0.5295	0.4705
2	476, 384,...	0.0487	0.0974	1	1337	0.3275	0.6725
3	1568,351, ...	0.0485	0.0971	1	1444	0.2744	0.7256

TABLE 4.23 – Statistiques de déstabilisation par la méthode Lien-clé

Pour ce qui est de la précision, nous avons maintenant suffisamment de données pour affirmer que la méthode *ProxiRank* est la meilleure pour les réseaux *petit-monde*. Cette dernière a supprimé en moyenne 47% des nœuds des réseaux testés, soit environ 20% moins de nœuds que les deux autres méthodes.

Regardons à présent les réseaux *aléatoires* dont les résultats sont donnés dans le tableau 4.24. Pour ce type de réseau, les méthodes *ProxiRank* et *Lien-clé* éliminent approximativement le même nombre de nœuds c'est-à-dire 67% et 64% respectivement. Cependant, nous remarquons une fois de plus que la méthode *Lien-clé* est de loin la plus rapide.

Moyenne globale pour les réseaux aléatoires				
Méthode	Nb. nœuds	Densité	% nœuds retirés	Temps (sec)
ProxiRank	100	0.05	53	0.000
	200	0.02, 0.04, 0.05, 0.06, 0.08	70	0.014
	1000	0.05, 0.1, 0.2, 0.3	70	3.243
	2000	0.05	70	7.378
	3000	0.05	70	6.708
		Moyenne	67	
Lien-clé	100	0.05	45	0.000
	200	0.02, 0.04, 0.05, 0.06, 0.08	67	0.000
	1000	0.05, 0.1, 0.2, 0.3	69	0.063
	2000	0.05	70	0.090
	3000	0.05	70	0.534
		Moyenne	64	
Clique	100	0.05	51	0.008
	200	0.02, 0.04, 0.06	70	0.168
	1000	0.05, 0.1, 0.2, 0.3	non testé	non testé
	2000	0.05	non testé	non testé
	3000	0.05	non testé	non testé
		Moyenne	61	

TABLE 4.24 – Statistiques globales pour les réseaux aléatoires

Chapitre 5

Conclusions et travaux futurs

5.1 Conclusion

En se basant sur les résultats observés et discutés dans la section précédente, nous nous permettons de tirer trois conclusions.

La première conclusion concerne la méthode *Clique*. Que ce soit pour les réseaux de type *petit-monde* ou *aléatoire*, cette méthode est moins optimale et pourtant plus coûteuse que les deux autres méthodes. En effet, celle-ci doit supprimer un plus grand nombre de nœuds pour atteindre l'objectif de déstabilisation. De plus, les tableaux 4.16 et 4.17 confirment qu'elle exige plus de temps de traitement. Un autre inconvénient de cette méthode est qu'elle est inapplicable pour des réseaux ne comportant aucune clique.

Pour ce qui est de la méthode *ProxiRank*, elle pourrait être avantageuse pour les réseaux petit-monde qui sont 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, il vaut mieux utiliser la méthode *Lien-clé*. Mais dans les faits, le temps de déstabilisation d'un réseau social réel est plutôt indépendant du temps de traitement lors de la simulation. Dans la réalité, la déstabilisation d'un réseau social n'est pas nécessairement exécutée par un programme informatique. En général, cela exige une intervention humaine de la part des personnes concernées. Le

programme servirait alors uniquement à identifier les nœuds cibles à retirer du réseau. Ce qui fait qu'en dépit de son temps de calcul désavantageux, la méthode *ProxiRank* est certainement très utile pour aider à la déstabilisation des réseaux petit-monde de grande taille puisqu'elle donne une meilleure précision que les autres méthodes.

Finalement, nous constatons que la méthode *Lien-clé* est remarquablement avantageuse pour les réseaux de type aléatoire. En effet, elle supprime en moyenne le même nombre de nœuds que la méthode *ProxiRank* mais dans un délai nettement plus court. Une autre observation concernant cette méthode est que les liens ont des valeurs de centralité très rapprochées, surtout dans le cas des réseaux *aléatoires*. En effet, nous avons observé qu'avec un seuil de groupement situant au 96^e centile, la désintégration est accomplie en une seule itération pour plusieurs cas. Pour cette raison, lors de l'expérimentation des réseaux de trois mille nœuds, nous avons choisi le 100^e centile pour regrouper les liens et le 96^e centile pour regrouper les nœuds cibles lors de la déstabilisation. Le résultat a démontré que le temps d'exécution pour la méthode *ProxiRank* s'est amélioré. Toutefois, comme il y a beaucoup de liens qui ont la même valeur de centralité d'intermédiarité, la méthode *Lien-clé* déstabilise le réseau en très peu d'itérations.

En ce qui concerne les mesures globales du réseau, nous avons constaté que le degré moyen, la densité et le nombre de composants ne permettent pas de déterminer le degré de déstabilisation d'un réseau. En effet, les expérimentations ont démontré que le degré moyen et la densité du réseau fluctuent indépendamment du nombre de nœuds retirés. Pour ce qui du nombre de composants, il ne peut à lui seul mesurer le niveau de désintégration car le réseau peut être déstabilisé au seuil désiré tout en restant connecté. Pour cette raison, cette mesure a été intégrée dans la nouvelle formule pour calculer l'efficacité du réseau.

5.2 Travaux futurs

Dans ce mémoire, nous avons proposé une nouvelle formule de calcul d'efficacité d'un réseau social. À l'aide cette formule, nous avons étudié et comparé trois méthodes pour déstabiliser un réseau social. Ces méthodes focalisent sur la centralité des nœuds et des liens ainsi que l'appartenance des nœuds aux communautés cliques.

Cependant, l'analyse de réseaux sociaux offre une variété de possibilités d'analyse qui méritent que l'on y porte un intérêt particulier.

Dans un projet futur, il serait sans doute intéressant d'expérimenter l'équivalence structurelle ou régulière des nœuds. Ainsi, au lieu de s'intéresser uniquement à la centralité, il conviendrait de rechercher les nœuds importants et ceux qui leur sont équivalents. Puisque les nœuds équivalents sont totalement substituables [7], on peut anticiper que la déstabilisation aurait un impact minimal sur le réseau si les nœuds retirés sont aussitôt remplacés. Cette approche nécessitera l'identification des nœuds clés et les nœuds qui leur sont équivalents structurellement ou régulièrement. Une des options serait de supprimer les nœuds clés puis leurs équivalents. On pourrait également envisager de les supprimer tous en même temps. La recherche permettra d'étudier l'impact de la prise en compte des nœuds équivalents sur le processus de déstabilisation.

À part la centralité des nœuds, nous nous sommes intéressés aux communautés cliques en laissant de côté les factions. Nous avons observé que la méthode *Clique* est significativement moins performante en terme de précision et en terme de temps par rapport aux méthodes *ProxiRank* et *Lien-clé*. Dans une future recherche, la comparaison entre les deux dernières méthodes et une méthode utilisant les factions permettra alors de confirmer s'il y a ou non intérêt à considérer les communautés dans le processus de déstabilisation.

Dans la présente recherche, nous avons utilisé une condition d'arrêt basée sur un seuil préétabli. Dans un travail futur, il serait certainement utile de développer une formule qui permet de calculer automatiquement ce seuil.

Bibliographie

[1] Allsup, R., Thomas, E., Monk, B., Frank, R., and Bouchard, M. Networking in child exploitation : Assessing disruption strategies using registrant information. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015* (2015), ACM, pp. 400–407.

[2] Arquilla, J., and Ronfeldt, D. *Networks and netwars : The future of terror, crime, and militancy*. Rand Corporation, 2001.

[3] Barabási, A.-L. Scale-free networks : a decade and beyond. *science 325*, 5939 (2009), 412–413.

[4] Berzinji, A., Kaati, L., and Rezine, A. Detecting key players in terrorist networks. In *Intelligence and Security Informatics Conference (EISIC), 2012 European* (2012), IEEE, pp. 297–302.

[5] Bonacich, P. Power and centrality : A family of measures. *American journal of sociology* (1987), 1170–1182.

[6] Borgatti, S. P., Everett, M. G., and Freeman, L. C. Ucinet for windows : Software for social network analysis.

[7] Borgatti, S. P., Everett, M. G., and Johnson, J. C. *Analyzing social networks*. SAGE Publications Limited, 2013.

[8] Carley, K. M., Lee, J.-S., and Krackhardt, D. Destabilizing networks. *Connections 24*, 3 (2002), 79–92.

[9] Chatterjee, J., and Gendarmerie royale du Canada. Direction des services de police communautaires, c. e. a. S.-d. d. l. r. e. d. l. *La transformation de la structure des groupes du crime organisé*. Sous-direction de la recherche et de l'évaluation, Direction des services de police communautaires, contractuels et autochtones, Gendarmerie royale du Canada, 2005.

[10] Chaurasia, N., and Tiwari, A. Efficient algorithm for destabilization of terrorist networks. *International Journal of Information Technology and Computer Science (IJITCS) 5*, 12 (2013), 21–30.

[11] Duan, B., Liu, J., Zhou, M., and Ma, L. A comparative analysis of network robustness against different link attacks. *Physica A : Statistical Mechanics and its Applications 448* (2016), 144–153.

[12] Durland, M. M., and Fredericks, K. A. An introduction to social network analysis. *New Directions for Evaluation 2005*, 107 (2005), 5–13.

[13] Erdős, P., and Rényi, A. On random graphs. *Publicationes Mathematicae Debrecen 6* (1959), 290–297.

[14] Franc, L. H. Social network analysis, rapport technique, laboratoire larim, uqo.

[15] Freeman, L. C. Centrality in social networks conceptual clarification. *Social networks 1*, 3 (1978), 215–239.

[16] Freeman, L. C., and White, D. R. Using galois lattices to represent network data. *Sociological methodology 23* (1993).

[17] Hanneman, R. A., and Riddle, M. Introduction to social network methods, 2005.

[18] Hopkins, A. Graph theory, social networks and counter terrorism. *Social Network and Counterterrorism Analysis* (2010).

[19] Katz, L. A new status index derived from sociometric analysis. *Psychometrika 18*, 1 (1953), 39–43.

[20] Knoke, D., and Yang, S. *Social network analysis, Second Edition*, vol. 154. Sage, 2008.

[21] Lagarrigue Adrien, Sarr Idrissa, M. R. Déstabilisation parcimonieuse d'un réseau social par effet cascade, manuscrit.

[22] Lindelauf, R., Borm, P., and Hamers, H. The influence of secrecy on the communication structure of covert networks. *Social Networks 31*, 2 (2009), 126–137.

[23] Marquetoux, N., Stevenson, M. A., Wilson, P., Ridler, A., and Heuer, C. Using social network analysis to inform disease control interventions. *Preventive Veterinary Medicine 126* (2016), 94 – 104.

[24] Memon, N., and Larsen, H. L. Practical algorithms for destabilizing terrorist networks. In *Intelligence and Security Informatics, IEEE International Conference on Intelligence and Security Informatics, ISI 2006, San Diego, CA, USA, May 23-24, 2006, Proceedings* (2006), pp. 389–400.

[25] Newman, M. E. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical review E 64*, 1 (2001), 016132.

[26] Newman, M. E. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences 98*, 2 (2001), 404–409.

[27] Newman, M. E. J. *Networks : an introduction*. Oxford university press, 2010.

[28] R Development Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[29] Ressler, S. Social network analysis as an approach to combat terrorism : Past, present, and future research. *Homeland Security Affairs 2*, 2 (2006).

[30] sans crainte, I. Rapport 2013 france digital. `http://www.internetsanscrainte.fr/s-informer/qu-est-ce-qu-reseau-social`, octobre 2015.

[31] Scott, J., and Carrington, P. J. *The SAGE handbook of social network analysis*. SAGE publications, 2011.

[32] Sun, S.-w., Ma, Y.-l., Li, R.-q., Wang, L., and Xia, C.-y. Tabu search enhances network robustness under targeted attacks. *Physica A : Statistical Mechanics and its Applications 446* (2016), 82–91.

[33] Watts, D. J., and Strogatz, S. H. Collective dynamics of 'small-world' networks. *nature 393*, 6684 (1998), 440–442.

[34] Wiil, U. K., Gniadek, J., and Memon, N. Measuring link importance in terrorist networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on* (2010), IEEE, pp. 225–232.

[35] Wikipedia. Pagerank. `https://en.wikipedia.org/wiki/PageRank`, juillet 2016.