

N° d'ordre 3330

THESE

En vue de l'obtention du : **DOCTORAT**

Centre de Recherche : Rabat IT center

Structure de Recherche : Laboratoire de Recherche en Informatique et
Télécommunications

Discipline : Sciences de l'ingénieur

Spécialité : Informatique et télécommunications

Présentée et soutenue le 12/09/2020 par :

Zakariya GHALMANE

**Les stratégies d'immunisation dans les réseaux complexes avec structure
communautaire**

JURY

Salma MOULINE	PES, Faculté des Sciences de Rabat, Université Mohammed V, Rabat	Présidente
Mohammed EL HASSOUNI	PES, Faculté des Lettres et des Sciences Humaines de Rabat, Université Mohammed V, Rabat	Directeur de thèse
Mohamed BOUHDADI	PES, Faculté des Sciences de Rabat, Université Mohammed V, Rabat	Rapporteur/Examineur
Dounia LOTFI	PH, Faculté des Sciences de Rabat, Université Mohammed V, Rabat	Rapporteur/Examineur
Mohamed NEMICHE	PES, Faculté des Sciences d'Agadir, Université Ibn Zohr, Agadir	Rapporteur/Examineur
Abdelkrim HAQIQ	PES, Faculté des Sciences et Techniques de Settat, Université Hassan 1, Settat	Examineur
Hocine CHERIFI	PES, IUT de Dijon, Université de Bourgogne, Dijon, France	Examineur
Chantal CHERIFI	PA, IUT Lumière, Université de Lyon 2, France	Examineur

Année Universitaire : 2019-2020

*À mes parents
À mes frères
À mes amis*

Twenty years from now you will be more disappointed by the things that you didn't do than by the ones you did do. So throw off the bowlines. Sail away from the safe harbor. Catch the trade winds in your sails. Explore. Dream. Discover.
Mark Twain



REMERCIEMENTS

Cette thèse de doctorat a été menée pour l'obtention du grade de Docteur de l'Université Mohammed V. Les travaux présentés dans ce mémoire ont été effectués au Laboratoire de Recherche en Informatique et Télécommunications (LRIT) de la Faculté des Sciences de Rabat (FSR) au Maroc sous la direction du Professeur **Mohammed EL HASSOUNI** et en collaboration avec le Laboratoire d'Informatique de Bourgogne (LIB) de l'Université de Bourgogne sous le co-encadrement du Professeur **Hocine CHERIFI**.

En premier lieu, je tiens à remercier mon directeur de thèse le Professeur **Mohammed EL HASSOUNI**, Professeur d'Enseignement Supérieur à la faculté des lettres et sciences humaines de Rabat, ses conseils précieux ont bien mener à bien cette thèse. Au-delà de ses qualités scientifiques, il est avant tout un homme généreux, avec qui il est agréable de discuter, d'échanger des idées, et de plaisanter. J'espère que ces immenses qualités continueront à m'influencer pour longtemps.

Je suis également reconnaissant au Professeur **Hocine CHERIFI**, Professeur à l'Université de Bourgogne de Dijon, France, pour l'encadrement de mes travaux de recherche, ses conseils, sa disponibilité. Les nombreuses discussions et échanges ont grandement contribué à la réussite de cette thèse. Je le remercie aussi d'avoir accepté de participer en tant qu'examineur dans le jury final.

Je suis particulièrement redevable au Professeur **Chantal CHERIFI**, Professeur Associé à l'université de Lyon 2, pour son engagement dans le co-encadrement de ma thèse. Cette thèse doit énormément à sa grande disponibilité, son dynamisme, ses encouragements, son ouverture d'esprit et bien sûr ses qualités scientifiques exceptionnelles. Je tiens à la remercier également d'avoir accepté de participer en tant qu'examineur dans le jury final.

Je tiens à remercier le Professeur **Salma MOULINE**, Professeur d'Enseignement Supérieur à la faculté des sciences de Rabat, d'avoir accepté de présider le jury de ma thèse.

Mes remerciements vont également au Professeur **Mohamed BOUHDADI**, Professeur d'Enseignement Supérieur à la faculté des sciences de Rabat, d'avoir accepté de rapporter et examiner ce mémoire de thèse.

Je voudrais également remercier le Professeur **Dounia LOTFI**, Professeur Habilité à la faculté des sciences de Rabat, qui m'a fait l'honneur de rapporter et examiner cette thèse.

Je tiens aussi à remercier le Professeur **Mohamed NEMICHE**, Professeur d'Enseignement Supérieur à la faculté des sciences d'Agadir, d'avoir accepté de juger la qualité de mon travail en tant que rapporteur et examinateur.

Mes remerciements vont aussi au Professeur **Abdelkrim HAQIQ**, Professeur d'Enseignement Supérieur à la faculté des sciences techniques de Settat, d'avoir accepté d'examiner ce mémoire de thèse.

Je remercie mes amis qui m'ont apporté leur soutien et leurs encouragements tout au long de cette aventure. Dans le désordre, Abdellah, Ayoub, Youssef, Yasser, Omar, Mouad, Belkacem, Anas, Saad.

Je garde le meilleur pour la fin, ma famille qui a supporté toutes les difficultés morales et matérielles pour me soutenir au terme de mes études. J'adresse ma profonde gratitude et mon immense reconnaissance à mes raisons d'être, mes parents, qui m'ont éduqué et orienté. Merci de m'avoir encouragé et soutenu dans mes choix. Nul mot et nulles expressions refléteront le grand amour et la profonde gratitude que je porte pour vous. Mes plus sincères remerciements vont aussi à mes frères. Aucun mot ne saurait retranscrire ici sans l'affaiblir le bonheur qu'ils m'ont toujours apporté, ni l'ampleur de ce que je leur dois.



RÉSUMÉ

L'immunisation de réseaux complexes vise à les fragmenter en petites parties avec un petit nombre de noeuds supprimés (immunisés). Ses applications incluent la prévention de la propagation des épidémies, la protection contre les attaques intentionnelles sur les réseaux, etc. Elle utilise les propriétés dynamiques de l'infection ainsi que les propriétés topologiques de la structure du réseau à travers les mesures de centralité. Ceci afin d'immuniser les noeuds les plus influents. Cependant, bien que la plupart des réseaux du monde réel aient une structure communautaire, peu de travaux ont été consacrés à l'étude de l'impact de cette propriété sur la propagation de l'épidémie. Dans cette thèse, nous nous concentrons sur le développement de nouvelles stratégies d'immunisation en utilisant plus de connaissances et de propriétés sur la structure de la communauté afin de mieux identifier et immuniser les noeuds influents. Dans un premier temps, nous visons à comprendre comment la structure communautaire affecte la dynamique épidémique. Ensuite, trois stratégies d'immunisation sont proposées en utilisant différents niveaux d'informations sur la structure de la communauté. Elles sont conçues pour des réseaux avec différentes forces de structure communautaire. Dans un deuxième temps, nous avons introduit un framework pour adapter toutes les mesures de centralité classiques proposées pour les réseaux sans structure communautaire aux réseaux modulaires. Dans ces types de réseaux, un noeud a deux types d'influences : une influence locale sur les noeuds appartenant à sa propre communauté à travers les liens intra-communautaires, et une influence globale sur les noeuds des autres communautés à travers les liens inter-communautaires. Ainsi, au lieu de représenter la centralité par une simple valeur scalaire, nous avons proposé de la représenter par un vecteur bidimensionnel, où chaque dimension représente un type d'influence différent que les noeuds peuvent exercer dans le réseau. Sa première composante est mesurée en calculant la centralité classique sur le réseau local, formé uniquement de liaisons intra-communautaires du réseau d'origine. Sa deuxième composante est quantifiée en calculant la centralité classique sur le réseau global, formé uniquement des liens inter-communautaires du réseau d'origine. Selon la force de la structure communautaire, ces deux composantes sont plus ou moins influentes. Dans notre troisième contribution, nous avons étendu ce framework aux réseaux avec des modules qui se chevauchent. Nous prenons en compte l'influence locale et globale des noeuds qui se chevauchent. Les performances des méthodes proposées ont été évaluées à l'aide de réseaux synthétiques et empiriques en utilisant le modèle épidémiologique Susceptible-Infecté-Rétabli.

Mots clés : Réseaux complexes, stratégie d'immunisation, mesure de centralité, structure communautaire, modules qui se chevauchent, modèle Susceptible-Infecté-Rétabli.



ABSTRACT

The immunization of complex networks aims at fragmenting them into small parts with a small number of removed (immunized) nodes. Its applications include prevention of epidemic spreading, protection against intentional attacks on networks, etc. It uses the dynamic properties of the infection as well as the topological properties of the network structure through the centrality measures in order to immunize the most influential nodes. However, although most real-world networks have a community structure, little work has been devoted to studying the impact of this property on the epidemic spread. In this thesis, we focus on developing new immunization strategies using more knowledge and features about the community structure in order to better identify and immunize the influential nodes. Initially, we aim to understand how community structure affects epidemic dynamics. Then, three immunization strategies are proposed using different levels of information about the community structure. They are designed for networks with different community structure strengths. In a second step, we have introduced a framework to adapt all the classic centrality measures proposed for networks without a community structure to modular networks. In these type of networks, a node has two types of influences : A local influence on the nodes belonging to its own community through the intra-community links, and a global influence on the nodes of the other communities through the inter-community links. Thus, instead of representing the centrality by a simple scalar value, we have proposed to represent it by a two-dimensional vector, where each dimension stands for a different type of influence that the nodes can exert in the network. Its first component is measured by calculating classical centrality on the local network. This network is formed only from intra-community links of the original network. In addition, its second component is quantified by calculating classical centrality on the global network. This network is formed solely from the inter-community links of the original network. Depending on the strength of the community structure, these two components are more or less influential. In our third contribution, we extended this framework to networks with overlapping modules. We take into account the local and global influence of overlapping nodes. The performances of the proposed methods have been evaluated using synthetic and empirical networks using the Susceptible-Infected-Recovered epidemiological model.

Keywords : Complex networks, immunization strategy, centrality measure, community structure, overlapping modules, Susceptible-Infected-Recovered epidemiological model.



TABLE DES MATIÈRES

Résumé	vii
Abstract	ix
Liste des figures	xxi
Liste des tableaux	xxiii
Liste des abréviations	xxiii
Chapitre 1 : Introduction générale	1
1.1 Contexte et problématique	1
1.2 Contributions de la thèse	4
1.3 Publications dans le cadre de la thèse	5
1.4 Organisation du manuscrit	6
Chapitre 2 : État de l'art sur les stratégies d'immunisation	9
2.1 Introduction	9
2.2 Stratégies d'immunisation	9
2.3 Modèle épidémique	10
2.4 L'immunisation des réseaux complexes non modulaires	11
2.4.1 Stratégies stochastiques	11
2.4.2 Stratégies déterministes	12
2.4.2.1 Stratégies d'immunisation basées sur le voisinage	13
2.4.2.2 Stratégies d'immunisation basées sur les voies	15
2.4.2.3 Stratégies d'immunisation basées sur le raffinement itératif	19
2.5 L'immunisation des réseaux complexes modulaire	23
2.5.1 Structure communautaire	23
2.5.1.1 Détection de communautés sans chevauchement	24
2.5.1.2 Détection de communautés avec chevauchement	26
2.5.1.3 Mesures d'évaluation	27
2.5.2 Stratégies stochastiques	29

2.5.3	Stratégies déterministes	33
2.5.3.1	Structure communautaire sans chevauchement	33
2.5.3.2	Structure communautaire avec chevauchement	43
2.6	Conclusion	45

Chapitre 3 : Stratégies d’immunisation pour les réseaux avec structure communautaire 47

3.1	Introduction	47
3.2	Stratégies proposées	49
3.2.1	Nombre de Communautés Voisines	50
3.2.2	Community Hub-Bridge	51
3.2.3	Weighted Community Hub-Bridge	52
3.2.4	Exemple	53
3.3	Cadre expérimental	56
3.3.1	Données utilisées	56
3.3.1.1	Réseaux synthétiques	56
3.3.1.2	Réseaux réels	56
3.3.2	Paramètres du modèle SIR	58
3.3.3	Schéma d’immunisation	58
3.3.4	Critère d’évaluation	59
3.3.4.1	Différence relative de la taille de l’épidémie	59
3.3.4.2	Le plus grand composant connecté	60
3.4	Résultats et discussion	60
3.4.1	Réseaux synthétiques	62
3.4.1.1	Influence de la force de la structure communautaire	62
3.4.1.2	Effets de la taille des communautés	64
3.4.1.3	Comparaison avec les méthodes alternatives	66
3.4.2	Réseaux réels	68
3.4.2.1	Efficacité de diffusion des méthodes proposées	69
3.4.2.2	Comparaison avec les méthodes alternatives	73
3.4.2.3	Influence des paramètres du processus de diffusion	75
3.4.3	Influence des algorithmes de détection de communautés	76
3.4.3.1	Comparaison des algorithmes de détection de communautés	77
3.4.3.2	Influence des algorithmes de détection de communautés sur les méthodes proposées	78
3.5	Conclusion	79

Chapitre 4 : Framework pour le calcul des centralités dans les réseaux avec structure communautaire 83

4.1	Introduction	83
4.2	Centralité modulaire	86
4.2.1	Définitions	87
4.2.1.1	Composante locale de la centralité modulaire	87

4.2.1.2	Composante globale de la centralité modulaire	87
4.2.1.3	Centralité modulaire	88
4.2.2	Algorithme	88
4.2.3	Extensions modulaires des mesures de centralité standard	90
4.2.3.1	Centralité d'intermédiarité modulaire	90
4.2.3.2	Centralité de proximité modulaire	90
4.2.3.3	Centralité des vecteurs propres modulaire	91
4.2.4	Exemple	92
4.2.5	Stratégies de classement basées sur la centralité modulaire	93
4.3	Cadre expérimental	95
4.3.1	Données utilisées	95
4.3.1.1	Réseaux synthétiques	95
4.3.1.2	Réseaux réels	96
4.3.2	Simulations SIR	98
4.3.3	Critère d'évaluation	98
4.4	Résultats des réseaux synthétiques	100
4.4.1	Évaluation de la composante locale et globale de la centralité modulaire	100
4.4.1.1	Structure communautaire bien définie	100
4.4.1.2	Structure communautaire à cohésion moyenne	102
4.4.1.3	Structure communautaire lâche	102
4.4.2	Évaluation des méthodes de classement basées sur la centralité modulaire	103
4.5	Résultats des réseaux du monde réel	104
4.5.1	Évaluation de la composante locale et globale de la centralité modulaire	108
4.5.1.1	Structure communautaire bien définie	108
4.5.1.2	Structure communautaire à cohésion moyenne	109
4.5.1.3	Structure communautaire lâche	110
4.5.2	Évaluation des méthodes de classement basées sur la centralité modulaire	110
4.5.3	Comparaisons avec les mesures alternatives	111
4.5.4	Influence des algorithmes de détection de communautés	113
4.6	Conclusion	115

Chapitre 5 : Le calcul des centralités dans les réseaux complexes avec une structure communautaire qui se chevauche 117

5.1	Introduction	117
5.2	Centralité modulaire chevauchante	118
5.2.1	Définitions	120
5.2.1.1	Composante locale de la centralité modulaire chevauchante	120
5.2.1.2	Composante globale de la centralité modulaire chevauchante	121
5.2.1.3	Centralité modulaire chevauchante	121

5.2.2	Algorithme	121
5.2.3	Schéma général	122
5.2.4	Combiner les composantes de la centralité modulaire chevauchante	123
5.3	Cadre expérimental	126
5.3.1	Réseaux synthétiques	126
5.3.2	Réseaux réels	127
5.4	Résultats des réseaux synthétiques	128
5.4.1	Influence de la force de la structure communautaire	129
5.4.1.1	Évaluation de la composante locale et globale de la centralité modulaire chevauchante	129
5.4.1.2	Évaluation des méthodes de classement basées sur la combinaison des composantes de la centralité modulaire chevauchante	132
5.4.2	Influence du nombre de noeuds qui se chevauchent	134
5.4.3	Influence du degré d'appartenance aux communautés qui se chevauchent	136
5.4.4	Synthèse	137
5.5	Résultats des réseaux du monde réel	138
5.5.1	Évaluation des composantes de la centralité modulaire chevauchante	138
5.5.2	Évaluation des méthodes de classement sur la base d'une combinaison des composantes de la centralité modulaire chevauchante	141
5.5.3	Comparaison avec les méthodes modulaires alternatives	142
5.5.4	Influence des algorithmes de détection de communautés	143
5.5.5	Synthèse	147
5.6	Conclusion	148
	Conclusion et perspectives	151
	Annexes	155
	Bibliographie	175



TABLE DES FIGURES

2.1	Le mécanisme d'infection du modèle classique (a) SI (b) SIR.	10
2.2	Diagramme schématique de la décomposition k-core.	15
2.3	Illustration d'un exemple de réseau, où le noeud le plus influent devrait être v_6 , mais selon l'indice d'excentricité, v_7 est le plus influent avec $ECC(7) = 2$	16
2.4	Méthodes d'immunisation stochastiques.	30
2.5	Méthodes d'immunisation déterministes.	34
3.1	Le rang des noeuds selon (a) la mesure du Nombre de Communautés Voisines β_{NVC} (b) la mesure du Community Hub-Bridge β_{HB} (c) la mesure du Weighted Community Hub-Bridge β_{WHB} (d) Mesure du degré (e) Mesure de l'intermédiarité. Les noeuds sont classés de plus influent (noeuds ayant la valeur de mesure la plus élevée) au noeud le moins influent (noeuds ayant la valeur de mesure la plus faible) du réseau.	55
3.2	Les principales étapes de schéma d'immunisation.	59
3.3	Influence de la force de la structure communautaire sur la taille épidémique des méthodes proposées. Chaque point représente la taille de l'épidémie par rapport à la fraction des noeuds immunisés. Les simulations sont effectuées sur des réseaux générés par LFR avec différentes valeurs de paramètres de mélange μ . Chaque valeur de taille épidémique est la moyenne de 600 simulations S.I.R.	61
3.4	Effet de la taille des communautés sur la propagation épidémique des méthodes proposées. Chaque point représente la taille de l'épidémie en fonction de la fraction des noeuds immunisés. Les simulations sont effectuées sur des réseaux LFR avec différentes structures communautaires. Chaque valeur est la moyenne de 600 simulations par réseau et par méthode d'immunisation.	64

-
- 3.5 Différence relative de la taille de l'épidémie $\Delta R/R_0$ en fonction de la fraction de noeuds immunisés. Les figures de gauche montrent la différence entre la méthode Community Hub-Bridge et les méthodes alternatives. Les figures du milieu montrent la différence entre la méthode du Nombre de communautés voisines et les méthodes alternatives, tandis que les figures de droite montrent la différence entre la méthode du Weighted Community Hub-Bridge et les méthodes alternatives. Nous notons qu'une valeur positive de $\Delta R/R_0$ signifie une performance plus élevée de la méthode proposée. Les simulations sont effectuées sur le réseau LFR avec différentes structures communautaires. Les valeurs finales sont obtenues en exécutant 600 simulations indépendantes par réseau, par couverture vaccinale et méthode d'immunisation. 67
- 3.6 La taille épidémique des méthodes d'immunisation effectuées sur six réseaux réels de types différents à savoir le réseau facebook de quatre universités (a) Caltech (b) Princeton (c) Georgetown (d) Oklahoma, et (e) Réseau de collaboration (f) Réseau du Power-grid. Les valeurs finales sont obtenues en exécutant 600 simulations indépendantes par réseau, couverture et méthode d'immunisation. 69
- 3.7 Différence relative de la taille de l'épidémie $\Delta R/R_0$ en fonction de la fraction de noeuds immunisés. Les figures de gauche illustrent la différence entre la méthode Community Hub-Bridge et les méthodes alternatives. Les figures du milieu et de droite montrent respectivement la différence entre la méthode du Nombre de communautés voisines, la Weighted Community Hub-Bridge et les méthodes alternatives. Nous notons qu'une valeur positive de $\Delta R/R_0$ signifie une performance plus élevée de la méthode proposée. Les simulations sont effectuées sur différents types de réseaux du monde réel. Les valeurs finales sont obtenues en exécutant 600 simulations indépendantes par réseau, couverture et méthode d'immunisation. 71
- 3.8 Effet de différentes méthodes d'immunisation sur la taille de l'épidémie lors des simulations SIR effectuées sur le réseau (a) Power-grid et (b) Georgetown, avec $\lambda = 0.9$ et $\gamma = 0.2$ 72
- 3.9 La taille du plus grand composant connecté (LCC) pour diverses stratégies d'immunisation effectuées sur le réseau (a) Power-grid et (b) Georgetown. Chaque point est le résultat de la taille *LCC* en fonction de la proportion des noeuds immunisés. 72
- 3.10 Effet de divers algorithmes de détection communautaire sur les performances des méthodes communautaires proposées. Chaque point représente la taille de l'épidémie par rapport à la proportion des noeuds immunisés. . 73

3.11	Différence relative de la taille de l'épidémie en fonction de la proportion des noeuds immunisés. Les figures de gauche montrent la différence entre la méthode Community Hub-Bridge et les méthodes alternatives. Les figures du milieu représentent la différence entre la méthode du Nombre de communautés voisines et les méthodes alternatives. Les figures de droite montrent la différence entre la méthode du Weighted Community Hub-Bridge et les méthodes alternatives. Les méthodes d'immunisation sont effectuées sur le réseau Power-grid en (a) et (b) et le réseau de Georgetown en (c) et (d) pour les algorithmes WalkTrap et Infomap.	81
4.1	Un exemple de réseau représentant le réseau local (G_l) et le réseau global (G_g) associé à un réseau modulaire (G) composé de quatre communautés qui ne se chevauchent pas.	92
4.2	La différence relative de la taille de l'épidémie Δr utilisant différentes mesures en fonction de la fraction des épandeurs initiaux f_0 est représentée pour les réseaux générés par l'algorithme LFR avec diverses valeurs μ . Pour chaque valeur μ , cinq exemples de réseaux sont générés. Une valeur de Δr plus élevé signifie une efficacité plus élevée des mesures locales, globales ou de combinaison par rapport à la définition de centralité standard. Les tailles épidémiques finales sont obtenues en calculant la moyenne de 200 simulations de modèle SIR par réseau, mesure de centralité et couverture initiale de propagation.	101
4.3	Différence relative de la taille de l'épidémie Δr de diverses mesures de centralité en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité sont calculées sur des réseaux du monde réel avec une forte structure communautaire (réseaux ego-Facebook, Power-grid et ca-GrQc avec des valeurs de paramètres de mélange estimées égales respectivement à 0.03, 0.034 et 0.095).	105
4.4	Différence relative de la taille de l'épidémie Δr de diverses mesures de centralité en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité sont calculées sur des réseaux du monde réel avec une structure communautaire moyenne (réseaux Princeton, Email-Eu-core et Caltech avec des valeurs de paramètres de mélange estimées respectivement égales à 0.354, 0.42 et 0.44).	106
4.5	Différence relative de la taille de l'épidémie Δr de diverses mesures de centralité en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité sont calculées sur un réseau avec une structure communautaire non cohésive (Georgetown) avec une valeur de paramètre de mélange estimée égale à 0.522.	107
5.1	Exemple de réseau illustrant comment former le réseau local pour les noeuds non chevauchants (G_l^{no}), le réseau local pour les noeuds chevauchants (G_l^o) et le Réseau global (G_g) à partir d'un réseau (G) avec une structure de communauté qui se chevauchent.	119

5.2	Les principales étapes des méthodes de classement basées sur la centralité modulaire chevauchante proposée.	122
5.3	La différence relative de la taille de l'épidémie Δr en fonction de la fraction des épandeurs initiaux f_0 , où $\Delta r = (R_c - R_s/R_s)$, R_c et R_s sont le nombre final de noeuds rétablis pour la mesure de centralité testée et la centralité standard, respectivement. Les mesures de centralité du degré (a), de l'intermédiarité (b), de la proximité (c) et des vecteurs propres (d) dérivées de la centralité modulaire chevauchante sont comparées à leur homologue standard conçue pour les réseaux sans structure communautaire. Des réseaux synthétiques générés par l'algorithme LFR avec différentes forces de structure de communauté sont utilisés (la valeur de leur coefficient de mélange est égale à 0.1, 0.4 et 0.6). Nous avons défini la proportion de noeuds qui se chevauchent on à 10% de la taille du réseau et le degré d'appartenance om à 10% du nombre total de communautés. Chaque valeur sur les courbes est obtenue en faisant la moyenne des résultats de 200 simulations SIR par méthode et fraction de noeuds initialement infectés. Une valeur positive de Δr signifie que la centralité testée est plus efficace que la centralité standard.	130
4	Influence du nombre de noeuds qui se chevauchent en terme de la taille épidémique. La différence relative de la taille de l'épidémie Δr entre la mesure standard et la mesure locale en fonction de la fraction des épandeurs initiaux f_0 . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés.	157
5	Influence du nombre de noeuds qui se chevauchent en terme de la taille épidémique. La différence relative de la taille de l'épidémie Δr entre la mesure standard et la mesure globale en fonction de la fraction des épandeurs initiaux f_0 . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés.	158
6	Influence du nombre de noeuds qui se chevauchent en terme de la taille épidémique. La différence relative de la taille de l'épidémie Δr entre la mesure standard et la mesure de module en fonction de la fraction des épandeurs initiaux f_0 . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés.	159
7	Influence du nombre de noeuds qui se chevauchent en terme de la taille épidémique. La différence relative de la taille de l'épidémie Δr entre la mesure standard et la mesure locale en fonction de la fraction des noeuds chevauchants on . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés.	160
8	Influence du nombre de noeuds qui se chevauchent en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre la mesure standard et la mesure globale en fonction de la fraction des noeuds qui se chevauchent on . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés.	161

-
- 9 Influence du **nombre de noeuds qui se chevauchent** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure de module** en fonction de la fraction des noeuds qui se chevauchent on . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés. 162
- 10 Influence du **degré d'appartenance aux communautés** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure locale** en fonction de la fraction des épandeurs initiaux f_0 . Nous définissons le nombre de noeuds qui se chevauchent on à 10% du nombre total de noeuds. 163
- 11 Influence du **degré d'appartenance aux communautés** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure globale** en fonction de la fraction des épandeurs initiaux f_0 . Nous définissons le nombre de noeuds qui se chevauchent on à 10% du nombre total de noeuds. 164
- 12 Influence du **degré d'appartenance aux communautés** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure de module** en fonction de la fraction des épandeurs initiaux f_0 . Nous définissons le nombre de noeuds qui se chevauchent on à 10% du nombre total de noeuds. 165
- 13 Influence du **degré d'appartenance aux communautés** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure locale** en fonction du degré d'appartenance om . Nous avons défini le nombre de noeuds qui se chevauchent on à 10% du nombre total de communautés. 166
- 14 Influence du **degré d'appartenance aux communautés** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure globale** en fonction du degré d'appartenance om . Nous avons défini le nombre de noeuds qui se chevauchent on à 10% du nombre total de communautés. 167
- 15 Influence du **degré d'appartenance aux communautés** en terme de la taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure de module** en fonction du degré d'appartenance om . Nous avons défini le nombre de noeuds qui se chevauchent on à 10% du nombre total de communautés. 168
- 16 La différence relative de la taille de l'épidémie Δr par rapport à la fraction des épandeurs initiaux f_0 . Les mesures de centralité du degré (a), d'intermédiarité (b), de la proximité (c) et du vecteur propre (d) sont comparées à leurs extensions du centralité modulaire chevauchante. Des réseaux du monde réel avec une structure communautaire bien définie (réseaux ego-Facebook, Netscience et ca-GrQc) sont utilisés. Les valeurs estimées de leur coefficient de mélange sont respectivement égales à 0.02, 0.03 et 0.11. 169

-
- 17 La différence relative de la taille de l'épidémie Δr en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité du degré (a), d'intermédiarité (b), de proximité (c) et du vecteur propre (d) dérivées de la centralité modulaire chevauchante sont comparées à leurs homologues standard conçu pour les réseaux sans structure communautaire. Des réseaux du monde réel avec une structure communautaire de force moyenne (réseaux de Princeton, Caltech et Georgetown) sont utilisés. Les valeurs estimées de leur coefficient de mélange sont respectivement égales à 0.22, 0.26 et 0.3. 170
- 18 La différence relative de la taille de l'épidémie Δr en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité du degré (a), d'intermédiarité (b), de proximité (c) et du vecteur propre (d) dérivées de la centralité modulaire chevauchante sont comparées à leurs homologues standard conçu pour les réseaux sans structure communautaire. Un réseau du monde réel (YeastS) avec une structure communautaire lâche est utilisé. La valeur estimée de son coefficient de mélange est égale à 0.47. 171
- 19 La taille de l'épidémie en fonction de la fraction d'épandeurs initiaux f_0 . La mesure standard et les variations modulaires de centralité du degré, d'intermédiarité, de proximité et du vecteur propre ainsi que certaines mesures alternatives (Membership, RWOS et OverlapNeighborhood) sont testées sur deux réseaux du monde réel de force de structure communautaire différente. 172
- 20 La différence relative de la taille de l'épidémie Δr en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité de degré, d'intermédiarité dérivées de la centralité modulaire chevauchante sont comparées à leurs homologues standard. Les mesures sont effectuées sur le réseau ego-Facebook en (a) et le réseau Yeast- protein interaction en (b) pour les algorithmes SLPA, LINKC et Louvain. 173



LISTE DES TABLEAUX

3.1	Paramètres du réseau LFR	57
3.2	Les propriétés topologiques de base de six réseaux du monde réel. N et E sont respectivement le nombre total de noeuds et de liens. Q est la modularité. N_c est le nombre de communautés. λ_{th} est le seuil épidémique.	58
3.3	Le paramètre de mélange estimé μ des réseaux du monde réel.	73
3.4	L'information mutuelle normalisée (Normalized Mutual Information NMI) dans les réseaux Power-grid et Georgetown.	76
3.5	Le paramètre de mélange estimé μ , le nombre de communautés N_c et la modularité Q dans les réseaux Power-grid et Georgetown.	76
4.1	Centralité du degré standard, composante globale et locale de la centralité du degré modulaire des noeuds dans un réseau d'exemple.	92
4.2	Paramètres du réseau LFR	96
4.3	Description des propriétés structurelles des réseaux du monde réel. N est le nombre total de noeuds, E est le nombre d'arêtes. $\langle k \rangle$, k_{max} sont respectivement le degré moyen et le degré max. C est le coefficient de regroupement moyen (average clustering coefficient). α_{th} est le seuil épidémique du réseau.	97
4.4	Les valeurs estimées du paramètre de mélange μ et la modularité Q des réseaux du monde réel.	107
4.5	Le paramètre de mélange estimé μ et la modularité Q dans les réseaux Power-grid et Georgetown.	113
5.1	Paramètres du réseau LFR	126
5.2	Propriétés des réseaux du monde réel. N est le nombre total de noeuds, E est le nombre d'arêtes. $\langle k \rangle$, k_{max} sont respectivement le degré moyen et le degré max. C est le coefficient de regroupement moyen (average clustering coefficient). λ_{th} est le seuil épidémique du réseau.	128
5.3	Le paramètre de mélange estimé μ , la proportion des noeuds qui se chevauchent on et le nombre moyen d'appartenance aux communautés par noeud m des réseaux empirique. Ces valeurs sont calculées sur la structure de la communauté découverte à l'aide de l'algorithme SLPA.	139

- 5.4 Le paramètre de mélange estimé μ , la proportion des noeuds qui se chevauchent on et le nombre moyen du degré d'appartenance aux communautés m dans les réseaux ego-Facebook et Yeast-protein interaction. 143
- 5 L'analyse de la variance *ANOVA* avec un niveau de signification $\alpha = 0.05$. *SS* est la somme des carrés. *df* est le degré de liberté. *MS* est le carré moyen. *F* est la statistique de test. *P - value* est la valeur de probabilité et *F_{critique}* est la valeur critique de F. 156



LISTE DES ABRÉVIATIONS

SI	<i>Susceptible-Infected</i>
SIR	<i>Susceptible-Infected-Recovered</i>
PR	<i>PageRank</i>
LFM	<i>Lancichinetti Fortunato Method</i>
SLPA	<i>Speaker-listener Label Propagation Algorithm</i>
DEMON	<i>Democratic Estimate of the Modular Organization of a Network</i>
CBF	<i>Community Bridge Finder</i>
DCBF	<i>Degree Community Bridge Finder</i>
BHD	<i>Bridge Hub Detector</i>
RWOS	<i>Random Walk Overlap Selection</i>
NNC	<i>Number of Neighboring Communities</i>
HB	<i>Community Hub-Bridge</i>
WHB	<i>Weighted Community Hub-Bridge</i>
NMI	<i>Normalized Mutual Information</i>
BVA	<i>Boundary Vicinity Algorithm</i>
IVD	<i>Inverse Vector Density</i>
CbC	<i>Community-based Centrality</i>
CbM	<i>Community-based Mediator</i>
GLR	<i>Gateway Local Rank</i>
DCL	<i>Degree and Clustering coefficient and Location</i>

OC *Overlapping constraint coefficient*

IM-PLA *Inuence Maximization-based on Label Propagation Algorithm*

LFR *Lancichinetti Fortunato et Radicchi*

LCC *Largest Connected Component*

LINKC *Link Communities*

YeastS *Yeast-protein interaction*

ca-GrQc *General Relativity and Quantum Cosmology collaboration*

1.1 Contexte et problématique

En général, un système complexe est littéralement un système dans lequel il existe de multiples interactions entre de nombreux composants différents. Une variété de systèmes complexes peut être modélisé et analysé par des réseaux complexes Havlin *et al.* (2012); Jia-sheng *et al.* (2011). Les réseaux complexes sont un ensemble de nombreux noeuds connectés qui interagissent de différentes manières. Dans la nature, nous pouvons trouver de nombreux réseaux différents, ce qui signifie différents types de noeuds et de connexions. Par exemple, dans un réseau social, les noeuds sont des personnes et les connexions peuvent être des relations d'amitié. Il en va de même pour le réseau d'interaction protéine-protéine où les noeuds représentent les protéines et les connexions représentent les interactions entre eux. en tant que bords. D'un point de vue mathématique et informatique, les réseaux complexes sont des structures des graphes décrivant des systèmes larges et complexes du monde réel. Un graphe est un outil mathématique permettant de représenter de manière synthétique des objets (les noeuds ou les sommets) et leurs relations (les liens ou les arêtes). Les réseaux complexes ont totalement certaines propriétés telles que la faible densité (sparseness Boccaletti *et al.* (2006), le coefficient de clustering élevé Boccaletti *et al.* (2006), le phénomène du petit monde (small-world) Watts et Strogatz (1998), etc. Ces réseaux sont des réseaux sans échelle (scale-free) avec une distribution de degré de loi de puissance Barabási (2009); Barabási et Albert (1999). Par conséquent, ils ont un grand nombre de noeuds de bas degré et un petit nombre de noeuds de haut degré Barabási (2009). Les réseaux complexes sont analysés pour comprendre la topologie, l'évolution

et le phénomène de diffusion, etc. Parmi tous ces sujets, la diffusion des épidémies ou d'informations est un sujet très populaire dont l'objectif principal est de détecter les meilleurs épandeurs pour limiter la diffusion dans un réseau Pei *et al.* (2014).

L'épidémie de maladies infectieuses menace gravement la vie des personnes et entraîne également de graves pertes économiques pour les pays victimes, comme nous l'avons constaté lors des épidémies de SRAS (syndromes respiratoires aigus sévères) en 2003 et du virus H1N1 de la grippe A en 2009 Jones et Handcock (2003); Grenfell et Bjørnstad (2005); Keeling et Rohani (2011) et coronavirus en 2019. Il est donc très important de découvrir les règles de propagation dans les groupes sociaux afin de prévenir les épidémies ou tout au moins de contrôler sa propagation. La vaccination permet de protéger les personnes et de les empêcher de transmettre des maladies parmi leurs contacts. La vaccination de masse n'étant pas toujours réalisable, en raison des ressources de vaccination limitées, d'où le besoin de mettre en oeuvre une stratégie d'immunisation. Ces stratégies sont d'un intérêt primordial pour la santé publique. Bien que l'immunisation et la quarantaine soient les deux mesures de base Ferguson *et al.* (2005, 2006); Wang *et al.* (2012), trouver un moyen efficace d'identifier les cibles à immuniser ou à mettre en quarantaine afin de supprimer efficacement une infection reste un problème urgent Madar *et al.* (2004). Beaucoup de travail est fait dans cette direction par divers chercheurs Boccaletti *et al.* (2006); Gong *et al.* (2013); Barthélemy *et al.* (2004); Halloran *et al.* (2008); Pastor-Satorras et Vespignani (2001); Singh et Singh (2012b,a, 2013).

La recherche d'épandeurs spécifiques pour diffuser les maladies infectieuses (ou l'information) dans le réseau est basée sur des mesures de centralité. Dans ces centralités, l'importance des noeuds est mesurée sur la base de certaines structures topologiques de réseau pour considérer les noeuds ayant la meilleure centralité comme des noeuds influents. Étant donné que les noeuds avec des valeurs de centralité élevées peuvent propager l'épidémie à l'ensemble du réseau plus facilement que les autres noeuds avec des mesures de centralité faibles. Ainsi, ce sont eux qui doivent être ciblés pour l'immunisation. Malheureusement, il n'y a pas de consensus général sur la définition de la centralité et de nombreuses mesures sont proposées Borgatti (2005); Borgatti et Everett (2006). La centralité du degré et la centralité d'intermédiarité (Betweenness) sont les méthodes les plus

populaires Iyer *et al.* (2013). Elles sont étudiées de manière approfondie pour comprendre leur effet sur la dynamique épidémique. La centralité du degré d'un noeud est son nombre de liens vers d'autres noeuds, tandis que la centralité d'intermédiarité mesure sa contribution aux chemins les plus courts entre chaque paire de noeuds du réseau. Comme la plupart des réseaux du monde réel suivent une distribution de loi de puissance, il est très efficace d'immuniser préférentiellement les hubs (c'est-à-dire les noeuds ayant des degrés élevés). Cependant, de nombreux réseaux du monde réel sont plus structurés que la simple distribution de degrés hétérogènes. Parfois, aucune information n'est disponible sur la structure globale des réseaux du monde réel. Par conséquent, les mesures de centralité basées sur les informations de réseau disponibles localement sont d'un intérêt primordial. De telles stratégies reposent uniquement sur des informations locales autour des noeuds sélectionnés. La stratégie la plus élémentaire est l'immunisation aléatoire où les noeuds cibles sont choisis au hasard quelle que soit la topologie du réseau.

La majorité des centralités proposées ne prennent pas en considération la structure modulaire des réseaux. Cependant, la structure des réseaux influence les modes de propagation des épidémies. Selon des recherches récentes, la structure communautaire joue un rôle crucial dans la propagation des épidémies Chakraborty *et al.* (2016); Chu *et al.* (2009); Gupta *et al.* (2015, 2016); Hébert-Dufresne *et al.* (2013); Huang *et al.* (2006); Huang et Li (2007); Liu et Hu (2005); Salathé et Jones (2010); Sun et Gao (2007); Wu et Liu (2008). Bien qu'il n'y ait pas de définition claire d'une structure modulaire, elle est très souvent appréhendée comme la division des noeuds du réseau en sous-groupes (communautés), où les noeuds intra-communautaires sont densément connectés, mais les noeuds inter-communautaires sont peu connectés les uns aux autres. Dans les réseaux avec une structure communautaire sans chevauchement, un noeud appartient à une seule communauté, alors qu'il peut être partagé par différentes communautés dans des structures de communautés qui se chevauchent. Dans cette thèse, nous visons à comprendre comment la structure communautaire affecte la dynamique épidémique et le contrôle des maladies infectieuses. Notre objectif est également de proposer de nouvelles stratégies d'immunisation adaptées pour les réseaux avec structure communautaire.

1.2 Contributions de la thèse

Dans ce qui suit, nous présentons les principales contributions scientifiques de cette thèse. Notre objectif est de comprendre comment la structure communautaire peut affecter la dynamique épidémique et comment on peut contrôler la propagation des épidémies dans ce type de réseau. A cette fin, trois contributions sont introduites. Ces contributions ont abouti à la publication de plusieurs articles scientifiques que nous citons dans la section suivante.

- La première contribution concerne la proposition de trois stratégies d'immunisation. Elles sont destinées à différents types de réseaux avec une force de structure communautaire allant de bien définie à une structure communautaire non cohésive. Notre objectif est d'étudier l'impact de la force de la structure communautaire sur le choix d'une stratégie d'immunisation appropriée. Les stratégies proposées exploitent des informations telles que la taille des communautés, le nombre de communautés attachées à chaque noeud et la densité d'interconnexion entre les communautés afin de caractériser et de classer les noeuds influents. Notre objectif est de montrer qu'engager davantage de propriétés topologiques de la communauté peut également améliorer les performances des stratégies d'immunisation. Dans ce travail, nous limitons notre attention aux réseaux dotant d'une structure communautaire sans chevauchement, c'est-à-dire qu'un noeud appartient à une seule communauté.
- Dans notre deuxième contribution, nous avons introduit un framework pour adapter toutes les mesures de centralité classiques proposées pour les réseaux sans structure communautaire aux réseaux modulaires sans chevauchement. Nous avons proposé un vecteur bidimensionnel (dit "centralité modulaire"), où chaque dimension représente un type d'influence différent que les noeuds peuvent exercer dans le réseau. Sa première composante est mesurée en calculant la centralité classique sur le réseau local. Ce réseau est formé uniquement à partir des liens intra-communautaires du réseau d'origine. De plus, sa deuxième composante est quantifiée en calculant la centralité classique sur le réseau global. Ce réseau est formé uniquement à partir des liens inter-communautaires du réseau d'origine. Selon la force de la structure

communautaire, ces deux composantes sont plus ou moins influentes.

- Dans la troisième contribution, nous avons étendu ce framework aux réseaux avec des modules qui se chevauchent. En effet, c'est un scénario fréquent dans les réseaux du monde réel où les noeuds appartiennent généralement à plusieurs communautés, en particulier pour les réseaux sociaux. La "centralité modulaire chevauchante" est une mesure bidimensionnelle qui quantifie l'influence locale et globale des noeuds (qui se chevauchent et qui ne se chevauchent pas). La composante globale de ce vecteur est définie de la même manière que la centralité modulaire. Elle est calculée sur le réseau global obtenu en supprimant tous les liens intra-communautaires du réseau d'origine. Cependant, le calcul de la composante locale dépend de la nature des noeuds. Pour un noeud qui ne se chevauchant pas, comme précédemment, seuls les membres de sa communauté sont pris en compte. Pour un noeud qui se chevauche, toutes les communautés auxquelles appartient le noeud sont fusionnées en une seule communauté.

1.3 Publications dans le cadre de la thèse

a) Revues internationales

- Z. Ghalmane, C. Cherifi, H. Cherifi, M. EL Hassouni "Centrality in Complex Networks with Overlapping Community Structure", Scientific Reports Journal (2019), vol. 9, no 1.
- Z. Ghalmane, M. EL Hassouni, C. Cherifi, H. Cherifi "Centrality in Modular Networks", EPJ Data Science Journal (2019), vol. 8, no 1, p. 15.
- Z. Ghalmane, M. EL Hassouni, H. Cherifi "Immunization of networks with non-overlapping community structure", Social Network Analysis and Mining, 2019, vol. 9, no 1, p. 45.

b) Conférences internationales

- Z. Ghalmane, C. Cherifi, H. Cherifi, M. EL Hassouni “An empirical study of the relation between the overlapping nodes and hubs in networks with modular structure”, The 8th International Conference on Complex Networks and Their Applications, December 10-12, 2019 Lisbon, Portugal.
- Z. Ghalmane, C. Cherifi, H. Cherifi, M. EL Hassouni “Searching for the hubs ? Look at the overlapping nodes !”, The 10th Conference on Network Modeling and Analysis (Modèles & Analyse des Réseaux : Approches Mathématiques & Informatiques MARAMI), November 6-8 2019 Dijon, France.
- Z. Ghalmane, C. Cherifi, H. Cherifi, M. EL Hassouni “Influence of the community structure on centrality measures”, Netsci 2019, May 27-31 2019, Burlington, Vermont USA.
- Z. Ghalmane, C. Cherifi, H. Cherifi, M. EL Hassouni “Centrality in Networks with Overlapping Communities”, The 7th International Conference on Complex Networks and Their Applications 2019.
- Z. Ghalmane, M. EL Hassouni, C. Cherifi, H. Cherifi “k-Truss Decomposition for Modular Centrality”, The 9th International Symposium on Signal, Image, Video and Communications ISIVC 2018.
- Z. Ghalmane, M. EL Hassouni, H. Cherifi “Betweenness Centrality for Networks with Non-Overlapping Community Structure”, 2018 IEEE Workshop on Complexity in Engineering (COMPENG), 1-5.
- Z. Ghalmane, M. EL Hassouni, H. Cherifi “Targeted immunization in networks with non-overlapping community structure”, The 6th International Conference on Complex Networks and Their Applications, Book of abstracts p104.

1.4 Organisation du manuscrit

Ce manuscrit est composé de trois chapitres en complément de cette introduction qui constitue le chapitre 1. Dans le chapitre 2, nous définissons tout d’abord les principaux

concepts relatifs au sujet de la thèse. Nous définissons l'immunisation des réseaux complexes ainsi que le modèle épidémique. Le reste du chapitre est consacré à l'état de l'art sur les stratégies d'immunisations des réseaux complexes. Le chapitre 3 se focalise sur les trois stratégies d'immunisation proposées pour les réseaux avec structure communautaire ainsi que l'impacte de cette dernière sur le dynamique de diffusion des épidémies. Dans le 4, on va présenter le framework proposé pour redéfinir toutes les mesures de centralités standard afin de les adapter aux réseaux modulaires. Dans le dernier chapitre (chapitre 5), ce framework est étendu aux réseaux avec une structure communautaire qui se chevauche. Nous concluons avec une conclusion générale, qui présentera notamment les perspectives que nous donnons à ces travaux. Enfin figure en annexe les résultats complémentaires que nous avons effectués pour évaluer et comparer les différentes approches des contributions qui nous ont être utiles dans le cadre de nos recherches.

2.1 Introduction

Ce chapitre vise à présenter une vue panoramique sur les différentes stratégies d'immunisation des réseaux complexes. Nous commençons tout d'abord par la définition de la notion d'immunisation. Nous introduisons également les modèles épidémiques qui servent à l'évaluation de ces méthodes. Nous présentons par la suite l'état de l'art sur les stratégies d'immunisation. Nous mettons l'accent premièrement sur les stratégies dédiées pour les réseaux non modulaire. Ensuite, nous décrivons les stratégies d'immunisation les plus récentes désignées pour les réseaux ayant une structure communautaire afin de les comparer et de positionner notre travail vis-à-vis de ces derniers.

2.2 Stratégies d'immunisation

L'objectif d'une stratégie d'immunisation Giabbanelli et Peters (2011) est de révéler l'ensemble des épandeurs les plus influents dans un réseau donné. Selon la quantité d'informations dont ils ont besoin sur la structure globale du réseau, ils peuvent être classés en deux catégories : les stratégies déterministes et les stratégies stochastiques. Le premier type de stratégies nécessite des informations sur l'ensemble de la topologie du réseau, tandis que le deuxième groupe de stratégies n'a besoin que de la connaissance de la structure du réseau au niveau du noeud.

Les stratégies d'immunisation déterministes sont basées sur un ordre de tous les noeuds du réseau afin de les immuniser en fonction de leur rang. Pour ce faire, une mesure dite centralité est calculée pour chaque noeud du réseau. Elle quantifie sa capacité

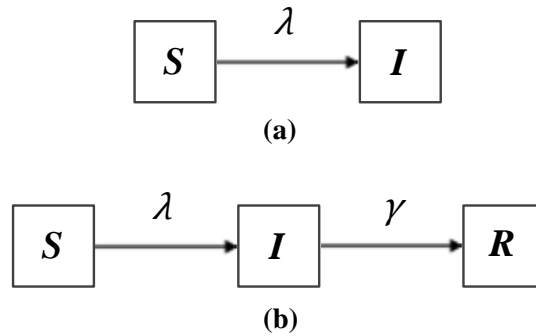


FIGURE 2.1 – Le mécanisme d’infection du modèle classique (a) SI (b) SIR.

à diffuser la maladie à l’intérieur du réseau. Le degré et l’intermédiarité sont les mesures de centralité les plus couramment utilisées pour classer les noeuds. Les noeuds sont ensuite ciblés dans l’ordre décroissant de leur rang du noeud le plus central au noeud le moins central. Puisque tous les noeuds étant impliqués dans ce processus, la connaissance de l’ensemble du réseau est alors requise pour ces stratégies. En revanche, les stratégies d’immunisation stochastique sont agnostiques quant à la structure globale du réseau. Elles peuvent fonctionner avec une quantité très limitée d’informations sur un noeud. La stratégie stochastique la plus simple est l’immunisation uniforme qui cible les noeuds de manière totalement aléatoire sans aucune information. La connaissance (Acquaintance) cite acq est une autre stratégie d’immunisation stochastique populaire qui sélectionne des voisins aléatoires de noeuds sélectionnés au hasard et les immunise s’ils ont été sélectionnés n fois. Généralement, les stratégies déterministes fonctionnent mieux que les stratégies stochastiques car elles peuvent utiliser plus d’informations sur les propriétés topologiques des réseaux. Cependant, les stratégies stochastiques sont généralement plus efficaces en terme de complexité.

2.3 Modèle épidémique

Les modèles Susceptible-Infecté (SI) et Susceptible-Infecté-Rétabli (SIR) sont largement utilisés pour la diffusion des infections et la diffusion d’informations dans différents domaines. Dans cette thèse, nous utilisons le modèle SIR pour estimer les capacités de diffusion des noeuds.

Le modèle Susceptible-Infecté (SI) Hurley *et al.* (2006) est considéré comme la forme la plus simple de tous les modèles épidémiques. Dans ce modèle, un noeud n'a que deux états possibles : un état Susceptible (S) ou Infecté (I). Le modèle peut être représenté par le diagramme de compartiment montré dans Figure 4 (a). Au début, tous les noeuds sont mis à l'état susceptible (les individus sont sans immunité). Après cela, l'état d'une petite proportion de noeuds sélectionnés par une stratégie d'immunisation donnée est défini sur l'état infecté. À chaque pas de temps, un noeud infecté peut infecter ses voisins susceptibles avec le taux de transmission λ . Ce processus se termine lorsqu'il n'y a aucun noeud sensible dans le réseau.

Le modèle épidémique Susceptible-Infecté-Rétabli (SIR) Newman (2002); Moreno *et al.* (2002) est utilisé pour simuler le processus de propagation dans les réseaux. Dans ce modèle, il existe trois états pour chaque noeud : susceptible (S), infecté (I) et rétabli (R). Le mécanisme d'infection du modèle SIR est montré dans Figure 4 (b). Initialement, les noeuds ciblés sont choisis en fonction d'une stratégie d'immunisation donnée jusqu'à ce qu'une couverture vaccinale souhaitée de la population soit atteinte, et que leur état soit réglé sur l'état rétabli ou résistant R. Tous les noeuds restants sont à l'état S. Après cette configuration initiale, l'infection commence à partir d'un noeud susceptible aléatoire. Son état passe à I. À chaque pas de temps, l'épidémie se propage d'un noeud infecté à un noeud susceptible voisin en fonction du taux de transmission de l'infection λ . De plus, les noeuds infectés récupèrent avec un taux γ , c'est-à-dire que la probabilité de rétablissement d'un noeud infecté par pas de temps est γ . Si le rétablissement se produit, l'état du noeud rétabli passe de infecté à résistant. Le processus de propagation de l'épidémie se termine lorsqu'il n'y a pas de noeud infecté dans le réseau. Après chaque simulation, nous enregistrons le nombre total de noeuds rétablis (la taille de l'épidémie).

2.4 L'immunisation des réseaux complexes non modulaires

2.4.1 Stratégies stochastiques

Les stratégies d'immunisation stochastique se concentrent sur l'utilisation des informations au niveau du noeud. Elles identifient les noeuds cibles sans connaître la structure

complète du réseau. Cela les rend plus efficaces et plus pratiques en termes de calcul dans les grands réseaux par rapport aux stratégies déterministes. Les stratégies stochastiques n'ont pas besoin d'informations sur la structure globale du réseau. La stratégie uniforme est la stratégie d'immunisation la plus élémentaire. Elle cible les noeuds d'une manière aléatoire sans utiliser d'informations sur la topologie du réseau.

La stratégie de connaissance (Acquaintance strategy) est un algorithme d'immunisation décrit par Cohen *et al.* (2003). Cette stratégie fonctionne comme suit :

Étape 1. choisissez un noeud aléatoire v_0 .

Étape 2. choisissez une connaissance aléatoire v_1 , c'est-à-dire un noeud voisin choisi au hasard de v_0 .

Étape 3. Immunisez les noeuds qui ont été appelés connaissances au moins n fois jusqu'à ce que la couverture vaccinale souhaitée soit atteinte.

Dans le cas $n = 1$, chaque connaissance sera immédiatement immunisée. Il a été démontré que la stratégie de connaissance identifie des individus hautement connectés, en particulier dans les réseaux à queue épaisse (tels que les réseaux dits sans échelle ou scale-free).

2.4.2 Stratégies déterministes

Les stratégies déterministes ciblent les noeuds en les classant selon une mesure de centralité donnée. La centralité d'un noeud reflète sa capacité à propager la maladie. La procédure des stratégies déterministes peut être spécifiée comme suit :

Étape 1. Sélectionnez une mesure de centralité donnée.

Étape 2. Calculez la centralité de chaque noeud du réseau.

Étape 3. Classez les noeuds par ordre décroissant, du plus grand au moins central.

Étape 4. Ciblez une proportion de noeuds de rang élevé pour l'immunisation.

Ces stratégies nécessitent la connaissance de l'ensemble du réseau car tous les noeuds sont impliqués dans le processus.

2.4.2.1 Stratégies d'immunisation basées sur le voisinage

Degré : Dans un réseau simple non dirigé $G(V, E)$ avec V et E étant respectivement l'ensemble des noeuds et l'ensemble des liaisons. Le degré d'un noeud v_i , noté k_i , est défini comme le nombre de voisins directement connectés de v_i . Mathématiquement, $k_i = \sum_j a_{ij}$, où $A = \{a_{ij}\}$ est la matrice d'adjacence, c'est-à-dire $a_{ij} = 1$ si v_i et v_j sont connectés et 0 sinon. La centralité des degrés est l'indice le plus simple pour identifier les influences des noeuds : plus il y a de connexions, plus l'influence du noeud est importante.

Dans un réseau dirigé $D(V, E)$ chaque lien est associé à une direction, et nous devrions alors considérer respectivement le degré extérieur et le degré inférieur d'un noeud. Par exemple, en pensant à twitter.com où un lien dirigé existe du noeud v_j au noeud v_i si v_j suit v_i , alors, le degré d'un noeud v_i (c'est-à-dire le nombre de noeuds ayant un lien dirigé pointant vers v_i) reflète la popularité de v_i tandis que le degré excessif de v_i (c'est-à-dire le nombre de liens de v_i à d'autres noeuds) représente dans une certaine mesure l'activité sociale de v_i Wasserman *et al.* (1994). Dans les réseaux pondérés, la centralité des degrés est généralement remplacée par la force, définie comme la somme des poids des liens associés.

LocalRank : La centralité des degrés pourrait être moins précise dans l'évaluation de l'influence des noeuds car elle utilise très limitées informations Chen *et al.* (2012); Kitsak *et al.* (2010). Chen *et al.* (2012) ont proposé l'extension de la centralité des degré. LocalRank est un algorithme efficace basé sur l'information locale des noeuds. Elle tient pleinement compte de l'information contenue dans les voisins de quatrième ordre de chaque noeud. Le score LocalRank du noeud v_i est défini comme suit :

$$LR(i) = \sum_{j \in \Gamma_i} Q(j) \quad (2.1)$$

$$Q(j) = \sum_{k \in \Gamma_j} R(k) \quad (2.2)$$

où Γ_i est l'ensemble des voisins les plus proches de v_i et $R(k)$ est le nombre des voisins les plus proches et les plus proches de v_k . L'algorithme LocalRank a une complexité beau-

coup plus faible que les centralités basées sur des chemins. En fait, la complexité de calcul de l'algorithme LocalRank est $O(n \langle k \rangle^2)$ qui croît presque linéairement avec l'échelle des réseaux. L'algorithme LocalRank peut également être étendu aux réseaux pondérés Gao *et al.* (2013).

Coreness : La centralité des degrés ne prend en compte que le nombre de voisins les plus proches et affirme que les noeuds de même degré ont la même influence dans le réseau. Kitsak *et al.* (2010) ont fait valoir que l'emplacement d'un noeud est plus important que le nombre de ses voisins immédiats pour évaluer son influence de propagation. C'est-à-dire que si un noeud est situé dans la partie centrale du réseau, l'influence du noeud sera plus élevée que celle qui est située en périphérie. Par conséquent, Kitsak *et al.* (2010) ont proposé la coreness comme un meilleur indicateur de l'influence de la propagation d'un noeud, qui peut être obtenue en utilisant la décomposition k-core (également appelée k-shell) Dorogovtsev *et al.* (2006) dans les réseaux.

Étant donné un réseau simple non dirigé G , initialement, la coreness c_i de chaque noeud isolé v_i (c'est-à-dire $k_i = 0$) est définie comme $c_i = 0$, et ces noeuds sont supprimés avant la décomposition en k-core. Ensuite, dans la première étape de la décomposition k-core, tous les noeuds de degré $k = 1$ seront supprimés. Cela entraînera une réduction des valeurs de degré aux noeuds restants. Supprimez continuellement tous les noeuds dont le degré résiduel $k \leq 1$, jusqu'à ce que tous les degrés résiduels restants sont supérieur à 1. Tous les noeuds supprimés dans la première étape de la décomposition forment le 1-shell et leur coreness k_s sont tous égaux à 1. Dans la deuxième étape, tous les noeuds restants dont les degrés $k = 2$ seront supprimés en premier lieu. Ensuite, supprimez de manière itérative tous les noeuds dont les degrés résiduels $k \leq 2$ jusqu'à ce que tous les noeuds restants dont les degrés résiduels $k > 2$. Les noeuds supprimés dans la deuxième étape de la décomposition forment le 2-shell et leur coreness k_s sont deux. Le processus de décomposition se poursuivra jusqu'à ce que tous les noeuds soient supprimés. Enfin, la coreness d'un noeud v_i est égale à sa couche correspondante. La figure 1 est un diagramme schématique simple de la décomposition en k-core. Apparemment, un noeud avec une plus grande coreness signifie que le noeud est situé dans une position plus centrale et est peut-

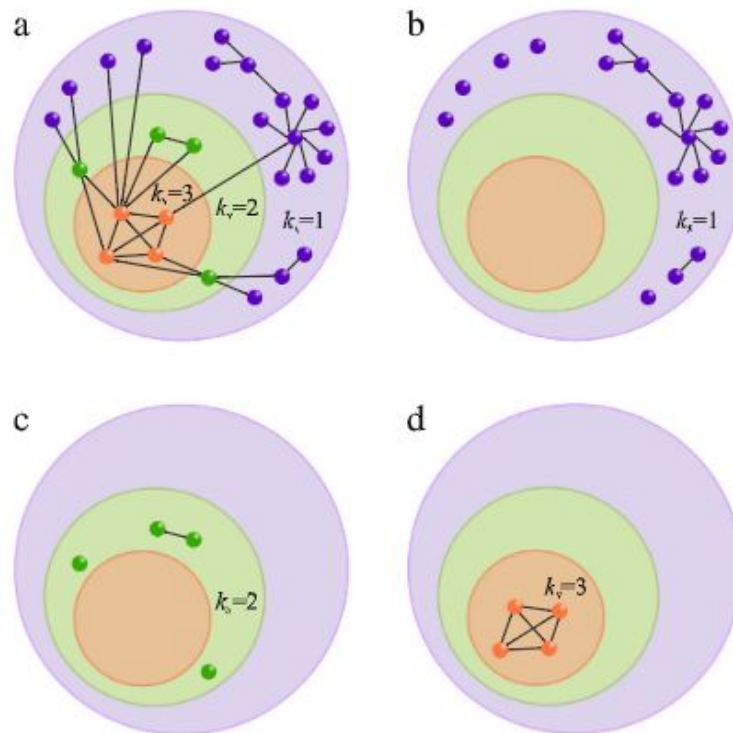


FIGURE 2.2 – Diagramme schématique de la décomposition k-core.

être plus important dans le réseau.

2.4.2.2 Stratégies d'immunisation basées sur les voies

Excentricité : Dans un réseau connecté $G(V, E)$, nous définissons d_{ij} comme la plus courte longueur de chemin entre le noeud v_i et v_j . On pense que plus la distance entre un noeud v_i et tous les autres noeuds est courte, plus le noeud est centré dans le réseau. Ainsi, l'excentricité d'un noeud v_i est définie comme la distance maximale entre tous les chemins les plus courts vers les autres noeuds Hage et Harary (1995) :

$$ECC(i) = \max_{v_j \neq v_i} \{d_{ij}\} \quad (2.3)$$

où v_j s'exécute sur tous les noeuds autres que v_i . Le noeud avec une excentricité plus petite est supposé être le plus influent. Notez que la distance maximale peut être affectée par quelques trajets inhabituellement longs, alors l'excentricité peut ne pas refléter l'importance des noeuds (voir un exemple sur la Fig. 3).

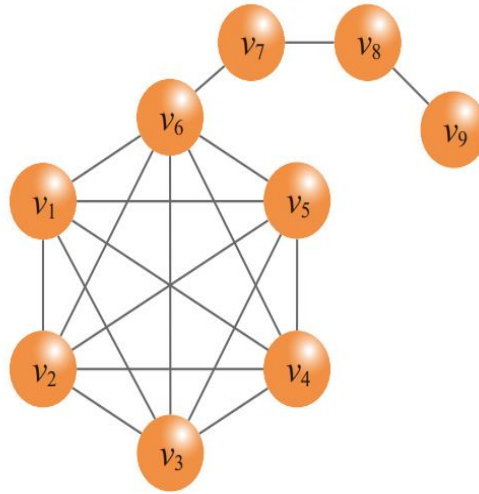


FIGURE 2.3 – Illustration d'un exemple de réseau, où le noeud le plus influent devrait être v_6 , mais selon l'indice d'excentricité, v_7 est le plus influent avec $ECC(7) = 2$.

La proximité (Closeness) : La centralité de proximité (closeness) élimine la perturbation en sommant toutes les distances entre le noeud cible et tous les autres noeuds. Pour un réseau connecté $G(V, E)$, la centralité de proximité d'un noeud v_i est définie comme l'inverse de la distance géodésique moyenne de v_i à tous les autres noeuds Sabidussi (1966); Freeman (1978) :

$$CC(i) = \frac{n - 1}{\sum_{j \neq i} d_{ij}} \quad (2.4)$$

De toute évidence, plus la proximité est grande, plus le noeud est central. La centralité de proximité peut également être comprise comme l'inverse de la longueur de propagation moyenne d'information dans le réseau. En général, le noeud avec la valeur de proximité la plus élevée a la meilleure vision du flux d'informations. Malheureusement, la définition originale a un inconvénient majeur : lorsque le réseau n'est pas connecté (dans les réseaux dirigés, le réseau doit être fortement connecté), il existe des paires de noeuds avec $d_{ij} = \infty$. Ainsi, une approche très populaire consiste à calculer la centralité de proximité en termes

d'inverse des distances moyennes harmoniques entre les noeuds :

$$CC(i) = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d_{ij}} \quad (2.5)$$

où $1/\infty = 0$. La centralité de proximité d'un noeud reflète l'efficacité avec laquelle il échange des informations avec les autres.

Katz : Contrairement à la centralité de proximité qui ne prend en compte que les longueurs de chemin les plus courtes entre les paires de noeuds, la centralité de Katz calcule les influences des noeuds en considérant tous les chemins du réseau. La centralité de Katz soutient l'idée qu'un chemin plus court joue un rôle plus important. En supposant que la longueur d'un chemin entre le noeud v_i et v_j est $d_{ij} = p$, la contribution du chemin à l'importance du noeud v_i et v_j est s^p , où $s \in (0, 1)$ est un paramètre réglable. Évidemment, la contribution des chemins plus longs sera largement supprimée si s est petit. En supposant que l_{ij}^p est le nombre de chemins entre v_i et v_j de longueur p , nous avons $(l_{ij}^p) = A^p$ où A est la matrice d'adjacence du réseau. Par conséquent, la matrice d'interaction qui décrit la relation entre n'importe quelle paire de noeuds est obtenue comme suit Katz (1953) :

$$K = (k_{ij}) = sA + s^2A^2 + \dots + s^pA^p + \dots = (IsA)^{-1}I \quad (2.6)$$

où I est la matrice d'identité. Afin d'assurer la mise en place du côté droit de l'Eq. (15), s doit être inférieur à l'inverse de la plus grande valeur propre de A . La valeur de k_{ij} est également connue sous le nom de similitude de Katz entre le noeud v_i et v_j en science des réseaux. La centralité de Katz d'un noeud v_i est définie comme suit Katz (1953) :

$$Katz(i) = \sum_{j \neq i} k_{ij} \quad (2.7)$$

La centralité de Katz a une structure mathématique gracieuse. Cependant, la grande complexité de calcul de la centralité de Katz rend difficile son utilisation dans les réseaux à grande échelle.

Index d'information : La centralité de l'indice de l'information (également appelé centralité de l'indice S-Z) mesure l'importance des noeuds sur la base des informations contenues dans tous les chemins possibles entre les paires de noeuds du réseau Stephenson et Zelen (1989). La définition de "l'information" vient de la théorie de l'estimation statistique. Il est supposé qu'il existe du bruit dans chaque lien, provoquant une perte lors de chaque transition de l'information. Plus le chemin est long, plus la perte est importante. Mathématiquement, la quantité totale d'informations transmises sur un chemin est égale à l'inverse de la longueur du chemin. La quantité d'informations pouvant être transmises entre une paire de noeuds v_i et v_j est égale à la somme des informations transmises par tous les chemins possibles entre eux, notée q_{ij} . L'index d'informations prend en compte tous les chemins possibles dans le réseau connecté mais ne nécessite pas de les énumérer. En fait, le q_{ij} s'est avéré être équivalent à la conductance électrique dans un réseau électrique Altmann (1993); Poulin *et al.* (2000). Sur la base de la théorie du réseau électrique, la quantité totale d'informations entre v_i et v_j peut être obtenue comme suit :

$$q_{ij} = (r_{ii} + r_{jj} - 2r_{ij})^{-1} \quad (2.8)$$

avec r_{ij} étant un élément de la matrice :

$$R = (r_{ij}) = (DA + F)^{-1} \quad (2.9)$$

où D est une matrice diagonale à n dimensions dont les éléments sont les degrés des noeuds correspondants. F est une matrice à n dimensions dont les éléments sont tous égaux à 1. L'indice d'information de v_i est alors défini comme la moyenne harmonique de q_{ij} Poulin *et al.* (2000) :

$$INF(i) = \left[\frac{1}{n} \sum_j \frac{1}{q_{ij}} \right] \quad (2.10)$$

L'indice d'information de v_i peut également être calculé par la moyenne arithmétique Altmann (1993). Il s'agit d'une variation de la centralité de la proximité d'un point de vue mathématique, avec une manière différente de mesurer la contribution de chaque chemin.

Cet indice peut être facilement étendu aux réseaux pondérés Stephenson et Zelen (1989).

Intermédiarité (Betweenness) : La centralité de l'intermédiarité (Betweenness) a d'abord été proposée par Bavelas en 1948 Bavelas (1948), puis reformulée par Shimbel Shimbel (1953) et Shaw Shaw (1954) en vue de la puissance potentielle d'un noeud pour contrôler le flux d'informations dans un réseau. En 1977, Freeman Freeman (1977) a généralisé la notion théorique de l'intermédiarité et l'a étendue aux réseaux connectés et non connectés, en lui montrant la façon dont nous utilisons aujourd'hui. Généralement, il existe plus d'un chemin le plus court commençant à partir du noeud v_s et se terminant à v_t . La contrôlabilité du flux d'informations de v_i peut être calculée en comptant tous les chemins les plus courts passant par v_i . Ainsi, la centralité d'intermédiarité du noeud v_i peut être définie comme suit :

$$BC(i) = \sum_{i \neq s, i \neq t, s \neq t} \frac{g_{st}^i}{g_{st}} \quad (2.11)$$

où g_{st} est le nombre de chemins les plus courts entre v_s et v_t et g_{st}^i est le nombre de chemins qui passent par v_i parmi tous les chemins les plus courts entre v_s et v_t .

2.4.2.3 Stratégies d'immunisation basées sur le raffinement itératif

L'influence d'un noeud est non seulement déterminée par le nombre de ses voisins, mais également déterminée par l'influence de ses voisins, connue sous le nom d'effet d'amélioration mutuelle Wittenbaum *et al.* (1999). Dans ce chapitre, nous allons sélectionner quelques centralités de raffinement itératif typiques dans lesquelles chaque noeud obtient le support de ses voisins. Parmi ces algorithmes, la centralité des vecteurs propres (Eigenvector) Bonacich (1972) et l'algorithme de nomination cumulative Poulin *et al.* (2000) sont conçus dans les réseaux non dirigés, tandis que le PageRank Brin et Page (1998) et ses variantes sont principalement utilisés dans les réseaux dirigés. PageRank était à l'origine utilisé pour classer les pages Web et était l'algorithme de base du moteur de recherche Google. PageRank introduit un facteur de saut aléatoire qui est un paramètre réglable dont la valeur optimale dépend à la fois de la structure du réseau et de la fonction objectif.

Centralité des vecteurs propres (Eigenvector) : La centralité des vecteurs propres suppose que l'influence d'un noeud est non seulement déterminée par le nombre de ses voisins, mais également déterminée par l'influence de chaque voisin Bonacich (1972, 1987). La centralité d'un noeud est proportionnelle à la somme des centralités des noeuds auxquels il est connecté Bonacich (2007). L'importance d'un noeud v_i , noté x_i , est :

$$x_i = c \sum_{j=1}^n a_{ij} x_j \quad (2.12)$$

qui peut être écrit sous forme matricielle :

$$\vec{x} = cA\vec{x} \quad (2.13)$$

où c est une constante de proportionnalité. Généralement, $c = 1/\lambda$ dans lequel λ est la plus grande valeur propre de A . La centralité des vecteurs propres peut être calculée efficacement par la méthode d'itération de puissance Hotelling (1936). Au début de l'itération de puissance, le score de chaque noeud est initialisé à 1. Chaque noeud partage ensuite son score uniformément avec ses voisins connectés et reçoit de nouvelles valeurs à chaque tour d'itération. Ce processus se répète jusqu'à ce que les valeurs des noeuds atteignent l'état d'équilibre. Du point de vue de cette méthode d'itération, l'algorithme PageRank est une variante de la centralité des vecteurs propres. Cependant, dans les réseaux dirigés, il est courant que de nombreux noeuds aient uniquement un degré externe, ce qui conduit à un état nul après le premier tour de l'itération de puissance. Pour résoudre ce problème, Bonacich et Lloyd (2001) ont proposé une variante de la centralité des vecteurs propres, appelée centralité alpha :

$$\vec{x} = \alpha A\vec{x} + \vec{e} \quad (2.14)$$

où \vec{e} est un vecteur des sources de statut exogènes et α un paramètre qui reflète l'importance relative des facteurs endogènes par rapport aux facteurs exogènes. En réf. Bonacich et Lloyd (2001), \vec{e} était supposé être un vecteur des uns, aboutissant essentiel-

lement à la même solution avec la centralité de Katz.

Nomination cumulative : La centralité des vecteurs propres n'est pas toujours une méthode idéale pour sa lenteur de convergence et parfois son enfermement dans une boucle sans fin. La nomination cumulative suppose que plus d'individus centraux seront nommés plus souvent dans les réseaux sociaux Poulin *et al.* (2000), et tient compte des valeurs de nomination de chaque noeud et ses voisins. Soit $\tilde{p}_i(t-1)$ le nombre cumulé de nominations du noeud v_i après $t-1$ itérations. Au début, la valeur de nomination de chaque noeud est initialement fixée à un. Ensuite, chaque noeud obtient des nominations de tous ses voisins et la valeur mise à jour est égale à sa valeur d'origine plus la somme des valeurs des voisins à chaque itération. La nomination cumulative du noeud v_i après t itérations est :

$$\tilde{p}_i(t) = \tilde{p}_i(t-1) + \sum_j a_{ij} \tilde{p}_j(t-1) \quad (2.15)$$

La nomination cumulative normalisée de v_i peut être calculée comme suit :

$$p_i(t) = \frac{\tilde{p}_i(t)}{\sum_j \tilde{p}_j(t)} \quad (2.16)$$

L'itération s'arrêtera si les nominations cumulatives normalisées de tous les noeuds atteignent l'état stationnaire. La nomination cumulative a une formule similaire avec la centralité d'alpha. Bien que la différence soit que le vecteur \vec{e} dans la centralité d'alpha est un vecteur fixe, cependant, l'élément correspondant dans la nomination cumulative, disons \tilde{p}_i^t , est égal à la valeur du noeud dans la dernière itération, ce qui améliore le taux de convergence.

PageRank : L'algorithme PageRank Brin et Page (1998) est une variante célèbre de la centralité des vecteurs propres. Il est utilisé pour classer les sites Web dans les moteurs de recherche Google et d'autres scénarios commerciaux Langville et Meyer (2011). Les algorithmes traditionnels de classement des sites Web basés sur des mots clés sont vulnérables à une attaque malveillante qui améliore l'influence d'un site Web en augmentant la densité

des mots clés non pertinents. PageRank distingue l'importance des différents sites Web en se basant sur une marche aléatoire sur le réseau construit à partir des relations des pages Web. Semblable à la centralité des vecteurs propres, PageRank suppose que l'importance d'une page Web est déterminée à la fois par la quantité et la qualité des pages qui lui sont liées. Initialement, chaque noeud (c'est-à-dire la page) obtient une valeur PR unitaire. Ensuite, chaque noeud distribue uniformément la valeur PR à ses voisins ainsi que ses liens sortants. Mathématiquement, la valeur PR du noeud v_i à l'étape t est :

$$PR_i(t) = \sum_{j=1}^n a_{ji} \frac{PR_j(t-1)}{k_j^{out}} \quad (2.17)$$

où n est le nombre total de noeuds dans le réseau et k_j^{out} est le degré extérieur du noeud v_j . L'itération ci-dessus s'arrêtera si les valeurs PR de tous les noeuds atteignent l'état stationnaire. Un inconvénient majeur du processus de marche aléatoire ci-dessus est que la valeur PR d'un noeud avec zéro degré externe ne peut pas être redistribuée, puis Eq. (37) ne peut pas garantir la convergence (il existe d'autres cas où l'équation (37) ne converge pas Guan-Rong *et al.* (2012)). Pour résoudre ces problèmes, un facteur de saut aléatoire a été introduit en supposant que l'internaute parcourra les pages Web avec les liens avec la probabilité s , quittera la page actuelle et ouvrira une page aléatoire avec la probabilité $1 - s$. En conséquence, l'Eq. (37) est modifié comme suit :

$$PR_i(t) = s \sum_{j=1}^n a_{ji} \frac{PR_j(t-1)}{k_j^{out}} + (1-s) \frac{1}{n} \quad (2.18)$$

Eq. (38) retournera à l'Eq. (37) lorsque $s = 1$. La probabilité de saut aléatoire s est généralement fixée à environ 0.85, mais doit en effet être testée dans différents scénarios Brin et Page (1998).

Le PageRank a été utilisé bien au-delà de son intention et de sa conception initiales Gleich (2015). Il a été appliqué pour classer une large gamme d'objets via leur structure de réseau : pour classer des images Jing et Baluja (2008) et des livres Meng (2009), pour classer des gènes Morrison *et al.* (2005) et des protéines Freschi (2007) en Biologie et Bio-informatique, pour classer des molécules Mooney *et al.* (2012) en chimie, pour classer les

régions du cerveau Zuo *et al.* (2012) et les neurones Crofts et Higham (2011) en neurosciences, pour classer les noms d'hôtes Arasu *et al.* (2002), les noyaux Linux Chepelianskii (2010) et les interfaces de programmation Kim *et al.* (2013) dans les systèmes d'information complexes, pour classer les leaders Lü *et al.* (2011); Java *et al.* (2006); Weng *et al.* (2010) dans les réseaux sociaux, pour classer les scientifiques Liu *et al.* (2005); Ding *et al.* (2009), les articles Chen *et al.* (2007); Ma *et al.* (2008); Sayyadi et Getoor (2009) et les revues Bollen *et al.* (2006); Butler (2008); West *et al.* (2010) en bibliométrie, pour classer les joueurs Radicchi (2011) et les équipes Govan *et al.* (2008) dans les sports, etc. Ghoshal et Barabasi Ghoshal et Barabási (2011) ont étudié l'émergence de noeuds super-stables dans des réseaux perturbés lors de l'évaluation de la propagation capacités des noeuds avec PageRank. Ils ont étudié le classement des différentes propriétés topologiques et ont constaté que le PageRank est sensible aux perturbations dans les réseaux aléatoires alors qu'il est stable dans les réseaux sans échelle (scale-free).

2.5 L'immunisation des réseaux complexes modulaire

2.5.1 Structure communautaire

De nombreux réseaux du monde réel présentent une structure communautaire, c'est-à-dire que leurs noeuds sont organisés en modules, appelés communautés. Les premières définitions de la structure communautaire ont été proposées par les analystes des réseaux sociaux. Ils ont étudié la structure des sous-graphes. La clique est le concept le plus répandu Bomze *et al.* (1999). Une clique est un sous-graphe complet tel que chacun de ses noeuds est associé à tous les autres. En général, les communautés ne sont pas des graphes complets. De plus, dans une clique, tous les noeuds ont un rôle identique, tandis que certains noeuds sont plus importants que d'autres dans les communautés, en raison de leurs schémas de liaison hétérogènes. Ainsi, cette notion ne peut pas être considérée comme un candidat approprié pour la définition de la communauté. Une définition informelle très répandue du concept de communauté le considère comme un groupe de noeuds fortement interconnectés par rapport aux autres noeuds Fortunato (2010). Une communauté est alors un sous-ensemble cohérent de noeuds faiblement connectés au reste du réseau. Ce point

de vue a été contesté, les travaux récents Palla *et al.* (2005); Leskovec *et al.* (2009); Jeub *et al.* (2015); Yang et Leskovec (2014) ont montré que les communautés peuvent également se chevaucher. Certains noeuds peuvent être partagés par plusieurs communautés. Dans les réseaux sociaux, par exemple, les individus peuvent participer à différents groupes en même temps, comme les collègues de travail, les amis ou la famille.

L'identification des communautés dans les réseaux peut offrir une idée claire de la façon dont le réseau est organisé. Nous pouvons en fait distinguer les noeuds qui sont totalement intégrés à l'intérieur de leurs communautés et les noeuds qui sont situés à la frontière des communautés. Ces noeuds peuvent jouer le rôle de ponts entre les communautés des réseaux et pourraient jouer un rôle majeur dans la dynamique des processus de diffusion à travers le réseau. La détection de communauté dans les réseaux, également appelée partitionnement de réseau ou clustering, est un problème mal caractérisé. Les définitions formelles peuvent différer dans la façon dont elles considèrent ces aspects de la cohésion et de la séparation des communautés. Il n'y a donc pas de définition universelle des modules que l'on devrait rechercher. Une telle ambiguïté laisse beaucoup de liberté pour proposer différents algorithmes de détection de communauté mettant en oeuvre différemment la notion de structure de communauté. Dans cette sous-section, nous présentons un ensemble représentatif de méthodes et les classons selon l'approche qu'elles appliquent pour découvrir les communautés.

2.5.1.1 *Détéction de communautés sans chevauchement*

a- Algorithmes basés sur la modularité

La modularité est une mesure répandue introduite par Newman (2006b); Zhang *et al.* (2009), qui mesure la qualité d'une structure communautaire. Elle évalue la connectivité interne des communautés identifiées à travers le nombre de liens intra et inter-communautaires. Les algorithmes basés sur l'optimisation de la modularité ont tendance à identifier la meilleure structure de communauté en termes de modularité.

FastGreedy proposé par Newman (2004) est basé sur une approche d'optimisation gourmande. Cela commence par un état dans lequel chaque noeud constitue sa propre communauté. L'algorithme fusionne à plusieurs reprises des paires de communautés pour

en obtenir de plus grandes. À chaque étape, les communautés jointes sont sélectionnées en considérant la plus grande augmentation (ou la plus petite diminution) de la modularité. FastGreedy produit une hiérarchie de structures communautaires. Le meilleur est celui qui obtient la modularité maximale.

Louvain proposé par Blondel *et al.* (2008) est un autre algorithme d'optimisation. Il s'appuie sur un processus d'optimisation gourmand. Il comprend une phase d'agglomération supplémentaire pour améliorer l'approche d'optimisation. Au départ, comme pour FastGreedy, chaque noeud constitue sa propre communauté. Après cela, un algorithme d'optimisation gourmand est appliqué pour identifier les communautés. La deuxième étape consiste à former un nouveau réseau, où les noeuds représentent les communautés trouvées lors de la première phase. Les liens inter-communautaires sont agrégés et représentés sous forme de liens entre les nouveaux noeuds, tandis que les liens intra-communautaires sont représentés par des auto-boucles. La première phase est répétée sur le nouveau réseau et le processus se termine lorsque des communautés stables sont atteintes.

b- Algorithmes à marche aléatoire

Divers algorithmes utilisent des marches aléatoires de différentes manières afin d'identifier les communautés dans les réseaux. Dans ce paragraphe, nous avons conservé l'un des algorithmes les plus influents de cette classe.

WalkTrap proposé par Pons et Latapy (2006) utilise une approche de regroupement aggloméré hiérarchique comme pour FastGreedy, mais avec un critère de fusion différent. Il utilise une mesure de distance basée sur des marches aléatoires. Cet algorithme est basé sur l'idée que les promenades aléatoires ont tendance à être piégées dans une communauté. Si deux noeuds i et j sont dans la même communauté, la probabilité d'arriver à un troisième noeud k situé dans la même communauté par une marche aléatoire ne devrait pas être différente pour les deux. La distance est construite en additionnant ces différences sur tous les noeuds, avec une correction pour le degré.

c- Algorithmes basés sur l'information

Le but de ces algorithmes est d'utiliser la structure communautaire de façon à représenter

le réseau en utilisant moins d'informations. Nous avons retenu deux algorithmes de cette classe.

InfoMod a été proposé par Rosvall et Bergstrom (2007), qui utilise une matrice de communauté et un vecteur d'appartenance comme représentation simplifiée du réseau en se concentrant sur la structure de la communauté. Le premier est une matrice d'adjacence représentant des communautés au lieu de noeuds, tandis que le second est un vecteur associant chaque noeud à une communauté. Cet algorithme utilise la mesure d'information mutuelle (présentée dans la sous-section suivante) afin de mesurer la quantité d'informations du réseau d'origine contenues dans la représentation simplifiée. La meilleure affectation parmi toutes les affectations possibles de noeuds aux communautés est celle associée à l'information mutuelle maximale.

InfoMap est un autre algorithme proposé par les mêmes auteurs Rosvall et Bergstrom (2008). Il tend à trouver l'ensemble de noeuds (communautés) contenant un flux d'informations intra-module élevé et un flux d'informations inter-module faible. L'algorithme InfoMap est basé sur une équation de carte. Elle est basée sur le flux d'informations utilisé pour trouver une représentation compressée d'un ensemble de parcours aléatoires dans un graphe. Les partitions de haute qualité sont trouvées en minimisant la quantité d'informations nécessaires pour représenter une marche aléatoire dans le réseau. En effet, la marche restera probablement plus longtemps à l'intérieur des communautés dans une partition contenant peu de liens inter-communautaires.

2.5.1.2 *Détéction de communautés avec chevauchement*

Dans cette sous-section, trois algorithmes de détection de communautés avec chevauchement très populaires appartenant à deux classes différentes sont présentés. La Méthode de Lancichinetti Fortunato (LFM) est un algorithme local d'expansion et d'optimisation. L'algorithme de propagation de l'étiquette du locuteur-écoutateur (Speaker-listener Label Propagation Algorithm SLPA) et l'estimation démocratique de l'organisation modulaire d'un réseau (Democratic Estimate of the Modular Organization of a Network DEMON) appartiennent à la catégorie de propagation des étiquettes. Plus de détails sont donnés dans des articles récents sur les méthodes de détection communautaire Jebabli *et al.*

(2018); Xie *et al.* (2013).

Méthode de Lancichinetti Fortunato (LFM) est proposé par Lancichinetti *et al.* (2009). Elle prend un noeud de départ aléatoire et lui ajoute des noeuds jusqu'à ce qu'une fonction de fitness soit localement maximale. Après avoir assemblé une communauté, le même processus est appliqué sur un autre noeud de départ qui n'est pas encore affecté à une communauté afin de développer une nouvelle communauté. La fonction fitness contrôle la force et la taille des communautés. La complexité la plus défavorable est $O(n^2)$ où n est le nombre de noeuds.

Algorithme de propagation d'étiquette haut-parleur (Speaker-listener Label Propagation Algorithm SLPA) est une extension de l'algorithme de propagation d'étiquettes (Label Propagation Algorithm LPA) proposé par Xie *et al.* (2011). Alors que dans LPA, chaque noeud ne contient qu'une seule étiquette qui est mise à jour de manière itérative en adoptant l'étiquette majoritaire dans le voisinage, dans SLPA, chaque noeud possède une mémoire contenant plusieurs étiquettes. En partant d'un noeud sélectionné comme écouteur, ses voisins envoient une étiquette suivant certaines règles vocales. L'auditeur sélectionne une étiquette selon une règle d'écoute et l'ajoute à sa mémoire. Une fois que tous les noeuds ont été visités, les communautés sont extraites de la mémoire du noeud converties en une distribution de probabilité d'étiquettes qui définit le degré d'appartenance aux communautés.

Estimation démocratique de l'organisation modulaire d'un réseau (Democratic Estimate of the Modular Organization of a Network DEMON) proposé par Coscia *et al.* (2012) a tendance à affecter un noeud à la communauté la plus fréquente par l'application d'un algorithme de propagation d'étiquette sur ses sous-graphes voisins. En d'autres termes, pour chaque noeud, leurs voisins votent pour son appartenance à la communauté. Tous les votes sont ensuite combinés pour construire la structure communautaire qui se chevauche.

2.5.1.3 Mesures d'évaluation

Dans le but d'évaluer la performance des différents algorithmes de détéctions, on utilise deux types de mesures. D'une part, les mesures de qualité tendent à répondre à la

question : Qu'est-ce qu'une bonne structure communautaire ? Ils sont généralement basés sur les propriétés locales des communautés. La connaissance de l'appartenance à la communauté de vérité terrain n'est pas nécessaire dans ce cas. La modularité est parmi les mesures de qualité les plus utilisées pour quantifier la qualité d'un partitionnement. Elle a été introduite par Newman (2006b) afin de formuler le fait qu'un sous-graphe est une communauté si le nombre de connexions entre ses noeuds est supérieur à ce qui serait attendu si des liens étaient assignés au hasard. Elle est décrite comme la proportion d'arêtes incidentes sur un sous-graphique donné moins le nombre d'arêtes disposées de manière aléatoire sur le même sous-graphique. Une modularité élevée signifie que les connexions des noeuds au sein des communautés sont plus denses que celles entre les noeuds dans différents modules. La modularité d'un partitionnement c (où c_i indique la classe attribuée au noeud i), s'exprime de la manière suivante :

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i * k_j}{2m} \right] \delta(c_i, c_j) \quad (2.19)$$

où A_{ij} est la valeur de la matrice d'adjacence entre les sommets i et j , k_i est la somme des arêtes adjacentes à i , m est le nombre d'arêtes du graphe et δ est le delta de Kronecker qui vaut 1 si ses arguments sont égaux et 0 sinon.

D'autre part, les métriques de clustering comparent les communautés découvertes par les algorithmes à celles de la vérité terrain. De nombreuses mesures ont été proposées dans la littérature. L'information mutuelle normalisée (Normalized Mutual Information NMI) est utilisé dans ce travail. Cette mesure est couramment utilisée dans la littérature de la détection des communautés. Elle a été définie dans le contexte du clustering classique Hubert et Arabie (1985) pour comparer deux partitions différentes d'un même ensemble de données, en mesurant la quantité d'informations qu'ils ont en commun. Danon *et al.* (2005) l'ont utilisé pour évaluer les performances des algorithmes de détection communautaire, et la mesure a ensuite été utilisée par divers autres auteurs Lancichinetti *et al.* (2008). Si les communautés estimées correspondent parfaitement aux communautés de référence, la mesure prend la valeur 1, alors qu'elle est nulle lorsqu'elles sont indépendantes.

Dans cette section nous avons défini la structure communautaire des réseaux complexes ainsi que quelques algorithmes populaires de différentes catégories pour les détecter. Nous présentons dans la prochaine section les stratégies d'immunisation les plus récentes proposées pour les réseaux dotés d'une structure communautaire avec et sans chevauchement.

2.5.2 Stratégies stochastiques

Les stratégies stochastiques sont totalement agnostiques quant à la topologie du réseau. Certaines méthodes profitent de la structure communautaire des réseaux en utilisant l'information locale au niveau des noeuds. Des chercheurs supposent que les ponts (bridge nodes) sont les épandeurs les plus influents. Ces noeuds peuvent propager l'épidémie à l'ensemble du réseau en raison de leur connectivité avec divers modules. Ils ont alors une grande influence globale sur l'ensemble du réseau. Par conséquent, leur immunisation peut achever la diffusion effective vers les différentes parties du réseau. La stratégie Community Bridge Finder *CBF* Salathé et Jones (2010) est une stratégie d'immunisation visant à cibler les noeuds qui jouent le rôle des ponts entre les communautés. Elle est basée sur un algorithme à marche aléatoire. D'un autre côté, les hubs communautaires auraient également un impact important sur les noeuds appartenant à leurs mêmes modules. Ainsi, ils ont une énorme influence locale dans leurs communautés. À cette fin, le Degree Community Bridge Finder *DCBF* Gong (2014) et le Bridge-Hub Detector *BHD* Gong *et al.* (2013) sont deux stratégies d'immunisation, qui ciblent les noeuds de pont avec des connexions élevées pour l'immunisation. Les ponts sélectionnés, dans ce cas, jouent également le rôle de hubs. La première stratégie est une variation du *CBF*, tandis que la dernière est basée sur l'élargissement des cercles d'amitié lors d'une marche aléatoire. D'autres chercheurs tentent de mettre en évidence l'importance des noeuds qui se chevauchent en termes de dynamique de propagation de l'épidémie. La stratégie Random-Walk Overlap Selection *RWOS* Taghavian *et al.* (2017) est proposée pour sélectionner les noeuds qui se chevauchent selon un algorithme aléatoire. Ces noeuds clés peuvent jouer un rôle majeur dans la diffusion des épidémies en raison de leur appartenance à plusieurs communautés. Dans cette section, nous présentons un bref aperçu des trois stratégies stochastiques basées sur une

structure de communauté sans chevauchement et celle adaptée aux réseaux dotés d'une structure communautaire qui se chevauchent (voir Figure 7).

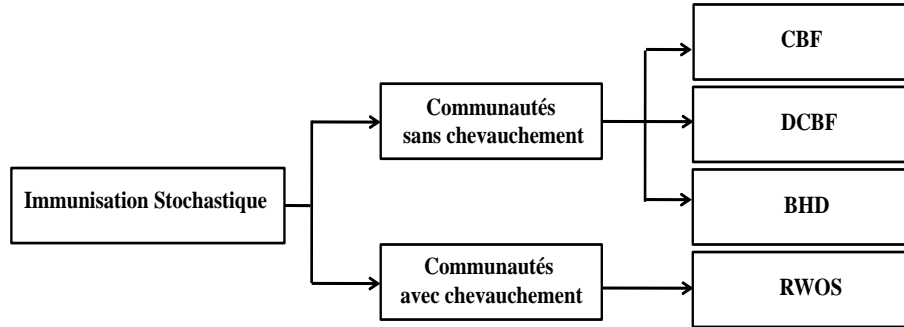


FIGURE 2.4 – Méthodes d'immunisation stochastiques.

Community bridge finder (CBF) : Les interventions d'immunisation d'individus hautement connectés ne suffisent pas toujours à protéger les réseaux contre les épidémies à grande échelle. En effet, cibler des individus reliant des communautés est parfois plus efficace que de simplement immuniser des noeuds avec une grande quantité de connexions. L'objectif de la stratégie CBF Salathé et Jones (2010) est d'identifier les noeuds agissant comme des ponts entre les communautés. Cette stratégie est basée sur des marches aléatoires. Cela fonctionne comme suit :

Étape 1. Sélectionnez un noeud aléatoire $v_{i=0}$.

Étape 2. Suivez une marche aléatoire avec la condition qu'un noeud n'a pas été visité par le chemin aléatoire auparavant.

Étape 3. À chaque noeud $v_{i \geq 2}$, vérifiez s'il est connecté à plusieurs noeuds visités. S'il n'y a qu'une seule connexion, v_{i-1} est considéré comme un pont potentiel.

Étape 4. Sélectionnez deux noeuds voisins aléatoires de v_i autres que v_{i-1} . Si les deux noeuds n'ont aucune connexion avec les noeuds précédemment visités, le noeud v_{i-1} est alors marqué comme un pont et il est immunisé. sinon, une marche aléatoire est reprise à v_{i-1} .

Cette stratégie a été comparée à la stratégie de connaissance (acquaintance). Des expériences approfondies ont été menées sur des réseaux synthétiques et du monde réel en utilisant le modèle épidémique SIR. Les résultats montrent que CBF surpasse principa-

lement la stratégie de connaissance. Ses meilleures performances sont obtenues dans des réseaux à forte structure communautaire.

Degree Community Bridge Finder (DCBF) DCBF Gong (2014) est une variante de la stratégie CBF. Le but de cette stratégie est de cibler les ponts avec un plus grand nombre de connexions. Comme CBF, c'est une stratégie basée sur la marche aléatoire. De plus, il réduit le temps d'exécution du CBF en ajoutant deux vérifications supplémentaires aux marches aléatoires. Cette stratégie suit les mêmes étapes décrites dans l'algorithme CBF. La différence est que les noeuds ne sont pas choisis au hasard parmi tous les noeuds possibles lors de la marche aléatoire, mais en fonction de leur degré de haut en bas. Deux contrôles supplémentaires sont également implémentés dans DCBF pour diminuer le temps de calcul de l'algorithme. Dans un premier temps, le nombre de noeuds visités dans un chemin en cours d'exécution est maintenu à une longueur de dix. De plus, le nombre de visites par tous les chemins aléatoires est enregistré pour chaque noeud. Le noeud est immunisé lorsque le nombre de visites k est égal à un certain nombre ($k = 2$). DCBF est testé sur des réseaux générés avec différentes modularités. Après avoir exécuté les simulations du modèle épidémique SIR, les résultats démontrent que le DCBF fonctionne mieux que l'algorithme CBF pour contrôler les épidémies. Ses performances augmentent dans les réseaux à forte structure communautaire (lorsque la modularité est très élevée $Q > 0,84$). Dans ce cas, les épidémies sont limitées localement à l'intérieur des communautés. DCBF pourrait alors cibler des noeuds hautement connectés dans les communautés locales, tandis que CBF est capable d'identifier uniquement les noeuds de pont.

Bridge-Hub Detector (BHD) Les communautés se caractérisent par l'hétérogénéité des connexions entre les noeuds reliant différentes communautés. Sur la base de cette idée, BHD Gong *et al.* (2013) vise à identifier les ponts hub comme cibles pour l'immunisation. Le Bridge-Hub Detector est une autre alternative à la stratégie CBF. Il est basé sur l'expansion des cercles d'amitié des noeuds visités. La stratégie BHD fonctionne comme suit :

Étape 1. Sélectionnez un noeud aléatoire $v_{i=0}$.

Étape 2. Suivez une marche aléatoire avec la condition qu'un noeud n'a pas été visité par le chemin aléatoire précédent.

Étape 3. Soit $v_{i>=2}$ le noeud visité après i étapes, et f_i est l'ensemble de ses voisins. Le noeud v_i est considéré comme une cible d'immunisation s'il y a au moins un noeud qui ne fait pas partie de l'ensemble F_{i-1} et qu'il n'est lié à aucun noeud dans F_{i-1} , où $F_{i-1} = f_0 \cup f_1 \cup f_2 \cup \dots \cup f_{i-1}$. Sinon, la marche aléatoire passe de v_i et le cercle d'amitié sera mis à jour à $F_i = F_{i-1} \cup f_i$.

Étape 4. Parmi les noeuds de f_i , un noeud v_H qui n'appartient pas et ne peut pas être reconnecté à F_{i-1} est choisi au hasard pour l'immunisation.

À la fin de cette procédure, une paire de noeuds, un pont et un pont hub sont sélectionnés pour l'immunisation. Cela est via des cercles d'amitié de noeuds visités aléatoirement. BHD a été testé sur des données simulées et empiriques construites à partir du réseau Facebook de cinq universités américaines en utilisant le modèle SIR. Il en résulte une taille épidémique plus petite par rapport aux stratégies Connaissance et CBF. En termes de complexité de calcul, l'Acquaintance est l'algorithme le plus rapide, suivi par CBF et BHD est le plus lent.

Random-Walk Overlap Selection (RWOS) Les noeuds qui se chevauchent n'ont pas nécessairement des valeurs de centralité élevées. Pourtant, ils peuvent avoir un effet majeur sur la propagation des épidémies d'une communauté à l'autre. Ces noeuds ont accès à plusieurs communautés du réseau en raison de leur grande membership (nombre de communautés auxquelles ils appartiennent). La stratégie RWOS Taghavian *et al.* (2017) est conçue pour cibler les noeuds qui se chevauchent pour l'immunisation selon une marche aléatoire. Il peut être spécifié comme suit :

Étape 1. Exécutez un algorithme de détection de communauté locale pour découvrir les communautés qui se chevauchent dans le réseau.

Étape 2. Définissez la liste des noeuds qui se chevauchent à partir des communautés extraites.

Étape 3. Sélectionnez au hasard un noeud du réseau et exécutez une marche aléatoire.

Étape 4. Chaque noeud visité est désigné comme cible pour l'immunisation s'il appartient à l'ensemble de noeuds qui se chevauchent. Ce processus se poursuit jusqu'à atteindre la couverture d'immunisation souhaitée.

Cette stratégie cible les noeuds chevauchants hautement connectés pour l'immunisation. Elle est basé sur l'idée que la probabilité de visiter n'importe quel noeud dans un chemin aléatoire est proportionnelle au degré du noeud. Le RWOS a été appliqué sur des réseaux synthétiques et réels. Le modèle épidémique SIR standard a été exécuté sur ces réseaux. Les résultats montrent que RWOS surpasse les stratégies CBF et BHD en termes de taille épidémique. Elle fonctionne parfois encore mieux que la stratégie membership (où les noeuds sont immunisés en fonction du nombre de communautés auxquelles ils appartiennent). De plus, ses performances s'améliorent dans les réseaux dotés d'une structure communautaire solide et de valeurs de membership plus élevées.

2.5.3 Stratégies déterministes

Les stratégies déterministes nécessitent la topologie de l'ensemble du réseau. Nous donnons maintenant un aperçu de certaines stratégies déterministes récentes conçues pour les réseaux modulaires. Elles sont classées en différentes catégories en fonction de leurs objectifs d'immunisation (voir Figure 5).

2.5.3.1 *Structure communautaire sans chevauchement*

Une pléthore de stratégies d'immunisation déterministes sont développées pour identifier les noeuds vitaux dans les réseaux à structure communautaire. Elles peuvent être classés en trois catégories dans les réseaux avec structure non chevauchante. Le premier type de stratégies met en évidence des noeuds avec des connexions externes vers des communautés étrangères. Elles ciblent les noeuds de pont, ce qui peut avoir une influence globale significative sur les autres noeuds du réseau. La deuxième catégorie tend à identifier les noeuds ayant la plus forte influence locale dans leurs propres communautés. Certaines stratégies ciblent les hubs pour l'immunisation en raison de leur grande influence sur les noeuds de leurs voisinages, tandis que d'autres immunisent les noeuds situés au coeur de la communauté. De plus, le groupe de stratégies appartenant à la troisième catégorie

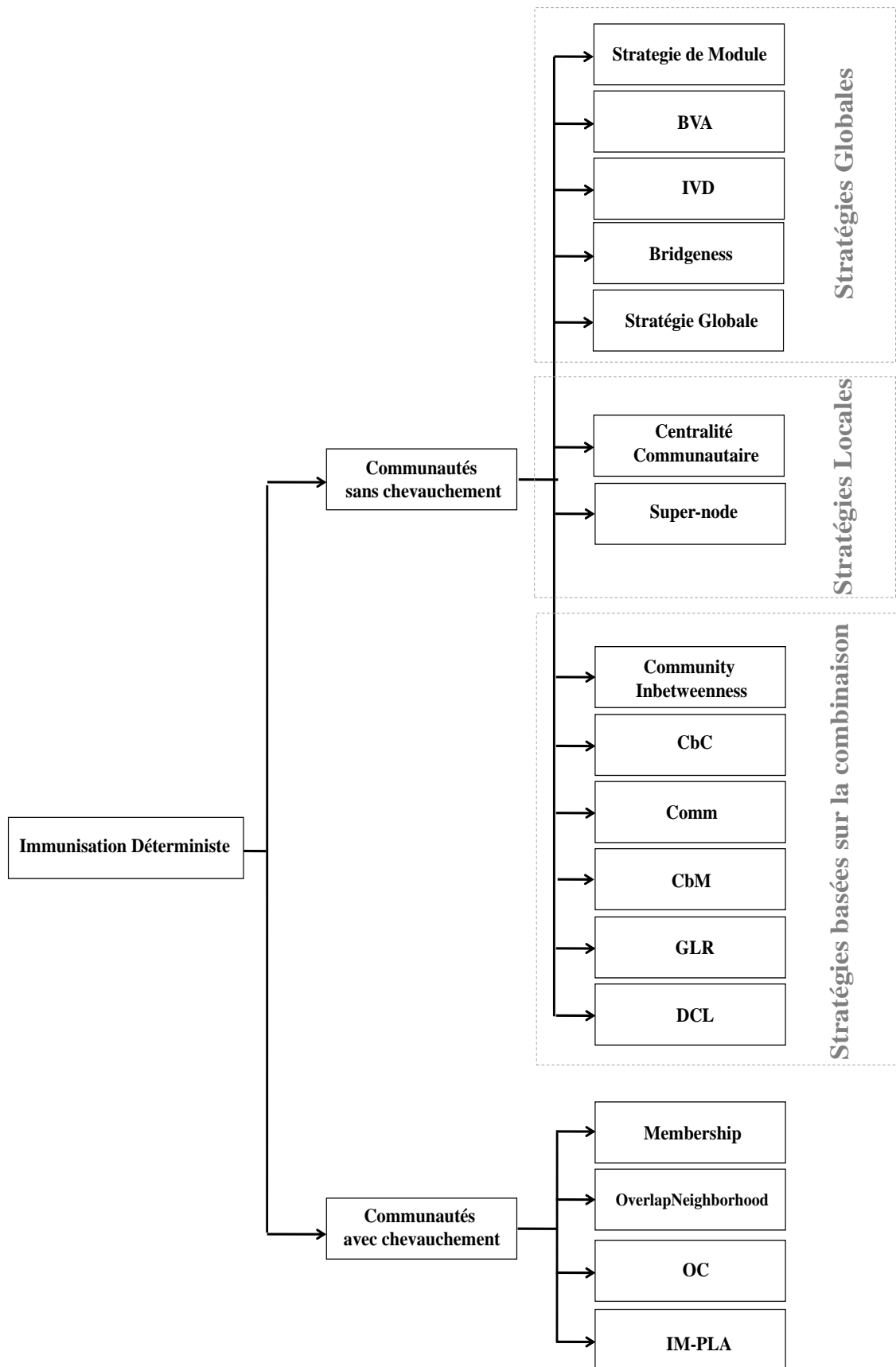


FIGURE 2.5 – Méthodes d'immunisation déterministes.

immunise les deux types de noeuds. Elles sélectionnent des noeuds ayant une influence à la fois locale et globale sur l'ensemble du réseau. Les méthodes déterministes présentées sont basées sur différentes approches. Certaines stratégies n'ont pas besoin de l'étiquette de communauté des noeuds tandis que d'autres utilisent différents niveaux d'informations sur la structure de la communauté.

a- Stratégies globales

Les ponts peuvent être considérés comme des individus qui connectent différents sous-groupes de noeuds dans les réseaux. Ils peuvent laisser les épidémies se déplacer d'un module à l'autre grâce à leurs connexions inter-communautaires. Par conséquent, ils ont une influence globale majeure sur l'ensemble du réseau. Une série de stratégies ont été proposées pour sélectionner ces noeuds critiques pour l'immunisation. La stratégie de Module (Module-based strategy *Mod strategy*) Masuda (2009) est proposée pour mettre en évidence les noeuds de pont entre les communautés. Elle est basé sur un calcul approximatif de la centralité des vecteurs propres du réseau à gros grains (appelé aussi méta-graphe). Où les communautés sont représentées simplement par des noeuds, et les connexions sont pondérées par le nombre de liens entre deux communautés. Différent de la méthode précédente, Mantzaris Mantzaris (2014) a proposé l'algorithme de voisinage de frontière (Boundary Vicinity Algorithm *BVA*). Il cible les noeuds d'immunisation en fonction de leur proximité avec les noeuds de pont. Ces deux méthodes peuvent être spécifiées comme suit :

Stratégie de Module (Mod strategy) : La stratégie de Module a été proposée par Masuda (2009). Compte tenu de la structure communautaire du réseau d'origine, cette stratégie est appliquée au réseau à gros grains. Où chaque communauté est représentée par un seul noeud, et les liens sont pondérés par le nombre de liens partagés par deux communautés voisines. Elle cible les noeuds en maximisant la mesure suivante :

$$Mod_i = 2\tilde{u}_K \sum_{I \neq K} d_{kI} \tilde{u}_I \quad (2.20)$$

Où \tilde{u}_K représente le vecteur propre correspondant à la K^{th} communauté. d_{kI} est le nombre de liens inter-communautaires qui existent entre le noeud k et la I^{th} communauté. Le premier terme de cette mesure (c.-à-d. $2\tilde{u}_K$) quantifie l'importance de la communauté à laquelle appartient le noeud k , tandis que la deuxième quantité (c.-à-d. $\sum_{I \neq K} d_{kI} \tilde{u}_I$) mesure sa connectivité avec d'autres communautés importantes. Après avoir immunisé tous les noeuds de pont, les noeuds restants sont classés en fonction de leur degré. Cette méthode cible préférentiellement les noeuds d'importance globale ayant des liens inter-communautaires importants plutôt que les hubs communautaires qui sont localement importants. L'efficacité de la stratégie de Module est testée en l'appliquant sur des réseaux générés et réels de diverses natures. Les résultats montrent qu'elle est dans la plupart des cas plus efficace que la stratégie Degree, Betweenness et Ress (une stratégie basée sur les vecteurs propres Restrepo *et al.* (2008)) dans les réseaux à structure modulaire.

Algorithme de voisinage de frontière (Boundary Vicinity Algorithm BVA) :

Il s'agit d'une autre stratégie globale proposée par Mantzaris *cite bva*. Cette stratégie classe les noeuds en fonction de leur proximité aux ponts (noeuds situés à la frontière) de chaque communauté. Il suit les étapes suivantes :

Étape 1. Définissez l'ensemble des communautés du réseau.

Étape 2. Extraire l'ensemble de ponts qui relie les communautés.

Étape 3. Exécutez un certain nombre de marches aléatoires d'un nombre de pas fixe choisi à partir de chaque pont. Ensuite, le nombre de visites à chaque noeud est compté.

Cette mesure quantifie la capacité d'un noeud donné à propager des épidémies à travers des ponts vers différentes communautés. En utilisant le modèle épidémique SI, la stratégie *BVA* surpasse la stratégie basée sur l'intermédiarité en termes de taille de l'épidémie.

Yoshida et Yamada (2017) a proposé la densité vectorielle inverse (Inverse Vector Density *IVD*). Il s'agit d'une autre stratégie d'immunisation qui ne nécessite pas les étiquettes communautaires des noeuds. Cette stratégie commence par la construction d'une représentation vectorielle des noeuds basée sur la mesure de la modularité. L'*IVD* immunise les

noeuds avec un petit nombre de vecteurs de noeuds voisins qui sont identifiés comme des ponts. Les résultats ont montré que cette stratégie est plus performante que la stratégie basée sur l'intermédiarité. La stratégie de Bridgeness est une autre méthode proposée par Jensen *et al.* (2015). Elle est basée sur la centralité de l'intermédiarité en ne considérant que les chemins les plus courts entre les noeuds appartenant à différentes communautés. Cette stratégie met en évidence les noeuds qui connectent différentes régions d'un réseau. En utilisant à la fois les réseaux synthétiques et réels, la stratégie de Bridgeness est globalement plus efficace que la stratégie basée sur l'intermédiarité pour identifier les ponts. Kitromilidis et Evans (2018) propose de redéfinir les mesures traditionnelles de centralité pour caractériser l'influence des artistes occidentaux. La stratégie globale (Global strategy) est basée sur le calcul des mesures de centralité standard en considérant uniquement les liens inter-communautaires des réseaux. Leur idée est basée sur le fait que les artistes influents ont généralement des liens au-delà de leur mouvement artistique. Les stratégies d'intermédiarité et de proximité globale sont comparées à leurs versions classiques. Elles ont été testées sur un réseau de collaboration de peintres. Les résultats expérimentaux montrent que les stratégies globales permettent de mettre en évidence certains noeuds influents qui auraient pu être manqués car ils ne se classent pas nécessairement en bonne place dans les stratégies basées sur des mesures standard.

b- Stratégies locales

Les hubs représentent les noeuds de haut degré avec une plus grande quantité de connexions qui dépassent largement la moyenne. Ils sont une propriété conséquente des réseaux sans échelle (scale-free networks). Dans les réseaux présentant la propriété modulaire, l'existence de tels noeuds peut être constatée dans chaque communauté. Ils ont alors une grande influence locale dans leurs propres communautés. Newman a proposé la centralité communautaire (Community centrality) Newman (2006a) pour identifier les noeuds qui jouent un rôle central au sein des communautés en termes de nombre de connexions. Ces noeuds sont responsables du flux maximal d'informations au sein de leurs communautés. He *et al.* (2015) a proposé la stratégie Super node qui peut immuniser les noeuds avec les liens intra-communautaires les plus élevés (ou avec l'indice k-core le plus élevé)

appartenant à diverses communautés. Les deux stratégies sont décrites comme suit :

Centralité communautaire (Community centrality) (CC) : Newman a proposé une formulation légèrement différente de la modularité. La centralité communautaire (Community centrality) Newman (2006a) est dérivée des vecteurs propres de la matrice de modularité. Où la matrice de modularité est divisée en deux projections. La première dimension représente les vecteurs propres positifs de la matrice de modularité tandis que la deuxième dimension représente les vecteurs négatifs. Ainsi, la modularité peut être écrite en termes de ces vecteurs comme suit :

$$Q = \sum_{k=1}^c |X_k|^2 - \sum_{k=1}^c |Y_k|^2 \quad (2.21)$$

Où c est le nombre de communautés. X et Y sont les vecteurs propres de la communauté dans les deux dimensions. Le i^{th} noeud dans la communauté k est représenté par deux vecteurs x_i et y_i (les i^{th} lignes de X_k et Y_k respectivement).

La magnitude d'un vecteur de noeud $|x_i|$ spécifie à quel point le noeud i est central dans sa communauté en termes de nombre de connexions. Ainsi, le noeud i a une grande contribution positive à la modularité lorsque cette mesure est grande. D'un autre côté, une valeur plus élevée de $|y_i|$ signifie que le noeud i a de nombreuses connexions à d'autres noeuds de communautés étrangères. Par conséquent, la centralité de la communauté est définie comme étant égale à la magnitude vectorielle $|x_i|$. Elle mesure la force avec laquelle un noeud donné i est affecté à sa communauté. Cette mesure a été testée dans le réseau de co-auteurs entre scientifiques. Les résultats montrent qu'il n'est pas très corrélé avec la centralité des degrés. De plus, certains noeuds avec une mesure de centralité communautaire élevée ont un degré relativement faible. Cependant, ils ont plus de connexions avec les noeuds de leurs communautés. Ainsi, les noeuds à forte valeur de centralité communautaire jouent un rôle central dans le processus de diffusion dans leur quartier local.

Super node : Cette stratégie commence par classer les communautés par ordre décroissant en fonction de leur taille. Après cela, le noeud avec le plus grand degré interne

(in-degree) est sélectionné dans la plus grande communauté. Ensuite, le noeud avec le degré interne le plus élevé dans la deuxième plus grande communauté et qui n'a pas de connexion avec les communautés précédentes est sélectionné comme deuxième écarteur. Notez qu'il n'y a qu'une seule communauté précédente pour le deuxième épandeur. Après avoir visité toutes les communautés du réseau, ce processus est redémarré jusqu'à atteindre le nombre souhaité de noeuds immunisés. Le but de cette méthode est de sélectionner plusieurs épandeurs de différentes communautés de manière équilibrée. Les simulations du modèle SIR sont effectuées dans des réseaux synthétiques et réels. Les résultats expérimentaux montrent que la stratégie Super node entraîne une taille épidémique plus petite que la stratégie basée sur les degrés. De plus, la stratégie Super node a également prouvé son efficacité lors de l'utilisation de la méthode de décomposition k-shell dans le processus de recherche des épandeurs influents dans chaque communauté.

c- Stratégies globales et locales

Les stratégies d'immunisation dans cette catégorie ont tendance à cibler les noeuds qui ont une influence à la fois locale et globales. Elles combinent les différents aspects des stratégies précédentes pour sélectionner les noeuds les plus influents du réseau. Ces noeuds sont les principaux diffuseurs de leurs communautés qui peuvent également diffuser les épidémies vers d'autres modules du réseau. La stratégie Community Inbetweenness Chan *et al.* (2009) ainsi que la stratégie *CbC* Zhao *et al.* (2015) sélectionnent les noeuds Hub-bridge pour l'immunisation. Elles peuvent être définis comme suit :

Community Inbetweenness : L'intermédialité (betweenness) classique doit résoudre le problème de chemin le plus court de toutes les paires, ce qui le rend impossible dans les grands réseaux. La stratégie Community Inbetweenness Chan *et al.* (2009) est proposée pour résoudre ce problème. Elle est basé sur une mesure basée sur l'entropie qui représente une approximation de la centralité de l'intermédialité. Elle classe les noeuds uniquement sur la base des informations de la communauté. Cette stratégie évalue l'importance du noeud en fonction de la proportion de ses liens environnants ou interne en plus des liens externes le connectant avec des communautés étrangères. La centralité de Community

Inbetweenness C_{CI} est définie comme suit :

$$C_{CI}(i) = k_i \sum_{c \in C} p_{i \rightarrow c} \log \left(\frac{1}{p_{i \rightarrow c}} \right) \quad (2.22)$$

Où k_i est le degré du noeud i . $p_{i \rightarrow c}$ est la proportion de liens reliant le noeud i à la communauté $c \in C$. C est l'ensemble des communautés qui ne se chevauchent pas. La Community Inbetweenness tend à sélectionner des noeuds avec une connectivité élevée et avec plus de liens vers différentes communautés. Il est basé sur l'idée que les noeuds avec une mesure d'intermédiarité élevée sont généralement situés entre des modules densément connectés. Ces noeuds sont également ciblés par la centralité d'intermédiarité standard. Les résultats de simulation sur des réseaux réels montrent que cette stratégie est plus efficace que la stratégie basée sur l'intermédiarité en terme de performances de calcul. Les deux stratégies sont également testées avec le modèle SIR dans Ghalmane *et al.* (2018) pour comparer leur taille épidémique. Les résultats montrent que la Community Inbetweenness fonctionne presque aussi bien que l'intermédiarité dans les réseaux à forte structure communautaire. Il est cependant plus efficace dans les réseaux à structure communautaire lâche.

Community-based Centrality (CbC) : Cette stratégie sélectionne les noeuds pour l'immunisation en fonction de leurs caractéristiques de liens et de la taille de leurs communautés. Elle cible les noeuds qui ont un grand impact dans leurs communautés et qui peuvent aussi propager des épidémies aux noeuds d'autres communautés. Il est basé sur une mesure qui évalue l'importance du noeud i via la formule suivante :

$$CbC_i = \sum_{c=1}^m d_{ic} \frac{S_c}{N} \quad (2.23)$$

Où d_{ic} est le nombre de liens entre le noeud i et les autres noeuds de la communauté c , m est le nombre de communautés dans le réseau, S_c est le nombre de noeuds dans communauté c et N est la taille du réseau. Les résultats de simulation utilisant le modèle SIR montrent que CbC surpasse certaines mesures traditionnelles telles que le degré et le K-shell. De plus, CbC peut également mieux refléter l'importance des noeuds par rapport

aux centralités de proximité, d'intermédiarité et des vecteurs propres, avec une complexité de calcul beaucoup plus faible.

Stratégie Comm : La stratégie Comm (Comm strategy) a été proposée par Gupta *et al.* (2016). Le but de cette stratégie est de cibler des noeuds qui sont à la fois des hubs dans leurs communautés et des ponts vers d'autres communautés. Elle classe les noeuds selon une mesure basée sur les degrés. Cette mesure est une combinaison pondérée du nombre de liens intra-communautaires et du nombre de liens inter-communautaires élevés à la puissance de deux. Il donne alors plus d'importance aux ponts. Les résultats sur les réseaux synthétiques et réels montrent que la stratégie Comm est plus efficace ou au moins fonctionne ainsi que la stratégie Module (Mod strategy), les stratégies basées sur le degré et l'intermédiarité en termes de propagation épidémique.

Community-based Mediator (CbM) : Cette méthode d'immunisation est basée sur la mesure du Community-based Mediator (CBM) Tulu *et al.* (2018). L'idée derrière cette stratégie est que si un individu a de nombreux liens dans plusieurs communautés, il peut alors jouer un rôle important pour diffuser des informations autour de son cercle. Cette méthode sélectionne les noeuds les plus intermédiaires qui reçoivent et diffusent des informations à travers les communautés que les autres noeuds. Il combine l'influence du degré et de l'intermédiarité des noeuds du réseau. La mesure CBM est définie comme suit :

$$CbM_i = H_i \frac{d_i}{\sum_{i=1}^N d_i} \quad (2.24)$$

Où H_i est l'entropie du noeud i . Il est défini par la formule suivante :

$$H_i = \left[- \sum \rho_i^{in} \log(\rho_i^{in}) \right] + \left[- \sum \rho_i^{ex} \log(\rho_i^{ex}) \right] \quad (2.25)$$

Où ρ_i^{in} représente la fraction des liens connectés à i à l'intérieur de sa communauté, tandis que ρ_i^{ex} indique la fraction des liens sortants du noeud i vers des noeuds appartenant à d'autres communautés. L'entropie est utilisée pour trouver des noeuds qui ont un équilibre entre la capacité de diffuser les informations dans le réseau. Les résultats

expérimentaux ont montré que les noeuds avec une valeur CbM élevée ont un impact plus important sur la diffusion des informations dans le réseau que les noeuds ayant une valeur élevée du Degré, intermédiarité, CbC, PageRank ou Eigenvector.

Gateway Local Rank (GLR) : Salavati *et al.* (2019) ont proposé une version améliorée de la stratégie basée sur la proximité conçue pour les réseaux modulaires. Elle diminue également la grande complexité de calcul de la méthode de proximité standard. La stratégie dite Gateway Local Rank GLR commence par ignorer les connexions entre les communautés. Ensuite, dans chaque communauté, un noeud critique est extrait à l'aide de la centralité d'intermédiarité. Après cela, le noeud avec le plus grand nombre de liens inter-communautaires est aussi extrait de chaque communauté. Dans la dernière étape, les noeuds sont classés en fonction de la somme de leurs chemins les plus courts avec les noeuds extraits (le noyau et le pont de chaque communauté) au lieu de calculer leurs chemins les plus courts en utilisant tous les noeuds du réseau. Des expériences sur des réseaux synthétiques et du monde réel utilisant le modèle de diffusion SIR démontrent l'efficacité de la stratégie GLR en comparaison avec les stratégies basées sur la proximité, le degré, l'intervalle et k-shell.

Degree and Clustering coefficient and Location (DCL) : Berahmand *et al.* (2018) a proposé la stratégie du degré et le coefficient de clustering et de localisation (Degree and Clustering coefficient and Location DCL). Elle immunise les meilleurs épandeurs sur la base d'une combinaison du degré et du coefficient de cluster inverse d'un noeud donné. Ces deux dernières mesures sont également combinées avec le degré de ses voisins et les liens communs entre le noeud et ses voisins pour définir l'emplacement d'un noeud (qu'il soit au coeur ou à la périphérie de la communauté). Cette stratégie est également capable d'identifier des ponts à faible degré et certains noeuds de concentrateur critiques. Les comparaisons basées sur les modèles SIR et SI révèlent que la méthode proposée surpasse les stratégies bien connues telles que le degré, l'intermédiarité, les vecteurs propres, le pagerank et les stratégies basées sur k-shell.

2.5.3.2 Structure communautaire avec chevauchement

Les communautés peuvent souvent se chevaucher dans des réseaux du monde réel. Dans ce cas, les noeuds peuvent appartenir à plusieurs communautés à la fois. L'identification de ces noeuds qui se chevauchent est cruciale pour contrôler la propagation de l'épidémie. Ces noeuds peuvent étendre la diffusion de l'épidémie dans toutes les communautés auxquelles ils appartiennent. Certaines stratégies sélectionnent ces noeuds pour l'immunisation. Hébert-Dufresne *et al.* (2013) a proposé la stratégie du membership. C'est une stratégie simple qui compte directement le degré d'appartenance de chaque noeud du réseau. OverlapNeighborhood (ON) Kumar *et al.* (2018) est une autre stratégie qui cible les voisins des noeuds qui se chevauchent pour l'immunisation. Elle est basé sur l'idée que les noeuds qui se chevauchent sont connectés avec de nombreux hubs situés dans les différentes communautés auxquelles ils appartiennent. Les stratégies membership et l'OverlapNeighborhood sont définies comme suit :

Membership (appartenance) : Cette stratégie Hébert-Dufresne *et al.* (2013) est appliquée aux réseaux dont la structure modulaire se chevauche. Il est basé sur une mesure qui compte simplement le nombre de communautés auxquelles appartient un noeud. Si le degré d'appartenance d'un noeud i est supérieure à 1, cela veut dire que ce noeud appartient à une région chevauchante du réseau. Les résultats expérimentaux utilisant le modèle SIR ont montré que cette stratégie surpasse les stratégies basées sur le degré, la coreness et l'intermédiarité dans les réseaux avec des communautés plus denses et en utilisant un taux d'infection plus élevé.

OverlapNeighborhood (ON) : Cette méthode Kumar *et al.* (2018) sélectionne les voisins immédiats des noeuds qui se chevauchent comme les principaux diffuseurs influents. Son objectif principal est de sélectionner les noeuds les plus connectés en utilisant une quantité limitée d'informations au niveau de la communauté. En effet, il y a une forte probabilité que les noeuds avec des connexions très élevées soient voisins des noeuds qui se chevauchent car ils font partie de plus d'une communauté. Cela est également dû à la distribution des degrés de loi de puissance dans les réseaux du monde réel. Les résultats

des simulations ont révélé que cette méthode surpasse les méthodes CBF, BHD et RWOS. Elle fonctionne mieux ou aussi bien que les méthodes basées sur la centralité du degré et de l'intermédiarité en utilisant moins d'informations sur la topologie du réseau global.

Overlapping constraint coefficient (OC) : La stratégie Overlapping constraint coefficient (*OC*) Wei *et al.* (2018) est une stratégie d'immunisation qui met en évidence les noeuds influents sur la base de la multiplication de deux mesures. La première mesure représente la membership d'un noeud donné qui quantifie sa capacité de propagation. Ainsi, plus un noeud appartient à des communautés, plus il peut influencer des communautés. La deuxième mesure représente le coefficient de contrainte réseau (network constraint coefficient) du noeud, qui quantifie sa vitesse de propagation dans les communautés. Les simulations SIR démontrent que la stratégie Overlapping constraint coefficient peut obtenir les meilleurs résultats par rapport aux stratégies basées sur le degré, l'interdépendance, la proximité et le k-shell.

Influence Maximization-based on Label Propagation Algorithm (IM-LPA) : La stratégie Influence Maximization based on Label Propagation Algorithm (*IM-LPA*) Zhao *et al.* (2016) est une autre méthode conçue pour les réseaux avec des communautés qui se chevauchent. Elle est basée sur une version améliorée de la méthode Label Propagation Algorithm (*LPA*) Raghavan *et al.* (2007). Elle est basée sur deux phases : la phase d'ensemencement et la phase de propagation des étiquettes. Au début de la phase d'ensemencement, l'ensemble des noeuds graines est vide et tous les noeuds du réseau sont considérés comme des noeuds candidats. Après cela, le noeud avec le degré le plus élevé est ajouté à cet ensemble et tous ses voisins sont supprimés de l'ensemble de noeuds candidat. Ce processus est répété jusqu'à ce que l'ensemble de noeuds candidats devient vide. Cette phase garantit que les noeuds graines sélectionnés sont indépendants les uns des autres. Dans la phase de propagation d'étiquettes, chaque noeud graine est associé à une étiquette unique. Ensuite, les étiquettes se développent à partir des noeuds d'origine jusqu'à couvrir tous les autres noeuds du réseau. Les noeuds peuvent avoir plusieurs étiquettes. Ainsi, ils peuvent appartenir à plusieurs communautés. À la fin de ce processus, la centralité de

chaque noeud peut être mesurée par le nombre de noeuds associés à son étiquette. Les noeuds avec la mesure la plus élevée peuvent propager les épidémies à un large ensemble de noeuds de leurs communautés. Le modèle de diffusion en cascade indépendant (Independent cascade diffusion model) (*IC*) a été réalisé sur des réseaux synthétiques et réels. Les résultats démontrent l'efficacité de la stratégie *IM-LPA* pour identifier les épandeurs influents par rapport aux stratégies basées sur le degré, l'intermédiarité, la proximité, le k-shell et le pagerank.

2.6 Conclusion

Dans ce chapitre, la notion de stratégie d'immunisation dans les réseaux complexes a été introduite. Les modèles épidémiques qui servent à évaluer l'efficacité de diffusion de chaque stratégie sont aussi définis. Ensuite, un état de l'art sur les approches d'immunisation destinées aux réseaux avec et sans structure communautaire ont été également présentés. Nous pouvons distinguer deux approches principales pour résoudre le problème d'immunisation. La première est l'approche stochastique bénéfique lorsque l'on sait peu de choses sur la structure complète du réseau ou lorsque les réseaux sont trop grands pour calculer les caractéristiques de chaque noeud. Cependant, en général, la deuxième approche de l'utilisation des stratégies déterministes surpasse les stratégies stochastiques. Nous présenterons dans la partie suivante (c.f. chapitre 3), trois stratégies d'immunisation déterministes dédiées pour les réseaux avec structure communautaire, nous étudions aussi l'impact de la force de la structure communautaire sur le choix d'une stratégie d'immunisation appropriée. Cela est suivi d'une présentation de notre framework proposé afin d'adapter toutes les mesures de centralité classiques proposées pour les réseaux sans structure communautaire aux réseaux modulaires (c.f. chapitre 4 et 5).

STRATÉGIES D'IMMUNISATION POUR LES RÉSEAUX AVEC STRUCTURE COMMUNAUTAIRE

Sommaire

1.1	Contexte et problématique	1
1.2	Contributions de la thèse	4
1.3	Publications dans le cadre de la thèse	5
1.4	Organisation du manuscrit	6

3.1 Introduction

Les épidémies représentent une menace énorme pour la vie humaine car nous vivons dans un monde toujours plus connecté. L'immunisation par la vaccination est une solution qui protège les individus et les empêche de transmettre des maladies infectieuses à d'autres personnes vivant dans le même groupe social. Cependant, la vaccination de chaque individu dans la société peut s'avérer impossible dans les cas où le temps ou les ressources sont limités. Pour résoudre ce problème, les stratégies d'immunisation sont des techniques essentielles pour réduire les risques de flambées épidémiques. Elles visent à immuniser quelques noeuds clés pour atteindre efficacement l'objectif de réduction ou d'arrêt de la propagation des maladies infectieuses. Les stratégies d'immunisation peuvent être classées en deux catégories : les stratégies d'immunisation déterministes ou stochastiques. Les stratégies d'immunisation déterministes nécessitent la connaissance de l'ensemble du réseau, d'où leur efficacité. Elles consistent à classer les noeuds selon une mesure de centralité spécifique comme la centralité de degré ou d'intermédiarité. Les noeuds avec une

mesure de centralité élevée sont ciblés pour la vaccination. Les stratégies d'immunisation stochastiques constituent un autre groupe de méthodes d'immunisation. Elles sont plus ou moins agnostiques quant à la structure topologique des réseaux. Dans ces stratégies, les noeuds cibles sont trouvés via la recherche locale. Ils ont besoin d'informations uniquement au niveau du noeud pour trouver les noeuds ciblés pour l'immunisation.

La structure des réseaux est cruciale pour expliquer les schémas épidémiologiques. Au cours des dernières années, de nombreuses stratégies d'immunisation ont été développées en utilisant diverses propriétés topologiques du réseau afin d'atténuer et de contrôler les épidémies. Malgré le fait qu'il existe des preuves évidentes que de nombreux réseaux sociaux manifestent une structure communautaire bien définie Porter *et al.* (2009); Ferrara (2012); Jebabli *et al.* (2015, 2014), cette propriété a besoin de plus de considération. Un réseau avec une structure communautaire forte se compose de sous-groupes cohésifs de sommets qui partagent de nombreuses connexions avec les membres de leur groupe et peu de connexions avec des sommets en dehors de leur groupe. Les noeuds de pont sont ceux qui relient différentes communautés. Ils créent une voie de propagation de la maladie en dehors de leur communauté. Leur influence sur la propagation des épidémies a été particulièrement étudiée dans les travaux précédents Liu et Hu (2005); Gupta *et al.* (2015); Ghalmane *et al.* (2018); Chakraborty *et al.* (2016). En effet, l'immunisation de ces noeuds permet de confiner la maladie dans la communauté où elle débute. Cependant, il ne faut pas négliger l'importance des noeuds hautement connectés intégrés dans leur communauté dans le processus de propagation de l'épidémie. Dans les réseaux du monde réel, la force de la structure communautaire peut aller d'une structure communautaire forte (peu de liens inter-communautaires) à une structure communautaire faible (forte proportion de liens inter-communautaires). Les stratégies d'immunisation proposées dans les études précédentes visent à cibler les principaux diffuseurs dans les réseaux à structure communautaire. Cependant, soit elles n'exploitent pas la force de la structure communautaire, soit elles ne l'utilisent pas correctement. C'est la raison pour laquelle elles ne conviennent pas à tous les types de réseaux. Pour résoudre ce problème, nous proposons de mieux utiliser les informations sur la structure communautaire afin de développer de nouvelles stratégies d'immunisation. Les trois stratégies d'immunisation présentées dans ce travail

sont destinées à différents types de réseaux avec une force de structure communautaire allant d'une structure communautaire bien définie à non cohésive. Notre objectif est donc de relier l'impact de la force de la structure communautaire au choix d'une stratégie d'immunisation appropriée. De plus, notre objectif est de montrer qu'engager davantage de propriétés topologiques de la communauté peut également améliorer les performances des stratégies d'immunisation.

Dans ce travail, nous limitons notre attention aux réseaux où chaque noeud appartient à une communauté unique. Nous utilisons également une approche déterministe. En d'autres termes, pour chaque noeud du réseau, une mesure d'influence est calculée et les noeuds sont classés et immunisés selon cette mesure. Ainsi, nous proposons et évaluons trois méthodes :

- i) La première méthode proposée cible les noeuds ayant une grande influence inter-communautaire. Elle est mesurée par le nombre de communautés voisines liées au noeud.
- ii) La deuxième méthode d'immunisation cible les noeuds qui pourraient avoir en même temps une forte influence à l'intérieur et à l'extérieur de leurs communautés. Une plus grande importance est accordée à ceux appartenant à de grandes communautés car ils pourraient affecter plus de noeuds. Cette stratégie est basée sur une combinaison pondérée du nombre de liaisons intra-communautaires et inter-communautaires de chaque noeud du réseau.
- iii) La troisième méthode a les mêmes objectifs que la précédente. Elle est conçu de manière à prendre également en compte la variation de densité des communautés.

3.2 Stratégies proposées

Afin de quantifier l'influence d'un noeud dans le processus de diffusion sur les réseaux modulaires, nous proposons trois mesures intégrant différents niveaux d'information.

Soit $G(V, E)$ un simple réseau non dirigé. V représente l'ensemble des noeuds et E est l'ensemble des arêtes. $C = \{C_1, \dots, C_k, \dots, C_m\}$ est l'ensemble des communautés qui ne se chevauchent pas tandis que m est le nombre de communautés du réseau ($G = \bigcup_{k=1}^m C_k$).

3.2.1 Nombre de Communautés Voisines

L'idée principale de cette mesure est de classer les noeuds en fonction du nombre de communautés qu'ils atteignent directement (via un lien). La raison du ciblage de ces noeuds est qu'ils sont plus susceptibles de contribuer à l'épidémie épidémique vers plusieurs communautés. Notez que tous les noeuds qui n'ont pas de liens inter-communautaires partagent la même valeur nulle pour cette mesure.

Pour un noeud donné i appartenant à une communauté $C_k \subset C$, le Nombre de Communautés Voisines (Number of Neighboring Communities NNC) $\beta_{NNC}(i)$ est donné par :

$$\beta_{NNC}(i) = \sum_{C_l \subset C \setminus \{C_k\}} \bigvee_{j \in C_l} a_{ij} \quad (3.1)$$

Où a_{ij} est égal à 1 lorsqu'un lien entre les noeuds i et j existe, et zéro sinon.

\bigvee représente l'opérateur logique de disjonction, c'est-à-dire $\bigvee_{j \in C_l} a_{ij}$ est égal à 1 lorsque le noeud i est connecté au moins à un des noeuds $j \in C_l$.

Certains noeuds de pont peuvent être connectés à une seule communauté voisine avec un nombre élevé de liaisons inter-communautaires. D'autres noeuds de pont peuvent avoir moins de liaisons inter-communautaires mais ces liaisons permettent d'atteindre plusieurs communautés. Cette stratégie permet de cibler des noeuds liés à un nombre élevé de communautés externes. Ainsi, elle peut cibler les ponts les plus influents du réseau. En effet, ces noeuds peuvent diffuser des épidémies dans de nombreuses communautés à travers le réseau. Cependant, cette stratégie présente certains inconvénients. Lorsque le réseau a peu de communautés, de nombreux noeuds ont le même rang. Les noeuds sont immunisés au hasard dans ce cas. De plus, elle ne cible pas les hubs pour l'immunisation. Ces noeuds peuvent affecter un grand nombre de noeuds dans leurs communautés. Ils ont alors une influence majeure dans leurs communautés locales. Dans l'ensemble, cette stratégie convient aux réseaux à structure communautaire de force moyenne ayant un nombre élevé de ponts. Pourtant, elle est inappropriée dans les réseaux avec une structure communautaire bien définie ou lâche, où les hubs peuvent jouer un rôle plus important dans la propagation de l'épidémie.

3.2.2 Community Hub-Bridge

Chaque noeud du réseau partage ses liens avec des noeuds à l'intérieur de sa communauté (liens intra-communautaires) et des noeuds à l'extérieur de sa communauté (liens inter-communautaires). En fonction de la répartition de ces liens, il peut propager l'épidémie plus ou moins dans sa communauté ou dans les communautés voisines. Par conséquent, il peut être considéré comme un hub dans sa communauté et un pont avec ses communautés voisines. C'est la raison pour laquelle nous appelons cette mesure la mesure Community Hub-Bridge. De plus, l'influence du hub dépend de la taille de la communauté, tandis que l'influence du pont dépend du nombre de ses communautés voisines.

Pour un noeud donné i appartenant à une communauté $C_k \subset C$, la mesure Community Hub-Bridge $\beta_{HB}(i)$ est donnée par :

$$\beta_{HB}(i)_{i \in C_k} = h_i(C_k) + b_i(C_k) \quad (3.2)$$

Où :

$$h_i(C_k) = \text{Card}(C_k) * k_i^{\text{intra}}(C_k) \quad (3.3)$$

$$b_i(C_k) = \beta_{NNC}(i) * k_i^{\text{inter}}(C_k) \quad (3.4)$$

$k_i^{\text{intra}}(C_k)$ et $k_i^{\text{inter}}(C_k)$ sont respectivement le degré intra-communautaire et le degré inter-communautaire du noeud i . $\text{Card}(C_k)$ est la taille de sa communauté. $\beta_{NNC}(i)$ représente le nombre de ses communautés voisines.

$h_i(C_k)$ tend à immuniser préférentiellement les hubs au sein des grandes communautés. En effet, ils peuvent infecter plus de noeuds que ceux appartenant à de petites communautés. $b_i(C_k)$ permet de cibler des noeuds qui ont plus de liens avec différentes communautés. Ces noeuds ont une grande influence inter-communautaire.

La stratégie Community Hub-Bridge cible les noeuds qui ont un bon équilibre entre les liens intra-communautaires et inter-communautaires. Elle sélectionne des noeuds jouant simultanément le rôle de hubs dans leurs communautés et des ponts vers d'autres communautés. Cette stratégie donne la priorité aux hubs situés dans les grandes communautés

en raison de leur forte influence locale. Ces noeuds peuvent infecter un grand nombre de noeuds du réseau s'ils sont contaminés. De plus, elle cible les ponts avec la connectivité la plus élevée liée au nombre maximal de communautés externes. Cela permet de cibler les noeuds avec l'influence globale la plus élevée du réseau. Cependant, cette méthode donne de l'importance aux hubs ainsi qu'aux les ponts, quelle que soit la force de la structure de communauté du réseau. Dans certaines situations, il faudrait accorder plus de poids à l'un des deux. Par exemple, dans les réseaux avec des communautés non cohésives, le réseau agit comme une grande communauté, dans ce cas, plus d'importance doit être accordée aux hubs car ils peuvent infecter plusieurs noeuds du réseau.

3.2.3 Weighted Community Hub-Bridge

La mesure Community Hub-Bridge cible en priorité les hubs des grandes communautés et les ponts reliés à plusieurs communautés. Cependant, aucune importance n'est accordée à la force de la structure communautaire. Lorsque la structure communautaire est bien définie, une plus grande importance devrait être accordée aux ponts. En effet, dans ce cas, briser le réseau dans plusieurs communautés permet de contenir l'épidémie qui se propage là où elle a commencé. Au contraire, lorsque la structure communautaire est très lâche, il est primordial d'immuniser les hubs dans les grandes communautés. La pondération de chaque composante du Community Hub-Bridge permet donc d'accorder plus ou moins d'importance aux ponts ou hubs en fonction de la solidité de la structure communautaire.

Pour un noeud donné i appartenant à une communauté $C_k \subset C$, la mesure pondérée du pont communautaire Hub $\beta_{WHB}(i)$ est donnée par :

$$\beta_{WHB}(i)_{i \in C_k} = \rho_{C_k} * h_i(C_k) + (1 - \rho_{C_k}) * b_i(C_k) \quad (3.5)$$

Où ρ_{C_k} représente la densité d'interconnexion entre la communauté C_k et les autres communautés du réseau. Elle est donnée par :

$$\rho_{C_k} = \frac{\sum_{i \in C_k} k_i^{inter} / (k_i^{inter} + k_i^{intra})}{Card(C_k)} \quad (3.6)$$

Si les communautés sont très cohésives, alors plus d'importance est accordée aux ponts afin d'isoler les communautés. Sinon, une plus grande importance est accordée aux hubs au sein des grandes communautés.

La diffusion épidémique d'un noeud dépend de sa position dans sa communauté en plus de la relation que sa communauté entretient avec les autres communautés du réseau. Dans cette perspective, la stratégie Weighted Community Hub-Bridge est conçue pour pouvoir s'adapter avec des noeuds appartenant à des communautés ayant des structures de différentes forces. Elle est très similaire à la stratégie Community Hub-Bridge. Pourtant, elle donne plus de poids aux ponts lorsque le réseau a une structure communautaire bien définie afin d'isoler les communautés. Considérons qu'une épidémie commence au coeur d'une communauté. Si la communauté est isolée, l'épidémie y reste confinée et ne se déplace pas vers d'autres parties du réseau. Cette stratégie donne également plus de poids aux hubs dans le cas de réseaux avec une structure communautaire très lâche, car le réseau agit dans ce cas comme une seule grande communauté.

3.2.4 Exemple

Afin d'illustrer le comportement de ces mesures, un exemple est donné dans Figure 3.1. Les noeuds sont classés selon la mesure du nombre de communautés voisines dans Figure 3.1 (a). Prenons l'exemple des noeuds $n5$ et $n10$ qui sont tous les deux des ponts communautaires et qui ont le même nombre de liens internes et externes. Selon la mesure de centralité de degré dans Figure 3.1 (d), les deux noeuds ont le même rang car cela ne dépend que de leur nombre de voisins. Cependant, ils ont des rangs différents selon la mesure du nombre de communautés voisines. La mesure proposée accorde plus d'importance au noeud $n5$ qui est lié à trois communautés externes, donc s'il est contaminé, il peut transmettre la maladie épidémique d'abord à sa propre communauté $C1$ et également aux communautés voisines $C2$, $C3$ et $C4$. Alors que la maladie épidémique pourrait être transmise aux noeuds appartenant aux communautés $C1$ et $C2$ en cas de contamination par le noeud $n10$. De plus, les noeuds $n15$ et $n12$ sont classés parmi les noeuds les moins influents selon la mesure de l'intermédiarité comme indiqué dans Figure 3.1 (e), bien que les deux soient des ponts communautaires susceptibles de contribuer à la flambée

épidémique dans les communautés externes. Par conséquent, la mesure Nombre de communautés voisines cible les ponts les plus influents qui peuvent propager les épidémies à plusieurs communautés.

Figure 3.1 (b) montre le rang des noeuds selon la mesure Community Hub-Bridge. Même si $n6$ et $n16$ ont tous les deux quatre liens internes à l'intérieur de leurs propres communautés, $n6$ est considéré comme plus influent car il est situé dans la communauté $C1$ qui est la plus grande communauté de le réseau. Par conséquent, il peut constituer une menace pour plusieurs noeuds du réseau s'il est infecté. Contrairement à la mesure de degré dans Figure 3.1 (d) qui classe les noeuds $n6$ et $n16$ dans le même rang en fonction de leur nombre de connexions sans tenir compte de leur emplacement dans le réseau. On constate également dans Figure 3.1 (a) que de nombreux noeuds ont le même rang car ils ont le même nombre de communautés voisines. Donc, si nous considérons les noeuds $n10$ et $n12$, ils sont tous deux connectés à une seule communauté voisine (respectivement $C1$ et $C3$), par conséquent, ils ont le même rang. Cependant, $n10$ a une plus grande connectivité avec $C1$ en termes de nombre de liens externes. La raison pour laquelle nous avons introduit la quantité de liaisons externes en tant que nouveau paramètre dans le deuxième terme de la mesure Community Hub-Bridge. Il s'agit de faire la distinction entre les ponts ayant une grande connectivité et ceux ayant une connectivité plus faible avec les communautés externes. Basé sur la mesure Community Hub-Bridge, $n10$ est plus influent que le noeud $n12$ comme on peut le voir dans Figure 3.1 (b). Le noeud $n10$ a trois connexions externes vers la communauté $C1$ alors que $n12$ n'a qu'une seule connexion vers $C3$. Par conséquent, l'influence des noeuds selon cette mesure est liée à deux facteurs : l'importance des noeuds au sein de leurs communautés en donnant la priorité à ceux situés dans les grandes communautés, et la connectivité des noeuds vers diverses communautés.

Les noeuds sont classés en fonction de la mesure Weighted Community Hub-Bridge dans Figure 3.1 (c). Le réseau donné dans cet exemple a une structure communautaire bien définie. Comme nous pouvons le voir clairement, si nous prenons l'exemple de la communauté $C1$, la densité des liens inter-communautaires est égale à $\rho_{C_1} \approx 0.15$. Par conséquent, 15% d'importance est donnée à $h_i(C_1)$ (l'expression qui cible les hubs) et 85% d'importance est donnée à $b_i(C_1)$ (l'expression qui cible les ponts). Cela explique pourquoi

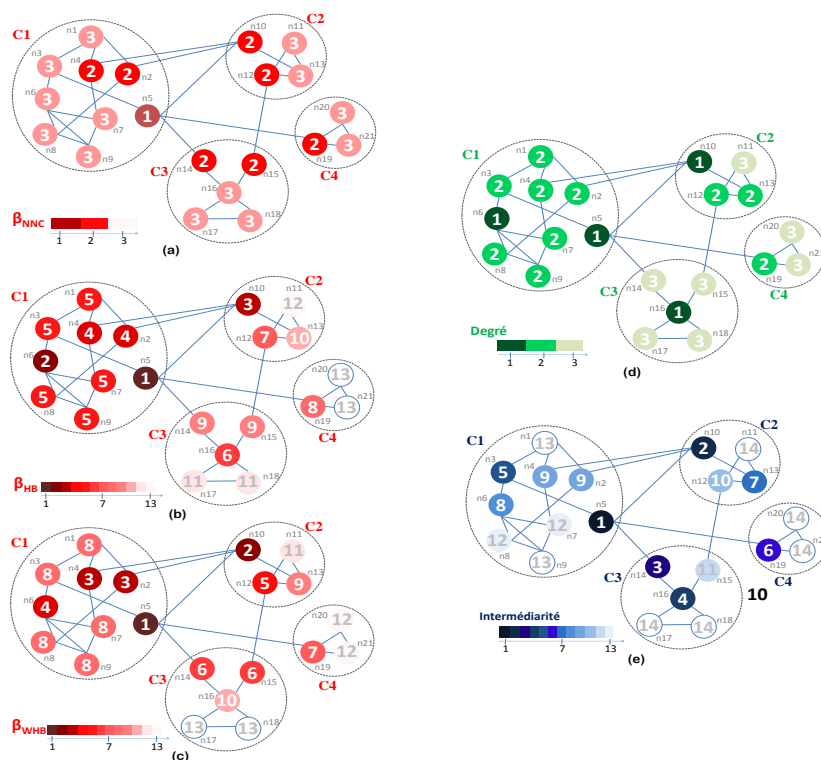


FIGURE 3.1 – Le rang des noeuds selon (a) la mesure du Nombre de Communautés Voisines β_{NNC} (b) la mesure du Community Hub-Bridge β_{HB} (c) la mesure du Weighted Community Hub-Bridge β_{WHB} (d) Mesure du degré (e) Mesure de l'intermédiarité. Les noeuds sont classés de plus influent (noeuds ayant la valeur de mesure la plus élevée) au noeud le moins influent (noeuds ayant la valeur de mesure la plus faible) du réseau.

tous les ponts de communauté ($n5$, $n2$ et $n4$) sont immunisés avant les autres noeuds de la communauté $C1$. Cela permet d'isoler cette communauté et d'empêcher la diffusion de l'épidémie de se déplacer de $C1$ vers les autres communautés du réseau. Ainsi, la mesure Weighted Community Hub-Bridge a la capacité de s'adapter à la force de la structure communautaire. Elle donne plus de poids aux ponts lorsque le réseau a une structure communautaire bien définie afin d'isoler les communautés, alors qu'il donne plus de poids aux hubs dans le cas de réseaux avec une structure communautaire faible puisque le réseau agit dans ce cas comme un seule grande communauté.

3.3 Cadre expérimental

Dans cette section, nous présentons les données et les méthodes utilisées dans l'évaluation empirique des différentes stratégies d'immunisation présentées ci-dessus.

3.3.1 Données utilisées

Afin d'évaluer la sous-étude des différentes mesures, des réseaux synthétiques aux propriétés topologiques contrôlées, ainsi que des réseaux du monde réel ont été utilisés.

3.3.1.1 Réseaux synthétiques

Les réseaux synthétiques sont générés en utilisant l'algorithme LFR (Lancichinetti, Fortunato et Radicchi) Lancichinetti *et al.* (2008). Il génère des échantillons aléatoires de réseaux avec un degré et taille de communauté distribué selon la loi de puissance. Par conséquent, l'algorithme LFR garantit des réseaux avec des fonctionnalités réalistes Orman *et al.* (2013). Cet algorithme permet de contrôler différents paramètres lors de la génération de réseaux. Les paramètres d'entrée du modèle sont le nombre de noeuds, le degré moyen et maximum souhaité, les exposants pour la distribution du degré et la taille de communauté, et le coefficient de mélange. Le paramètre de mélange (mixing parameter) μ détermine principalement le rapport entre le nombre de voisins externes d'un noeud (appartenant à une différente communauté) et le degré total du noeud. Sa valeur contrôle la force de la structure communautaire. Pour les petites valeurs de μ , les communautés sont bien séparées car elles partagent peu de liens, tandis que lorsque μ augmente, la proportion de liens inter-communautaires augmente, ce qui rend l'identification de la communauté difficile. Des études expérimentales ont montré que pour un réseau sans échelle, l'exposant de distribution des degrés α varie généralement de 2 à 3, et le degré maximal est estimé à $k_{max} \sim n^{1/(\alpha-1)}$ Albert et Barabási (2002); Boccaletti *et al.* (2006); Newman (2003). Les valeurs des paramètres utilisés dans nos expériences sont données dans Table 5.1.

3.3.1.2 Réseaux réels

Des réseaux du monde réel de nature diverse (réseaux sociaux en ligne, réseau technologique et réseau de collaboration) sont utilisés pour tester les stratégies d'immunisation.

TABLEAU 3.1 – Paramètres du réseau LFR

Nombre de noeuds N	15 000
Degré moyen $\langle k \rangle$	7
Degré maximum k_{max}	122
Exposant de la distribution des degrés α	3
Exposant pour la distribution de taille de communauté σ	2.5
Paramètre de mélange μ	0.1, 0.4, 0.7, 0.9
Plages de taille de communauté	[50 250],[100 500]

- **Facebook** : Nous utilisons un réseau rassemblé par Traud *et al.* (2012) à partir du réseau social Facebook¹. Ces données incluent le réseau d'amitié de cinq universités américaines. Il fournit également des informations sur les individus tels que le dortoir, le domaine de spécialisation et l'année de cours.

- **Power-grid** : Ce réseau technologique est un réseau non orienté et non pondéré contenant des informations sur la topologie du réseau Power grid du Western States des États-Unis. Un lien représente une ligne d'alimentation. Un noeud est soit un générateur, un transformateur ou une sous-station. Ces données² sont compilées par D. Watts et S. Strogatz Watts et Strogatz (1998).

- **Relativité Générale et Cosmologie Quantique (GR-QC)** : GR-QC³ est un réseau de collaboration collecté à partir de l'arXiv e-print. Il couvre les collaborations scientifiques entre les auteurs d'articles soumis à la catégorie Relativité Générale et Cosmologie Quantique. Les noeuds représentent les auteurs et il existe un lien entre deux noeuds s'ils sont co-auteurs d'un article. Ces données sont disponibles dans le référentiel SNAP compilé par Yang et Leskovec (2012).

La structure communautaire de ces réseaux étant inconnue, nous utilisons un algorithme de détection communautaire. Nous avons choisi d'utiliser l'algorithme de Louvain qui s'est avéré efficace dans les réseaux synthétiques et réels Blondel *et al.* (2008); Orman *et al.* (2011). De plus, les propriétés topologiques des communautés découvertes sont également réalistes Orman *et al.* (2012).

Les propriétés topologiques de base de ces réseaux sont données dans Table 5.2.

1. <http://code.google.com/p/socialnetworksimulation/>
 2. <http://www-personal.umich.edu/~mejn/netdata/>
 3. <http://snap.stanford.edu/data/ca-GrQc.html>

TABLEAU 3.2 – Les propriétés topologiques de base de six réseaux du monde réel. N et E sont respectivement le nombre total de noeuds et de liens. Q est la modularité. N_c est le nombre de communautés. λ_{th} est le seuil épidémique.

Réseau	N	E	Q	N_c	λ_{th}
Caltech	620	7255	0.788	13	0.012
Princeton	5112	28684	0.753	21	0.006
Georgetown	7423	162982	0.521	42	0.006
Oklahoma	10386	88266	0.914	67	0.031
Power grid	4941	6594	0.92	41	0.092
CR-QC	5242	14496	0.86	396	0.059

3.3.2 Paramètres du modèle SIR

La valeur du taux de transmission λ est choisie supérieure au seuil épidémique du réseau λ_{th} afin de mieux caractériser la capacité de propagation, définie comme Wang *et al.* (2016) :

$$\lambda_{th} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \quad (3.7)$$

Où $\langle k \rangle$ et $\langle k^2 \rangle$ sont respectivement les premier et deuxième moments de la distribution des degrés. Les valeurs de seuil épidémique λ_{th} de tous les réseaux utilisés dans cet article sont rapportées dans Table 5.2. La même valeur de taux de transmission ($\lambda = 0.1$) est utilisée dans toutes les expériences. Elle est supérieur des valeurs du seuil épidémique λ_{th} de toutes les collectes de données utilisées dans ce travail. Nous fixons également la valeur du taux de rétablissement γ à 0.2. Cette petite valeur est choisie afin de donner à chaque noeud infecté de nombreuses chances d'infecter ses voisins avec la probabilité γ avant de passer au statut rétabli.

3.3.3 Schéma d'immunisation

Pour étudier la propagation d'une maladie infectieuse sur un réseau de contacts, nous utilisons la méthodologie décrite dans Figure 7. Pour les stratégies déterministes, l'influence de chaque noeud du réseau est calculée selon une mesure de centralité donnée. Ensuite, les noeuds sont triés par ordre décroissant de leurs valeurs d'influence. Ensuite, les noeuds avec la centralité la plus élevée sont supprimés du réseau (ou leur état est défini sur résistant) jusqu'à ce qu'une couverture d'immunisation souhaitée soit atteinte.

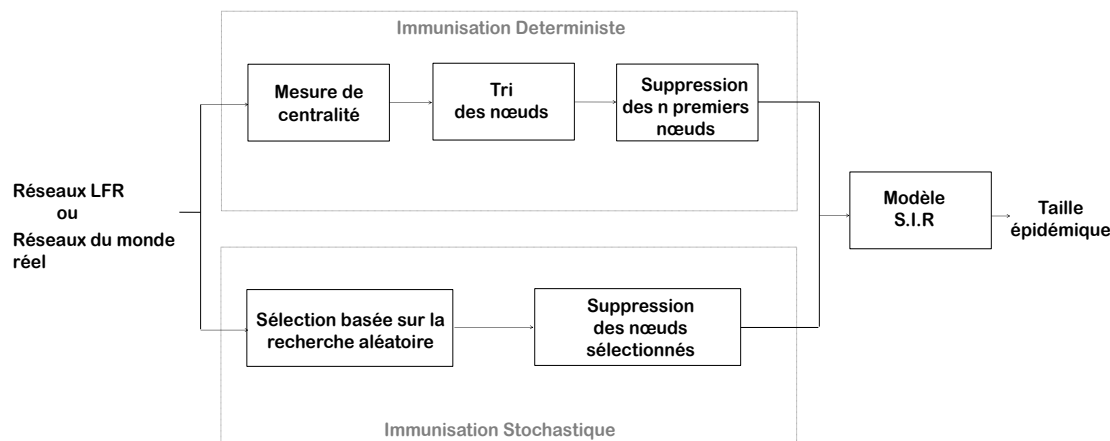


FIGURE 3.2 – Les principales étapes de schéma d'immunisation.

Pour l'immunisation stochastique, les noeuds sont ciblés et supprimés selon une stratégie aléatoire lancée à partir de noeuds choisis au hasard dans le réseau. Dans les deux cas, le réseau obtenu après l'immunisation ciblée est utilisé pour simuler le processus de propagation, en exécutant les simulations du modèle épidémique SIR. Après une simulation, nous enregistrons le nombre total de cas rétablis (la taille de l'épidémie). Afin d'assurer l'efficacité de l'évaluation du modèle de propagation SIR, les résultats sont moyennés sur 600 réalisations indépendantes. Enfin, nous calculons la taille moyenne de l'épidémie pour évaluer l'efficacité des méthodes proposées.

3.3.4 Critère d'évaluation

3.3.4.1 Différence relative de la taille de l'épidémie

Pour comparer les performances de différentes stratégies d'immunisation, nous utilisons la fraction de la taille de l'épidémie. Nous utilisons également la différence relative de taille de l'épidémie (Relative difference of outbreak size) $\Delta r_{\beta_0, \beta_c}$ définie par :

$$\Delta r_{\beta_0, \beta_c} = \frac{R_{\beta_0} - R_{\beta_c}}{R_{\beta_0}} \quad (3.8)$$

Où R_{β_0} et R_{β_c} sont respectivement le nombre final de noeuds rétablis pour la stratégie alternative et la stratégie proposée après les simulations SIR. Si la différence relative de la taille de l'épidémie est positive, l'épidémie se propage moins avec la stratégie propo-

sée. C'est donc la plus efficace. Sinon, l'épidémie se propage davantage avec la stratégie proposée et la stratégie alternative est plus efficace.

3.3.4.2 *Le plus grand composant connecté*

Nous utilisons également la taille du plus grand composant connecté (Largest Connected Component *LCC*) pour tester l'efficacité des stratégies proposées quels que soient les modèles épidémiologiques utilisés. Il s'agit du plus grand sous-graphe restant après le processus d'immunisation. La taille du plus grand composant connecté est utilisée pour mesurer la limite maximale à laquelle une épidémie peut se propager. Le *LCC* est l'une des mesures permettant de quantifier les performances de stratégies d'immunisation. Il se concentre sur les changements dans la structure du composant géant après la suppression de certains noeuds. En effet, la taille du plus grand composant connecté est calculée après suppression d'une certaine proportion de noeuds sélectionnés selon une stratégie d'immunisation. De toute évidence, plus le *LCC* est petit, meilleure est la stratégie d'immunisation.

3.4 Résultats et discussion

Dans cette section, nous rapportons les résultats de deux séries d'expériences. La première série d'expériences est réalisée avec des réseaux synthétiques à structure communautaire contrôlée. Elle vise à mieux comprendre la relation entre la structure communautaire et les mesures de centralité. Ces expériences sont menées sur des réseaux générés avec l'algorithme LFR. En effet, cet algorithme permet de contrôler diverses propriétés topologiques de la structure communautaire. Nous étudions l'influence de la force de la structure communautaire. L'effet de l'intervalle de taille de la communauté est également étudié. Enfin, les stratégies d'immunisation proposées sont comparées à la fois aux stratégies déterministes (stratégies du Degré, intermédiarité et Comm) et aux stratégies stochastiques (stratégies du Community Bridge Finder Salathé et Jones (2010) et Bridge Hub Detector Gong *et al.* (2013)).

La deuxième série d'expériences concerne les réseaux du monde réel. Des réseaux sociaux en ligne, un réseau technologique et un réseau de collaboration sont utilisés. Rappelons

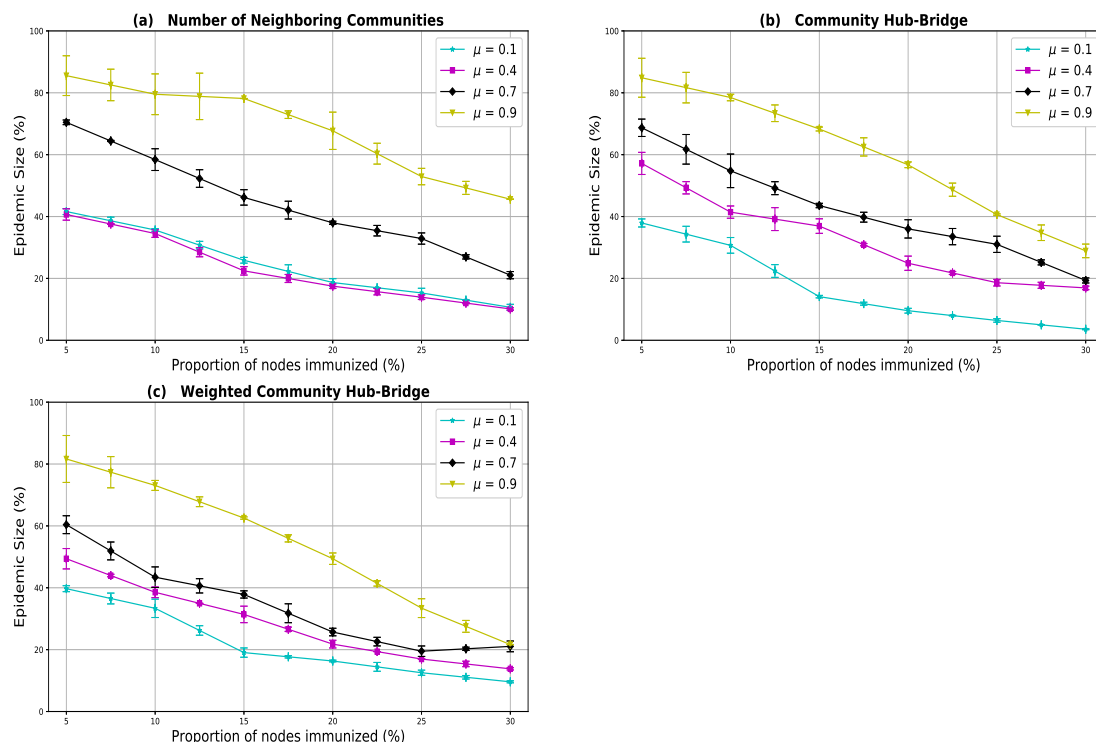


FIGURE 3.3 – Influence de la force de la structure communautaire sur la taille épidémique des méthodes proposées. Chaque point représente la taille de l'épidémie par rapport à la fraction des noeuds immunisés. Les simulations sont effectuées sur des réseaux générés par LFR avec différentes valeurs de paramètres de mélange μ . Chaque valeur de taille épidémique est la moyenne de 600 simulations S.I.R.

que, comme il n'y a pas de données de vérité terrain pour ces réseaux, la structure de la communauté est découverte à l'aide de l'algorithme de Louvain. En effet, des études antérieures sur les réseaux synthétiques ont montré qu'il réussit à identifier les communautés quelque soit la force de la structure communautaire du réseau Orman *et al.* (2011). Tout d'abord, les stratégies d'immunisation proposées sont comparées et discutées, puis leur évaluation est effectuée par rapport à des stratégies alternatives stochastiques et déterministes.

3.4.1 Réseaux synthétiques

3.4.1.1 Influence de la force de la structure communautaire

Dans le modèle LFR, la valeur du paramètre de mélange μ varie de 0 à 1. Celui-ci permet de contrôler la force de la structure communautaire, de communautés bien séparées avec peu de liens inter-communautaires (faibles valeurs de μ) à un réseau sans structure communautaire (valeurs élevées de μ). Afin d'étudier l'effet de la force de la structure communautaire sur la performance des méthodes proposées, cinq réseaux ont été générés pour chaque valeur du paramètre de mélange ($\mu = 0.1, 0.4, 0.7$ et 0.9). Figure 3.3 rapporte la fraction moyenne de la taille de l'épidémie par rapport à la proportion de noeuds immunisés pour chaque valeur de μ . Selon les résultats rapportés dans cette figure, les performances des stratégies Community Hub-Bridge et Weighted Community Hub-Bridge diminuent tout en augmentant la valeur du paramètre de mélange. Quelle que soit la fraction des noeuds immunisés, les méthodes fonctionnent mieux lorsque les communautés sont bien séparées. Lorsque la fraction des interconnexions entre les communautés augmente, les performances diminuent progressivement. En effet, avec des communautés bien séparées, les épidémies sont localisées dans quelques communautés, alors qu'elles ont tendance à se propager davantage lorsque les interconnexions augmentent. La stratégie Nombre de communautés voisines manifeste les meilleures performances pour une valeur de force de communauté moyenne ($\mu = 0.4$). Son efficacité diminue légèrement dans le cas d'une structure communautaire bien définie, et elle empire encore lorsqu'elle est très lâche. Passons maintenant aux comparaisons des méthodes proposées entre elles. Nous pouvons distinguer trois cas en fonction de la force de la structure communautaire.

Dans les réseaux présentant une structure communautaire très forte, nous pouvons voir dans Figure 3.3 que la stratégie Community Hub-Bridge est la plus efficace. Cela est dû au fait que les deux méthodes alternatives (nombre de communautés voisines et Weighted Community Hub-Bridge) ciblent préférentiellement les ponts. En fait, ce n'est pas la meilleure solution dans un réseau où prédominent les liens intra-communautaires. Comme il y a peu de connexions externes par rapport au nombre total de connexions (les liens intra-communautaires sont considérés comme 90% du total des liens du réseau lorsque

$\mu=0.1$), les épidémies locales peuvent s'éteindre avant d'atteindre d'autres communautés. Par conséquent, l'immunisation des hubs communautaires semble être plus efficace que l'immunisation des ponts dans les réseaux à forte structure communautaire. C'est la raison pour laquelle la méthode Community Hub-Bridge qui cible les noeuds ayant un bon équilibre des connexions internes et externes est plus efficace.

Dans les réseaux avec une structure communautaire faible comme on peut le voir dans Figure 3.3, la méthode Weighted Community Hub-Bridge est la plus efficace. En effet, lorsque μ a une valeur élevée, le réseau n'a pas de structure communautaire bien définie. Dans ce cas, la stratégie Weighted Community Hub-Bridge peut mieux s'adapter à la structure de la communauté. Cela donne plus de poids aux hubs communautaires car ce sont les noeuds les plus influents dans les réseaux avec une structure communautaire lâche. N'oubliez pas que le réseau agit comme une communauté unique dans le cas extrême où $\mu \geq 0.9$. C'est la raison pour laquelle il fonctionne mieux que les autres méthodes proposées.

Dans les réseaux avec une structure communautaire de force moyenne, le nombre de communautés voisines surpasse toutes les autres méthodes proposées comme indiqué dans Figure 3.3. Dans ce type de réseaux, les noeuds ont de nombreuses connexions externes tout en conservant une structure de communauté bien préservée. Par conséquent, il y a beaucoup plus d'options pour que l'épidémie se propage facilement aux communautés voisines. Comme la stratégie Nombre de communautés voisines cible les ponts communautaires les plus influents, elle empêche l'épidémie de se propager à plusieurs communautés. C'est la raison pour laquelle cette méthode d'immunisation montre ses meilleures performances dans ce cas.

Pour résumer, la stratégie du pont Community-Hub est bien adaptée aux situations où les communautés sont bien définies (communautés denses avec peu de liens entre les communautés). La stratégie du Weighted Hub-Bridge est recommandée lorsque la structure communautaire est très lâche. Pour les situations intermédiaires, la stratégie Nombre de communautés voisines est plus efficace.

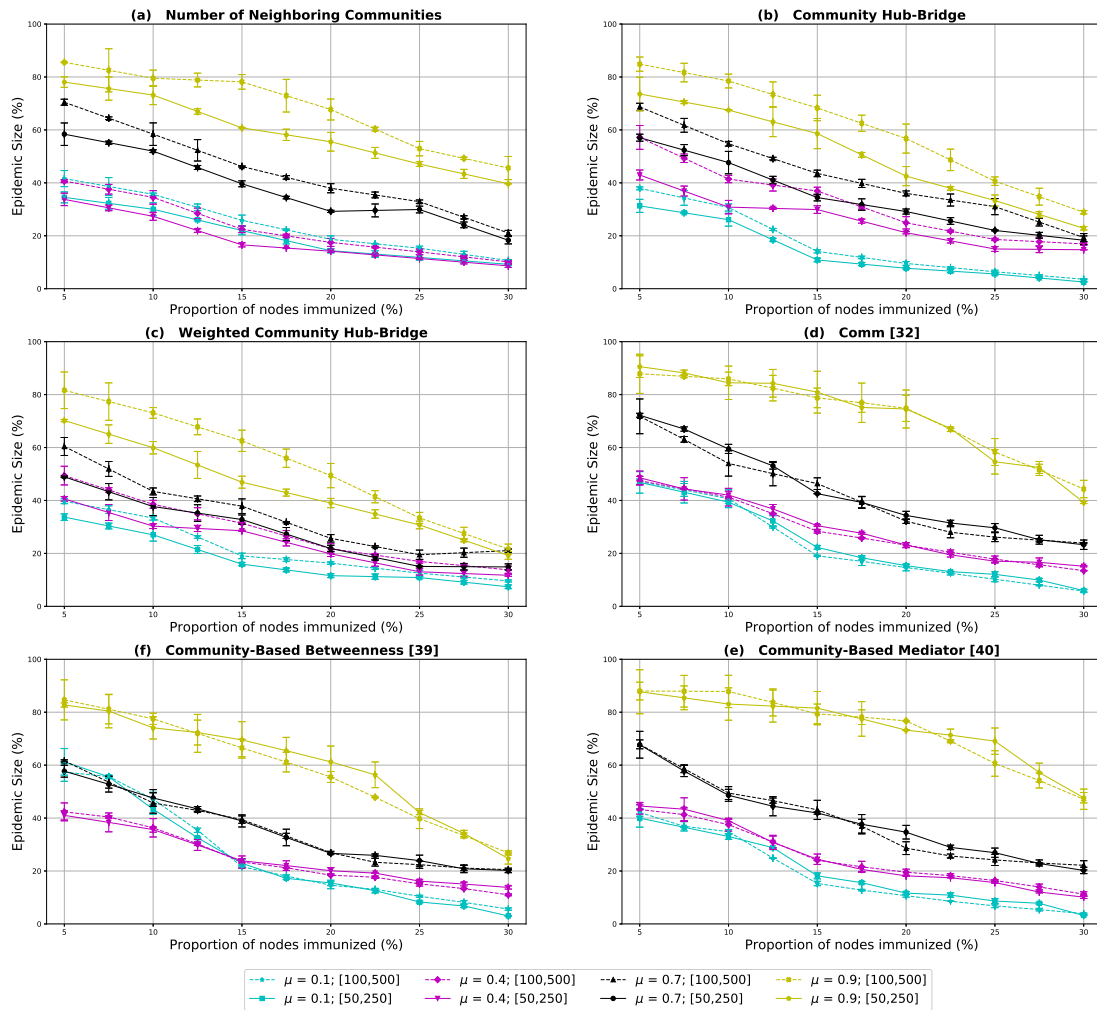


FIGURE 3.4 – Effet de la taille des communautés sur la propagation épidémique des méthodes proposées. Chaque point représente la taille de l'épidémie en fonction de la fraction des noeuds immunisés. Les simulations sont effectuées sur des réseaux LFR avec différentes structures communautaires. Chaque valeur est la moyenne de 600 simulations par réseau et par méthode d'immunisation.

3.4.1.2 Effets de la taille des communautés

Le but de cette enquête est de montrer l'impact de la taille des communautés sur la performance des méthodes proposées. Les études décrites ci-dessus ont été réalisées avec une taille de communautés comprise entre 100 et 500. Dans ce paragraphe, elles sont également évaluées dans des réseaux dont la taille de la communauté est égale à [50, 250]. Figure 3.4 rapporte la taille de l'épidémie par rapport au pourcentage de noeuds immunisés

pour des valeurs du paramètre de mélange μ allant de $\mu = 0.1$ à $\mu = 0.9$, et avec les deux gammes de tailles de communautés étudiées. On peut voir que toutes les stratégies d'immunisation présentent le même comportement. Elles fonctionnent toujours mieux dans les réseaux avec une gamme de taille de communauté plus petite. De plus, les différences entre les tailles épidémiques dans les deux situations diminuent lorsque la proportion de noeuds immunisés augmente. Dans les réseaux avec des communautés de grande taille, il y a un petit nombre de communautés. Par conséquent, la mesure du Nombre de communautés voisines a également de petites valeurs, et de nombreux noeuds ont les mêmes valeurs (comme le montre l'exemple donné dans la section 3.2.4). Cela rend le classement des noeuds moins efficace. Dans les réseaux dont la taille des communautés est plus petite, il y a beaucoup plus de communautés. Dans ce cas, plus de noeuds ont différents nombres de communautés voisines, ainsi le classement des noeuds est plus efficace. C'est la raison pour laquelle le nombre de communautés voisines obtient de meilleurs résultats dans ce dernier cas. En ce qui concerne les mesures Community Hub-Bridge et Weighted Community Hub-Bridge, les deux sont pondérés en fonction du nombre de communautés voisines. Ce poids devient plus discriminant à mesure que la taille de la communauté diminue. Cela explique pourquoi ces deux centralités fonctionnent également mieux dans les réseaux avec des communautés de petite taille. De plus, il est également illustré dans Figure 3.4 (d)-(f) que les autres stratégies basées sur la structure communautaire (ex. Comm, CBB et CBM) ne sont pas affectées par la taille de la communauté. Pour confirmer cela, nous effectuons l'analyse de variance ANOVA sur la performance de ces stratégies tout en faisant varier à la fois la plage de taille de la communauté et le paramètre de mélange du réseau. À partir des résultats rapportés dans Table 5, on peut remarquer que la valeur estimée de F est toujours inférieure à la valeur critique de F ($F < F_{critic}$). Ainsi, nous pouvons conclure que Comm, CBB et CBM affichent les mêmes performances tout en changeant la plage de taille de la communauté, et ce dans des réseaux avec diverses forces de structure communautaire. Ces méthodes sont basées sur des mesures qui ne tiennent pas compte non plus de la taille et du nombre de communautés. D'où la stabilité de leurs performances malgré le changement de la taille des communautés.

À partir de Figure 3.4, on peut également voir que la méthode Community Hub-Bridge

est toujours la meilleure méthode d'immunisation dans les réseaux avec une structure communautaire bien définie, et que la méthode du nombre de communautés voisines surpasse les autres méthodes d'immunisation proposées (lorsque $\mu = 0.4$). De plus, la Weighted Community Hub-Bridge est toujours la stratégie la plus efficace dans les réseaux à structure communautaire non cohésive.

3.4.1.3 Comparaison avec les méthodes alternatives

Figure 3.5 rapporte la différence relative de la taille de l'épidémie entre les stratégies proposées et les alternatives stochastiques (Connaissance, CBF, BHD) et déterministes (Degré, intermédiarité, Comm, Community-based Betweenness et Community-based Mediator) en fonction de la fraction des noeuds immunisés. Community Hub-Bridge est pris comme référence en (a), (d) et (g), Nombre de communautés voisines en (b), (e) et (h) et Weighted Community Hub-Bridge en (c), (f) et (i). Les valeurs du paramètre de mélange ($\mu = 0.1$, $\mu = 0.4$, $\mu = 0.7$) couvrent les trois situations en termes de force communautaire (structure communautaire forte, moyenne et faible).

Figure 3.5 (a), (d) et (g) montre que $\Delta R/R_0$ a généralement une valeur positive. Ainsi, Community Hub-Bridge donne une taille épidémique plus petite par rapport à toutes les méthodes alternatives quelle que soit la fraction des valeurs des noeuds immunisés, et cela vaut pour les différentes forces de la structure communautaire. Les figures du milieu de Figure 3.5 rendent compte des résultats de l'évaluation comparative de la stratégie du Nombre de communautés voisines. Dans l'ensemble, elle est plus efficace que les méthodes alternatives testées. Cependant, l'intermédiarité et Community-based Mediator réussissent mieux dans les réseaux à forte structure communautaire ($\mu = 0.1$). En effet, la différence relative est négative dans ce cas. Par conséquent, cibler les ponts communautaires n'est pas la meilleure solution d'immunisation dans les réseaux avec une structure communautaire très bien définie. Elle est également moins performante que la Community-based Betweenness dans les réseaux à structure communautaire lâche ($\mu = 0.7$). On peut également remarquer à partir de Figure 3.5 (c), (f) et (i) que la méthode Weighted Community Hub-Bridge résulte en une taille épidémique plus faible par rapport aux autres méthodes. Pour résumer, si l'on exclut le cas de la stratégie du Nombre de communautés

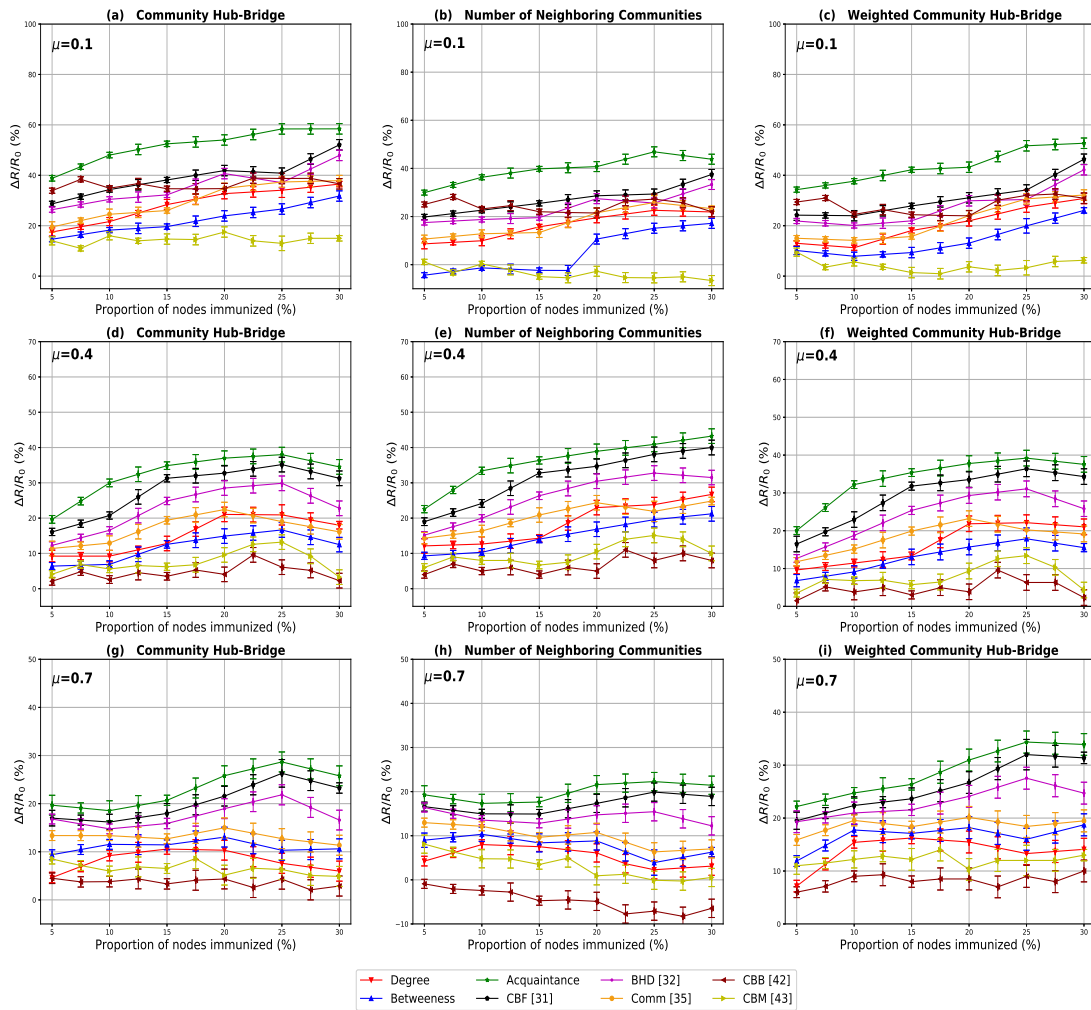


FIGURE 3.5 – Différence relative de la taille de l'épidémie $\Delta R/R_0$ en fonction de la fraction de noeuds immunisés. Les figures de gauche montrent la différence entre la méthode Community Hub-Bridge et les méthodes alternatives. Les figures du milieu montrent la différence entre la méthode du Nombre de communautés voisines et les méthodes alternatives, tandis que les figures de droite montrent la différence entre la méthode du Weighted Community Hub-Bridge et les méthodes alternatives. Nous notons qu'une valeur positive de $\Delta R/R_0$ signifie une performance plus élevée de la méthode proposée. Les simulations sont effectuées sur le réseau LFR avec différentes structures communautaires. Les valeurs finales sont obtenues en exécutant 600 simulations indépendantes par réseau, par couverture vaccinale et méthode d'immunisation.

voisins dans la situation où le réseau a une structure communautaire forte ($\mu = 0.1$), dans toutes les autres situations, la différence relative de l'épidémie est toujours positive. Cela indique que toutes les stratégies proposées surpassent les alternatives. Passons

maintenant à des comparaisons plus détaillées. Tout d'abord, ces résultats démontrent clairement la supériorité des méthodes déterministes sur les méthodes stochastiques. En effet, dans tous les cas, les écarts les plus importants sont observés avec la stratégie de Connaissance (Acquaintance) suivie de CBF et BHD. En fait, leur rang est corrélé avec le niveau d'information qu'elles utilisent sur la topologie du réseau. En fait, la méthode de connaissance est totalement agnostique quant à la topologie du réseau, CBF cible les ponts entre les communautés tandis que BHD cible à la fois les ponts et les hubs. Même si CBF et BHD sont des méthodes communautaires, elles n'utilisent les informations qu'au niveau des noeuds choisis au hasard, d'où leurs faibles performances. L'efficacité comparée des cinq stratégies déterministes alternatives dépend de la force de la structure communautaire. Pour une structure communautaire forte ($\mu = 0.1$), la stratégie degré et Comm sont très proche tandis que les méthodes d'intermédiarité et la Community-based Mediator sont légèrement plus performantes quelle que soit la valeur de la fraction des noeuds immunisés. Pour une force de structure communautaire moyenne ($\mu = 0.4$), les résultats sont plus mitigés, même si la méthode du Community-based Betweenness est encore un peu plus efficace. Pour une structure communautaire faible ($\mu = 0.7$), les cinq stratégies sont bien séparées. La méthode de la Community-based Betweenness arrive en tête, suivi de la stratégie de Community-based Mediator, de degré, d'intermédiarité et Comm en termes d'efficacité.

3.4.2 Réseaux réels

Notre objectif étant de couvrir un large éventail de situations, les données du monde réel proviennent de différents domaines (réseaux sociaux, techniques et de collaboration). Afin de relier les résultats de cet ensemble d'expériences à ceux effectués sur des données synthétiques, nous estimons le paramètre de proportion de mélange de la structure de communauté découverte par l'algorithme de Louvain. En effet, des expériences réalisées avec des réseaux synthétiques ont montré que la force de la communauté est un paramètre majeur pour expliquer l'efficacité des stratégies d'immunisation proposées. Les valeurs estimées rapportées dans Table 5.3 montrent que les réseaux couvrent un large éventail de forces communautaires.

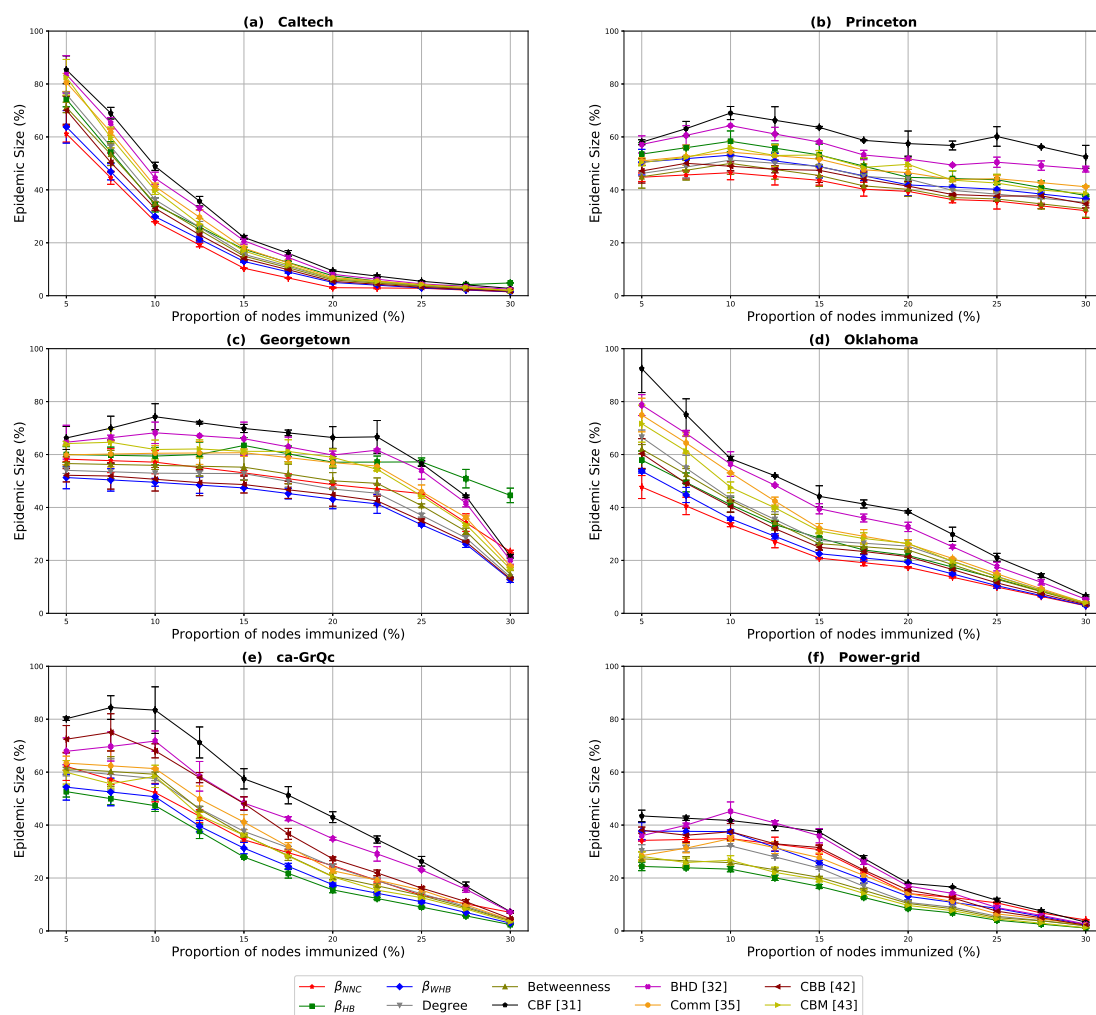
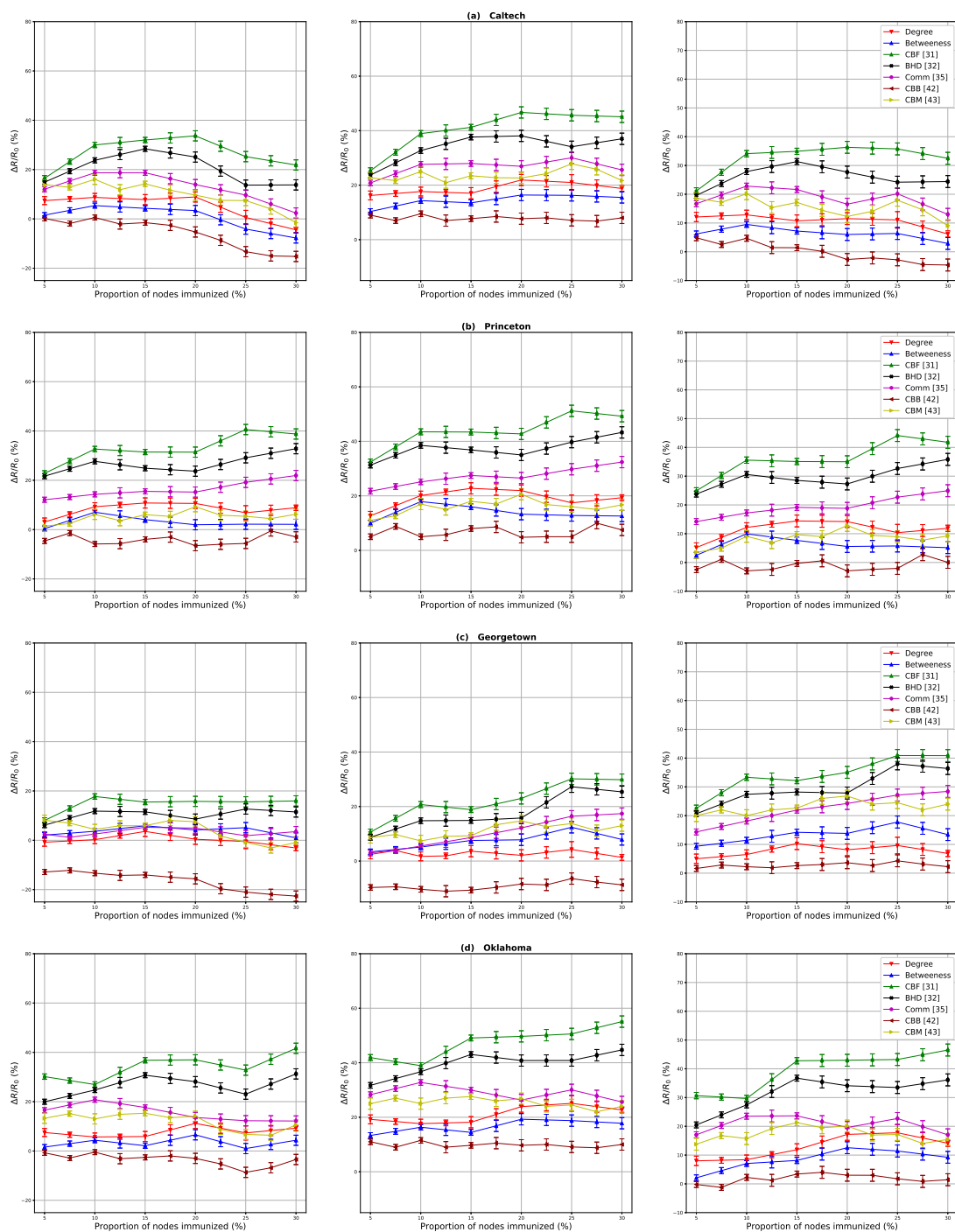


FIGURE 3.6 – La taille épidémique des méthodes d’immunisation effectuées sur six réseaux réels de types différents à savoir le réseau facebook de quatre universités (a) Caltech (b) Princeton (c) Georgetown (d) Oklahoma, et (e) Réseau de collaboration (f) Réseau du Power-grid. Les valeurs finales sont obtenues en exécutant 600 simulations indépendantes par réseau, couverture et méthode d’immunisation.

3.4.2.1 Efficacité de diffusion des méthodes proposées

Figure 3.6 montre la taille de l’épidémie en fonction de la fraction de noeuds immunisés obtenue après les simulations SIR pour les méthodes d’immunisation proposées. Ces résultats confirment les conclusions que nous avons tirées avec les réseaux synthétiques. On peut observer sur Figure 3.6 (e) et (f) que dans les réseaux à forte structure communautaire, la Community Hub-Bridge est la méthode d’immunisation la plus efficace. En



effet, la valeur estimée du paramètre de mélange μ est égale à 0.03 et 0.09 respectivement pour le réseau Power-grid et le réseau de collaboration. Les communautés sont très bien

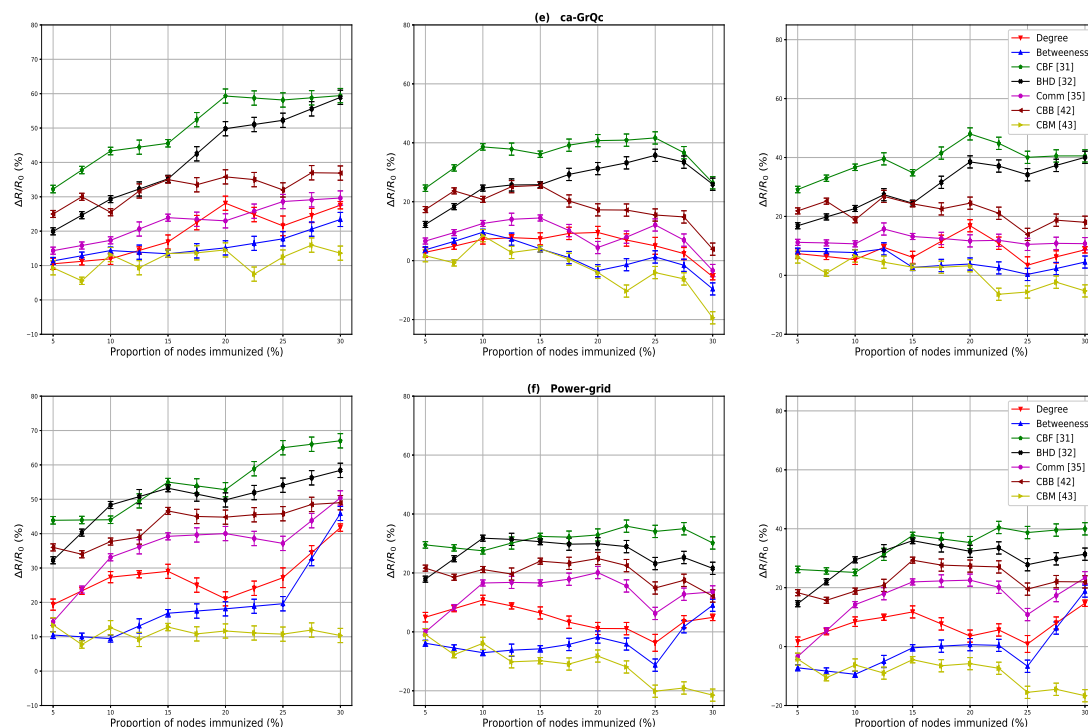


FIGURE 3.7 – Différence relative de la taille de l'épidémie $\Delta R/R_0$ en fonction de la fraction de noeuds immunisés. Les figures de gauche illustrent la différence entre la méthode Community Hub-Bridge et les méthodes alternatives. Les figures du milieu et de droite montrent respectivement la différence entre la méthode du Nombre de communautés voisines, la Weighted Community Hub-Bridge et les méthodes alternatives. Nous notons qu'une valeur positive de $\Delta R/R_0$ signifie une performance plus élevée de la méthode proposée. Les simulations sont effectuées sur différents types de réseaux du monde réel. Les valeurs finales sont obtenues en exécutant 600 simulations indépendantes par réseau, couverture et méthode d'immunisation.

séparées, et la méthode Community Hub-Bridge cible les noeuds avec un bon équilibre de liens intra et inter-communautaires. C'est là que réside sa supériorité.

Dans les réseaux avec une force de structure communautaire moyenne indiquée dans Figure 3.6 (a), (b) et (d), le Nombre de communautés voisines surpasse les autres méthodes proposées. Elle cible les ponts connectés à plusieurs communautés, ce qui facilite la propagation des épidémies à travers l'ensemble du réseau. C'est donc la méthode la plus efficace dans les réseaux de Caltech, Princeton et Oklahoma.

La Weighted Community Hub-Bridge est la méthode la plus efficace pour le réseau de Georgetown (où $\mu = 0.522$) comme indiqué dans Figure 3.6 (c). Cette méthode dépend de

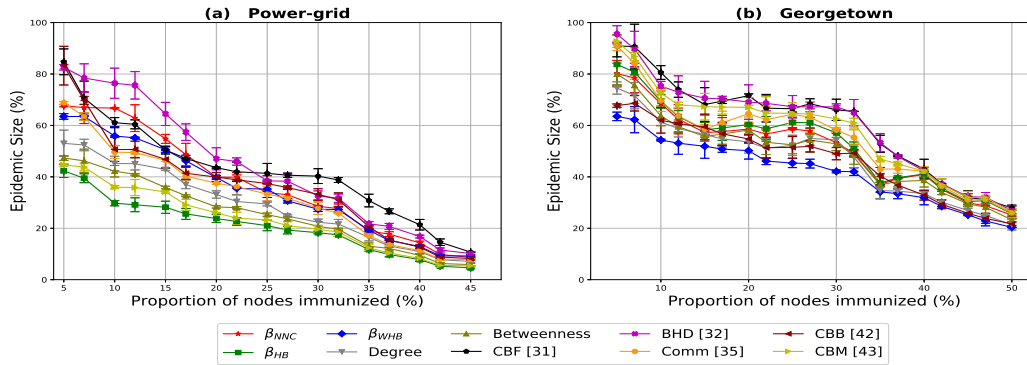


FIGURE 3.8 – Effet de différentes méthodes d’immunisation sur la taille de l’épidémie lors des simulations SIR effectuées sur le réseau (a) Power-grid et (b) Georgetown, avec $\lambda = 0.9$ et $\gamma = 0.2$.

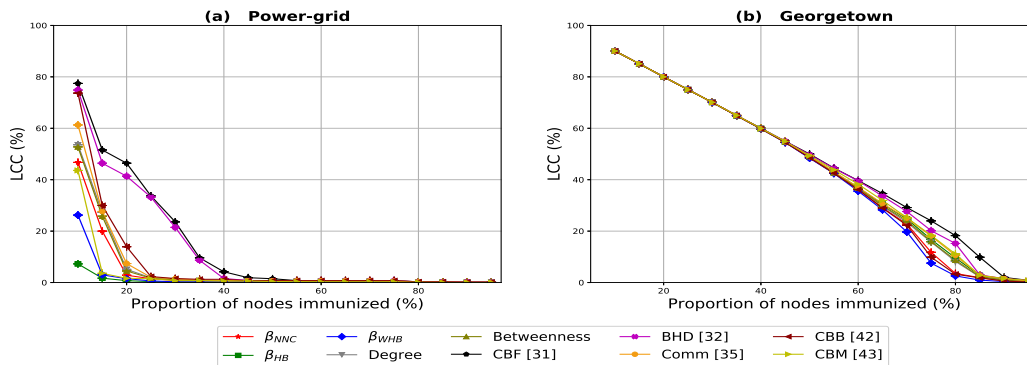


FIGURE 3.9 – La taille du plus grand composant connecté (LCC) pour diverses stratégies d’immunisation effectuées sur le réseau (a) Power-grid et (b) Georgetown. Chaque point est le résultat de la taille LCC en fonction de la proportion des noeuds immunisés.

la fraction des liens inter-communautaires pour chaque communauté au sein du réseau, ce qui nous permet de donner la pondération appropriée pour favoriser soit l’influence inter-communautaire ou intra-communautaire. C’est la raison pour laquelle elle surpasse les autres méthodes proposées dans le réseau de Georgetown qui n’a pas une structure communautaire solide. Enfin, ces résultats confirment l’influence primordiale du paramètre de proportion de mélange afin de choisir la stratégie la plus appropriée dans une situation donnée. Sur la base des résultats ci-dessus obtenus après avoir utilisé des réseaux du monde réel avec différentes structures, tailles et types, ce qui importe le plus, c’est la force de la structure communautaire.

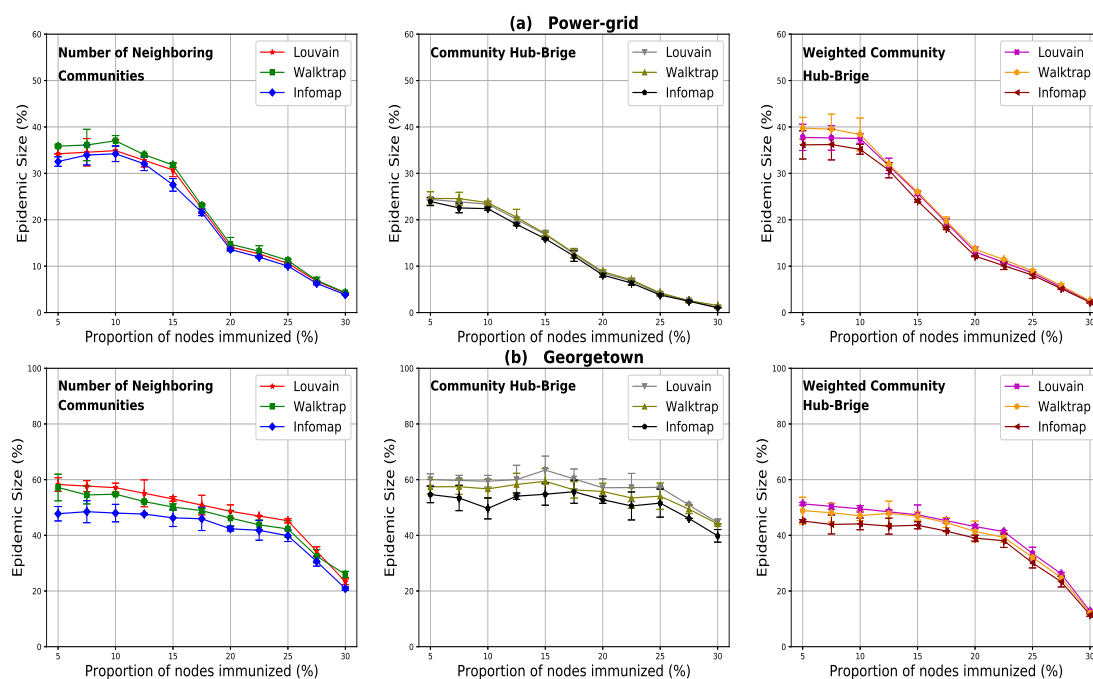


FIGURE 3.10 – Effet de divers algorithmes de détection communautaire sur les performances des méthodes communautaires proposées. Chaque point représente la taille de l'épidémie par rapport à la proportion des noeuds immunisés.

TABLEAU 3.3 – Le paramètre de mélange estimé μ des réseaux du monde réel.

Réseau	Power-grid	ca-GrQc	Princeton	Oklahoma	Caltech	Georgetown
μ	0.034	0.095	0.354	0.441	0.448	0.522

3.4.2.2 Comparaison avec les méthodes alternatives

La différence relative de la taille de l'épidémie entre les stratégies déterministes et stochastiques et les stratégies proposées est rapportée dans Figure 3.7. De même, avec les réseaux synthétiques, les stratégies stochastiques (CBF et BHD) fonctionnent mal par rapport aux stratégies déterministes. En effet, il apparaît clairement que ces deux types de méthodes sont bien séparés. Les résultats de l'évaluation comparative des stratégies déterministes sont tout à fait cohérents avec ce à quoi on aurait pu s'attendre. Les stratégies proposées sont globalement plus efficaces que leurs concurrents. Cela est d'autant plus vrai lorsqu'ils sont utilisés de manière appropriée.

Les figures de gauche de Figure 3.7 montrent la comparaison entre la Community Hub-Bridge et les méthodes alternatives. Elle surpasse les méthodes déterministes dans

les réseaux avec une structure communautaire forte (Power Grid et ca-GrQc) avec un gain minimal d'environ 10% sur les meilleures alternatives (Community-based Mediator et l'intermédierité). Ses avantages diminuent lorsque la force de la structure communautaire se relâche. Elle est toujours supérieure à l'intermédierité pour les réseaux de Princeton et d'Oklahoma, malgré leur proportion de mixage moyenne portée. Cependant, lorsque la structure de la communauté s'affaiblit (Caltech et Georgetown), elle est moins performante que la méthode de Community-based Mediator et l'intermédierité lorsque la fraction de noeuds immunisés est supérieure à 20%.

Les figures du milieu de Figure 3.7 représentent la comparaison entre le Nombre de communautés voisines et son alternative. Elle montre ses meilleures performances pour les réseaux avec des valeurs de proportions de mixage moyennes (Princeton, Oklahoma, Caltech) avec des gains supérieurs à 10% par rapport à l'alternative la plus performante (Community-based Betweenness). Cependant, elle réalise moins que la Community-based Betweenness alors qu'elle est toujours meilleur que les autres alternatives pour Georgetown (comme la stratégie des degré avec des gains inférieurs à 10%). Cependant, elle fonctionne dans certains cas moins que la Community-based Mediator, le degré et l'intermédierité dans les réseaux à forte structure communautaire (Power-grid et les réseaux de collaboration).

Les figures de droite de Figure 3.7 illustrent la comparaison entre la méthode du Weighted Community Hub-Bridge et les stratégies alternatives. Comme prévu, elle surpasse ses concurrents dans les réseaux avec une force de structure communautaire moyenne et élevée. Cependant, cela peut être pire que la Community-based Mediator et l'intermédierité pour les réseaux avec une forte structure communautaire.

Pour résumer, ces expériences révèlent que les algorithmes proposés sont très efficaces pour identifier les noeuds influents à sélectionner pour l'immunisation. Lorsqu'elles sont utilisées sur les réseaux appropriés en termes de force communautaire, elles surpassent les stratégies disponibles, simplement en utilisant des informations pertinentes sur la structure de la communauté.

3.4.2.3 Influence des paramètres du processus de diffusion

Pour tester la robustesse des résultats face à la variation des paramètres du modèle SIR, les résultats de simulations avec $\lambda = 0.9$ et $\gamma = 0.2$ sont rapportés dans Figure 3.8. Les résultats montrent qu'en augmentant le taux d'infection λ , une plus grande proportion de noeuds immunisés est nécessaire pour atténuer la propagation de l'épidémie. Ceci est valable pour toutes les méthodes d'immunisation testées. Par exemple, seulement 30% des noeuds doivent être supprimés (immunisés) pour toutes les stratégies visant à arrêter la propagation de l'épidémie dans le réseau Power-grid lorsque λ est égal à 0.1 (Figure 3.6 (f)), alors qu'environ 50% des noeuds doivent être immunisés lorsque λ est égal à 0.9 (Figure 3.8 (a)). Dans le réseau de Georgetown, environ 40% des noeuds doivent être immunisés lorsque λ est égal à 0.1 (Figure 3.6 (c)), tandis qu'un taux d'immunisation de 60% des noeuds est requis dans le cas d'une valeur élevée du taux d'infection (Figure 3.8 (b)), et cela vaut pour toutes les stratégies. Par conséquent, la probabilité qu'un noeud infecté contamine ses voisins augmente avec l'augmentation du taux d'infection λ . Ainsi, l'épidémie se propage à un rythme plus élevé. Par conséquent, une plus grande proportion de noeuds immunisés est nécessaire pour empêcher la propagation de l'épidémie.

Nous utilisons également la taille du plus grand composant connecté (Largest Connected Component *LCC*) pour confirmer l'efficacité des stratégies proposées. Figure 3.9 rapporte le *LCC* de diverses méthodes d'immunisation calculées sur deux réseaux du monde réel avec une force de structure communautaire différente (Power-grid et Georgetown network). Pour les deux réseaux dans Figure 3.9, on peut voir qu'en augmentant la proportion de noeuds immunisés, la taille du plus grand composant connecté diminue. Dans le réseau Power-grid, la courbe de la stratégie Community Hub-Bridge diminue plus rapidement que toutes les autres stratégies déterministes et stochastiques alternatives, comme indiqué dans Figure 3.9 (a). Ainsi, le réseau peut être décomposé efficacement en sélectionnant les noeuds influents selon cette stratégie. Elle est suivie par la Weighted Community Hub-Bridge et la stratégie de Community-based Mediator. Les méthodes stochastiques (Community Bridge Finder et Bridge-Hub Detector) fonctionnent mal pour diviser le réseau. La stratégie Weighted Community Hub-Bridge est la stratégie la plus efficace du réseau

TABLEAU 3.4 – L'information mutuelle normalisée (Normalized Mutual Information NMI) dans les réseaux Power-grid et Georgetown.

Réseau	Power-grid			Georgetown		
	Louvain	WalkTrap	Infomap	Louvain	WalkTrap	Infomap
Louvain	-	0.872	0.751	-	0.287	0.181
WalkTrap	-	-	0.818	-	-	0.429
Infomap	-	-	-	-	-	-

TABLEAU 3.5 – Le paramètre de mélange estimé μ , le nombre de communautés N_c et la modularité Q dans les réseaux Power-grid et Georgetown.

Réseau	Mesure	Algorithme de détection		
		Louvain	WalkTrap	Infomap
Power-grid	μ	0.034	0.036	0.039
	N_c	41	45	53
	Q	0.92	0.907	0.931
Georgetown	μ	0.522	0.515	0.487
	N_c	42	193	272
	Q	0.521	0.546	0.604

de Georgetown, comme indiqué dans Figure 3.9 (b). Sa courbe diminue plus rapidement que toutes les autres stratégies d'immunisation avant d'atteindre l'état d'équilibre. Par conséquent, l'immunisation des noeuds selon la stratégie du Weighted Community Hub-Bridge divise rapidement le réseau en plusieurs modules indépendants, ce qui conduit à son effondrement. Comme dans le cas du modèle SIR, la stratégie Community Hub-bridge est la stratégie la plus efficace dans les réseaux avec une structure communautaire bien définie (par exemple, un réseau Power-grid), tandis que la stratégie Weighted Community Hub-Bridge surpasse toutes les autres méthodes dans les réseaux présentant une structure communautaire lâche (par exemple, le réseau de Georgetown).

3.4.3 Influence des algorithmes de détection de communautés

Dans cette section, nous rapportons un ensemble d'expériences sur les réseaux Power-grid et Georgetown utilisant les algorithmes de détection de communauté WalkTrap et Infomap. Nous choisissons ces deux réseaux parce que Power-Grid a une structure communautaire bien définie tandis que Georgetown a une structure communautaire lâche.

Le but de ces expériences est d'obtenir une image claire des variations de la structure communautaire.

3.4.3.1 Comparaison des algorithmes de détection de communautés

Pour comparer la structure communautaire découverte par WalkTrap, Infomap et Louvain, nous utilisons l'information mutuelle normalisée (Normalized Mutual Information *NMI*) car cette mesure est couramment utilisée dans la littérature sur la détection communautaire Danon *et al.* (2005). Sa valeur estimée pour chaque réseau est reportée dans Table 3.4. Dans le réseau avec une structure de communauté bien définie (par exemple, le réseau Power-grid), les valeurs NMI sont élevées. Cela signifie que les structures communautaires découvertes par les trois algorithmes sont très similaires. Dans le réseau de Georgetown, les valeurs NMI sont inférieures à 0.5. Cela indique que les structures communautaires sont assez différentes.

Nous rapportons également la proportion de liens inter-communautaires, le nombre de communautés détectées et la valeur de modularité dans Table 4.5. Pour le réseau avec une structure de communauté bien définie, les trois algorithmes détectent presque le même nombre de communautés avec un nombre relativement plus grand pour Infomap. Cela confirme la similitude de la structure communautaire. Pour le réseau de Georgetown, nous observons une grande variation de ce paramètre. Ceci est un autre signe que les structures communautaires sont très différentes. La modularité mesure la qualité de la structure communautaire. Ses valeurs sont très élevées lorsque la structure de communauté du réseau est bien définie et relativement faibles pour le réseau avec une structure de communauté lâche. Selon ce paramètre, les performances des trois algorithmes sont comparables pour les réseaux avec des communautés bien définies. Infomap est l'algorithme le plus précis, suivi de WalkTrap puis Louvain lorsque la structure de la communauté est lâche.

Pour résumer, lorsque les structures communautaires sont bien définies (faibles valeurs de la proportion de liens inter-communautaires), les algorithmes découvrent les mêmes communautés, tandis que lorsque la structure communautaire est lâche, leurs résultats peuvent être très différents. De plus, les valeurs des paramètres de mixage utilisant les différents algorithmes sont très proches. Globalement, les trois algorithmes de détection

ont les mêmes performances dans les réseaux avec des communautés bien définies avec une légère préférence pour Infomap.

3.4.3.2 Influence des algorithmes de détection de communautés sur les méthodes proposées

Figure 3.11 montre la différence relative de la taille de l'épidémie entre les stratégies proposées et les alternatives en fonction de la fraction des noeuds immunisés. Les stratégies proposées sont évaluées sur le réseau Power-grid en (a) et (b) et le réseau Georgetown en (c) et (d) pour les algorithmes de détection WalkTrap et Infomap. Les figures de gauche représentent la comparaison entre la méthode Community Hub-bridge et les méthodes alternatives. Les figures du milieu représentent la différence entre le Nombre de communautés voisines et les méthodes alternatives. D'un autre côté, les figures de droite représentent la différence entre la méthode Weighted Community Hub-Bridge et les méthodes alternatives.

Dans les réseaux à forte structure communautaire, les performances du Community Hub-bridge sont toujours meilleures que les méthodes alternatives avec un gain moyen de 13% par rapport à la meilleure alternative (Community-based Mediator) pour Infomap, alors que le gain est d'environ 10% pour les algorithmes WalkTrap et Louvain. Cette méthode a un gain minimal de 3% pour Infomap. Les figures du milieu et de droite de Figure 3.11 (a) et (b) montrent que $\Delta R/R_0$ présente parfois une valeur négative à la fois pour les stratégies du Nombre de communautés voisines et du Weighted Community Hub-Bridge. Dans toutes les figures, elles sont moins performantes que CBM et l'intermédiarité dans le cas de l'algorithme de Louvain. Par conséquent, pour les trois algorithmes, la Community Hub-Bridge est la stratégie la plus efficace dans les réseaux avec une structure communautaire bien définie. De plus, elle montre ses meilleures performances après avoir détecté des communautés via l'algorithme Infomap.

Dans les réseaux avec une structure communautaire lâche, la Community Hub-Bridge est moins performant que la Community-Based Betweenness pour WalkTrap et Infomap, comme le montrent les figures de gauche de Figure 3.11 (c) et (d). Ces résultats sont similaires à ceux obtenus avec l'algorithme de Louvain. En revanche, la méthode du Nombre

de communautés voisines surpasse ses concurrents avec un gain moyen de 3% et 6% par rapport à sa meilleure alternative (Community-Based Betweenness) pour WalkTrap et Infomap respectivement (voir les courbes du milieu de Figure 3.11 (c) et (d)), alors qu'elle est moins performante que CBB pour Louvain. En effet, les algorithmes WalkTrap et Infomap détectent un plus grand nombre de communautés par rapport à l'algorithme de Louvain, qui découvre de grandes communautés. Cela rend le classement utilisant le Nombre de communautés voisines plus efficace dans le cas de WalkTrap et d'Infomap. Les figures de droite de Figure 3.11 (c) et (d) montrent que $\Delta R/R_0$ (en prenant la Weighted Community Hub-bridge comme stratégie de référence) a toujours une valeur positive. Ses performances sont meilleures que les méthodes alternatives avec un gain moyen de 10% et 15% par rapport à la meilleure alternative (CBB) pour WalkTrap et Infomap respectivement, alors que le gain est d'environ 6% pour Louvain. Par conséquent, la Weighted Community Hub-Bridge est la stratégie la plus efficace dans les réseaux avec une structure communautaire peu claire pour tous les algorithmes de détection testés. Pourtant, comme prévu, ses meilleures performances sont obtenues avec l'algorithme Infomap.

3.5 Conclusion

L'adoption d'une stratégie d'immunisation appropriée a suscité beaucoup d'intérêt parmi les chercheurs visant à contrôler toute menace de propagation de maladies infectieuses. Malgré la présence d'une structure communautaire dans tous les réseaux sociaux, cette propriété a été largement ignorée par les stratégies d'immunisation existantes. Dans cet article, trois stratégies communautaires sont proposées. Elles utilisent plus d'informations topologiques liées aux réseaux avec une structure communautaire qui ne se chevauchent pas. Les stratégies proposées sont évaluées dans différents réseaux synthétiques et réels. Pour vérifier leur efficacité, le modèle épidémique SIR est utilisé. Tout d'abord, les résultats montrent que les stratégies stochastiques sont moins performantes que les stratégies déterministes. En effet, comme ils n'ont pas accès à l'ensemble de la structure du réseau, il n'est pas facile d'exploiter leurs propriétés.

L'enquête approfondie montre également que, d'une manière générale, les stratégies

d'immunisation proposées ont une taille épidémique plus petite que les stratégies déterministes d'immunisation les plus influentes (Community-based Mediator et Community-based Betweenness) et la stratégie Comm conçue pour les réseaux avec une structure communautaire qui ne se chevauchent pas. La méthode Community Hub-Bridge est particulièrement adaptée aux réseaux à forte structure communautaire. La méthode du Nombre de communautés voisines montre son meilleur avec une structure communautaire de force moyenne tandis que la Weighted Community Hub-Bridge est plus efficace dans les réseaux à structure communautaire faible. De plus, la gamme de tailles des communautés joue un rôle important dans le processus de diffusion. Les stratégies d'immunisation sont plus efficaces lorsque la taille des communautés est petite. Les résultats des variations des paramètres du modèle SIR montrent que les stratégies d'immunisation affichent le même type de performances. Cependant, en augmentant le taux d'infection, une plus grande proportion de noeuds immunisés est nécessaire pour atténuer la propagation de l'épidémie. Pour tester l'efficacité des stratégies proposées indépendamment des modèles épidémiologiques, nous calculons la taille de la plus grande composante connectée LCC . Les résultats montrent que les méthodes proposées sont encore plus efficaces que les alternatives. De plus, nous rapportons également un ensemble d'expériences utilisant des algorithmes de détection Walktrap et Infomap pour découvrir des communautés. Les résultats des investigations montrent que les performances des méthodes proposées présentent le même comportement dans les réseaux avec une structure communautaire bien définie, c'est le cas pour les trois algorithmes de détection communautaire. Leurs performances sont différentes dans les réseaux ayant une structure communautaire n'est pas claire. Dans ce cas, les meilleurs résultats sont obtenus grâce à l'algorithme Infomap.

L'un des principaux avantages de ce travail est de montrer que des gains significatifs peuvent être réalisés en utilisant mieux les informations de de la structure communautaire. Il peut être étendu dans plusieurs directions. Premièrement, ces mesures peuvent être améliorées en utilisant des poids plus fins afin de les rendre plus robustes aux variations de la structure de la communauté. Maintenant que l'impact de la force de la structure communautaire a été clairement identifié, des versions stochastiques des stratégies proposées doivent être conçues.

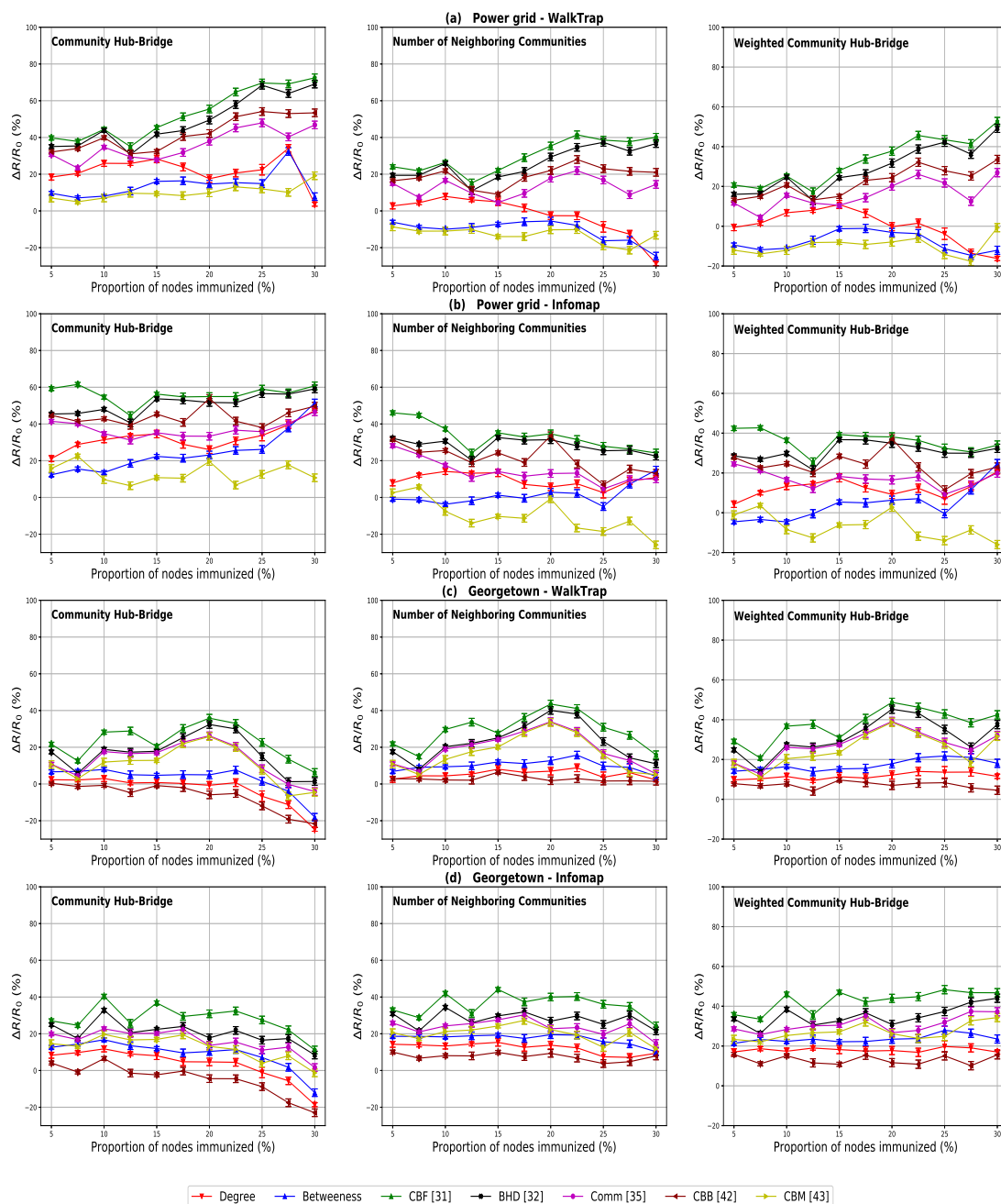


FIGURE 3.11 – Différence relative de la taille de l'épidémie en fonction de la proportion des noeuds immunisés. Les figures de gauche montrent la différence entre la méthode Community Hub-Bridge et les méthodes alternatives. Les figures du milieu représentent la différence entre la méthode du Nombre de communautés voisines et les méthodes alternatives. Les figures de droite montrent la différence entre la méthode du Weighted Community Hub-Bridge et les méthodes alternatives. Les méthodes d'immunisation sont effectuées sur le réseau Power-grid en (a) et (b) et le réseau de Georgetown en (c) et (d) pour les algorithmes WalkTrap et Infomap.

**FRAMEWORK POUR LE CALCUL DES CENTRALITÉS DANS LES RÉSEAUX
AVEC STRUCTURE COMMUNAUTAIRE****Sommaire**

2.1	Introduction	9
2.2	Stratégies d'immunisation	9
2.3	Modèle épidémique	10
2.4	L'immunisation des réseaux complexes non modulaires	11
2.5	L'immunisation des réseaux complexes modulaire	23
2.6	Conclusion	45

4.1 Introduction

L'identification des noeuds les plus influents d'un réseau a retenu beaucoup d'attention des chercheurs ces dernières années en raison de ses nombreuses applications Mantzaris (2014); Zhao *et al.* (2015); Luo *et al.* (2016); Tulu *et al.* (2018); Salavati *et al.* (2019); Berahmand *et al.* (2018). En effet, ces noeuds clés jouent un rôle majeur dans le contrôle de l'épidémie Wang *et al.* (2017), l'augmentation de la publicité sur un nouveau produit Medo *et al.* (2009) ou le contrôle de la propagation des rumeurs ou des épidémies Zhang *et al.* (2016). L'approche la plus populaire pour découvrir ces noeuds centraux consiste à quantifier leur influence à l'aide de mesures de centralité. Diverses mesures de centralité ont été proposées pour quantifier l'influence des noeuds en fonction de leurs propriétés topologiques Lü *et al.* (2016). La centralité des degrés, la centralité de l'intertermédarité,

la centralité de la proximité sont parmi les mesures de centralité les plus élémentaires et les plus utilisées.

La majorité des réseaux du monde réel présentent l'organisation modulaire des noeuds, la soi-disant structure communautaire Coscia *et al.* (2011); Cherifi (2014); Fortunato et Hric (2016); Jebabli *et al.* (2015, 2014). Bien qu'il y ait eu un énorme effort concernant la définition de cette propriété, il n'y a pas de consensus formel sur une définition qui capture la topologie d'une communauté. Celle-ci est intuitivement appréhendée comme des groupes de noeuds densément connectés où les individus interagissent les uns avec les autres plus intensément qu'avec ceux du reste du réseau. Par conséquent, les communautés sont des groupes de noeuds partageant certaines propriétés communes et jouent des rôles similaires dans le phénomène d'interaction au sein des réseaux. Outre leurs diverses définitions, les communautés se sont également avérées présenter un certain nombre de fonctionnalités intéressantes telles que le chevauchement des modules Palla *et al.* (2005). Certains noeuds peuvent donc être partagés par plusieurs communautés. En effet, dans certains réseaux sociaux, les individus peuvent participer simultanément à différents groupes, comme les collègues de travail, les amis ou la famille.

Des travaux antérieurs ont montré que la structure communautaire a un effet important sur le processus de propagation dans les réseaux Saxena *et al.* (2018); Kumar *et al.* (2018); Chakraborty *et al.* (2016); Taghavian *et al.* (2017). Cependant, les mesures de centralité classiques ne prennent pas en compte l'influence de cette propriété topologique majeure sur la dynamique de propagation Lü *et al.* (2016). Dans un réseau modulaire, on peut distinguer deux types de liens qui soutiennent le processus de diffusion : les liens qui relient des noeuds appartenant à la même communauté (liens intra-communautaires ou liens forts) et les liens qui relient les communautés (liens inter-communautaires ou liens faibles) Ferrara *et al.* (2012); Granovetter (1977). Les premiers exercent une influence locale sur le processus de diffusion (c'est-à-dire au niveau communautaire), tandis que les seconds ont une influence globale (au niveau du réseau). Par conséquent, nous pensons que ces deux types de liens doivent être traités différemment. En effet, les liaisons intra-communautaires contribuent à la diffusion dans les zones localisées densément connectées des réseaux, tandis que les liaisons inter-communautaires permettent la propagation vers

les zones éloignées du réseau. Supposons qu'une épidémie commence dans une communauté, car elle est fortement connectée, les liens intra-communautaires auront tendance à confiner l'épidémie au sein de la communauté, tandis que les liens inter-communautaires auront tendance à la propager aux autres communautés. Comme leur rôle est assez différent, nous proposons de représenter la centralité des réseaux modulaires par un vecteur bidimensionnel où la première composante quantifie l'influence intra-communautaire (ou locale) et la seconde composante quantifie l'influence inter-communautaire (ou globale) de chaque noeud individuel du réseau. Pour calculer ces composantes, nous devons diviser le réseau d'origine en un réseau local et global.

Le réseau local est obtenu en supprimant tous les liens inter-communautaires du réseau d'origine. Le réseau global est obtenu en supprimant tous les liens intra-communautaires du réseau d'origine. Notez aussi que si le réseau d'origine est constitué d'un seul composant connecté, le réseau global et local se divisent en plusieurs composants connectés. Par conséquent, le temps d'exécution de centralité aux réseaux à plusieurs composants est plus rapide. Puisque les noeuds doivent être classés en fonction de leurs valeurs de centralité, il est nécessaire d'adopter une stratégie basée sur une combinaison des deux composantes de ce vecteur. Diverses stratégies, qui incluent différents niveaux d'informations sur la structure de la communauté, peuvent être utilisées. Comme notre principale préoccupation, dans ce chapitre, est de mettre en évidence la nature multidimensionnelle de la centralité dans les réseaux modulaires plutôt que de concevoir des méthodes de classement optimales, quelques stratégies élémentaires sont évaluées. Des stratégies de combinaison simples sont étudiées, par exemple le module du vecteur de centralité ainsi qu'une combinaison linéaire pondérée de ses deux composantes.

L'approche proposée peut être résumée comme suit :

- Sélectionner une mesure de centralité standard donnée.
- Former le réseau local en supprimant tous les liens inter-communautaires du réseau modulaire.
- Calculer la composante locale de la centralité modulaire à l'aide de la centralité standard.

- Former le réseau global en supprimant tous les liens intra-communautaires du réseau modulaire.
- Calculer la composante globale de la centralité modulaire à l'aide de la centralité standard.
- Trier les noeuds selon une mesure de classement donnée dérivée de ce vecteur bidimensionnel.

Nous avons mené une série d'expériences afin de tester la pertinence de la centralité modulaire par rapport à ses homologues classiques. Compte tenu des mesures de centralité les plus influentes (degré, intermédiarité, proximité, vecteurs propres), les composantes locale et globale ont été évaluées séparément. De plus, des combinaisons simples des deux composantes ont été testées. Des expériences ont été menées sur des réseaux synthétiques à structure communautaire contrôlée et sur des réseaux du monde réel dans un scénario de propagation d'épidémie utilisant le modèle SIR. Les résultats montrent que les épandeurs identifiés par l'approche proposée sont plus influents que ceux visés par les mesures de centralité standard. Notez que ce travail ne prend en compte que les réseaux non orientés et non pondérés par souci de simplicité, mais les résultats peuvent être facilement étendus à des situations plus générales.

4.2 Centralité modulaire

Dans ce chapitre, nous limitons notre attention à la structure de communauté qui ne se chevauche pas (c'est-à-dire qu'un noeud appartient à une seule communauté). Notre objectif principal est de prendre en compte la structure de la communauté afin d'identifier les noeuds influents. En effet, dans les réseaux modulaires, un noeud a deux types d'influence : une influence locale qui est liée à ses caractéristiques communautaires et une influence globale liée à ses interactions avec les autres communautés. En se basant sur cette hypothèse, nous fournissons une définition générale de la centralité dans les réseaux modulaires. Nous concevons un algorithme générique pour calculer la centralité d'un noeud sous cette définition générale. L'extension de centralité modulaire peut être

naturellement déduite des diverses définitions existantes de centralité conçues pour les réseaux sans structure communautaire. Pour illustrer ce processus, nous donnons les extensions modulaires des mesures de centralité les plus influentes (intermédiarité, proximité et vecteurs propres).

4.2.1 Définitions

4.2.1.1 Composante locale de la centralité modulaire

Considérons un réseau noté $G(V, E)$, où $V = \{v_1, v_2, \dots, v_n\}$ et $E = \{(v_i, v_j) \mid v_i, v_j \in V\}$ désigne respectivement l'ensemble des sommets et des arêtes. Sa structure de communauté sans chevauchement \mathcal{C} est partitionnée en un ensemble de communautés $\mathcal{C} = \{C_1, \dots, C_m\}$ où C_k est la k^{me} communauté et m est le nombre de communautés. Le réseau local G_l est formé par l'union de tous les modules disjoints du réseau $G_l = \bigcup_{k=1}^m C_k$. Ces composants sont obtenus en supprimant tous les liens inter-communautaires entre les modules du réseau d'origine G . Chaque module représente une communauté C_k notée $C_k(V_k, E_k)$. Où $V_k = \{v_i^k \mid v_i \in V\}$ et $E_k = \{(v_i^{k_1}, v_j^{k_2}) \mid v_i, v_j \in V \text{ et } k_1 = k_2\}$, tandis que v_i^k fait référence à tout noeud v_i appartenant à la communauté C_k .

Pour une mesure de centralité sélectionnée β , nous définissons $\beta_L(v_i^k)$ comme la centralité locale du noeud $v_i \in V_k$. Elle est calculé séparément dans chaque module C_k du graphe local G_l .

4.2.1.2 Composante globale de la centralité modulaire

Considérons le réseau $G(V, E)$, le réseau global G_g est formé par l'union de tous les composants connectés du graphe qui sont obtenus après avoir supprimé tous les liens intra-communautaires du réseau d'origine $G(V, E)$. Supposons que $\mathcal{S} = \{S_1, \dots, S_p\}$ est l'ensemble des composants connectés révélés et $p = |\mathcal{S}|$ est la taille de l'ensemble \mathcal{S} , le réseau global est défini par $G_g = \bigcup_{q=1}^p S_q$. Chaque composant S_q est noté $S_q(V_q, E_q)$. Où $V_q = \{v_i^q \mid v_i \in V\}$ et $E_q = \{(v_i^{q_1}, v_j^{q_2}) \mid v_i, v_j \in V \text{ et } q_1 = q_2\}$, tandis que v_i^q fait référence à tout noeud v_i appartenant au composant S_q . Dans ce réseau, il peut y avoir des noeuds isolés (c'est-à-dire des noeuds qui ne sont pas directement liés à une autre communauté). Ces noeuds sont supprimés de G_g afin d'obtenir un réseau coupé formé uniquement par

des noeuds liés à différentes communautés par un saut. Par conséquent, l'ensemble des noeuds de G_g est défini ensuite par $V_g = \{v_i \in V \mid |\mathcal{N}_{v_i}^1| \neq 0\}$. Où $\mathcal{N}_{v_i}^n$ est l'ensemble de noeuds de voisinage accessible en n sauts. Il est défini par $\mathcal{N}_{v_i}^n = \{v_j \in V \mid v_i \neq v_j \text{ et } d_G(v_i, v_j) \leq n\}$, d_G est la distance géodésique.

Pour une mesure de centralité sélectionnée β , nous définissons $\beta_G(v_i^q)$ comme la centralité globale du noeud $v_i \in V_g$. Elle est calculée sur chaque composant connecté S_q inclus dans le graphe global G_g . N'oubliez pas que la mesure de centralité globale des noeuds isolés supprimés est égal à 0.

4.2.1.3 Centralité modulaire

C'est un vecteur à deux composantes. La première composante quantifie l'influence locale des noeuds dans leur propre communauté via le graphe local G_l , tandis que la seconde composante mesure l'influence globale des noeuds sur les autres communautés via les composants connectés du graphe global G_g . La centralité modulaire d'un noeud v_i est donnée par :

$$B_M(v_i) = (\beta_L(v_i^k), \beta_G(v_i^q)) \quad k \in \{1, \dots, m\} \text{ et } q \in \{1, \dots, p\} \quad (4.1)$$

Où β_L et β_G représentent respectivement la centralité locale et globale du noeud v_i .

4.2.2 Algorithme

La centralité modulaire est calculée comme suit :

Étape 1. Choisir une mesure de centralité standard β .

Étape 2. Supprimer tous les liens inter-communautaires du réseau d'origine G pour obtenir l'ensemble des communautés \mathcal{C} formant le réseau local G_l .

Étape 3. Calculer la mesure locale β_L pour chaque noeud de sa propre communauté.

Étape 4. Supprimer tous les liens intra-communautaires du réseau d'origine pour révéler l'ensemble des composants connectés \mathcal{S} formés par les liens inter-communautaires.

Étape 5. Former le réseau global G_g en fonction de l'union de tous les composants connectés. Les noeuds isolés sont supprimés de ce réseau et leur valeur de centralité globale est

égal à 0.

Étape 6. Calculer la mesure globale β_G des noeuds reliant les communautés en fonction de chaque composant du réseau global.

Étape 7. Ajouter β_L et β_G au vecteur de centralité modulaire B_M .

Le pseudo-code de l'algorithme pour calculer la centralité modulaire est donné dans algorithme 4.1.

Algorithm 4.1 Calcul générique de la centralité modulaire

Input : Graph $G(V, E)$, Centrality measure β

Output: A map $M(\text{node} : \text{centrality vector})$

Supprimer tous les liens inter-communautaires de G pour former le réseau local G_l

Supprimer tous les liens intra-communautaires de G pour former le réseau global G_g

Créer et initialiser un dictionnaire vide $M(\text{noeud} : B_M)$

$B_M(v_i) = (\beta_L(v_i^k), \beta_G(v_i^q))$ représente le vecteur de centralité, où chaque noeud v_i du réseau doit être associé à sa valeur locale et globale selon la mesure de centralité sélectionnée β

for each $C_k \subset G_l$, où $k \in \{1, \dots, m\}$ **do**

for each $v_i^k \in V_k$ **do**

 Calculer $\beta_L(v_i^k)$

$B_M(v_i).add(\beta_L(v_i^k))$

end

end

for each $S_q \subset G_g$, où $q \in \{1, \dots, p\}$ **do**

for each $v_i^q \in V_q$ **do**

 Calculer $\beta_G(v_i^q)$

$B_M(v_i).add(\beta_G(v_i^q))$

end

end

for each $v_i \in V$ **do**

$M.add(v_i, B_M(v_i))$

end

Retourner le dictionnaire M

4.2.3 Extensions modulaires des mesures de centralité standard

Afin d'illustrer le processus permettant d'étendre une centralité donnée définie pour un réseau sans structure communautaire à un réseau modulaire, nous donnons comme exemples les définitions modulaires de la centralité de l'intermédiarité, de la proximité et des vecteurs propres.

4.2.3.1 Centralité d'intermédiarité modulaire

La centralité d'intermédiarité modulaire prend en compte séparément les chemins qui commencent et finissent dans la même communauté et ceux qui commencent et finissent dans différentes communautés. Pour un noeud donné v_i , elle est représentée par le vecteur suivant :

$$B_M(v_i) = (\beta_L(v_i^k), \beta_G(v_i^q)) \quad k \in \{1, \dots, m\} \text{ and } q \in \{1, \dots, p\} \quad (4.2)$$

Où :

$$\beta_L(v_i^k) = \sum_{v_s, v_t \in C_k} \frac{\sigma_{st}^l(v_i)}{\sigma_{st}} \quad (4.3)$$

$$\beta_G(v_i^q) = \sum_{v_s, v_t \in S_q} \frac{\sigma_{st}^g(v_i)}{\sigma_{st}} \quad (4.4)$$

β_L mesure la centralité entre les noeuds dans leur propre communauté et β_G mesure la centralité entre les noeuds reliant les communautés. σ_{st}^l et σ_{st}^g comptent le nombre de chemins les plus courts reliant les noeuds v_s et v_t dans le réseau local et global respectivement, tandis que $\sigma_{st}^l(v_i)$ et $\sigma_{st}^g(v_i)$ représentent le nombre de chemins les plus courts reliant les noeuds v_s et v_t et en passant par v_i dans le réseau local et global respectivement.

4.2.3.2 Centralité de proximité modulaire

La centralité de proximité modulaire considère séparément les distances les plus courtes des noeuds provenant de la même communauté ou d'une autre communauté que

le noeud de départ v_i . Elle est défini comme suit :

$$B_M(v_i) = (\beta_L(v_i^k), \beta_G(v_i^q)) \quad k \in \{1, \dots, m\} \text{ and } q \in \{1, \dots, p\} \quad (4.5)$$

Où :

$$\beta_L(v_i^k) = \frac{1}{\sum_{v_j \in C_k} d_{ij}^l} \quad (4.6)$$

$$\beta_G(v_i^q) = \frac{1}{\sum_{v_j \in S_q} d_{ij}^g} \quad (4.7)$$

β_L et β_G mesurent respectivement la composante locale et globale de la centralité de proximité modulaire. d_{ij}^l et d_{ij}^g indiquent la longueur de la géodésique du noeud v_i au noeud v_j basé sur le réseau local et global respectivement.

4.2.3.3 Centralité des vecteurs propres modulaire

La centralité des vecteurs propres modulaire prend en compte séparément à la fois le nombre et l'importance des voisins appartenant à la même communauté et ceux appartenant à différentes communautés pour mesurer sa centralité. La centralité des vecteurs propres modulaire d'un noeud v_i est défini par le vecteur suivant :

$$B_M(v_i) = (\beta_L(v_i^k), \beta_G(v_i^q)) \quad k \in \{1, \dots, m\} \text{ and } q \in \{1, \dots, p\} \quad (4.8)$$

Où :

$$\beta_L(v_i^k) = \frac{1}{\lambda} \sum_{v_j \in C_k} a_{ij} \beta_L(v_j^k) \quad (4.9)$$

$$\beta_G(v_i^q) = \frac{1}{\lambda} \sum_{v_j \in S_q} a_{ij} \beta_L(v_j^q) \quad (4.10)$$

β_L et β_G mesurent respectivement la composante locale et globale de la centralité des vecteurs propre modulaires. $A = (a_{i,j})$ est la matrice d'adjacence du réseau, c'est-à-dire $a_{i,j} = 1$ si le sommet v_i est lié au sommet v_j , 0 sinon et λ est une constante.

TABLEAU 4.1 – Centralité du degré standard, composante globale et locale de la centralité du degré modulaire des noeuds dans un réseau d'exemple.

ID	4	11	18	22	1	2	3	16	5	7	8
β	7	7	5	5	4	4	4	4	3	3	3
β_L	3	7	3	3	4	2	4	2	3	2	3
β_G	4	0	2	2	0	2	0	2	0	1	0
ID	9	10	14	15	17	19	21	6	12	20	13
β	3	3	3	3	3	3	3	2	2	2	1
β_L	2	3	2	3	3	3	3	1	2	2	1
β_G	1	0	1	0	0	0	0	1	0	0	0

4.2.4 Exemple

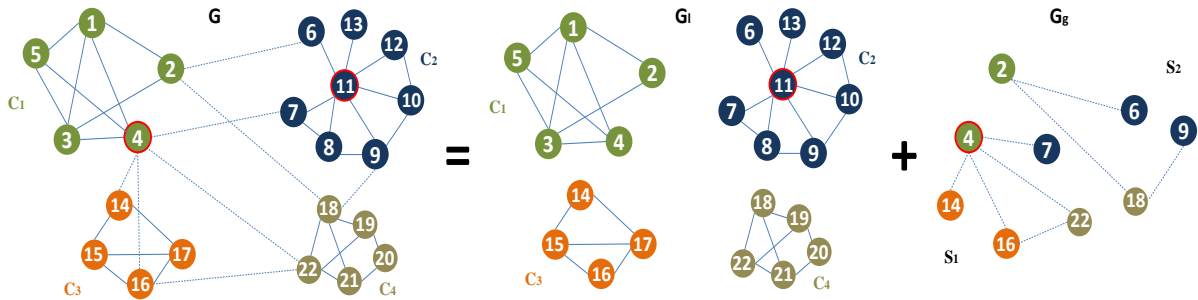


FIGURE 4.1 – Un exemple de réseau représentant le réseau local (G_l) et le réseau global (G_g) associé à un réseau modulaire (G) composé de quatre communautés qui ne se chevauchent pas.

L'exemple rapporté dans Figure 4.1 permet d'illustrer les deux types d'influence qui peuvent se produire dans un réseau modulaire. Dans un souci de simplicité, nous considérons la mesure de centralité du degré. Dans ce cas, v_4 et v_{11} sont les noeuds les plus influents car ils ont la valeur de degré la plus élevée ($\beta(v_{11}) = \beta(v_4) = 7$). Même s'ils partagent la même valeur de degré, l'influence qu'ils ont sur les autres noeuds du réseau n'est pas comparable. En effet, leur position dans le réseau est assez différente : v_{11} est intégré dans sa communauté, tandis que v_4 est à la frontière de sa communauté. L'inspection du réseau local et global nous donne une image claire de leurs différences. Comme le montre le réseau local G_l , le noeud v_{11} est le plus influent au niveau de la communauté car il est lié à tous les noeuds de sa communauté ($\beta_L(v_{11}^2) = 7$), tandis que v_4 n'est lié qu'à 3 noeuds de sa communauté ($\beta_L(v_4^1) = 3$). En fait, les deux noeuds v_1 et v_3 sont plus

influent que le noeud v_4 dans la communauté C_1 avec leurs valeurs de degré local plus élevées. En regardant le réseau global G_g , il apparaît clairement que le noeud v_4 est le noeud le plus influent au niveau du réseau car il est connecté à 4 noeuds à l'intérieur de son composant ($\beta_G(v_4^1) = 4$). Ces noeuds appartiennent à toutes les autres communautés du réseau (C_1 , C_2 et C_3). Par conséquent, v_4 est plus influent que le noeud v_{11} dans le réseau global G_g en raison de sa capacité à atteindre les différents modules du réseau par rapport à v_{11} qui n'a d'influence que localement (dans la communauté C_2).

Pour résumer, on peut remarquer à partir de cet exemple que lorsque nous considérons la centralité du degré, les centres communautaires sont les diffuseurs les plus influents localement en raison de leur capacité à atteindre un nombre élevé de noeuds dans leurs propres communautés. Les ponts qui sont reliés à différentes communautés sont les diffuseurs les plus influents à l'échelle global car ils permettent d'atteindre un nombre élevé de communautés sur tout le réseau.

4.2.5 Stratégies de classement basées sur la centralité modulaire

Afin de classer les noeuds en fonction de leur centralité, il est nécessaire de dériver une valeur scalaire du vecteur de centralité modulaire. Pour ce faire, nous pouvons procéder de différentes manières. Afin de mettre en évidence les caractéristiques essentielles de la centralité dans les réseaux modulaires, nous choisissons de considérer trois stratégies. Les deux premiers sont simples. En effet, un moyen simple de combiner les composantes de la centralité modulaire est d'utiliser le module et l'argument de ce vecteur. La troisième stratégie utilise plus d'informations sur la structure de la communauté afin de voir si cela peut être bénéfique.

Le module r du vecteur modulaire B_M d'un noeud v_i est défini par :

$$r(v_i) = \|B_M(v_i)\| = \sqrt{(\beta_L(v_i^k))^2 + (\beta_G(v_i^q))^2} \quad k \in \{1, \dots, m\} \text{ et } q \in \{1, \dots, p\} \quad (4.11)$$

L'argument φ du vecteur modulaire B_M d'un noeud v_i est défini comme suit :

$$\varphi(v_i) = \arctan\left(\frac{\beta_G(v_i^q)}{\beta_L(v_i^k)}\right) \quad k \in \{1, \dots, m\} \text{ et } q \in \{1, \dots, p\} \quad (4.12)$$

Nous proposons d'utiliser la tangente de l'argument car elle a une plage plus élevée que l'argument. Elle est définie par :

$$\tan(\varphi(v_i)) = \frac{\beta_G(v_i^q)}{\beta_L(v_i^k)} \quad k \in \{1, \dots, m\} \text{ et } q \in \{1, \dots, p\} \quad (4.13)$$

Notez que dans ces stratégies de classement, les informations utilisées sur la structure de la communauté sont très limitées. Comme nous prévoyons que l'intégration de plus de connaissances sur la structure de la communauté dans la stratégie de combinaison des composantes de la centralité modulaire peut améliorer l'efficacité de la méthode de classement, nous étudions également la soi-disant "mesure modulaire pondérée (Weighted Modular measure)". Elle est basé sur une combinaison linéaire des composantes du vecteur de centralité modulaire, pondérée par une mesure de la force des communautés.

La mesure modulaire pondérée α_W d'un noeud v_i est donnée par :

$$\alpha_W(v_i) = (1 - \mu_{C_k}) * \beta_L(v_i^k) + \mu_{C_k} * \beta_G(v_i^q) \quad (4.14)$$

Où $k \in \{1, \dots, m\}$, $q \in \{1, \dots, p\}$ et :

$$\mu_{C_k} = \frac{\sum_{v_i \in C_k} k^{inter}(v_i^k)}{\sum_{v_i \in C_k} k(v_i^k)} \quad (4.15)$$

Où μ_{C_k} est la fraction des liens inter-communautaires de la communauté C_k .

$k^{inter}(v_i^k)$ est le nombre de liens inter-communautaires du noeud v_i^k et $k(v_i^k)$ est le degré du noeud v_i^k .

La mesure modulaire pondérée fonctionne comme suit :

- Une communauté C_k , où prédominent les liens intra-communautaires, est fortement connectée et a donc une structure communautaire très bien définie. Si une épidémie débute dans une communauté aussi cohésive, elle a plus de chance de rester confinée que de se propager à travers les quelques liens qui permettent d'atteindre les autres communautés du réseau. Dans ce cas, la priorité doit être donnée à l'immunisation locale. Par conséquent, plus de poids est accordé à la composante locale de la centralité modulaire β_L pour cibler les noeuds les plus influents de la communauté car

elle est bien séparée des autres communautés du réseau.

- Une communauté C_k où prédominent les liens inter-communautaires a une structure communautaire non cohésive. Il est plus probable qu'une épidémie débutant dans cette communauté se diffuse aux autres communautés à travers les nombreux liens qu'elle partage avec les autres communautés. Par conséquent, plus de poids est accordé à la composante globale de la mesure de centralité modulaire β_G afin de cibler les noeuds qui peuvent propager l'épidémie plus facilement sur tout le réseau en raison de la structure communautaire lâche de C_k .

4.3 Cadre expérimental

Dans cette section, nous donnons quelques informations sur l'ensemble de données synthétiques et réelles utilisées dans l'évaluation empirique des mesures de centralité. Le processus de simulation SIR est rappelé, ainsi que la mesure des performances utilisée dans les expériences.

4.3.1 Données utilisées

4.3.1.1 Réseaux synthétiques

Afin de générer des réseaux modulaires artificiels aux propriétés topologiques contrôlées, le benchmark LFR est utilisé (voir subsection 3.3.1.1). Le paramètre de coefficient de mélange μ de cet algorithme vaut de 0 à 1. Il représente la proportion moyenne de liens entre un noeud et ceux situés en dehors de sa communauté. Ce paramètre permet de contrôler la force de la structure communautaire. Si sa valeur est faible, il existe peu de liens entre les communautés et elles sont bien séparées les unes des autres. Dans ce qui suit, nous désignons cette situation comme une "structure communautaire bien définie". Une valeur élevée de μ indique une structure communautaire très lâche. En effet, dans ce cas, un noeud partage plus de liens avec des noeuds en dehors de sa communauté qu'avec des noeuds à l'intérieur de sa communauté. Avec des valeurs μ allant de 0.2 à 0.45, la structure communautaire est appelée "structure communautaire à cohésion moyenne". Les réseaux peuvent avoir différentes tailles allant de dizaines à des millions de noeuds. De

TABLEAU 4.2 – Paramètres du réseau LFR

Nombre de noeuds	4000
Degré moyen	7
Degré maximum	80
Exposant de la distribution des degrés	2.5
Exposant de la distribution des tailles de communautés	2.5
Paramètre de mélange	0.1, 0.4, 0.7
Plage de taille des communautés	[15 200]

plus, il est également difficile de caractériser les degrés moyen et maximal car ils sont très variables. Par conséquent, nous choisissons pour ces paramètres des valeurs consensuelles tout en considérant également le temps de calcul des simulations. Ces paramètres sont rapportés dans Table 5.1.

4.3.1.2 Réseaux réels

Bien que le modèle LFR produise des réseaux assez réalistes, des propriétés incontrôlées telles que la transitivité et la corrélation de degré peuvent s’écarter considérablement de celles observées dans les réseaux du monde réel Orman *et al.* (2013). Par conséquent, il est nécessaire d’utiliser des réseaux du monde réel dans le processus d’évaluation. Afin de couvrir un large éventail de situations, nous avons sélectionné des réseaux d’origines diverses : réseaux sociaux en ligne, réseaux de collaboration, réseaux technologiques et des réseaux de communication. Tous les réseaux sont non orientés et non pondérés. L’algorithme de Louvain est utilisé pour dévoiler la structure communautaire de ces réseaux. Nous choisissons cette méthode d’optimisation gourmande pour sa simplicité. De plus, cet algorithme populaire s’est avéré être un bon compromis entre efficacité et complexité lorsqu’il est utilisé dans de nombreux types de réseaux différents Orman *et al.* (2011) Orman *et al.* (2012).

- **Réseaux sociaux** : Quatre échantillons du réseau Facebook sont utilisés. Le réseau ego-Facebook collecté auprès des participants à l’aide de l’application Facebook.¹ Leskovec et Mcauley (2012) et le réseau d’amitié Facebook de 3 universités américaines (Caltech, Princeton, Georgetown) collectées par Traud *et al.* (2012). Les noeuds représentent des individus (participants au sondage ou membres de l’Université) et les arêtes représentent

1. <http://snap.stanford.edu/data>

TABLEAU 4.3 – Description des propriétés structurelles des réseaux du monde réel. N est le nombre total de noeuds, E est le nombre d'arêtes. $\langle k \rangle$, k_{max} sont respectivement le degré moyen et le degré max. C est le coefficient de regroupement moyen (average clustering coefficient). α_{th} est le seuil épidémique du réseau.

Réseau	N	E	$\langle k \rangle$	k_{max}	C	λ_{th}
ego-Facebook	4039	88234	43.69	1045	0.605	0.009
Caltech	620	7255	43.31	248	0.443	0.012
Princeton	5112	28684	88.93	628	0.298	0.006
Georgetown	7423	162982	90.42	1235	0.268	0.006
Email-Eu-core	986	25552	33.24	347	0.399	0.013
Power-grid	4941	6594	2.66	19	0,107	0.092
CR-QC	4158	13428	5.53	81	0.529	0.059

des liens d'amitié en ligne entre deux individus. Dans le réseau universitaire, afin d'obtenir des données pertinentes pour la propagation des infections épidémiques, seule la relation des individus vivant dans le même dortoir ou étudiant la même majeure est prise en compte.

- **Réseau de communication** : Le réseau Email-Eu-core¹ a été généré à partir des données de messagerie d'une grande institution de recherche européenne. L'ensemble de données contient uniquement la communication entre les membres de l'institution. Chaque noeud correspond à une adresse e-mail et un bord est établi entre deux noeuds u et v , si au moins un e-mail a été échangé entre l'adresse u et l'adresse v .

- **Réseau technologique** : Power-Grid² est un réseau contenant des informations sur la topologie du Western States Power Grid des États-Unis. Un lien représente une ligne d'alimentation. Un noeud est soit un générateur, un transformateur ou une sous-station.

- **Réseau de collaboration** : GR-QC¹ (Relativité Générale et Cosmologie Quantique) a été collecté à partir de l'e-print arXiv et couvre les collaborations scientifiques entre les auteurs des articles soumis à la catégorie Relativité Générale et Cosmologie Quantique. Si un auteur i a co-écrit un article avec l'auteur j , le graphique contient une liaison de i à j .

Les propriétés topologiques de base de ces réseaux sont données dans Table 5.2.

2. <http://www-personal.umich.edu/~mejn/netdata/>

4.3.2 Simulations SIR

Pour évaluer l'efficacité des mesures de centralité, nous considérons un scénario de propagation d'épidémie utilisant le modèle Susceptible-Infecté-Rétabli (SIR) Moreno *et al.* (2002). Dans ce cadre, les noeuds peuvent être classés en trois classes : S(susceptible), I(infecté) et R(rétabli). Initialement, tous les noeuds sont définis comme noeuds susceptibles. Ensuite, une fraction donnée f_0 des noeuds les mieux classés selon la mesure de centralité testée est mise à l'état infecté. Après cette configuration initiale, à chaque itération, chaque noeud infecté affecte l'un de ses voisins susceptibles avec une probabilité λ . De plus, les noeuds infectés se transforment en noeuds rétablis avec la probabilité de rétablissement γ . Pour mieux caractériser la capacité de propagation, la valeur du taux de transmission λ est choisie supérieure au seuil épidémique du réseau λ_{th} . Les valeurs de seuil épidémique λ_{th} pour les réseaux utilisés dans cette étude sont rapportées dans Table 5.2. Dans toutes les expériences, nous utilisons la même valeur du taux de transmission ($\lambda = 0.1$). Naturellement, il est beaucoup plus grand que les valeurs du seuil épidémique λ_{th} de l'ensemble de données. La valeur de la probabilité de rétablissement est également constante ($\gamma = 0.2$). Nous choisissons cette petite valeur pour que chaque noeud infecté ait de nombreuses chances d'infecter ses voisins avec la probabilité de λ avant de passer au statut rétabli. Le processus se poursuit jusqu'à ce qu'il n'y ait plus de noeud infecté dans le réseau. Enfin, lorsque le processus de propagation s'arrête, le nombre de noeuds à l'état "Rétabli" R est utilisé pour mesurer l'efficacité de propagation de la fraction des noeuds initialement infectés. Plus la valeur de R est élevée, plus les noeuds initialement sélectionnés sont influents. Pour garantir l'efficacité du processus d'évaluation, les résultats des simulations SIR sont moyennés sur 200 réalisations indépendantes. Une description plus détaillée du processus de simulation SIR est donnée dans algorithm 4.2.

4.3.3 Critère d'évaluation

Afin de comparer l'efficacité du classement d'une mesure de centralité avec celle obtenue par une mesure de centralité de référence, nous calculons la différence relative de

Algorithm 4.2 Description du processus de simulation SIR

Input : Graphe $G(V, E)$,
 Mesure de centralité : β ,
 Fraction des épandeurs initiaux : f_0 ,
 Taux de transmission : λ ,
 Taux de rétablissement : γ ,
 Nombre de simulations : n

Output: Le nombre moyen de noeuds rétablis après les simulations SIR : r_{av} ,
 L'écart type : r_{dev}

Classer les noeuds selon la mesure de centralité β
 Définir tous les noeuds comme noeuds sensibles
 Calculer n_I le nombre de noeuds initialement infectés : $n_I \leftarrow \text{card}(V) * f_0$
 Sélectionner n_I des noeuds les mieux classés et changez leur état en état infecté
 Ajouter les noeuds initialement infectés à la liste des infectés $L_Infected$

Initialiser la liste de nombre des noeuds rétablis n_R obtenus après chaque simulation :

```

 $L\_nbrR \leftarrow EmptyList()$ 
for counter from 0 to  $n$  do
   $n_R \leftarrow 0$ 
  while  $L\_Infected \neq Null$  do
    Sélectionner un noeud infecté  $v$  dans la liste des infectés  $L\_Infected$ 
    for each noeud  $v'$  voisin de  $v$  do
      if  $v'$  est susceptible then
        Avec une probabilité  $\lambda$  définir le noeud  $v'$  comme infecté
         $L\_Infected.add(v')$ 
      end
      else
        Avec une probabilité  $\gamma$  définissez le noeud  $v'$  comme rétabli
         $L\_Infected.remove(v')$ 
         $n_R \leftarrow n_R + 1$ 
      end
    end
  end
   $n_R \leftarrow n_R - n_I$ 
   $L\_nbrR.add(n_R)$ 
end

```

Calculer le nombre moyen r_{av} et l'écart type r_{dev} des noeuds rétablis sur les n simulations en fonction de la liste L_nbrR

Renvoie r_{av}, r_{dev}

la taille de l'épidémie (expliqué dans subsection 3.3.4.1). Notez que le nombre final de noeuds rétablis après les simulations SIR pour la méthode standard est donné comme référence. Ainsi, une valeur positive de Δr indique une efficacité plus élevée de la méthode testée par rapport à la référence.

Des expériences approfondies ont été réalisées afin d'évaluer l'efficacité des extensions modulaires des centralités les plus populaires (degré, intermédiarité, proximité et centralité des vecteurs propres) par rapport à leur définition standard. Premièrement, la composante locale et globale des diverses mesures de centralité modulaire est comparée à leur contrepartie standard. Ensuite, les trois méthodes de classement basées sur la combinaison des composantes de la centralité modulaire sont également évaluées. Ces expériences sont menées sur des réseaux synthétiques et réels.

4.4 Résultats des réseaux synthétiques

Des réseaux avec différentes valeurs de paramètres de mélange ont été créés afin de mieux comprendre l'effet de la force de la structure communautaire sur la performance des différentes mesures de centralité. Figure 5.3 représente la différence relative de la taille de l'épidémie en fonction de la fraction de noeuds immunisés avec la mesure standard utilisée comme référence. Les valeurs des paramètres de mélange μ couvrent toute la gamme de la force de la structure communautaire.

4.4.1 Évaluation de la composante locale et globale de la centralité modulaire

4.4.1.1 Structure communautaire bien définie

Dans les réseaux avec une structure de communauté bien définie, la composante locale de la centralité modulaire surpasse toujours les mesures standard pour toutes les mesures de centralité comme indiqué dans les panneaux de gauche du Figure 5.3 (lorsque $\mu = 0.1$). Le gain est d'environ 20% par rapport à la mesure standard pour la centralité de proximité, de degré et des vecteurs propres. Le plus petit gain concerne la centralité de l'intermédiarité avec une valeur moyenne de 10%. Cependant, la composante globale de

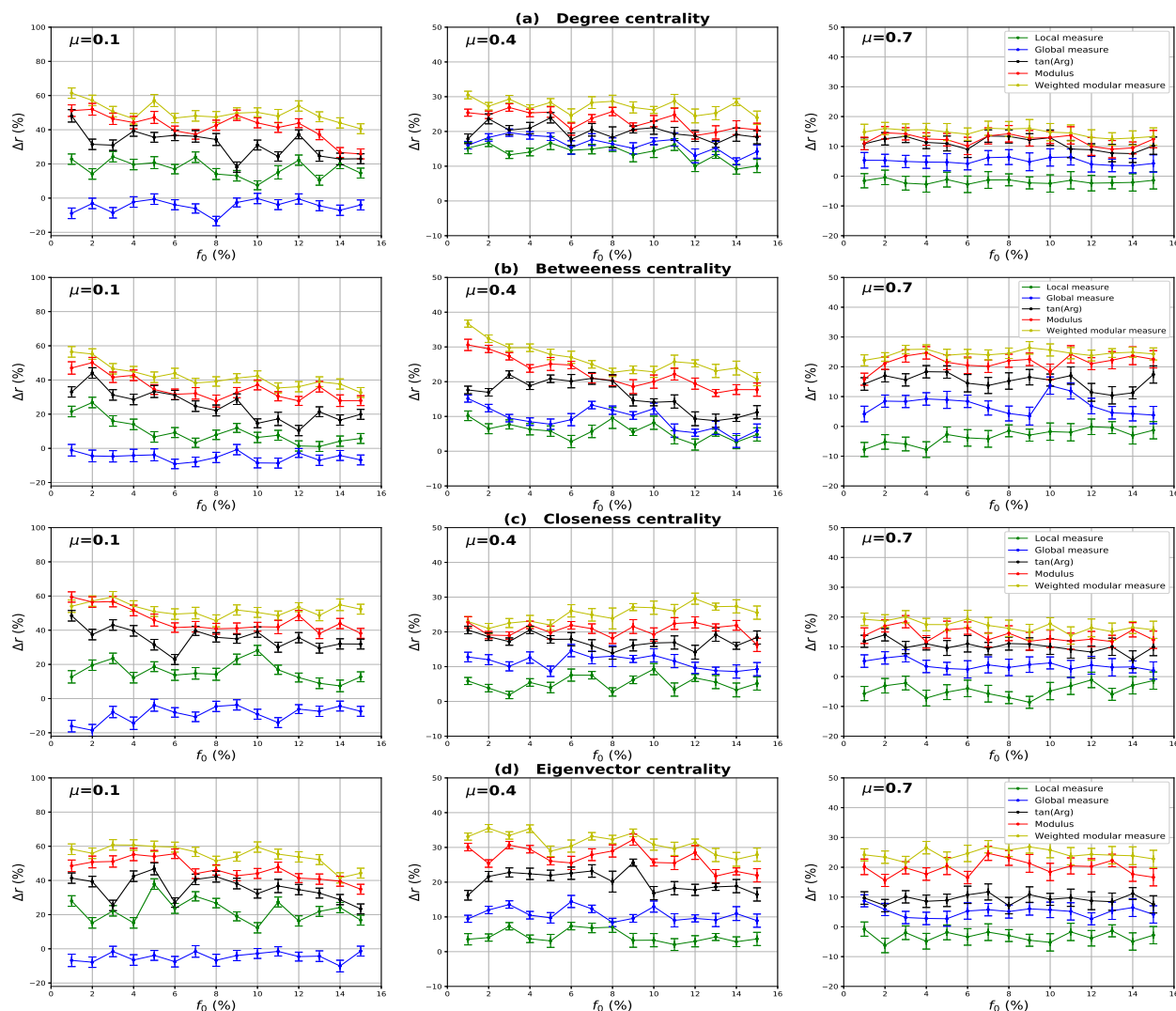


FIGURE 4.2 – La différence relative de la taille de l'épidémie Δr utilisant différentes mesures en fonction de la fraction des épandeurs initiaux f_0 est représentée pour les réseaux générés par l'algorithme LFR avec diverses valeurs μ . Pour chaque valeur μ , cinq exemples de réseaux sont générés. Une valeur de Δr plus élevée signifie une efficacité plus élevée des mesures locales, globales ou de combinaison par rapport à la définition de centralité standard. Les tailles épidémiques finales sont obtenues en calculant la moyenne de 200 simulations de modèle SIR par réseau, mesure de centralité et couverture initiale de propagation.

la centralité modulaire est toujours moins performante que les mesures standard. Ces résultats démontrent clairement qu'il est plus efficace d'immuniser les noeuds influents à l'intérieur des communautés lorsqu'il y a peu de liens inter-communautaires dans les réseaux. En effet, comme il y a peu de périphéries inter-communautaires, l'infection peut s'éteindre avant d'atteindre d'autres communautés. Ainsi, l'influence locale des noeuds est plus importante que l'influence globale dans les réseaux à forte structure communautaire.

4.4.1.2 *Structure communautaire à cohésion moyenne*

Les panneaux du milieu de Figure 5.3 montrent les performances des différentes méthodes de classement dans les réseaux à structure communautaire à force moyenne ($\mu = 0.4$). Dans ce cas, les composantes Globale et Locale de la centralité modulaire sont toujours plus efficaces que la centralité standard. Le gain de performance de la composante Globale de la centralité Modulaire est toujours supérieur à celui de la composante Locale. En effet, la composante globale surpasse la mesure standard avec un gain d'environ 12% pour l'intermédiarité, la proximité et la centralité des vecteurs propres. Le gain le plus important concerne la centralité du degré avec une valeur moyenne de 17%. La composante locale de la centralité modulaire fonctionne mieux que la mesure standard avec un gain d'environ 5% pour la centralité de l'intermédiarité, de la proximité et des vecteurs propres, et d'environ 12% pour la centralité des degrés. Ces résultats envoient un message clair : dans les réseaux avec une force de structure communautaire moyenne, l'influence globale est plus importante que l'influence locale. En effet, avec un plus grand nombre de liens inter-communautaires, il existe plus d'options pour propager les épidémies aux autres communautés du réseau.

4.4.1.3 *Structure communautaire lâche*

Le panneau de droite de Figure 5.3 rapporte la comparaison entre la centralité modulaire et les mesures de centralité traditionnelles dans les réseaux à structure communautaire non cohésive ($\mu = 0.7$). Il apparaît que la différence relative de la taille de l'épidémie entre la composante globale de la centralité modulaire et la centralité standard est toujours positive alors qu'elle est toujours négative pour la composante locale de la centralité

modulaire. Et cela est vrai pour toutes les mesures de centralité testées. En fait, il y a un gain d'environ 5% en utilisant la composante globale de la centralité modulaire, tandis que la composante locale est moins performante que la mesure traditionnelle avec une moyenne de 5% pour les mesures de centralité degré, intermédiarité, proximité et des vecteurs propres. Par conséquent, nous pouvons conclure que dans les réseaux à structure communautaire lâche, l'influence globale est dominante, même si la différence avec la mesure standard n'est pas aussi importante que pour les réseaux à structure communautaire moyenne. En effet, dans cette situation ($\mu = 0.7$), les liaisons inter-communautaires constituent la majorité des liaisons du réseau (environ 70% des liaisons se situent entre les communautés). En fait, comme la structure communautaire n'est pas bien définie, des différences mineures sont observées avec un réseau qui n'a pas de structure communautaire.

4.4.2 Évaluation des méthodes de classement basées sur la centralité modulaire

Figure 5.3 rapporte également la différence relative de la taille de l'épidémie Δr en fonction de la fraction des épandeurs initiaux f_0 pour les trois méthodes de classement (le module, la tangente de l'argument de la centralité modulaire et la centralité modulaire pondérée) et pour les différentes mesures de centralité et forces de structure communautaire. La première observation qui peut être faite à partir de ces résultats est que la combinaison des composantes de la centralité modulaire est toujours plus efficace que l'utilisation d'une seule composante ou de la centralité conventionnelle. Cette remarque vaut pour toutes les mesures de centralité étudiées et quelle que soit la force de la structure communautaire. De plus, le classement des trois méthodes de combinaison en termes d'efficacité est toujours le même. La mesure modulaire pondérée occupe le premier rang. Elle est suivie du module puis de la tangente de l'argument de la centralité modulaire. Nous pensons que cela est dû au fait que la mesure modulaire pondérée utilise plus d'informations sur la topologie de la structure communautaire que ses alternatives. En effet, les pondérations introduites dans cette mesure permettent d'ajuster localement l'importance relative de la composante locale et globale pour chaque communauté du réseau. Ainsi, la mesure modulaire pondérée peut s'adapter à la structure de chaque communauté du

réseau. Par conséquent, elle est plus efficace que les autres méthodes de classement proposées. L'un des principaux avantages de ce résultat est de mettre en évidence le fait que des gains significatifs peuvent résulter de l'amélioration de la combinaison des composantes locales et globales et qu'il peut encore y avoir des améliorations à faire dans cette direction. En d'autres termes, des mesures encore plus efficaces peuvent être obtenues si des informations supplémentaires pertinentes sur la structure de la communauté sont utilisées. De plus, on constate que les stratégies de classement affichent leurs meilleures performances dans des réseaux avec une structure communautaire bien définie. Par exemple, le module de la centralité modulaire surpasse la mesure standard avec un gain, en moyenne, de 40% dans les réseaux à forte structure communautaire, 25% dans les réseaux à structure communautaire de force moyenne et 20% dans les réseaux avec structure communautaire lâche pour toutes les mesures de centralité. Pour la mesure modulaire pondérée, le gain est d'environ 42% dans les réseaux à forte structure communautaire, 29% dans les réseaux à structure communautaire de force moyenne et 25% dans les réseaux à structure communautaire lâche pour toutes les mesures de centralité. Le gain diminue donc à mesure que la structure communautaire devient de plus en plus lâche. La raison derrière ce comportement est que la centralité locale est typique des réseaux avec une structure communautaire tandis que la centralité globale est une caractéristique des réseaux sans structure communautaire. À mesure que la proportion de mélange augmente, les différences avec les réseaux sans structure communautaire deviennent de moins en moins importantes. En effet, la taille du réseau global augmente jusqu'à ce qu'elle tende à représenter la majeure partie du réseau d'origine. Dans le cas limite, il s'agit d'un réseau sans structure communautaire et la centralité modulaire se réduit à sa composante globale qui est identique aux mesures de centralité classiques.

4.5 Résultats des réseaux du monde réel

Dans cette section, nous rapportons les résultats de l'ensemble d'expériences sur des réseaux du monde réel. Des expériences réalisées avec des réseaux synthétiques ont montré que la force de la structure communautaire joue un rôle majeur dans la détermination des performances des différentes mesures de centralité. Par conséquent, nous adoptons la

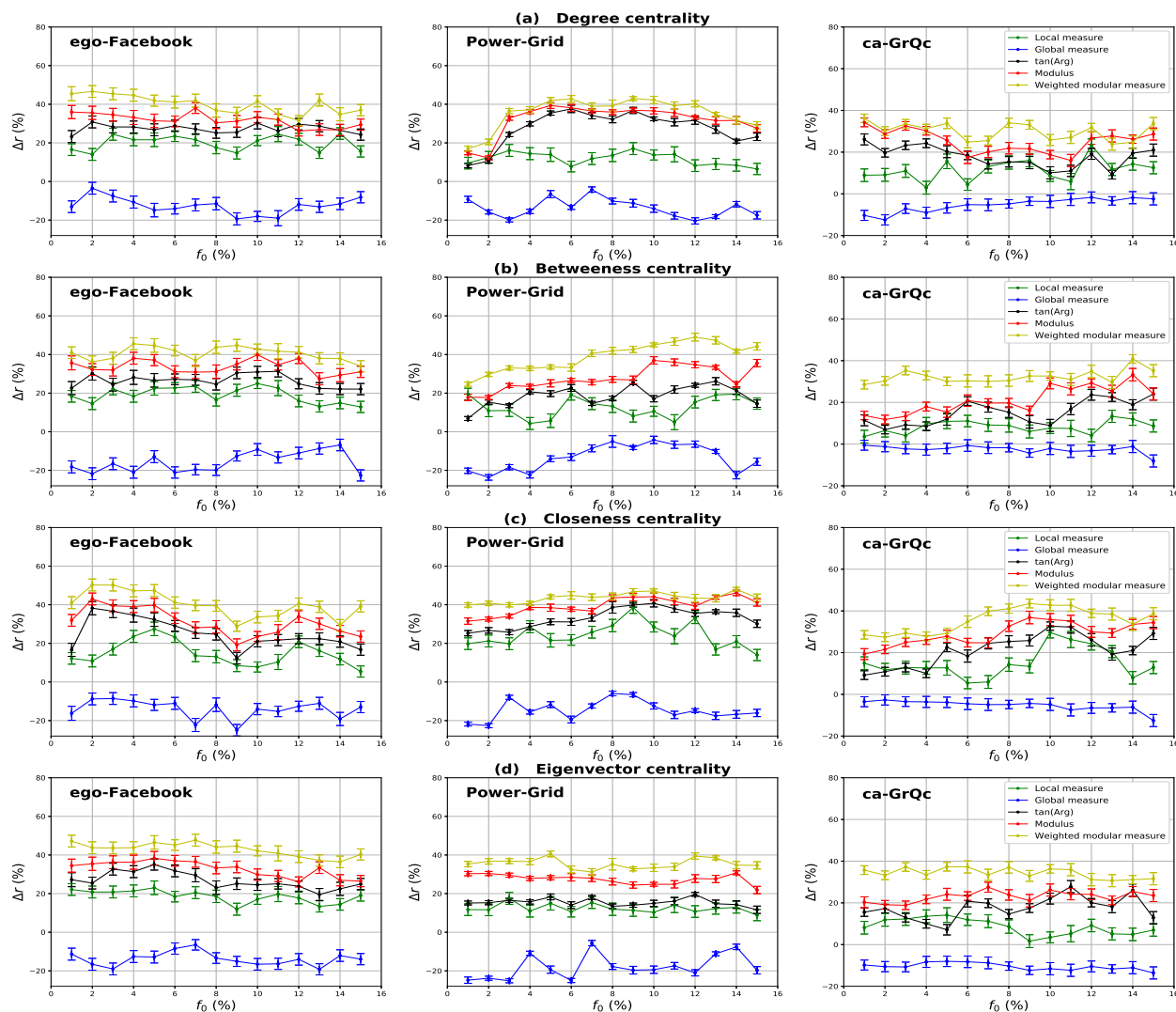


FIGURE 4.3 – Différence relative de la taille de l'épidémie Δr de diverses mesures de centralité en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité sont calculées sur des réseaux du monde réel avec une forte structure communautaire (réseaux ego-Facebook, Power-grid et ca-GrQc avec des valeurs de paramètres de mélange estimées égales respectivement à 0.03, 0.034 et 0.095).

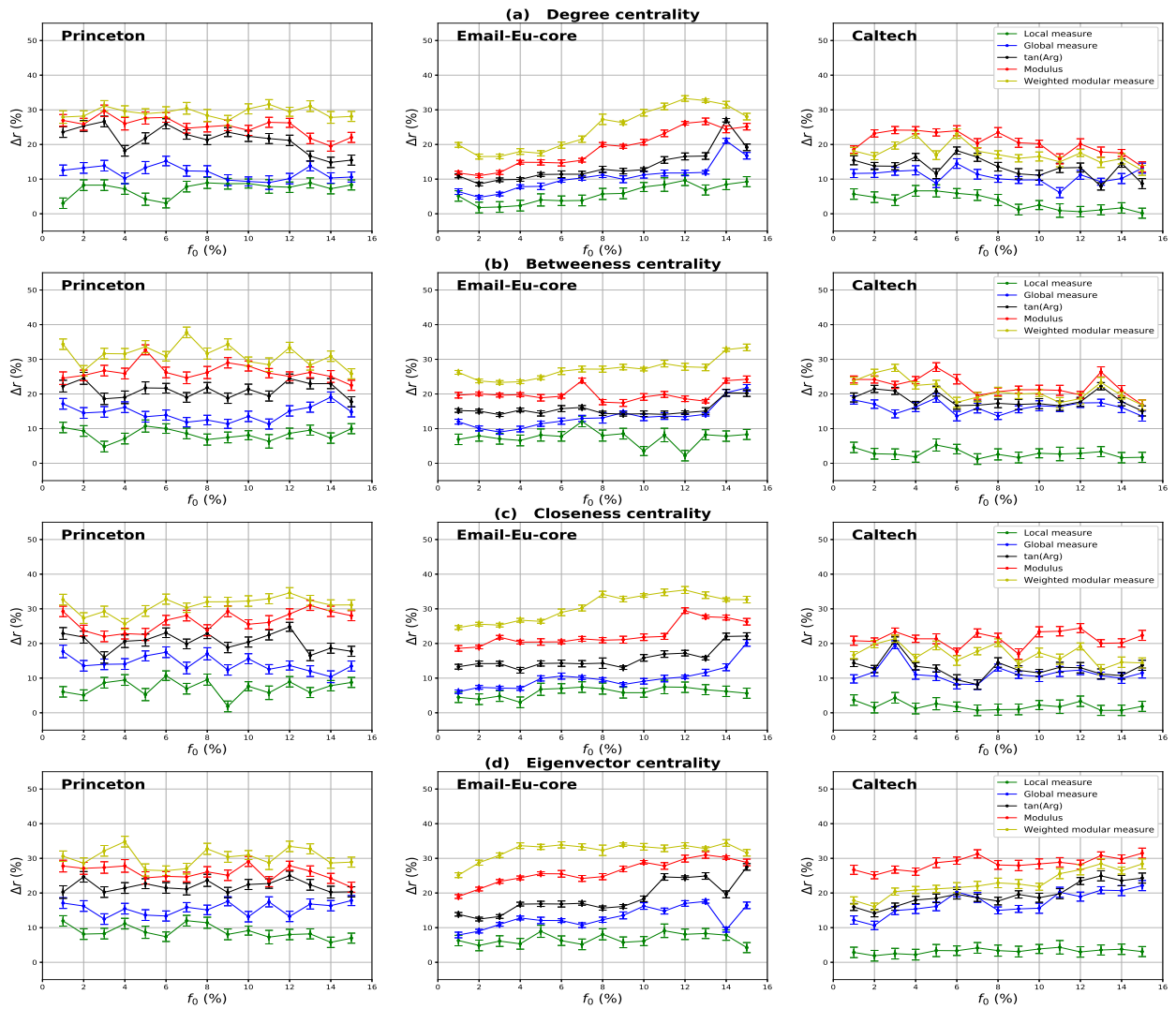
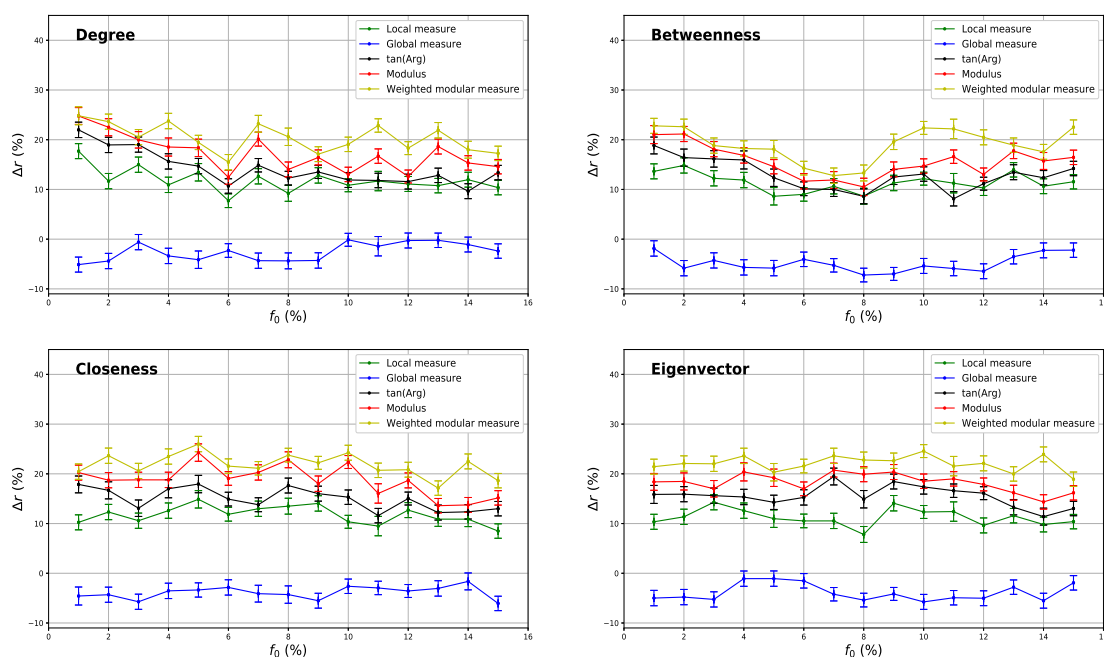


FIGURE 4.4 – Différence relative de la taille de l'épidémie Δr de diverses mesures de centralité en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité sont calculées sur des réseaux du monde réel avec une structure communautaire moyenne (réseaux Princeton, Email-Eu-core et Caltech avec des valeurs de paramètres de mélange estimées respectivement égales à 0.354, 0.42 et 0.44).

TABLEAU 4.4 – Les valeurs estimées du paramètre de mélange μ et la modularité Q des réseaux du monde réel.

Réseau	ego-Facebook	Power-grid	ca-GrQc	Princeton
μ	0.03	0.034	0.095	0.354
Q	0.834	0.934	0.86	0.753
Réseau	Email-Eu-core	Caltech	Georgetown	
μ	0.42	0.448	0.522	
Q	0.569	0.788	0.662	

FIGURE 4.5 – Différence relative de la taille de l'épidémie Δr de diverses mesures de centralité en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité sont calculées sur un réseau avec une structure communautaire non cohésive (Georgetown) avec une valeur de paramètre de mélange estimée égale à 0.522.

même présentation pour les réseaux du monde réel afin de relier les résultats de cet ensemble d'expériences avec ceux obtenus en utilisant des réseaux synthétiques. Une fois la structure de la communauté découverte à l'aide de l'algorithme de Louvain, le paramètre de proportion de mélange est calculé pour chaque réseau. Les valeurs estimées sont rapportées dans Table 5.3. D'après ces résultats, nous pouvons classer le réseau ego-Facebook, Power Grid et ca-GrQc comme des réseaux à forte structure communautaire. Princeton, Email-Eu-core et Caltech ont une structure communautaire de force moyenne tandis que Princeton a une structure communautaire faible.

4.5.1 Évaluation de la composante locale et globale de la centralité modulaire

4.5.1.1 Structure communautaire bien définie

La différence relative de la taille de l'épidémie entre les mesures de centralité modulaire et la mesure standard est indiquée dans Figure 4. Pour évaluer leurs performances dans des réseaux à forte structure communautaire, les réseaux ego-Facebook, Power grid et ca-GrQc sont utilisés. Pour ces réseaux, la valeur estimée des paramètres de mélange va de 0.03 à 0.095. Dans cette figure, nous remarquons que pour toutes les mesures de centralité testées, la mesure standard surpasse la composante globale de la centralité modulaire alors qu'elle est moins performante que sa composante locale. Prenons par exemple la centralité de l'intermédiarité. Avec une fraction des épandeurs initiaux égale à 8%, le gain en termes de taille d'épidémie pour la composante locale de la centralité modulaire par rapport à l'intervalle standard est de 19% pour le réseau ego-Facebook, de 14% pour Power-Grid et 9% pour ca-QrGc. Inversement, dans la même situation, la perte associée à l'utilisation de la composante globale de la centralité modulaire au lieu de l'intervalle standard varie de 4% à 11%.

Dans ces réseaux, les communautés sont densément connectées et il existe peu de liens entre les communautés. Par conséquent, dans la plupart des cas, des zones contagieuses se trouvent au coeur des communautés et la propagation de l'épidémie peut s'arrêter avant même d'atteindre le périmètre communautaire. Ainsi, il y a une faible probabilité qu'un pont (lien inter-communautaire) propage l'épidémie aux autres communautés. C'est la raison pour laquelle la composante locale de la centralité modulaire fonctionne toujours

mieux que la composante globale. De plus, nous pouvons également remarquer sur Figure 4 que lorsque la valeur du paramètre de mélange augmente (c'est-à-dire que la structure de la communauté s'affaiblit), la composante locale de la centralité modulaire devient moins efficace tandis que la composante globale fonctionne mieux. Cela est dû au fait que la composante globale augmente avec le nombre de liaisons inter-communautaires.

4.5.1.2 *Structure communautaire à cohésion moyenne*

Figure 7 montre la différence relative de la taille de l'épidémie entre les mesures de centralité modulaire et la mesure de centralité standard pour les réseaux du monde réel avec une force de structure communautaire moyenne. À cette fin, les réseaux Princeton, Email-Eu-core et Caltech sont utilisés car leurs valeurs de paramètres de mélange estimées vont de 0.354 à 0.448. Les résultats sont très clairs. Dans toutes les situations, la composante locale et la composante globale des mesures de centralité modulaire surpassent la mesure de centralité standard. De plus, il y a toujours un léger avantage pour la composante Globale sur la composante Locale. Pour placer ces idées sur un exemple simple, considérons la centralité de l'intermédierité avec une fraction initiale de noeuds infectés égale à 8%. La composante globale de la mesure l'intermédierité modulaire est plus efficace que l'intermédierité traditionnel avec un gain de 12% pour Princeton, 13% pour Email-Eu-core et 15% pour le réseau Caltech. Ces chiffres doivent être comparés à un gain de 8% pour Princeton, 7% pour Email-Eu-core et 3% pour Caltech en utilisant la composante locale de l'intermédierité modulaire au lieu de la centralité classique de l'intermédierité. On peut également remarquer que l'écart entre leurs performances respectives augmente à mesure que la valeur du paramètre de mélange augmente. En effet, à mesure que la structure communautaire s'affaiblit, l'influence relative de la composante globale de la centralité modulaire devient de plus en plus importante. Dans ces réseaux, les noeuds ont approximativement autant de liens internes qu'il y a de liens externes. Par conséquent, l'épidémie peut se propager facilement à toutes les communautés du réseau grâce au grand nombre de liens inter-communautaires. C'est la raison pour laquelle la composante Globale de la centralité modulaire surpasse toujours la composante Locale. De plus, la structure communautaire du réseau est toujours bien préservée, ce qui explique

que la centralité modulaire est plus efficace que la centralité classique.

4.5.1.3 *Structure communautaire lâche*

Figure 5 montre la différence relative de la taille de l'épidémie entre les mesures de centralité modulaire et la centralité standard pour le réseau de Georgetown. Avec une valeur de paramètre de mélange égale à 0.522, ce réseau est classé comme un réseau avec une structure communautaire faible. Dans tous les cas, la mesure standard fonctionne mieux que la composante locale de la centralité modulaire et elle est moins bonne que sa composante globale. En moyenne, il y a un gain d'environ 10% pour la composante globale contre une perte de 5% pour la composante locale des quatre mesures de centralité testées. Dans ce type de réseaux, les liens inter-communautaires prédominent, ce qui se traduit par une plus grande influence de la composante globale de la centralité modulaire. En effet, l'épidémie peut se propager plus facilement dans les différentes communautés du réseau grâce à la grande quantité de liens externes. De plus, nous remarquons que la différence relative de la taille de l'épidémie entre la composante globale de la mesure de centralité modulaire et la mesure standard diminue par rapport aux réseaux à structure communautaire moyenne. En effet, il existe de moins en moins de différences topologiques entre les réseaux à structure communautaire faible et les réseaux sans structure communautaire à mesure que la valeur du paramètre de mélange augmente.

4.5.2 **Évaluation des méthodes de classement basées sur la centralité modulaire**

Les figures 3 à 5 indiquent également la différence relative de la taille de l'épidémie entre les méthodes de classement basées sur la centralité modulaire et les mesures de centralité standard. Les résultats sont une preuve claire de l'efficacité de la centralité modulaire. Quelle que soit la stratégie de classement de la centralité modulaire adoptée, elle surpasse dans toutes les situations la composante locale et globale de la centralité modulaire et la centralité classique. Les améliorations en termes de performances par rapport à la centralité classique sont assez importantes. Par exemple, avec une fraction d'épancheurs initiaux égale à 8%, le module de la centralité de l'intermédialité modulaire permet un gain de 45% sur le réseau ego-Facebook, 28% sur Princeton et 24% sur Georgetown.

Étant donné que les stratégies de classement utilisent à la fois les informations locales et globales de chaque noeud, elles sont plus efficaces que les mesures reposant sur des informations locales ou globales prises séparément. De plus, la mesure modulaire pondérée est généralement la mesure la plus efficace dans la plupart des cas. Elle utilise la fraction des liens inter-communautaires comme information supplémentaire pour cibler les diffuseurs les plus influents de chaque communauté. Elle peut donner plus ou moins de poids à la composante locale et globale selon la force de la structure communautaire individuelle. Cela explique sa supériorité sur les autres mesures de classement. Pour résumer, ces expériences révèlent que la combinaison des composantes de la centralité modulaire permet de concevoir des méthodes de classement efficaces. De plus, l'utilisation d'informations plus pertinentes sur la structure de la communauté au niveau de la communauté permet de concevoir des méthodes de classement encore plus efficaces. À noter aussi que les mesures de classement affichent leurs meilleurs résultats dans des réseaux à forte structure communautaire.

4.5.3 Comparaisons avec les mesures alternatives

Figure 6 montre la fraction moyenne de la taille de l'épidémie par rapport à la proportion des épandeurs initiaux pour les composantes de centralité modulaire et leurs homologues standard. La performance des autres mesures de centralité modulaire présentées dans le chapitre précédent sont également indiquées dans cette figure (c.-à-d. Community centrality, Number of Neighboring Communities NNC and Community Hub-Bridge centrality CHB). La figure présente les résultats pour le réseau Power grid en (a) et le réseau de Georgetown en (b). Le premier a une structure communautaire bien définie tandis que le second a une structure communautaire lâche. Pour les deux réseaux dans Figure 6 (a) et (b) et pour toutes les mesures testées, on peut voir qu'en augmentant la proportion des épandeurs initiaux, la taille de l'épidémie augmente également. Cependant, cette évolution est plus lente dans Power-grid que dans Georgetown. En effet, à cause de ses modules bien séparés, les épidémies ne peuvent pas se déplacer facilement d'un module à l'autre dans Power-grid.

Globalement, les centralités de l'intermédiarité (la composante locale, globale, centra-

lité modulaire pondérée ainsi que l'indermédierité standard) surpassent toutes les autres alternatives. Notez que même la centralité standard de l'indermédierité est plus efficace que les autres alternatives. Ce résultat est indépendant de la force de la structure communautaire. Dans les réseaux avec une structure de communauté bien définie (c'est-à-dire le réseau Power-grid), on peut voir sur Figure 6 (a) que les centralités basées sur les vecteurs propres sont juste en dessous, suivies des centralités basées sur le degré. Les performances les plus faibles sont obtenues par les mesures de centralité basées sur la proximité. La Comm et le nombre de communautés voisines affichent une efficacité inférieure par rapport aux trois versions (mesures modulaires standard, locales et pondérées) des quatre centralités précédentes. Leur performance est cependant aussi bonne que les mesures globales avec une performance légèrement supérieure pour la mesure NNC. Ces deux méthodes ont tendance à cibler les ponts qui ont une forte influence globale sur le réseau. C'est la raison pour laquelle ils fonctionnent au même niveau que les mesures globales. De plus, la mesure Community Hub-Bridge a globalement les mêmes performances que l'indermédierité modulaire pondéré. Comme cette centralité intègre à la fois l'influence locale et globale des noeuds, elle fonctionne mieux que la plupart des autres mesures. Cela corrobore le fait que les deux dimensions doivent être prises en compte pour concevoir une mesure de centralité dans les réseaux modulaires.

Dans les réseaux avec une structure communautaire lâche (le réseau de Georgetown), on peut remarquer de Figure 6 (b) que les rangs des mesures de centralité en termes d'efficacité sont différents. Les mesures basées sur l'intermédierité occupent toujours le premier rang, les centralités basées sur le degré sont les deuxièmes mesures les plus performantes, suivies des mesures de centralité basées sur la proximité et les vecteurs propres. Dans ce type de réseaux, la centralité du Comm et le nombre de communautés voisines obtiennent de meilleurs résultats que la mesure Locale et les mesures de centralité standard. Leurs performances sont aussi bonnes que les mesures globales car elles mettent également en évidence des noeuds à forte influence globale. Ces noeuds peuvent jouer un rôle majeur dans le processus de diffusion dans les réseaux à structure communautaire non cohésive. Cela est dû à la grande quantité de liens inter-communautaires. D'où la meilleure performance de ces deux mesures communautaires. De plus, les courbes de la

TABLEAU 4.5 – Le paramètre de mélange estimé μ et la modularité Q dans les réseaux Power-grid et Georgetown.

Réseau	Mesure	Algorithme de détection	
		Louvain	Infomap
Power-grid	μ	0.034	0.038
	Q	0.92	0.93
Georgetown	μ	0.522	0.491
	Q	0.521	0.601

centralité Community Hub-Bridge sont généralement en haut de toutes les figures. Ainsi, il est, comme prévu, plus efficace que les composantes de centralité modulaire et standard de toutes les centralités testées. Elle présente des performances globales similaires à celles mesurées par la plupart des mesures modulaires pondérées (degré, proximité, vecteurs propres), à l'exception de l'intermédiarité modulaire pondérée qui fonctionne mieux. Les deux ciblent efficacement des noeuds avec une forte influence locale et mondiale dans le réseau.

4.5.4 Influence des algorithmes de détection de communautés

Dans cette section, nous rapportons un ensemble d'expériences menées sur les réseaux Power-grid et Georgetown en utilisant l'algorithme Infomap Rosvall et Bergstrom (2008) au lieu de Louvain. Ces deux réseaux du monde réel sont choisis en raison de la force différente de leur structure communautaire. Le réseau Power-grid a une structure communautaire solide tandis que le réseau de Georgetown a une structure communautaire non cohésive. L'objectif principal de ces expériences est d'obtenir une image claire de l'influence de l'algorithme de détection communautaire sur les performances des composantes de centralité modulaire. Les valeurs estimées de la proportion de liens inter-communautaires et de la modularité des algorithmes Infomap et Louvain sont rapportées dans Table 4.5.

Figure 8 illustre la différence relative de la taille de l'épidémie entre les extensions de centralité modulaire (du degré et d'intermédiarité) avec leurs homologues classiques. On peut déduire de cette figure que les méthodes locales, globales et combinées présentent globalement le même comportement que dans le cas de l'algorithme de Louvain. Dans les

réseaux avec une structure communautaire bien définie (par exemple, le réseau Powergrid), la mesure locale fonctionne toujours mieux que la classique avec un gain moyen de 23% et 20% pour les mesures de centralité de degré et d'intermédiarité respectivement. Alors que le gain moyen est de l'ordre de 20% et 19% dans le cas de l'algorithme de Louvain pour les deux mesures de centralité. La mesure standard, en revanche, surpasse la mesure globale pour toutes les proportions des épandeurs initiales. De plus, le gain global de la mesure modulaire pondérée est d'environ 45% et 42% pour les mesures de degré et d'intermédiarité lors de l'utilisation d'Infomap, alors qu'il est d'environ 40% pour les deux mesures de centralité lorsque l'algorithme de Louvain est utilisé. Ainsi, les extensions de centralité modulaire ont des performances légèrement meilleures pour l'algorithme Infomap. En effet, les algorithmes de détection d'Infomap et de Louvain ont à peu près les mêmes paramètres de mélange et mesures de modularité. Par conséquent, dans les réseaux avec une structure de communauté bien définie, les deux algorithmes de détection de communauté découvrent environ la même structure de communauté. Par conséquent, les performances de la centralité modulaire ont à peu près le même comportement.

Dans les réseaux avec une structure communautaire lâche (par exemple, le réseau de Georgetown), la mesure standard est toujours plus performante que la mesure locale. Cependant, la mesure globale donne de meilleurs résultats que la mesure standard avec un gain moyen de 17% et 16% pour les mesures de centralité de degré et d'intermédiarité respectivement. Alors que le gain moyen est d'environ 11% dans le cas de l'algorithme de Louvain pour les deux mesures de centralité. D'un autre côté, le gain global de la mesure modulaire pondérée est d'environ 29% et 25% pour les mesures de degré et d'intermédiarité lors de l'utilisation d'Infomap, alors qu'il est d'environ 20% et 19% pour les deux mesures de centralité respectivement lorsque l'algorithme de Louvain est employé. Dans ce réseau, l'algorithme Infomap a un paramètre de mélange relativement plus petit et une modularité plus élevée. Infomap est alors plus précis que l'algorithme de Louvain. Cela explique pourquoi les performances des composantes de centralité modulaire s'améliorent dans les réseaux avec une structure de communauté non cohésive lorsque l'algorithme de détection d'Infomap est utilisé.

Globalement, les résultats de cet ensemble d'expériences montrent que les variations

de la structure communautaire découverte ont un impact sur la performance des mesures de centralité. L'efficacité des mesures augmente avec la modularité de la structure communautaire.

4.6 Conclusion

Dans cette partie, nous proposons une définition générale des mesures de centralité dans les réseaux à structure communautaire qui ne se chevauchent pas. Elle est basée sur le fait que les liens intra-communautaires et inter-communautaires doivent être considérés différemment. En effet, les liaisons intra-communautaires contribuent à la diffusion dans les zones densément connectées du réseau, tandis que les liaisons inter-communautaires permettent la propagation globale vers les différentes communautés du réseau. Par conséquent, nous proposons de représenter la centralité des réseaux modulaires par un vecteur bidimensionnel, où la première composante mesure l'influence locale d'un noeud dans sa communauté et la seconde composante quantifie son influence globale sur les autres communautés. Sur la base de cette hypothèse, les mesures de centralité définies pour les réseaux sans structure communautaire peuvent être facilement étendues aux réseaux modulaires. En considérant les mesures de centralité les plus influentes comme des exemples typiques, nous avons défini leur extension modulaire. Des expériences basées sur un scénario de propagation épidémique utilisant à la fois des réseaux synthétiques et des réseaux réels ont été menées afin de mieux comprendre l'influence des deux composantes de la centralité modulaire. Tout d'abord, les résultats sur les réseaux synthétiques et réels sont assez cohérents. Il apparaît que la composante locale est plus efficace dans les réseaux avec une forte structure communautaire tandis que la composante globale prend les devants à mesure que la structure communautaire s'affaiblit. La comparaison avec la centralité classique tourne toujours à l'avantage de la centralité modulaire. Plus précisément, dans les réseaux à forte structure communautaire, la composante locale de la centralité modulaire surpasse la composante globale et la centralité standard, tandis que dans les réseaux à structure communautaire moyenne ou faible, la composante globale est plus performante que ses alternatives. De plus, on observe également que combiner les deux composantes

de la centralité modulaire pour classer les noeuds en fonction de leur influence est toujours plus efficace que d'utiliser une seule composante. En outre, un gain supplémentaire peut être obtenu si la stratégie de classement intègre plus d'informations sur la force de la structure communautaire. Nous effectuons également un ensemble d'expériences utilisant l'algorithme de détection d'Infomap pour découvrir des communautés. Les résultats montrent que les performances des variantes de centralité modulaire présentent le même comportement dans les réseaux avec une structure de communauté bien définie. Leur performance, cependant, est différente dans les réseaux avec une structure communautaire lâche. Dans ce cas, des résultats légèrement meilleurs sont obtenus avec l'algorithme Infomap.

**LE CALCUL DES CENTRALITÉS DANS LES RÉSEAUX COMPLEXES AVEC
UNE STRUCTURE COMMUNAUTAIRE QUI SE CHEVAUCHE**

Sommaire

3.1	Introduction	47
3.2	Stratégies proposées	49
3.3	Cadre expérimental	56
3.4	Résultats et discussion	60
3.5	Conclusion	79

5.1 Introduction

Dans le chapitre précédent, nous avons introduit un nouveau framework pour adapter les mesures de centralité qui sont agnostiques de la structure communautaire aux réseaux avec une structure communautaire non chevauchante. La "centralité modulaire" est un vecteur bidimensionnel où chaque dimension représente un type d'influence différent que les noeuds peuvent exercer dans le réseau. La première dimension quantifie leur influence locale dans leur propre communauté, tandis que la deuxième dimension quantifie leur influence globale sur les autres communautés auxquelles ils sont liés dans le réseau. Une analyse approfondie de la centralité modulaire à l'aide de données synthétiques et du monde réel a démontré sa supériorité sur les mesures classiques et les méthodes alternatives conçues pour les réseaux modulaires.

Dans ce chapitre, nous étendons ce framework aux réseaux dont la structure communautaire se chevauche. Dans ce cas, les noeuds qui se chevauchent introduisent un nouveau

défi. En effet, comme ils peuvent appartenir à plusieurs communautés, nous devons re-définir les notions locales et globales, notamment pour les noeuds qui se chevauchent. En effet, dans les réseaux à structure communautaire qui se chevauchent, les noeuds ont également deux types d'influence. Cependant, l'influence locale dépend de la nature des noeuds. Les noeuds qui ne se chevauchent pas ont une influence locale uniquement sur les noeuds qui appartiennent à leurs communautés via les liaisons intra-communautaires, tandis que les noeuds qui se chevauchent ont une influence locale sur les noeuds de toutes les communautés auxquelles ils appartiennent. De plus, les deux types de noeuds ont une influence globale sur les noeuds qui appartiennent à différentes communautés voisines via les liens inter-communautaires. Sur la base de cette idée, nous proposons ce que l'on appelle la "centralité modulaire chevauchante" (Overlapping Modular centrality), qui est une nouvelle représentation des mesures de centralité standard dédiées aux réseaux à structure communautaire chevauchante. Il s'agit d'un vecteur à deux dimensions, où chaque composante représente les différents types d'influence. De manière similaire au chapitre précédent, nous effectuons une analyse comparative de la composante locale et globale avec leur contrepartie classique conçue pour les réseaux non modulaires. De plus, nous proposons et étudions deux mesures de classement basées sur une combinaison des composantes locales et globales (le module du vecteur de centralité modulaire chevauchante et une combinaison linéaire pondérée des deux composantes). Comme il existe plusieurs façons de combiner les deux composantes, plutôt que de nous concentrer sur la méthode de classement la plus efficace, nous étudions les gains de performances qui peuvent être obtenus en utilisant plus d'informations sur la structure de la communauté.

5.2 Centralité modulaire chevauchante

Dans cette section, nous présentons les différents éléments qui constituent la base de l'approche proposée. La définition de la centralité modulaire chevauchante et l'algorithme pour calculer ses composantes sont donnés dans cette section.

Afin de calculer ces deux composantes, le réseau d'origine est décomposé en un réseau local qui capture les interactions à l'intérieur des communautés et un réseau global qui

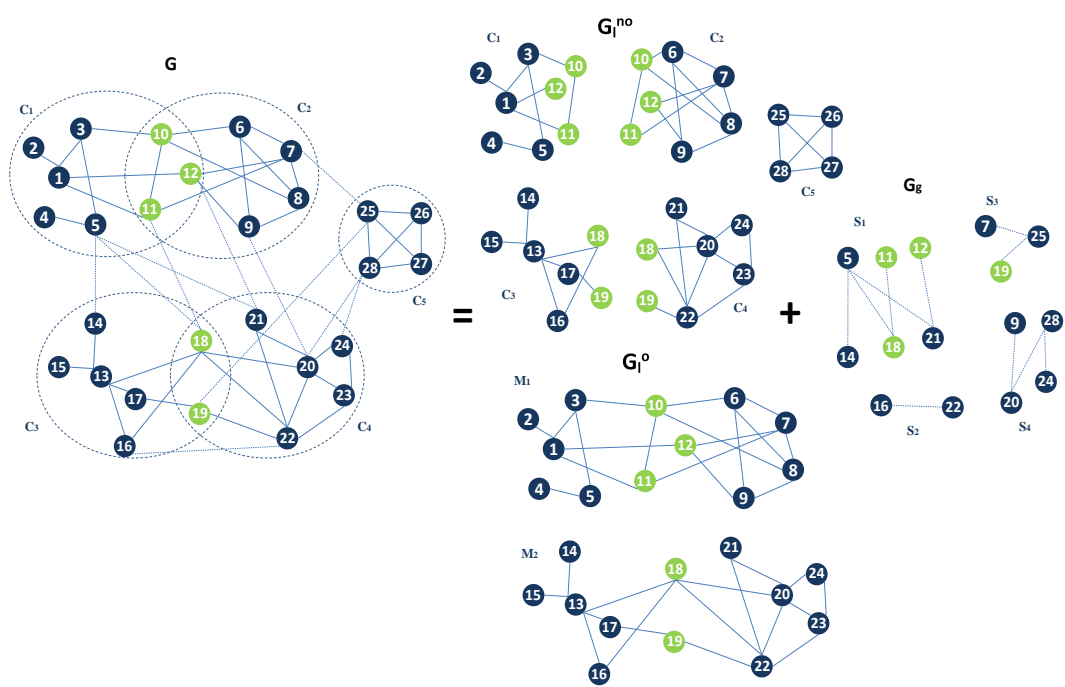


FIGURE 5.1 – Exemple de réseau illustrant comment former le réseau local pour les noeuds non chevauchants (G_l^{no}), le réseau local pour les noeuds chevauchants (G_l^o) et le Réseau global (G_g) à partir d'un réseau (G) avec une structure de communauté qui se chevauchent.

prend en compte des interactions entre les communautés. La définition du réseau local dépend du type de noeud considéré. Pour un noeud qui ne se chevauche pas, il se réduit à sa communauté unique, tandis que pour un noeud qui se chevauche, il inclut toutes les communautés auxquelles il appartient.

Soit $G(V, E)$ un simple réseau non dirigé. $V = V_o \cup V_{no}$ est l'ensemble des noeuds, où V_o et V_{no} représentent respectivement l'ensemble des noeuds chevauchants et non chevauchants, et E est l'ensemble des liens. $\mathcal{C} = \{C_1, ..C_k, C_m\}$ est l'ensemble des communautés et m est le nombre de communautés.

5.2.1 Définitions

5.2.1.1 Composante locale de la centralité modulaire chevauchante

Nous définissons le voisinage local d'un noeud qui chevauche M_y comme l'union des communautés auxquelles il appartient $M_y = \bigcup_{r=1}^s C_r$. Le réseau local G_l est l'ensemble formé par l'union des voisinages locaux $G_l = \bigcup_{y=1}^z M_y$, où $\mathcal{M} = \{M_1, ..M_y, .., M_z\}$ est l'ensemble des modules connectés obtenus et $z = |\mathcal{M}|$ est la taille de l'ensemble \mathcal{M} . Chaque module M_y est noté $M_y(V_y, E_y)$. Où $V_y = \{v_i^y \mid v_i \in V\}$ et $E_y = \{(v_i^{y_1}, v_j^{y_2}) \mid v_i, v_j \in V \text{ et } y_1 = y_2\}$, tandis que v_i^y fait référence à tout noeud v_i appartenant au module M_y . Pour une mesure de centralité donnée β , nous définissons $\beta_L(v_i^y)$ comme la composante locale de la centralité modulaire du noeud qui se chevauche $v_i \in M_y$. Elle est calculée sur tous les composants connectés du réseau local G_l .

Un noeud qui ne se chevauche pas est un cas particulier. Ce type de noeuds appartient à une seule communauté C_k . Afin de calculer la composante locale d'un noeud qui ne se chevauche pas, nous définissons le réseau local noté G_l comme l'union de tous les modules du réseau $G_l = \bigcup_{k=1}^m C_k$. Ces communautés isolées sont obtenues en supprimant tous les liens inter-communautaires du réseau d'origine G et en reproduisant tous les noeuds qui se chevauchent avec leurs liens intra-communautaires dans chacune de leurs communautés partagées. Chaque module représente une communauté C_k notée $C_k(V_k, E_k)$, où $V_k = \{v_i^k \mid v_i \in V\}$ et $E_k = \{(v_i^{k_1}, v_j^{k_2}) \mid v_i, v_j \in V \text{ et } k_1 = k_2 = k\}$, tandis que v_i^k fait référence à tout noeud v_i appartenant au communauté C_k . Prenons un noeud non

chevauchant $v_i^k \in C_k$ et une mesure de centralité donnée β . Sa composante de centralité locale $\beta_L(v_i^k)$ est calculée dans son voisinage local (sa communauté isolée) C_k en utilisant la définition habituelle de β .

5.2.1.2 Composante globale de la centralité modulaire chevauchante

Nous définissons le réseau global G_g comme l'union de tous les composants connectés du graphe qui sont obtenus après avoir supprimé tous les liens intra-communautaires du réseau d'origine $G_g = \bigcup_{q=1}^p S_q$, où $\mathcal{S} = \{S_1, \dots, S_p\}$ est l'ensemble des composants connectés révélés et $p = |\mathcal{S}|$ est la taille de l'ensemble \mathcal{S} . Chaque composant S_q est noté $S_q(V_q, E_q)$, où $V_q = \{v_i^q \mid v_i \in V\}$ et $E_q = \{(v_i^{q_1}, v_j^{q_2}) \mid v_i, v_j \in V \text{ and } q_1 = q_2\}$, tandis que v_i^q fait référence à tout noeud v_i appartenant au composant S_q .

Pour une mesure de centralité donnée β , nous définissons $\beta_G(v_i^q)$ comme la centralité globale du noeud $v_i \in S_q$. Elle est calculée sur tous les composants connectés S_q formant le graphe global G_g . Les noeuds qui ne font pas partie du réseau global reçoivent une valeur nulle.

5.2.1.3 Centralité modulaire chevauchante

Considérons une mesure de centralité β , définie pour un réseau sans structure communautaire (intermédiarité, degré, etc.), son extension modulaire chevauchante pour un noeud $v_i \in V$ est donnée par :

$$B_{OM}(v_i) = (\beta_L(v_i), \beta_G(v_i^q)) \quad q \in \{1, \dots, p\} \quad (5.1)$$

Où β_L est la composante locale de la centralité modulaire chevauchante d'un noeud v_i . Elle est égale à β_L^o si v_i est un noeud qui se chevauche $v_i \in V_o$ et β_L^{no} sinon ($v_i \in V_{no}$) β_G est la composante globale de la centralité modulaire chevauchante du noeud v_i .

5.2.2 Algorithme

La procédure de calcul de la centralité modulaire chevauchante peut être spécifiée comme suit :

Étape 1. Sélectionner une mesure de centralité standard donnée β .

Étape 2. Supprimer tous les liens inter-communautaires de G et répliquez tous les noeuds qui se chevauchent avec leurs liens intra-communautaires dans chacune de leurs communautés partagées pour former le réseau local pour les noeuds non chevauchants G_l^{no} comme illustré dans Figure 5.1.

Étape 3. Pour chaque communauté $C_k \in \mathcal{C}$, calculer la composante locale β_L^{no} de tous les noeuds qui se chevauchent pas $v_i^k \in V_k \setminus V_o$.

Étape 4. Former le réseau local des noeuds qui se chevauchent G_l^o défini comme l'ensemble obtenu par l'union des voisinages locaux comme illustré dans Figure 5.1.

Étape 5. Pour chaque module $M_y \in \mathcal{M}$, calculer la composante locale β_L^o pour tous les noeuds qui se chevauchent $v_i^y \in V_y \setminus V_{no}$.

Étape 6. Supprimer tous les liens intra-communautaires afin de former le réseau global G_g comme indiqué dans Figure 5.1.

Étape 7. Pour chaque module $S_q \in \mathcal{S}$, calculer la composante globale de la centralité modulaire chevauchante β_G de tous les noeuds $v_i^q \in V_q$. La centralité globale des noeuds isolés est égale à 0.

Étape 8. Ajouter β_L et β_G au vecteur de centralité modulaire superposé B_{OM} .

5.2.3 Schéma général

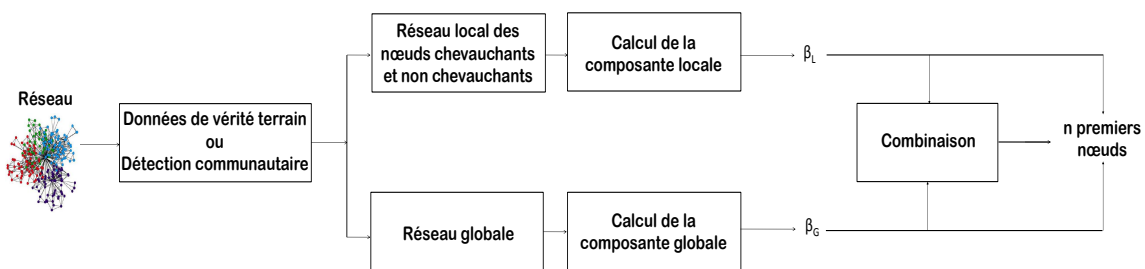


FIGURE 5.2 – Les principales étapes des méthodes de classement basées sur la centralité modulaire chevauchante proposée.

Figure 5.2 illustre la méthodologie utilisée pour classer les noeuds en fonction de leur importance dans les réseaux avec une structure de communauté qui se chevauchent. Tout d'abord, la structure communautaire du réseau doit être connue. Dans le cas où il n'y a pas de données de vérité terrain disponibles, un algorithme de détection de communauté peut

être utilisé pour découvrir la structure de la communauté Jebabli *et al.* (2018); Orman *et al.* (2012). Compte tenu de la structure communautaire, les réseaux locaux et globaux sont extraits. Le réseau local des noeuds non chevauchants est obtenu en supprimant tous les liens inter-communautaires du réseau d'origine et en répliquant tous les noeuds qui se chevauchent avec leurs liens intra-communautaires dans chacune de leurs communautés partagées, tandis que le réseau local des noeuds qui se chevauchent sont formés par l'union de leurs voisinages locaux. De plus, le réseau global est le résultat de la suppression de tous les liens intra-communautaires avec tous les noeuds qui n'ont aucune connexion en dehors de leurs communautés (les noeuds isolés). Étant donné une mesure de centralité telle que la centralité de degré, par exemple, sa composante locale de la centralité modulaire chevauchante est calculée sur le réseau local pour les noeuds non chevauchants et chevauchants. La composante globale de la centralité modulaire chevauchante est calculée sur le réseau global. Les noeuds qui ne sont pas dans le réseau global ont une valeur de centralité globale nulle. Enfin, les noeuds les plus influents sont sélectionnés en fonction de leur rang. Ils peuvent être classés selon la composante locale ou globale de la centralité modulaire chevauchante ou selon une mesure basée sur leur combinaison (reportez-vous à la sous-section suivante).

5.2.4 Combiner les composantes de la centralité modulaire chevauchante

Afin d'obtenir une liste de noeuds influents basée sur le vecteur bidimensionnel de la centralité modulaire chevauchante, une mesure de classement doit être dérivée. Nous considérons une combinaison simple et plus sophistiquée qui intègre des propriétés topologiques supplémentaires de la structure communautaire. Dans un premier temps, nous proposons de classer les noeuds en fonction du module des composantes de la centralité modulaire chevauchante. Le module r du vecteur de centralité modulaire chevauchante B_{OM} d'un noeud v_i est défini comme suit :

$$r(v_i) = \|B_{OM}(v_i)\| = \sqrt{(\beta_L(v_i))^2 + (\beta_G(v_i^q))^2} \quad q \in \{1, \dots, p\} \quad (5.2)$$

Où β_L est la mesure locale et β_G est la mesure globale.

Notez que les informations utilisées sur la structure de la communauté sont très limitées. Afin d'utiliser plus de connaissances sur la structure de la communauté, nous considérons ce que l'on appelle la "centralité modulaire chevauchante pondérée (Weighted Overlapping Modular Centrality)". Cette deuxième mesure est une combinaison linéaire pondérée des composantes locale et globale de la centralité modulaire chevauchante qui prend en considération des informations supplémentaires sur la structure de la communauté.

Considérons qu'une épidémie part du coeur d'une communauté. Si cette communauté est bien isolée, alors l'épidémie reste confinée, alors qu'elle se propage facilement aux autres communautés si la communauté est étroitement connectée aux autres communautés. Par conséquent, la capacité de diffusion d'un noeud dépend de sa position dans sa communauté en plus de la relation que sa communauté entretient avec les autres communautés du réseau. Motivés par cette idée, nous proposons une mesure de classement basée sur la centralité modulaire chevauchante qui peut s'adapter à l'environnement du noeud. Sa composante locale est pondérée par la taille de la communauté afin de prioriser les hubs situés dans les grandes communautés. La composante globale est également pondérée par le nombre de communautés voisines afin de favoriser les noeuds de pont partageant de nombreuses connexions avec plusieurs communautés. La mesure de centralité modulaire chevauchante pondérée pour un noeud donné v_i est alors donnée par :

$$\beta_w(v_i) = h(v_i) + b(v_i^q) \quad q \in \{1, \dots, p\} \quad (5.3)$$

Où :

$$h(v_i) = \begin{cases} (1 - \rho) * \frac{S_{C_k}}{N} * \beta_L^{no}(v_i^k) & \text{if } v_i \in V_{no} \text{ where } k \in \{1, \dots, m\}, \\ (1 - \rho) * \frac{S_{M_y}}{N} * \beta_L^o(v_i^y) & \text{if } v_i \in V_o \text{ where } y \in \{1, \dots, z\} \end{cases} \quad (5.4)$$

Et

$$b(v_i^q) = \rho * \frac{\alpha(v_i^q)}{n_c} * \beta_G(v_i^q) \quad \text{where } q \in \{1, \dots, p\} \quad (5.5)$$

Où S_{C_k} est la taille de la communauté C_k ($S_{C_k} = \text{card}(C_k)$), tandis que S_{M_y} est la taille

du voisinage local M_y ($S_{M_y} = \text{card}(M_y)$). N est la taille du réseau. $\alpha(v_i^q)$ est le nombre de communautés voisines connectées au noeud v_i^q tandis que n_c est le nombre total de communautés dans le réseau. ρ est la fraction des liens inter-communautaires. Elle est définie comme suit :

$$\rho = \begin{cases} \rho_{C_k} = \frac{\sum_{v_i^k \in C_k} k^{inter}(v_i^k)}{\sum_{v_i^k \in C_k} k(v_i^k)} & \text{if } v_i \in V_{no} \text{ where } k \in \{1, \dots, m\}, \\ \rho_{M_y} = \frac{\sum_{v_i^y \in M_y} k^{inter}(v_i^y)}{\sum_{v_i^y \in M_y} k(v_i^y)} & \text{if } v_i \in V_o \text{ where } y \in \{1, \dots, z\} \end{cases} \quad (5.6)$$

Où ρ_{C_k} est la fraction des liens inter-communautaires de la communauté C_k , tandis que ρ_{M_y} est la fraction des liens inter-communautaires du voisinage local M_y . $k(v_i)$ est le degré du noeud v_i et $k^{inter}(v_i)$ représente le nombre de liens inter-communautaires du noeud v_i .

Cette mesure fonctionne comme suit :

- Si le noeud v_i est situé à l'intérieur d'une communauté (ou à l'intérieur d'un voisinage local dans le cas d'un noeud qui se chevauche) avec une structure bien définie, alors ρ a une petite valeur. Ainsi, plus de poids est accordé à la mesure locale. Dans ce cas, la mesure modulaire chevauchante pondérée tend à cibler les hubs communautaires en raison de leur capacité à contribuer à la propagation de l'épidémie grâce au nombre élevé de liens intra-communautaires.
- Si le noeud v_i est situé à l'intérieur d'une communauté (ou à l'intérieur d'un voisinage local dans le cas d'un noeud qui se chevauche) avec une structure de communauté non cohésive, alors ρ a une valeur élevée. Ainsi, une plus grande importance est accordée à la mesure globale. Dans ce cas, la mesure modulaire chevauchante pondérée tend à cibler les noeuds qui peuvent diffuser l'épidémie plus facilement vers les communautés voisines sur tout le réseau via la quantité élevée de liens inter-communautaires.

TABLEAU 5.1 – Paramètres du réseau LFR

Nombre de noeuds	4000
Degré moyen	10
Degré maximum	20
Exposant pour la distribution des degré	2.7
Exposant pour la distribution des tailles de communautés	2.5
Paramètre de mélange	0.1, 0.4, 0.6
Proportion de noeuds qui se chevauchent	[5,50]%
Degré d'appartenance	[5,30]%

5.3 Cadre expérimental

Dans cet axe, des expériences approfondies ont été réalisées afin d'évaluer l'efficacité des composantes de la centralité modulaire chevauchante ainsi que les méthodes basées sur leurs combinaison (le module et la mesure pondérée). Ces extensions modulaires concerne les centralités les plus populaires (degré, intermédiarité, proximité et centralité des vecteurs propres). Des simulations SIR sont réalisées sur des réseaux synthétiques et du monde réel de différentes forces de structure de communauté. Afin de comparer les diverses extensions de la centralité modulaire chevauchante à leur contrepartie standard, nous calculons la différence relative de la taille épidémique (voir subsection 3.3.4.1). Le nombre final de noeuds rétablis après les simulations SIR pour la méthode standard est donné comme référence. Ainsi, on peut dire que la mesure de centralité testée est plus performante que la mesure standard si la valeur de Δr est positive.

5.3.1 Réseaux synthétiques

Dans cette partie de ce chapitre, nous utilisons également des réseaux LFR avec structure communautaire de différentes forces. Cela en variant la valeur du paramètre de mélange μ . Afin de générer des réseaux avec une structure de communauté qui se chevauchent, l'algorithme LFR permet aussi de contrôler le nombre de noeuds qui se chevauchent on et leur degré d'appartenance om (le nombre de communautés auxquelles ils appartiennent). Nous choisissons quelques valeurs consensuelles pour les paramètres qui sont rapportés dans Table 5.1.

5.3.2 Réseaux réels

Nous utilisons aussi des réseaux du monde réel pour valider les résultats obtenus avec les réseaux LFR. Afin de couvrir un large éventail de situations, ces réseaux proviennent de divers domaines tels que les réseaux sociaux, les réseaux de collaboration et les réseaux biologiques.

- **Réseaux sociaux** : Quatre échantillons du réseau Facebook sont utilisés. Le réseau ego-Facebook ainsi que le réseau Facebook de 3 universités américaines (Caltech, Princeton et Georgetown) sont utilisés (les détails sur ces réseaux sont rapportés dans la sous-section 4.3.1).

- **Réseau de collaboration** : Deux réseaux de collaboration sont utilisés. Le réseau de collaboration de la relativité générale et la cosmologie quantique GR-QC (voir la sous-section 4.3.1). Le deuxième réseau est Netscience¹. C'est un réseau de co-auteurs scientifiques dont la recherche se concentre sur la théorie des réseaux. Les nœuds représentent les auteurs et les arêtes relient chaque paire d'individus dont les noms apparaissent ensemble en tant qu'auteurs d'un article.

- **Réseau biologique** : Le réseau d'interaction de levure-protéine (Yeast-protein interaction²) est un réseau formé d'interactions de protéines contenues dans la levure. Un nœud représente une protéine et un lien représente une interaction métabolique entre deux protéines.

Les propriétés topologiques de base de ces réseaux sont rapportées dans Table 5.2. Afin de dévoiler la structure de la communauté qui se chevauchent de ces réseaux, l'algorithme Speaker-Listener Label Propagation Algorithm (SLPA) (sous-section 2.5.1.2) est utilisé. Nous choisissons cet algorithme car il s'agit de l'un des algorithmes les mieux classés selon la mesure des informations mutuelles normalisées (NMI) (voir sous-section 2.5.1.3). De plus, cet algorithme s'est révélé être un bon compromis entre l'efficacité et la complexité lorsqu'il est utilisé dans de nombreux types de réseaux différents Jebabli *et al.* (2018). Notons que certains réseaux utilisés avec la centralité modulaire (sans chevauchement) ne sont pas utilisés dans le cas de réseaux à structure qui se chevauchent.

1. <https://icon.colorado.edu>

2. <http://konect.uni-koblenz.de/networks>

TABLEAU 5.2 – Propriétés des réseaux du monde réel. N est le nombre total de noeuds, E est le nombre d’arêtes. $\langle k \rangle$, k_{max} sont respectivement le degré moyen et le degré max. C est le coefficient de regroupement moyen (average clustering coefficient). λ_{th} est le seuil épidémique du réseau.

Réseau	N	E	$\langle k \rangle$	k_{max}	C	λ_{th}
ego-Facebook	4039	88234	43.69	1045	0.605	0.009
Netscience	1589	2742	3.45	34	0.74	0.052
ca-GrQc	4158	13428	5.53	81	0.529	0.059
Princeton	5112	28684	88.93	628	0.298	0.006
Caltech	620	7255	43.31	248	0.443	0.012
Georgetown	7651	163225	90.42	1235	0.268	0.006
Yeast-protein interaction	1870	2277	2.43	56	0.051	0.095

Cela est dû au fait que l’algorithme SLPA ne détecte pas les noeuds qui se chevauchent dans ces réseaux.

5.4 Résultats des réseaux synthétiques

Le but de cet ensemble d’expériences est d’évaluer l’influence des différents paramètres de la structure communautaire (la force de la structure communautaire, la proportion des noeuds qui se chevauchent dans le réseau et leur degré d’appartenance aux communautés). Ils sont réalisés dans un scénario de propagation épidémique basé sur le modèle SIR sur des réseaux synthétiques générés à l’aide de l’algorithme LFR. Cet algorithme génère des réseaux avec des distributions de degrés de loi de puissance. Il permet de générer des réseaux avec une structure communautaire qui se chevauchent. La force de la structure de la communauté peut être ajustée via le paramètre de mélange μ . Il représente la fraction des liens inter-communautaires dans les réseaux et varie de 0 à 1. Il est défini sur une petite valeur pour générer des réseaux avec peu de liaisons entre les communautés. Ces réseaux ont une structure communautaire bien définie. Le paramètre de mélange est défini sur une valeur élevée pour générer des réseaux avec une structure de communauté lâche. Dans ces réseaux, les noeuds ont plus de connexions avec des noeuds à l’extérieur de leurs communautés qu’avec ceux à l’intérieur de leurs propres communautés. De plus, des réseaux avec une force de structure communautaire moyenne peuvent être générés avec des valeurs μ allant de 0.2 à 0.35. D’autres paramètres tels que on et om peuvent contrôler

la proportion des noeuds qui se chevauchent dans le réseau et leur degré d'appartenance respectivement. Nous comparons l'efficacité du classement associée aux extensions de centralité modulaire chevauchante du degré, d'intermédiarité, de proximité et de centralité des vecteurs propres avec ceux obtenus en utilisant leur homologue traditionnel.

5.4.1 Influence de la force de la structure communautaire

La valeur du paramètre de mélange μ contrôle la proportion de liens inter-communautaires et donc la force de la structure de la communauté ($0 < \mu < 1$). Pour les faibles valeurs de μ , les réseaux présentent une structure communautaire bien définie (peu de liens inter-communautaires). L'augmentation de la proportion de liens inter-communautaires brouille la structure communautaire. Afin de mieux comprendre l'effet de la force de la structure de la communauté sur la performance des différentes mesures de centralité, des réseaux avec trois valeurs de paramètres de mélange ($\mu = 0.1, 0.4, 0.6$) ont été générés et des simulations SIR ont été effectuées. Les courbes dans Figure 5.3 indiquent la différence relative de la taille de l'épidémie Δr par rapport à la fraction d'épandeurs initiaux obtenue avec la mesure standard ((a) Centralité du degré (b) Centralité de l'intermédiarité (c) Centralité de la proximité (d) Centralité des vecteurs propres) utilisée comme référence. Les valeurs négatives indiquent que la mesure standard est plus efficace, tandis que c'est le contraire pour les valeurs positives.

5.4.1.1 Évaluation de la composante locale et globale de la centralité modulaire chevauchante

Les résultats pour les réseaux avec une structure de communauté bien définie ($\mu = 0.1$) sont rapportés sur le panneau de gauche de Figure 5.3. On peut remarquer que la composante locale de la centralité modulaire chevauchante surpasse toujours la mesure standard. Le gain sur la mesure standard est d'environ 25% pour la centralité de degré, et il est d'environ 22% pour la centralité d'intermédiarité. Le gain le plus faible concerne les mesures de centralité de proximité et des vecteurs propres avec une valeur moyenne de 20%. Cependant, les mesures classiques sont plus efficaces que la composante globale de la centralité modulaire chevauchante. Dans cette situation, il y a peu de connexions

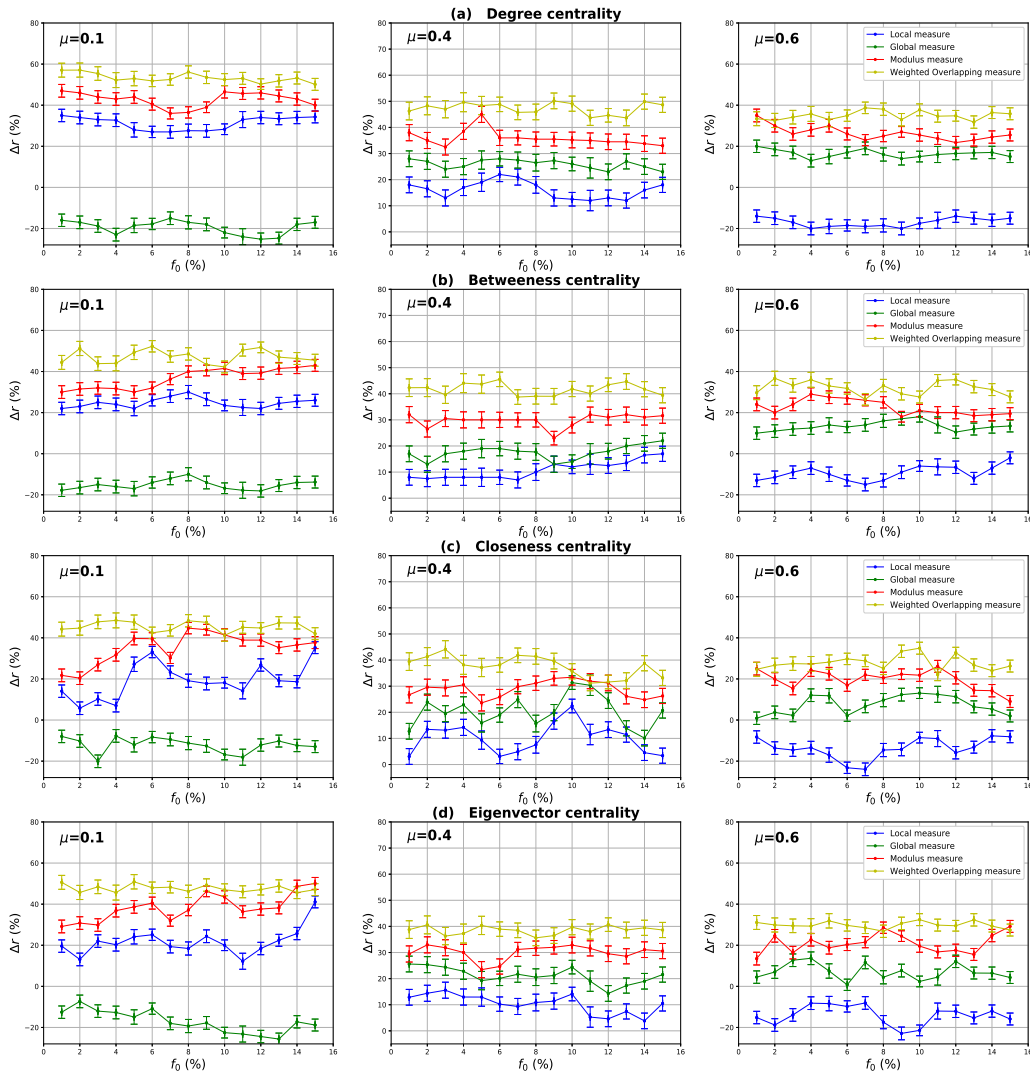


FIGURE 5.3 – La différence relative de la taille de l'épidémie Δr en fonction de la fraction des épandeurs initiaux f_0 , où $\Delta r = (R_c - R_s/R_s)$, R_c et R_s sont le nombre final de noeuds rétablis pour la mesure de centralité testée et la centralité standard, respectivement. Les mesures de centralité du degré (a), de l'intermédiarité (b), de la proximité (c) et des vecteurs propres (d) dérivées de la centralité modulaire chevauchante sont comparées à leur homologue standard conçue pour les réseaux sans structure communautaire. Des réseaux synthétiques générés par l'algorithme LFR avec différentes forces de structure de communauté sont utilisés (la valeur de leur coefficient de mélange est égale à 0.1, 0.4 et 0.6). Nous avons défini la proportion de noeuds qui se chevauchent on à 10% de la taille du réseau et le degré d'appartenance om à 10% du nombre total de communautés. Chaque valeur sur les courbes est obtenue en faisant la moyenne des résultats de 200 simulations SIR par méthode et fraction de noeuds initialement infectés. Une valeur positive de Δr signifie que la centralité testée est plus efficace que la centralité standard.

entre les communautés, et il est plus efficace d'immuniser les noeuds à forte influence locale au sein de leurs communautés. En effet, si une épidémie se déclare au coeur d'une communauté, l'infection peut s'éteindre avant d'atteindre d'autres communautés car il y a peu de périphéries inter-communautaires. Par conséquent, il est plus efficace de cibler des noeuds à forte influence locale dans des réseaux avec une structure communautaire forte que d'isoler les communautés.

Le panneau central de Figure 5.3 montre le même type de résultats pour les réseaux avec une force de structure de communauté moyenne ($\mu = 0.4$). Dans ces figures, Δr est toujours positif à la fois pour la composante locale et la composante globale de la centralité modulaire chevauchante. La composante locale de la centralité modulaire chevauchante donne de meilleurs résultats que la mesure standard avec un gain d'une valeur moyenne de 10% pour l'intermédiarité, la proximité et la centralité des vecteurs propres. Le gain le plus important concerne la centralité du degré avec une valeur moyenne de 15%. La mesure globale surpasse la mesure standard avec un gain d'environ 18% pour la centralité de intermédiarité, 20% pour la centralité du vecteur de proximité et des vecteurs propres. Le gain le plus important concerne également la centralité du degré, avec une valeur moyenne de 28%. Par conséquent, la mesure globale dans son ensemble est plus précise que la mesure locale. Dans cette situation, il y a presque autant de liens intra-communautaires qu'il y a de liens inter-communautaires. Ainsi, si une épidémie débute dans une communauté, elle peut se propager très facilement aux différentes communautés du réseau grâce aux nombreux liens inter-communautaires. C'est pourquoi l'influence globale est plus importante que l'influence locale dans les réseaux à structure communautaire de force moyenne.

Les résultats des expériences utilisant des réseaux avec une structure de communauté faible ($\mu = 0.6$) sont rapportés dans le panneau de droite de Figure 5.3. En observant ces courbes, nous remarquons que la différence relative de la taille de l'épidémie Δr entre la composante globale de la centralité modulaire chevauchante et la mesure traditionnelle est positive, alors qu'elle est toujours négative entre la composante locale de la centralité modulaire chevauchante et la mesure traditionnelle. Le gain moyen sur les mesures traditionnelles de centralité d'intermédiarité, de proximité et des vecteurs propres est d'environ 10% pour la composante globale de la centralité modulaire chevauchante. Le plus grand

gain est encore atteint pour la centralité du degré avec une valeur moyenne de 15%. La composante locale de la centralité modulaire chevauchante est moins performante que la mesure standard avec une valeur moyenne de 15% pour les mesures de centralité de degré, d'intermédiarité, de proximité et des vecteurs propres. Les liens inter-communautaires prédominent dans ce type de réseaux. Par conséquent, ces liens peuvent diffuser des épidémies à l'échelle globale sur tout le réseau. La meilleure stratégie dans cette situation est d'isoler les communautés.

5.4.1.2 *Évaluation des méthodes de classement basées sur la combinaison des composantes de la centralité modulaire chevauchante*

Afin d'évaluer l'impact de la combinaison d'influence locale et globale des noeuds, nous considérons deux mesures de classement. La première est le module des deux composantes de la centralité modulaire chevauchante. La seconde est une combinaison linéaire pondérée des deux composantes. Les poids sont calculés en utilisant des informations supplémentaires sur la structure de la communauté (taille de la communauté, nombre de communautés voisines). Nous nous attendons à ce que l'efficacité de la combinaison augmente avec une meilleure connaissance de la structure communautaire. Dans Figure 5.3, nous rapportons les résultats expérimentaux des mesures de classement basées sur le module et la combinaison linéaire pondérée des composantes locale et globale de la centralité modulaire chevauchante. Il apparaît clairement qu'elles surpassent la mesure classique ainsi que les deux composantes de la centralité modulaire chevauchante. Ce résultat est valable pour toutes les mesures de centralité testées. En effet, le module fonctionne mieux que la mesure standard avec un gain moyen de 37% dans les réseaux à forte structure communautaire, 30% dans les réseaux à structure communautaire de force moyenne, et 20% dans les réseaux à faible structure communautaire pour l'intermédiarité, la proximité et la centralité des vecteurs propres. Le gain le plus important concerne la centralité du degré avec une valeur moyenne de 42% dans les réseaux à forte structure communautaire, 35% dans les réseaux à structure communautaire de force moyenne et 25% dans les réseaux à faible structure communautaire. Par conséquent, la combinaison des deux composantes de la centralité modulaire chevauchante est beaucoup plus efficace que l'utilisation d'une

seule de ses composantes ou de la mesure de centralité classique.

De plus, on peut remarquer sur Figure 5.3 que les courbes de la combinaison linéaire pondérée des composantes de centralité modulaire chevauchante sont situées en haut sur toutes les figures. C'est donc la mesure la plus efficace quelle que soit la force de la structure communautaire. En fait, le gain est d'environ 47% dans les réseaux à structure communautaire bien définie, 40% dans les réseaux à structure communautaire de force moyenne et 30% dans les réseaux à structure communautaire peu claire pour l'intermédiarité, la proximité et la centralité des vecteurs propres. Le gain le plus important concerne la centralité du degré avec une valeur moyenne de 52% dans les réseaux à structure communautaire bien définie, 45% dans les réseaux à structure communautaire de force moyenne et 35% dans les réseaux à structure communautaire peu claire. Cela confirme notre intuition que d'autres gains peuvent être obtenus en utilisant plus d'informations sur les propriétés topologiques de la structure communautaire. En effet, dans cette mesure, nous introduisons la fraction des liens inter-communautaires de chaque communauté dans le cas de noeuds non chevauchants, et de chaque module de voisinage local dans le cas de noeuds qui se chevauchent. L'idée est de cibler les noeuds les plus influents des réseaux en prenant soin des propriétés topologiques de chaque communauté. Nous introduisons également la taille des communautés afin de sélectionner les hubs dans les grandes communautés pour leur forte influence locale. De plus, le nombre de communautés voisines est également utilisé pour sélectionner des noeuds de ponts communautaires connectés à plusieurs communautés. En effet, ils ont une forte influence globale. L'ajustement des pondérations permet de s'adapter à la structure communautaire en privilégiant soit la composante locale soit la composante globale. Comme cette mesure utilise plus d'informations sur la structure de la communauté, elle est plus efficace que le module et les composantes locale et globale de la centralité modulaire chevauchante. Pour résumer, l'utilisation d'une stratégie de combinaison est plus efficace par rapport aux mesures locale, globale ou traditionnelle. Ces résultats sont obtenus pour différentes mesures de centralité et pour des réseaux avec différentes forces de structure communautaire. Les meilleurs gains sont obtenus dans des réseaux avec une structure communautaire bien définie. En effet, l'augmentation de la valeur du paramètre de mélange brouille la structure de la communauté. Le réseau a tendance à

se comporter de plus en plus comme une seule grande communauté. Par conséquent, la performance des mesures modulaire diminue. Comme prévu, de meilleurs résultats sont obtenus en utilisant des informations plus pertinentes sur les propriétés topologiques de la structure communautaire. Ces résultats suggèrent que de nouvelles améliorations peuvent être atteintes en utilisant des stratégies de combinaison encore plus efficaces.

5.4.2 Influence du nombre de noeuds qui se chevauchent

Le but de cette étude est de montrer l'influence de la proportion des noeuds qui se chevauchent sur les performances de la centralité modulaire chevauchante. Des réseaux avec différentes forces de structure communautaire sont utilisés dans cette expérience (le paramètre de mélange varie de $\mu = 0.1$ à $\mu = 0.6$). De plus, la valeur du degré d'appartenance aux communautés est fixée à 10% du nombre total de communautés trouvées dans chaque réseau. La différence relative de la taille de l'épidémie Δr des mesures locale, globale et du module par rapport à la fraction des épandeurs initiaux f_0 est représentée dans figs. 4 to 6. Elle est aussi représentée en fonction de la proportion des noeuds qui se chevauchent on dans figs. 7 to 9.

On peut voir dans figs. 4 and 5 que les mesures locale et globale présentent le même comportement si la proportion de noeuds qui se chevauchent change. En fait, leurs performances augmentent à mesure que la proportion des noeuds qui se chevauchent augmente. Ces résultats sont valables pour toutes les mesures de centralité comme illustré dans figs. 7 and 8. Cependant, leurs performances diminuent lorsque la proportion des noeuds qui se chevauchent dépasse une valeur moyenne de 27% comme indiqué dans figs. 7 and 8. La performance de la mesure locale, par exemple dans les réseaux à forte structure communautaire, augmente avec un gain moyen de 11% pour la centralité de degré et d'intermédiarité, et de 7% pour les mesures de centralité de proximité et des vecteurs propres. Le gain moyen est d'environ 6% dans les réseaux avec une force de structure communautaire moyenne pour toutes les mesures de centralité, comme indiqué dans Figure 7. La performance de la mesure globale augmente avec un gain moyen de 5% dans les réseaux à forte structure communautaire, 4% dans les réseaux à structure communautaire de force moyenne. Le gain le plus important est obtenu avec la centralité du degré et de

d'intermédiation dans les réseaux à structure communautaire non cohésive. En effet, dans ce cas, nous observons un gain moyen de 7%, alors qu'il est de 3% avec la proximité et la centralité des vecteurs propres comme illustré dans Figure 8.

En augmentant la proportion des noeuds qui se chevauchent, la taille de leur réseau local augmente. Par conséquent, leur influence locale augmente puisque ces noeuds peuvent appartenir à plusieurs communautés voisines. C'est pourquoi le classement basé sur la composante locale de la centralité modulaire chevauchante devient plus efficace. De plus, lorsque la proportion de noeuds qui se chevauchent devient plus grande, il y a de grandes chances que la proportion des noeuds qui se chevauchent à l'intérieur du réseau global augmente également. Ainsi, ces noeuds peuvent diffuser l'épidémie à l'échelle globale dans de nombreuses communautés étrangères. Cela augmente donc l'influence globale de ces noeuds dans le réseau. Par conséquent, les performances de la mesure globale augmentent aussi tout en augmentant la proportion des noeuds qui se chevauchent. De plus, lorsque la proportion des noeuds qui se chevauchent devient très importante, la structure de la communauté devient lâche car la majorité des noeuds peuvent appartenir à plusieurs communautés. C'est la raison pour laquelle les performances de la mesure locale et globale diminuent lorsque la proportion des noeuds qui se chevauchent dépasse la valeur moyenne de 27%.

Le même type de résultats est indiqué dans figs. 6 and 9 pour la mesure basée sur le module. Dans Figure 6, on peut remarquer que la mesure de module présente le même comportement dans des réseaux avec des structures de communauté différentes. Ses performances augmentent tant que la proportion de noeuds qui se chevauchent augmente. Pourtant, ses performances cessent d'augmenter lorsque la proportion des noeuds qui se chevauchent dépasse une valeur moyenne de 27%. Comme cela est montré dans Figure 9, cela est vrai pour les réseaux avec différentes forces de structure communautaire et pour toutes les mesures de centralité testées. La performance de la mesure basée sur le module augmente d'un gain moyen de 13% dans les réseaux à structure communautaire bien définie, 10% dans les réseaux à force de structure communautaire moyenne et 8% dans les réseaux à structure communautaire peu claire. ce comportement est tout à fait indépendant des mesures de centralité comme indiqué dans les panneaux gauche, central et droit

de Figure 9. Comme prévu, l'augmentation de l'influence locale et globale des noeuds entraîne également une augmentation des performances de la mesure de module. Cela est dû au fait que cette mesure de classement est basée sur une combinaison des composantes locale et globale de la centralité modulaire chevauchante.

5.4.3 Influence du degré d'appartenance aux communautés qui se chevauchent

Notre objectif dans ces expériences est d'étudier l'influence du degré d'appartenance des noeuds qui se chevauchent sur les performances de la centralité modulaire chevauchante. Des réseaux ayant une structure communautaire différente sont utilisés. La proportion des noeuds qui se chevauchent est fixée à 10% de la taille du réseau. La différence relative de la taille de l'épidémie Δr des mesures locale, globale et du module par rapport à la fraction des épandeurs initiaux f_0 est représentée dans figs. 10 to 12. Elle est représentée en fonction du degré d'appartenance aux communautés om dans figs. 13 to 15.

Nous pouvons remarquer à partir de figs. 10 and 11 que les mesures locale et globale présentent le même comportement lorsque le degré d'appartenance aux communautés varie. Nous observons dans figs. 13 and 14 que les performances de toutes les mesures de centralité augmentent tant que le degré d'appartenance des noeuds qui se chevauchent augmente. La performance de la mesure locale augmente d'une valeur moyenne de 10% pour la centralité du degré dans les réseaux à structure communautaire forte et moyenne, et de 7% dans les réseaux à structure communautaire lâche. Comme indiqué dans Figure 13, le gain est également d'environ 7% pour l'intermédiarité, la proximité et la centralité des vecteurs propres dans tous les réseaux avec une force de structure communautaire variée. Figure 14 montre que globalement, le gain de performances de la mesure globale est d'environ 5%. En augmentant le degré d'appartenance, le voisinage local des noeuds qui se chevauchent devient plus grand. ces derniers deviennent plus influents dans le réseau local. C'est pourquoi la mesure locale fonctionne mieux si la valeur du degré d'appartenance des noeuds qui se chevauchent augmente. De plus, les noeuds qui se chevauchent situés dans le réseau global sont connectés à de nombreuses communautés étrangères. Ainsi, en augmentant leur degré d'appartenance, ces noeuds appartiennent à plus de communautés. Par conséquent, l'épidémie peut se propager d'une communauté à l'autre à travers ces

noeuds, affectant un très grand nombre de communautés. Ils peuvent aussi jouer le rôle de ponts très influents. C'est la raison pour laquelle la mesure globale devient plus efficace lorsque la valeur du degré d'appartenance des noeuds qui se chevauchent augmente.

Figures 12 and 15 montre les résultats du même type d'expériences pour le module de la mesure de centralité modulaire chevauchante. On peut voir dans Figure 12 que le module présente le même comportement pour les quatre mesures de centralité dans tous les réseaux testés. Ses performances augmentent tant que le degré d'appartenance des noeuds qui se chevauchent augmente comme indiqué dans Figure 15. Le gain est d'environ 15% pour la centralité des degrés et 11% pour les autres mesures de centralité testées, et c'est le cas pour les réseaux avec une structure communautaire forte et moyenne (voir le panneau de gauche et le milieu de Figure 15). Dans les réseaux à structure communautaire non cohésive, le gain est d'environ 10% pour toutes les mesures de centralité testées (voir le panneau de droite de Figure 15). Les stratégies de classement basées sur une combinaison des composantes de centralité modulaire chevauchante deviennent également plus efficaces si nous augmentons la valeur du degré d'appartenance des noeuds qui se chevauchent.

5.4.4 Synthèse

Dans cette section, les extensions de la centralité modulaire chevauchante de degré, d'intermédierité, de proximité et des vecteurs propres sont testées sur des réseaux synthétiques. Ces mesures sont appliquées dans un premier temps sur des réseaux de force de structure communautaire différente ayant un paramètre de mélange allant de 0.1 à 0.6 pour tester l'influence de la structure communautaire sur leur performance. Les résultats montrent que pour toutes les centralités, la mesure standard est plus performante que la mesure globale alors qu'elle est moins efficace que la mesure locale dans les réseaux à structure communautaire bien définie. Dans les réseaux avec une force de structure communautaire moyenne ou faible, les résultats montrent que la mesure globale surpasse la mesure standard, tandis que la mesure locale a une performance les plus faibles. Ainsi, les performances des composantes de la centralité modulaire chevauchante changent avec la variation de la proportion des liens inter-communautaires du réseau. De plus, les méthodes combinées sont les mesures les plus efficaces. Leurs performances sont encore plus élevées

si elles utilisent correctement plus d'informations sur la structure de la communauté. Les deuxièmes investigations visent à montrer l'influence de la proportion des noeuds qui se chevauchent sur la performance des variations de la centralité modulaire chevauchante. Les résultats montrent qu'en augmentant la proportion des noeuds qui se chevauchent, les performances des variations de centralité modulaire chevauchante augmentent. Leurs performances diminuent cependant lorsque cette proportion devient très importante. Dans ce cas, la majorité des noeuds peuvent appartenir à plusieurs communautés, ce qui fait que le réseau perd sa force de structure communautaire. La dernière partie des expériences vise à montrer l'influence du degré d'appartenance aux communautés des noeuds qui se chevauchent sur les performances de la centralité modulaire chevauchante. Les résultats expérimentaux montrent que quelle que soit la force de la structure communautaire du réseau, les performances des extensions de la centralité modulaire chevauchante augmentent, tout en augmentant le degré d'appartenance des noeuds qui se chevauchent.

5.5 Résultats des réseaux du monde réel

Dans cette section, nous rapportons les résultats d'expériences approfondies sur des réseaux du monde réel provenant de divers domaines (réseaux sociaux, collaboratifs et biologiques). Le but de cet ensemble d'expériences est de les comparer avec les résultats obtenus avec les réseaux synthétiques. Ainsi, des réseaux du monde réel avec une force de structure communautaire différente sont utilisés. Comme il n'y a pas de données de vérité sur le terrain pour ces réseaux, l'algorithme Speaker-Listener Label Propagation Algorithm SLPA est utilisé afin de découvrir les communautés qui se chevauchent. Les estimations des valeurs des paramètres de mélange sont calculées en fonction des communautés découvertes. Table 5.3 rapporte les valeurs estimées.

5.5.1 Évaluation des composantes de la centralité modulaire chevauchante

Les résultats pour les réseaux avec une structure communautaire bien définie (réseaux ego-Facebook, Netscience et ca-GrQc) sont rapportés dans Figure 16. La proportion de

TABLEAU 5.3 – Le paramètre de mélange estimé μ , la proportion des noeuds qui se chevauchent on et le nombre moyen d'appartenance aux communautés par noeud m des réseaux empirique. Ces valeurs sont calculées sur la structure de la communauté découverte à l'aide de l'algorithme SLPA.

Réseau	ego-Facebook	Netscience	ca-GrQc	Princeton	Caltech	Georgetown	Yeast-protein interaction
μ	0.075	0.094	0.11	0.25	0.27	0.31	0.47
$on(\%)$	8.33	7.84	23.25	7.47	4.55	15.25	33.75
m	1.89	2.01	1.65	1.74	1.066	1.41	1.078

mélange estimée pour ces réseaux est respectivement de 0.075, 0.094 et 0.11. Figure 16 (b) montre que les mesures classiques surpasse toujours la composante globale de la mesure de centralité modulaire chevauchante, alors qu'elle fonctionne moins bien que sa composante locale. Prenons par exemple la centralité de l'intermédiarité. Le gain de sa mesure locale par rapport à l'intermédiarité standard a une valeur moyenne de 25% pour le réseau ego-Facebook, 24% pour le réseau Netscience et 20% pour le réseau ca-GrQc. Cependant, l'intermédiarité globale est de 18%, 17% et 15% moins efficace que l'intermédiarité standard dans les réseaux ego-Facebook, Netscience et ca-GrQc, respectivement. Dans ces réseaux, les communautés sont peu connectées entre elles. Dans ce cas, il y a de fortes chances que l'épidémie reste confinée dans la même communauté et ne se propage pas aux communautés étrangères puisque la zone contagieuse se trouve dans la plupart des cas au coeur des communautés. Par conséquent, la propagation de l'épidémie peut s'arrêter avant même d'atteindre les noeuds situés aux bords de la communauté. Ainsi, il y a peu de chances que les ponts communautaires propagent l'épidémie à d'autres communautés. Cela explique pourquoi la mesure locale surpasse toujours la mesure globale. De plus, on peut remarquer sur Figure 16 que plus la valeur du paramètre de mélange est élevée, plus la composante globale de la mesure de la centralité modulaire chevauchante est performante. Cela est dû au fait que le nombre de liaisons inter-communautaires augmente tandis que le nombre de liaisons intra-communautaires diminue.

Figure 17 montre le même type de résultats pour les réseaux de Princeton, Caltech et Georgetown. Ces réseaux présentent une force de structure communautaire moyenne. En effet, leurs valeurs estimées des paramètres de mélange sont respectivement égales à 0.25, 0.27 et 0.31. On peut remarquer sur Figure 17 que les deux composantes de la centra-

lité modulaire chevauchante surpassent la mesure traditionnelle pour toutes les mesures de centralité testées. En effet, la mesure d'intermédiation locale est plus performante que l'interdépendance standard avec un gain moyen de 12%, 10% et 9% pour les réseaux de Princeton, Caltech et Georgetown respectivement comme cela est montré dans Figure 17 (b). La mesure d'intermédiation globale, en revanche, surpasse l'intermédiation traditionnelle avec un gain moyen de 14%, 17% et 18% pour Princeton, Caltech et Georgetown comme il est montré dans Figure 17 (b). Dans ces réseaux, il existe un grand nombre de liens inter-communautaires reliant des noeuds de différentes communautés. Ainsi, ces liens facilitent la transmission de l'épidémie entre les communautés. Dans ce cas, l'influence globale des noeuds devient plus importante que leur influence locale. Par conséquent, la composante globale est plus performante que les composantes locales pour toutes les mesures de centralité. On peut également voir sur Figure 17 que les performances du composante globale sont proches des performances du composante locale dans le réseau Princeton. Cette performance augmente dans les réseaux de Caltech et Georgetown qui ont des valeurs de paramètres de mélange plus élevées. Ainsi, l'influence globale des noeuds croît avec le nombre de liens inter-communautaires.

Les résultats expérimentaux avec le réseau Yeast-protein interaction caractérisé par une structure communautaire lâche sont rapportés dans Figure 18. On peut remarquer que la différence relative de la taille de l'épidémie Δr est toujours positive pour la mesure globale, alors qu'elle est toujours négative pour la mesure locale. Ainsi, la mesure standard donne de meilleurs résultats que la mesure locale et elle est moins performante que la mesure globale pour toutes les mesures de centralité. Le gain de la mesure globale de l'intermédiation par rapport à l'intermédiation standard est d'environ 16%. En revanche, la mesure locale de l'intermédiation est 17% moins efficace que l'intermédiation standard tel que rapporté dans Figure 18 (b). Dans ce réseau, il existe une grande quantité de liens inter-communautaires, les noeuds ont donc une grande influence globale. C'est la raison pour laquelle la mesure globale surpasse toujours la mesure locale. De plus, le gain de la mesure globale diminue dans ce réseau. Cela est dû à la structure communautaire lâche de ce réseau, ce qui entraîne une diminution de l'efficacité des mesures modulaires.

5.5.2 Évaluation des méthodes de classement sur la base d'une combinaison des composantes de la centralité modulaire chevauchante

Les résultats expérimentaux de l'évaluation des mesures de classement basées sur une combinaison des deux composantes de la centralité modulaire chevauchante sont rapportés dans figs. 16 to 18. Elles sont assez similaires à ceux observés avec les réseaux synthétiques pour toutes les mesures de centralité testées. En effet, les deux stratégies combinées surpassent toujours la mesure traditionnelle et chaque composante de la centralité modulaire chevauchante dans tous les cas. Le module de la centralité modulaire chevauchante de l'intermédiarité est plus performant que la mesure standard avec un gain moyen compris entre 39% et 47% dans les réseaux à structure communautaire bien définie (ego-Facebook, Netscience et cr-GrQc), entre 30% et 35% dans les réseaux à structure communautaire de force moyenne (Princeton, Caltech et Georgetown) et environ 24% dans le réseau Yeast-protein interaction qui a une structure communautaire faible. La mesure modulaire chevauchante pondérée de la centralité d'intermédiarité est globalement la plus efficace.

Comme prévu, l'utilisation d'une stratégie de combinaison simple permet des résultats plus efficaces. De plus, l'utilisation d'informations pertinentes sur les propriétés topologiques de la structure communautaire permet d'obtenir des résultats encore meilleurs, comme c'est le cas pour la mesure modulaire chevauchante pondérée. En effet, cette mesure utilise plus d'informations sur la structure de la communauté comme la proportion de liens inter-communautaires de chaque module du réseau. Elle prend également en considération la taille des communautés ainsi que le nombre de communautés étrangères attachées à chaque noeud. Cela explique l'effet de cette stratégie par rapport à la stratégie modulaire simple. En outre, nous pouvons également remarquer que les meilleurs résultats sont obtenus pour les réseaux avec une structure communautaire bien définie (ego-Facebook, Netscience et ca-GrQc). La performance des stratégies de combinaison diminue avec la force de la structure de la communauté, comme observé dans le cas du réseau Yeast-protein interaction.

5.5.3 Comparaison avec les méthodes modulaires alternatives

La taille de l'épidémie en fonction de la proportion des épandeurs initiaux pour la centralité modulaire chevauchante (module, composante locale, composante globale) et de leurs divers homologues classiques (centralité du degré, d'intermédiarité, de proximité et des vecteurs propres) est illustrée dans Figure 19. Les performances des méthodes alternatives (à savoir, la Membership, Random-Walk Overlap Selection *RWOS* and OverlapNeighborhood) sont également représentées dans cette figure. Les résultats pour le réseau ego-Facebook sont rapportés dans Figure 19 (a), alors qu'ils sont rapportés dans Figure 19 (b) pour le réseau Yeast-protein interaction. Nous choisissons ces deux réseaux car ego-Facebook a une structure communautaire bien définie tandis que Yeast-protein interaction a une structure communautaire lâche. Le but de cette section est de comparer les performances de toutes les variations de la centralité modulaire chevauchante avec les méthodes alternatives conçues pour les réseaux modulaires.

Les résultats montrent qu'en augmentant la proportion des épandeurs initiaux, la taille de l'épidémie augmente également. Ceci est valable pour toutes les mesures testées et pour les deux réseaux. Cependant, cette évolution est modérée dans le réseau ego-Facebook alors qu'elle est plus rapide dans le réseau Yeast-protein interaction. Dans le premier réseau, les communautés sont bien séparées, les épidémies ne peuvent alors pas se propager très facilement entre ses modules. Dans les deux réseaux, l'OverlapNeighborhood est aussi performante dans la plupart des cas que la centralité de degré standard. Elle surpasse même, dans certains cas, les performances de la centralité d'intermédiarité standard. En outre, Elle surpasse les centralités standard de proximité et des vecteurs propres et les autres méthodes alternatives (Membership et *RWOS*). La méthode OverlapNeighborhood sélectionne au hasard les voisins des noeuds qui se chevauchent. En effet, les noeuds qui se chevauchent sont plus susceptibles d'être connectés aux hubs de chaque communauté à laquelle ils appartiennent. C'est la raison pour laquelle elle fonctionne aussi bien que la centralité de degré standard. De plus, pour les deux réseaux, on peut voir dans Figure 19 que les courbes de *RWOS* et de la Membership sont toujours situées en bas de toutes les figures. Elles sont les moins efficace. Ces deux mesures n'utilisent pas autant d'informations sur la topologie du réseau que les autres mesures. C'est pourquoi ils présentent des

TABLEAU 5.4 – Le paramètre de mélange estimé μ , la proportion des noeuds qui se chevauchent on et le nombre moyen du degré d'appartenance aux communautés m dans les réseaux ego-Facebook et Yeast-protein interaction.

Réseau	Mesure	Algorithme de détection		
		SLPA	LINKC	Louvain
ego-Facebook	μ	0.075	0.096	0.03
	$on(\%)$	8.33	7.09	0
	m	1.89	1.37	0
Yeast-protein interaction	μ	0.47	0.51	0.49
	$on(\%)$	33.75	6.03	0
	m	1.078	1.64	0

performances inférieures.

De plus, dans le réseau ego-Facebook, la composante locale de la centralité modulaire chevauchante surpasse les mesures alternatives. La composante globale est cependant plus efficace que les méthodes alternatives du réseau Yeast-protein interaction. Le premier réseau a une structure communautaire bien définie. Les noeuds avec des connexions intra-communautaires élevées sont plus susceptibles d'infecter de nombreux noeuds du réseau. C'est ce qui rend la mesure locale très efficace dans ego-Facebook car elle cible des noeuds de nature différente (se chevauchent et ne se chevauchant pas) avec une forte influence locale. Le deuxième réseau, en revanche, a une structure communautaire lâche. Les noeuds avec des liens inter-communautaires élevés sont plus susceptibles d'avoir la plus grande influence sur ce type de réseaux. C'est la raison pour laquelle les performances de la mesure globale surpassent les méthodes alternatives. De plus, le module est la mesure la plus influente dans les deux réseaux et pour toutes les mesures de centralité testées. Cette méthode basée sur une combinaison prend en compte l'influence locale et globale des noeuds qui se chevauchent et ceux qui ne se chevauchent pas. Par conséquent, les méthodes de combinaison peuvent mettre en évidence les noeuds clés qui jouent un rôle majeur dans le processus de diffusion.

5.5.4 Influence des algorithmes de détection de communautés

Nous rapportons dans cette section un ensemble d'expériences sur les réseaux ego-Facebook et Yeast-protein interaction en utilisant les algorithmes Link Communities

(LINKC) Ahn *et al.* (2010) et Louvain Blondel *et al.* (2008); Orman *et al.* (2011) (voir subsection 2.5.1). Le premier algorithme est utilisé pour tester la robustesse des résultats si une autre méthode de détection de communautés qui se chevauchent est choisie. Le deuxième algorithme est utilisé pour comparer la centralité modulaire chevauchante à la centralité modulaire (présentée dans la première partie de ce chapitre) puisque l'algorithme de Louvain est conçu pour les communautés qui se chevauchent pas. Figure 20 représente la différence relative de la taille de l'épidémie entre les extensions de centralité modulaire chevauchante (centralité de degré et d'intermédiarité) et leurs homologues classiques. Ces mesures sont calculées sur deux réseaux du monde réel : le réseau ego-Facebook caractérisé par une structure communautaire bien définie et le réseau Yeast-protein interaction qui a une structure communautaire lâche. Les valeurs estimées du paramètre de mélange et la proportion des noeuds qui se chevauchent sont rapportées dans Table 5.4.

On peut déduire de Figure 20, que pour les deux algorithmes (LINKC et Louvain), la composante locale, la composante globale et le module de la centralité modulaire chevauchante présentent globalement le même comportement qu'avec l'algorithme SLPA. Dans les réseaux avec une structure communautaire bien définie (par exemple, le réseau ego-Facebook), la composante locale fonctionne toujours mieux que la classique avec un gain moyen de 33% et 20% pour la centralité du degré, alors que le gain est d'environ 23% et 19% pour la centralité de l'intermédiarité lors de l'utilisation des algorithmes LINKC et Louvain respectivement. Le gain moyen est d'environ 37% et 26% dans le cas de l'algorithme SLPA pour les deux mesures de centralité. La mesure standard, en revanche, surpasse la composante globale pour toutes les proportions d'épandeurs initiaux. De plus, le gain global du module de la centralité modulaire chevauchante est d'environ 44% et 32% pour la centralité du degré, alors qu'il est d'environ 38% et 34% pour la centralité de l'intermédiarité lors de l'utilisation des algorithmes LINKC et Louvain respectivement. Cependant, le gain moyen est d'environ 46% et 42% pour les deux mesures de centralité lors de l'utilisation de SLPA. Ainsi, si nous comparons les différentes méthodes de détection de communautés qui se chevauchent, il apparaît que l'extension de centralité modulaire chevauchante basée sur l'algorithme SLPA est légèrement plus efficace. En effet, pour les deux algorithmes, les valeurs de paramètres de mélange et la proportion de

noeuds qui se chevauchent sont assez similaires. Par conséquent, dans les réseaux avec une structure communautaire bien définie, ils découvrent à peu près la même structure communautaire. C'est pourquoi, les performances de la centralité modulaire chevauchante ont à peu près le même comportement. Cependant, les composantes de la centralité modulaire présentent les plus faibles performances lorsque l'algorithme de Louvain est utilisé. Cela montre l'importance des noeuds qui se chevauchent dans le processus de diffusion.

Dans les réseaux avec une structure communautaire lâche (par exemple, le réseau Yeast-protein interaction), la mesure standard surpasse toujours la composante locale de la centralité modulaire chevauchante. La composante globale, cependant, fonctionne mieux que la centralité standard avec un gain moyen de 5% et 3% pour la centralité du degré, tandis que le gain est d'environ 4% et 2% pour la centralité d'intermédiarité lors de l'utilisation des algorithmes LINKC et Louvain respectivement. Le gain moyen est d'environ 12% et 10% dans le cas de l'algorithme SLPA pour les deux mesures de centralité. Le gain global de la mesure du module est d'environ 14% et 10% pour la centralité du degré, tandis qu'il est d'environ 13% et 7% pour la centralité de l'intermédiarité lors de l'utilisation des algorithmes LINKC et Louvain respectivement. De plus, le gain est d'environ 20% et 19% pour les deux mesures de centralité respectivement lorsque l'algorithme SLPA est utilisé. Dans ce réseau, les performances des deux composantes de centralité modulaire chevauchante pour l'algorithme SLPA sont supérieures à celles obtenues lorsque l'algorithme LINKC est utilisé. En effet, l'algorithme SLPA a un paramètre de mélange relativement plus petit et une proportion plus élevée de noeuds qui se chevauchent. Comme prévu, les performances des variations de la centralité modulaire chevauchante augmentent tant que la proportion des noeuds chevauchants augmente. Ces résultats sont en concordance avec ceux basés sur des expériences de réseaux synthétiques. Ainsi, cela suggère que SLPA est plus précis que l'algorithme LINKC. Les performances les plus faibles, cependant, sont obtenues en utilisant l'algorithme de Louvain même s'il a un paramètre de mélange plus petit par rapport à l'algorithme LINKC. Cet algorithme ne prend pas en compte les noeuds qui se chevauchent. Ainsi, nous pouvons conclure que la prise en compte de l'influence des noeuds qui se chevauchent améliore les performances des composantes de centralité modulaire.

De plus, le panneau de droite de Figure 20 indique les performances des mesures modulaires pondérées pour la centralité modulaire et la centralité modulaire chevauchante des mesures de degré et d'intermédiarité. Les paramètres des mesures de centralité modulaire chevauchante pondérée sont calculés à l'aide de SLPA ou LINKC et sont respectivement appelés centralité modulaire chevauchante pondérée basée sur SLPA (Weighted Overlapping SLPA) et centralité modulaire chevauchante pondérée basée sur LINKC (Weighted Overlapping LINKC). Les paramètres de la centralité modulaire pondérée (sans chevauchement) sont calculés à l'aide de l'algorithme de Louvain. Dans le réseau ego-Facebook, la centralité modulaire chevauchante pondérée basée sur SLPA surpasse la centralité modulaire pondérée sans chevauchement avec un gain moyen de 10% et 8% respectivement pour les mesures de centralité de degré et d'intermédiarité. La centralité modulaire chevauchante pondérée basée sur LINKC donne également de meilleurs résultats que la mesure modulaire pondérée sans chevauchement avec un gain moyen de 7% et 6% pour les mesures de centralité du degré et de l'intermédiarité. Dans le réseau Yeast-protein interaction, la centralité modulaire chevauchante pondérée basée sur SLPA effectue 13% et 14% de plus que la mesure modulaire pondérée sans chevauchement pour les centralités de degré et d'intermédiarité, respectivement. De plus, la centralité modulaire chevauchante pondérée basée sur LINKC surpasse la mesure modulaire pondérée sans chevauchement avec un gain moyen de 7% pour les centralités de degré et d'intermédiarité. Comme prévu, dans les deux réseaux, les mesures de centralité modulaire chevauchante pondérée sont plus efficaces que la centralité modulaire pondérée sans chevauchement. De plus, la centralité modulaire chevauchante pondérée basée sur SLPA offre les meilleures performances. Par conséquent, la prise en compte des chevauchements entre les communautés est très important afin de mettre en évidence les diffuseurs influents dans les réseaux modulaires.

Dans l'ensemble, les résultats de cet ensemble d'expériences montrent que les performances de la centralité modulaire chevauchante sont influencées par les variations de la structure communautaire découverte. Quelle que soit la force de la structure communautaire, les algorithmes de détection SLPA et LINKC présentent dans la plupart du temps de meilleures performances par rapport à l'algorithme de Louvain. Par conséquent, la centralité proposée présente des performances plus élevées lorsqu'un algorithme de détection

de communautés qui se chevauchent est utilisée. Cela montre le rôle majeur qui joue les noeuds qui se chevauchent en termes de diffusion des épidémies entre les communautés.

5.5.5 Synthèse

Dans ce chapitre, nous avons également effectué une série d'expériences sur des réseaux empiriques de diverses natures. Des réseaux du monde réel de différentes forces de structure communautaire sont utilisés. L'algorithme SLPA est utilisé pour découvrir les communautés qui se chevauchent du réseau. Les résultats montrent que les performances des variations de la centralité modulaire chevauchante des quatre centralités testées présentent le même comportement que pour les réseaux synthétiques. En effet, la composante locale montre ses meilleures performances dans les réseaux avec une structure communautaire bien définie, tandis que la composante globale montre ses meilleures performances dans les réseaux avec une force de structure communautaire moyenne et faible. De plus, les mesures combinées surpassent toutes les autres mesures dans tous les différents types de réseaux. Ces résultats corroborent les conclusions que nous avons tirées avec les réseaux synthétiques. La deuxième partie des expériences vise à comparer la centralité modulaire chevauchante avec leurs homologues classiques avec quelques mesures alternatives (Membership, RWOS et OverlapNeighborhood). Les résultats expérimentaux montrent que, dans la plupart des cas, le OverlapNeighborhood est aussi performant que la centralité de degré standard et dépasse dans certains cas la centralité d'intermédiarité standard. Les autres mesures alternatives (RWOS et la Membership) ont une performance inférieure. Cela est dû à leur utilisation d'informations limitées sur la topologie du réseau par rapport aux autres mesures. Un autre ensemble d'expériences est effectué pour montrer l'influence du choix de l'algorithme de détection de communauté sur les performances de la centralité modulaire chevauchante. Nous utilisons l'algorithme LINKC pour tester la robustesse des résultats si un autre algorithme de détection des communautés qui se chevauchent est choisi. Nous avons également utilisé l'algorithme de Louvain qui est conçu pour découvrir des communautés non chevauchantes afin de comparer la centralité modulaire chevauchante avec la centralité modulaire. Dans les réseaux avec une structure de communauté bien définie, les résultats montrent que les performances de la centralité modulaire che-

vauchante ne varient pas de manière significative lorsque l'algorithme LINKC est utilisé. Cela est dû au fait que les deux algorithmes révèlent environ le même ensemble de communautés. Cependant, les performances des variations de centralité modulaire chevauchante diminuent pour l'algorithme LINKC dans les réseaux avec une structure de communauté lâche. Dans ce cas, les algorithmes de détection de communauté découvrent différents ensembles de communautés. De plus, quelle que soit la force de la structure communautaire, la centralité modulaire chevauchante (utilisant SLPA ou LINKC) surpasse la centralité modulaire (utilisant Louvain). Par conséquent, les noeuds qui se chevauchent jouent un rôle majeur en termes de diffusion des épidémies entre les communautés.

5.6 Conclusion

Un nouveau framework permettant d'étendre les mesures de centralité classiques aux réseaux modulaires qui se chevauchent est introduit. Notre approche est inspirée par l'idée qu'il faut considérer deux types d'influences pour un noeud dans un réseau modulaire : une influence locale sur les noeuds appartenant à leurs communautés exercée par les liens intra-communautaires, et une influence globale sur les noeuds de les communautés restantes qui passent par les liens inter-communautaires. Par conséquent, dans les réseaux modulaires qui se chevauchent, les mesures de centralité doivent être représentées par un vecteur bidimensionnel que nous appelons la "centralité modulaire chevauchante". Une composante mesure l'influence locale du noeud, tandis que l'autre mesure son influence globale. Une série d'expériences a été réalisée afin de tester l'efficacité des extensions de centralité modulaire chevauchante par rapport à leur homologue définie pour les réseaux sans structure communautaire. Compte tenu des mesures de centralité les plus influentes (degré, distance, proximité, vecteur propre), les composantes locale et globale ont été évaluées séparément. De plus, une combinaison simple des deux composantes (module du vecteur bidimensionnel) a été testée. Une autre méthode basée sur une combinaison (centralité modulaire chevauchante pondérée) qui utilise des connaissances supplémentaires sur la topologie de la structure de la communauté a également été évaluée. Des expériences ont été menées sur des réseaux synthétiques et du monde réel en utilisant le modèle épidémique SIR. Les résultats montrent que les épandeurs identifiés par l'approche

proposée sont plus influents que ceux ciblés par les mesures de centralité conçues pour les réseaux non modulaires. En outre, les enquêtes menées sur les réseaux synthétiques montrent que les performances de la centralité modulaire chevauchante augmentent avec la proportion des noeuds qui se chevauchent et leur degré d'appartenance aux communautés. Les comparaisons avec d'autres mesures de centralité conçues pour les réseaux modulaires sont en faveur du vecteur proposé. De plus, des expériences effectuées sur des réseaux du monde réel avec divers algorithmes de détection de communauté montrent que la centralité modulaire chevauchante est très robuste aux variations de la structure de la communauté. Ces investigations démontrent aussi qu'elle est plus efficace que la centralité modulaire (lorsqu'on ignore les chevauchements entre les communautés), d'où l'importance des noeuds qui se chevauchent dans le processus de propagation de l'épidémie.



CONCLUSION ET PERSPECTIVES

Comprendre le comportement d'une maladie infectieuse est essentiel car la maladie peut devenir une situation cruciale si elle n'est pas complètement éradiquée à temps. Le processus de diffusion de la maladie peut être modélisé par une représentation en réseau des personnes et des interactions entre elles. Une infection se propage d'un individu à un autre dans le réseau, et donc l'immunisation ou la vaccination de chaque personne l'exclut du processus de propagation. Un objectif important en épidémiologie est de trouver un moyen d'arrêter ou au moins de ralentir la vitesse de contagion. L'élimination d'une maladie nécessite une couverture vaccinale spécifique. Cette quantité de couverture est impossible à atteindre en raison des contraintes de temps, d'approvisionnement en vaccins, etc. Pour arrêter les épidémies dans les réseaux sociaux, trouver des diffuseurs influents est notre objectif principal. Par conséquent, la recherche de méthodes d'immunisation efficaces reste un domaine particulièrement intéressant pour les chercheurs. Ces stratégies d'immunisation se divisent en deux catégories : stochastiques et déterministes. Dans les stratégies stochastiques, les noeuds ciblés sont identifiés en collectant localement des informations à partir de noeuds sélectionnés au hasard dans le réseau. Ils sont totalement agnostiques quant à la structure complète du réseau. Les stratégies déterministes, en revanche, supposent la connaissance de l'ensemble du réseau. Ces stratégies déterminent la succession dans laquelle les noeuds d'un réseau doivent être immunisés pour atténuer la propagation de l'épidémie. Elles classent tous les noeuds selon une mesure de centralité donnée. De haut en bas, les noeuds sont ciblés en fonction de leur rang. Les stratégies déterministes se sont avérées très efficaces pour contrôler les épidémies. Leur seul inconvénient est leur forte exigence de la topologie globale du réseau. Cependant, en général,

les stratégies déterministes surpassent les stratégies stochastiques en termes de maîtrise d'une épidémie.

La structure des réseaux influence les modes de propagation des épidémies. De nombreux réseaux du monde réel sont constitués de communautés. La structure de communauté existe lorsque les connexions entre les membres d'un groupe de noeuds sont plus denses que les connexions entre les membres de différents groupes de noeuds. Malgré le fait que cette propriété soit bien reconnue, elle est très souvent ignorée lorsqu'il s'agit de l'utiliser pour élaborer de nouvelles stratégies d'immunisation. Dans ce travail, nous visons à comprendre comment la structure communautaire affecte la dynamique épidémique et le contrôle des maladies infectieuses. Nous examinons également de nouvelles stratégies d'immunisation conçues pour les réseaux modulaires. Dans le chapitre 2, nous avons introduit tout d'abord les modèles épidémiques qui servent à évaluer l'efficacité de diffusion de chaque stratégie d'immunisation. Nous avons aussi introduit la notion de la structure communautaire ainsi que les algorithmes les plus efficaces et de diverses catégories utilisés pour sa détection. Ensuite, nous avons présenté un état de l'art sur les approches d'immunisation stochastiques et déterministes destinées aux réseaux sans et avec structure communautaire. Dans le chapitre 3, nous proposons et étudions trois nouvelles stratégies d'immunisation déterministes. Ce chapitre est consacré aux réseaux avec une structure communautaire qui ne se chevauchent pas (c'est-à-dire qu'un noeud appartient à une seule communauté). Ces trois stratégies sont destinées à des réseaux avec des forces de structure communautaire différentes, allant d'une structure bien définie à une structure communautaire lâche. Notre objectif est d'étudier l'impact de la force de la structure communautaire sur le choix d'une stratégie d'immunisation appropriée. Afin de caractériser l'influence d'un noeud, diverses informations sont utilisées telles que le nombre de communautés voisines que le noeud peut atteindre, la nature des liens (liens intra-communautaires, liens inter-communautaires), la taille des communautés et la densité d'interconnexion entre les communautés. Le but de cela est de montrer qu'engager plus de propriétés topologiques de la communauté peut également améliorer les performances des stratégies d'immunisation. Des simulations numériques avec le modèle épidémiologique Susceptible-Infecté-Rétabli sont effectuées sur des réseaux réels et synthétiques. Les

résultats expérimentaux montrent que les stratégies proposées sont plus efficaces que les alternatives classiques agnostiques à la structure communautaire. De plus, elles surpassent les stratégies stochastiques et déterministes alternatives conçues pour les réseaux modulaires.

Dans un réseau modulaire, on peut distinguer deux types d'influences pour un noeud : une influence locale sur les noeuds appartenant à sa propre communauté à travers les liens intra-communautaires, et une influence globale sur les noeuds des autres communautés à travers les liens inter-communautaires. Par conséquent, les mesures de centralité ne devraient pas être représentées par une simple valeur scalaire. Dans le chapitre 4, nous avons introduit un framework pour adapter toutes les mesures de centralité classiques proposées pour les réseaux sans structure communautaire aux réseaux modulaires. Nous avons proposé un vecteur bidimensionnel. Sa première composante mesure l'influence locale du noeud, tandis que la deuxième composante mesure son influence globale. Sur la base de cette hypothèse, nous avons proposé d'étendre toutes les mesures de centralité classiques aux réseaux modulaires. Dans notre travail, nous avons considéré deux cas correspondant à la nature des communautés (avec et sans chevauchement). Ainsi, dans le chapitre 4, nous avons présenté la "Centralité modulaire" qui signifie le vecteur bidimensionnel dans les réseaux modulaires avec des communautés non chevauchantes, tandis que le chapitre 5 est consacré à la "Centralité modulaire superposée" faisant référence à ce vecteur de centralité dans les réseaux modulaires avec des communautés avec chevauchement. Nous avons mené une série d'expériences afin de tester la pertinence de la centralité modulaire et de la centralité modulaire superposée par rapport à leur homologue classique. Compte tenu des mesures de centralité les plus influentes (degré, intermédiarité, proximité, vecteurs propres), les composantes locale et globale ont été évaluées séparément. De plus, quelques combinaisons simples des deux composantes ont été testées. Des expériences ont été menées sur des réseaux synthétiques à structure communautaire contrôlée et sur des réseaux du monde réel dans un scénario de propagation d'épidémie utilisant le modèle SIR. Les résultats montrent que les épandeurs identifiés par l'approche proposée pour les deux types de réseaux sont plus influents que ceux visés par les mesures de centralité standard.

Tous ces contributions nous donnent un sens d'orientation pour la conception de nouvelles stratégies d'immunisation adaptées à la topologie du réseau. La structure communautaire ne peut être ignorée, et beaucoup plus de connaissances sur la formation des communautés et leurs principales caractéristiques doivent être découverts et intégrés dans les stratégies d'immunisation afin de mieux identifier les noeuds influents.



ANNEXES

TABLEAU 5 – L'analyse de la variance ANOVA avec un niveau de signification $\alpha = 0.05$. SS est la somme des carrés, df est le degré de liberté. MS est le carré moyen. F est la statistique de test. P – *value* est la valeur de probabilité et $F_{critique}$ est la valeur critique de F .

Source de variations	SS	df	MS	F	P -value	F_{crit}
Stratégie Comm						
$\mu = 0.1$; [100,500]	Entre groupes	4.256503573	1	4.256503573		
$\mu = 0.1$; [50,250]	Dans les groupes	4378.911918	20	218.9455959	0.019440919	0.890504334
$\mu = 0.4$; [100,500]	Entre groupes	3.25263213	1	3.25263213		
$\mu = 0.4$; [50,250]	Dans les groupes	2885.642407	20	144.2821203	0.022543556	0.882154017
$\mu = 0.7$; [100,500]	Entre groupes	14.15562603	1	14.15562603		
$\mu = 0.7$; [50,250]	Dans les groupes	5632.639949	20	281.6319974	0.050262847	0.824880951
$\mu = 0.9$; [100,500]	Entre groupes	1.420498127	1	1.420498127		
$\mu = 0.9$; [50,250]	Dans les groupes	5037.9883	20	251.899415	0.00563914	0.967814872
Stratégie CBB						
$\mu = 0.1$; [100,500]	Entre groupes	2.968469205	1	2.968469205		
$\mu = 0.1$; [50,250]	Dans les groupes	7796.175363	20	389.8087681	0.007615194	0.931328629
$\mu = 0.4$; [100,500]	Entre groupes	1.419799471	1	1.419799471		
$\mu = 0.4$; [50,250]	Dans les groupes	2176.587136	20	108.8293568	0.013046107	0.910202616
$\mu = 0.7$; [100,500]	Entre groupes	2.78891	1	2.78891		
$\mu = 0.7$; [50,250]	Dans les groupes	3787.848287	20	189.3924143	0.01472556	0.999564357
$\mu = 0.9$; [100,500]	Entre groupes	12.76013782	1	12.76013782		
$\mu = 0.9$; [50,250]	Dans les groupes	7569.738518	20	378.4869259	0.03371355	0.856166211
Stratégie CBM						
$\mu = 0.1$; [100,500]	Entre groupes	6.632987506	1	6.632987506		
$\mu = 0.1$; [50,250]	Dans les groupes	3579.629836	20	178.9814918	0.037059628	0.849285018
$\mu = 0.4$; [100,500]	Entre groupes	2.153661701	1	2.153661701		
$\mu = 0.4$; [50,250]	Dans les groupes	2810.5024	20	140.52512	0.01532581	0.973948282
$\mu = 0.7$; [100,500]	Entre groupes	1.349802017	1	1.349802017		
$\mu = 0.7$; [50,250]	Dans les groupes	4637.43895	20	231.8719475	0.005821325	0.939940411
$\mu = 0.9$; [100,500]	Entre groupes	0.54374064	1	0.54374064		
$\mu = 0.9$; [50,250]	Dans les groupes	3591.973201	20	179.59866	0.003027532	0.956666083

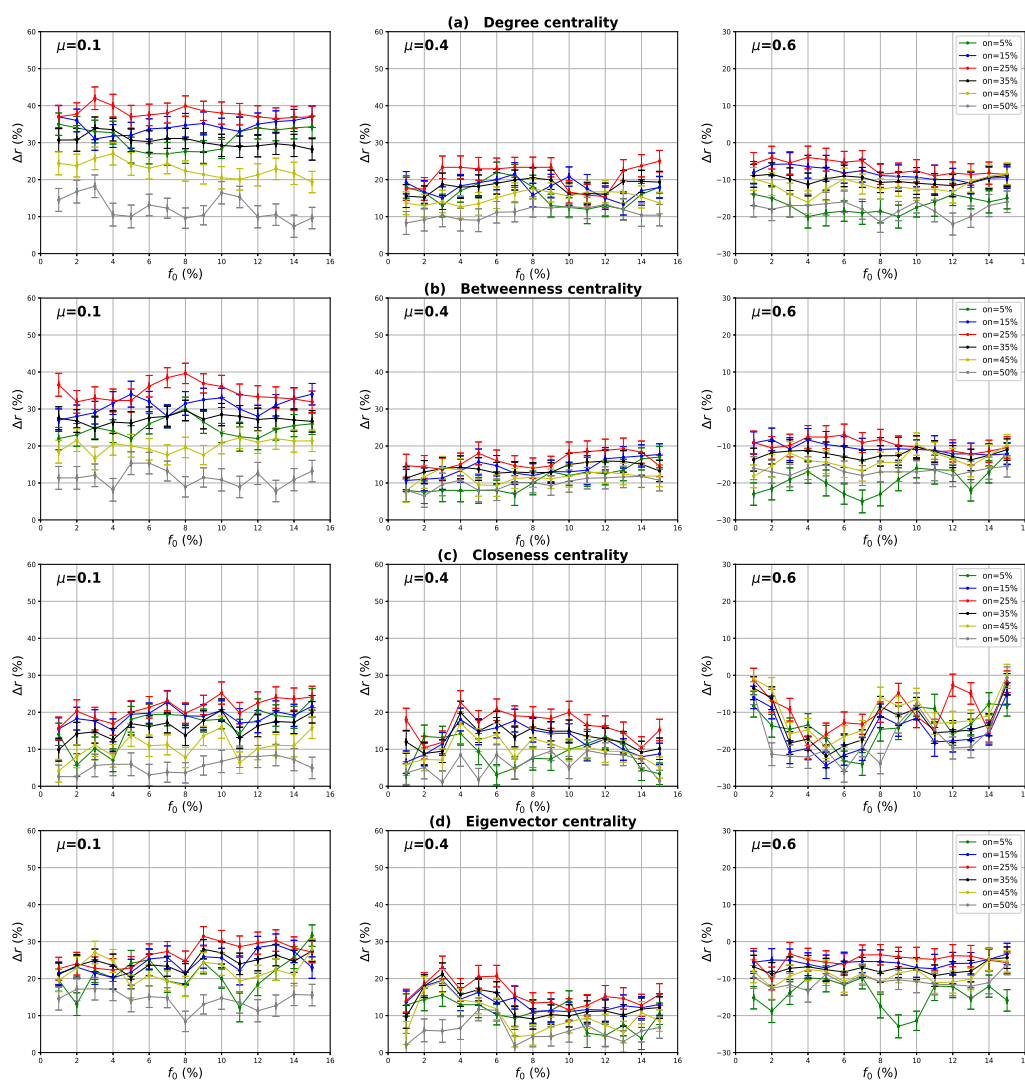


FIGURE 4 – Influence du nombre de noeuds qui se chevauchent en terme de la taille épidémique. La différence relative de la taille de l'épidémie Δr entre la mesure standard et la mesure locale en fonction de la fraction des épandeurs initiaux f_0 . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés.

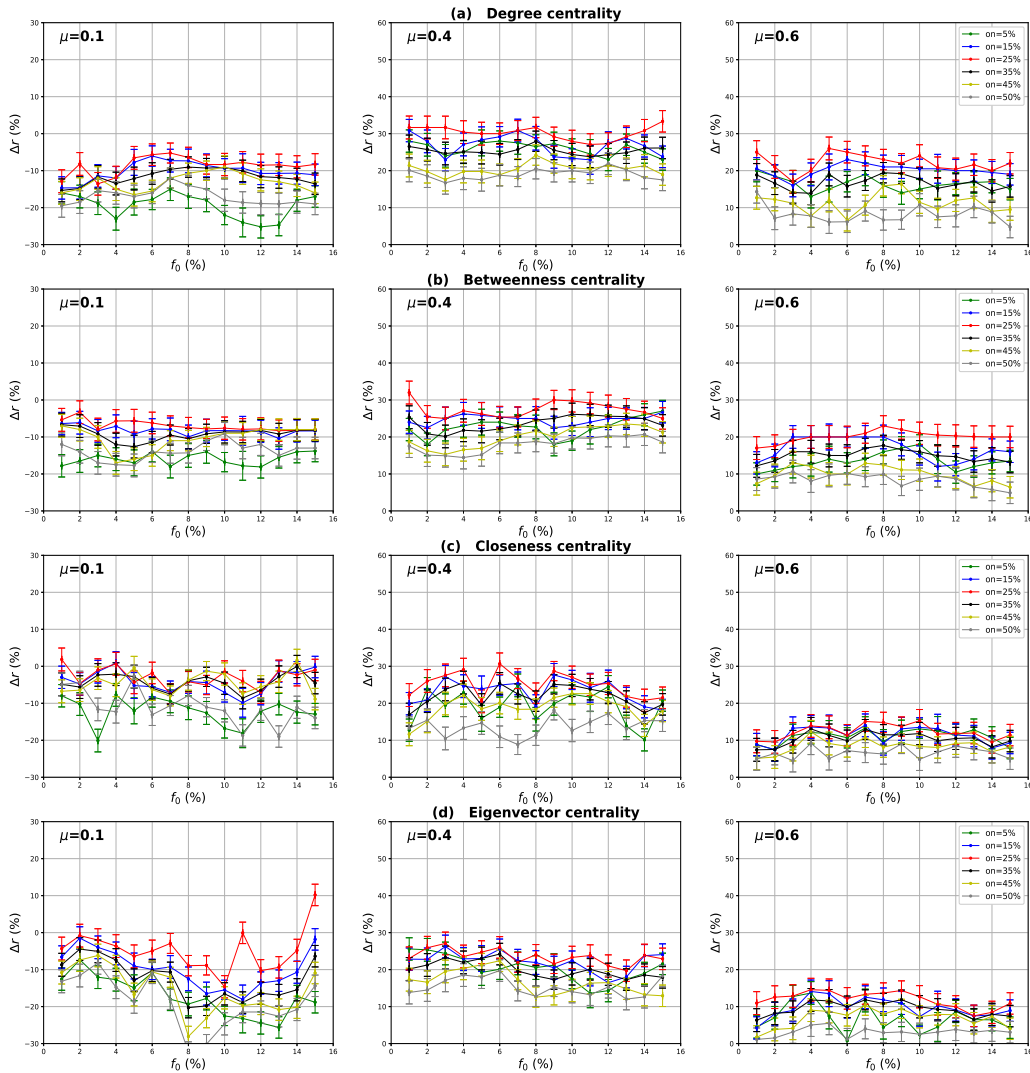


FIGURE 5 – Influence du nombre de noeuds qui se chevauchent en terme de la taille épidémique. La différence relative de la taille de l'épidémie Δr entre la mesure standard et la mesure globale en fonction de la fraction des épandeurs initiaux f_0 . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés.

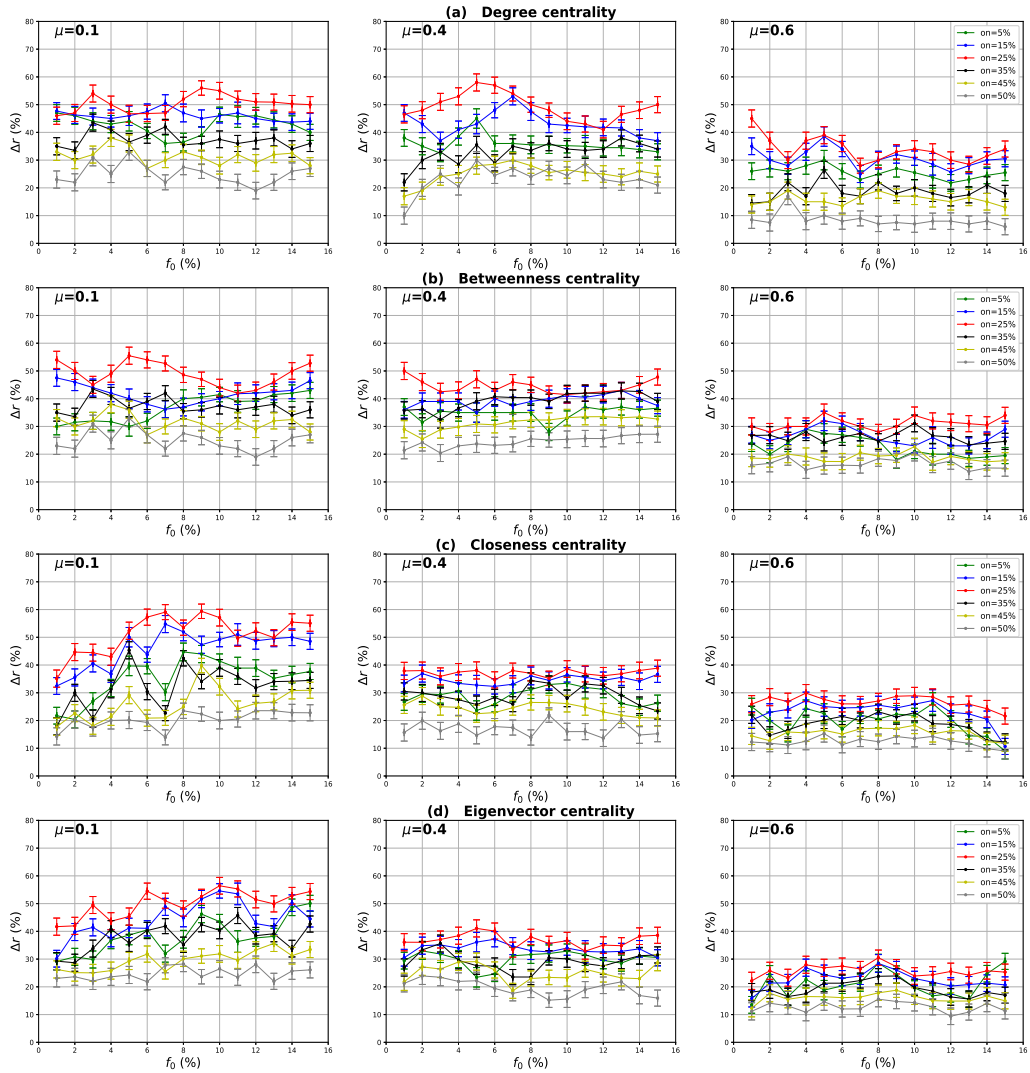


FIGURE 6 – Influence du nombre de noeuds qui se chevauchent en terme de la taille épidémique. La différence relative de la taille de l'épidémie Δr entre la mesure standard et la mesure de module en fonction de la fraction des épandeurs initiaux f_0 . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés.

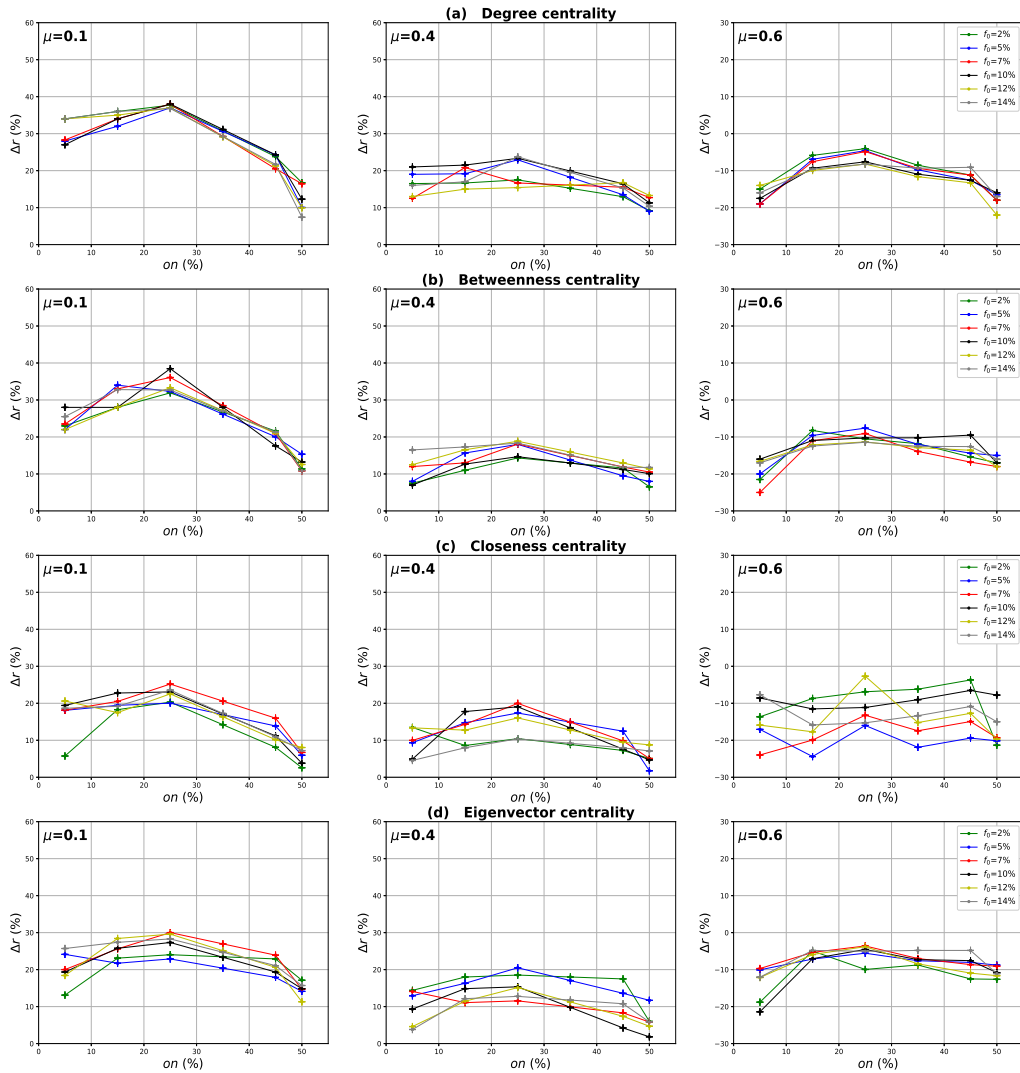


FIGURE 7 – Influence du nombre de noeuds qui se chevauchent en terme de la taille épidémique. La différence relative de la taille de l'épidémie Δr entre la mesure standard et la mesure locale en fonction de la fraction des noeuds chevauchants on . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés.

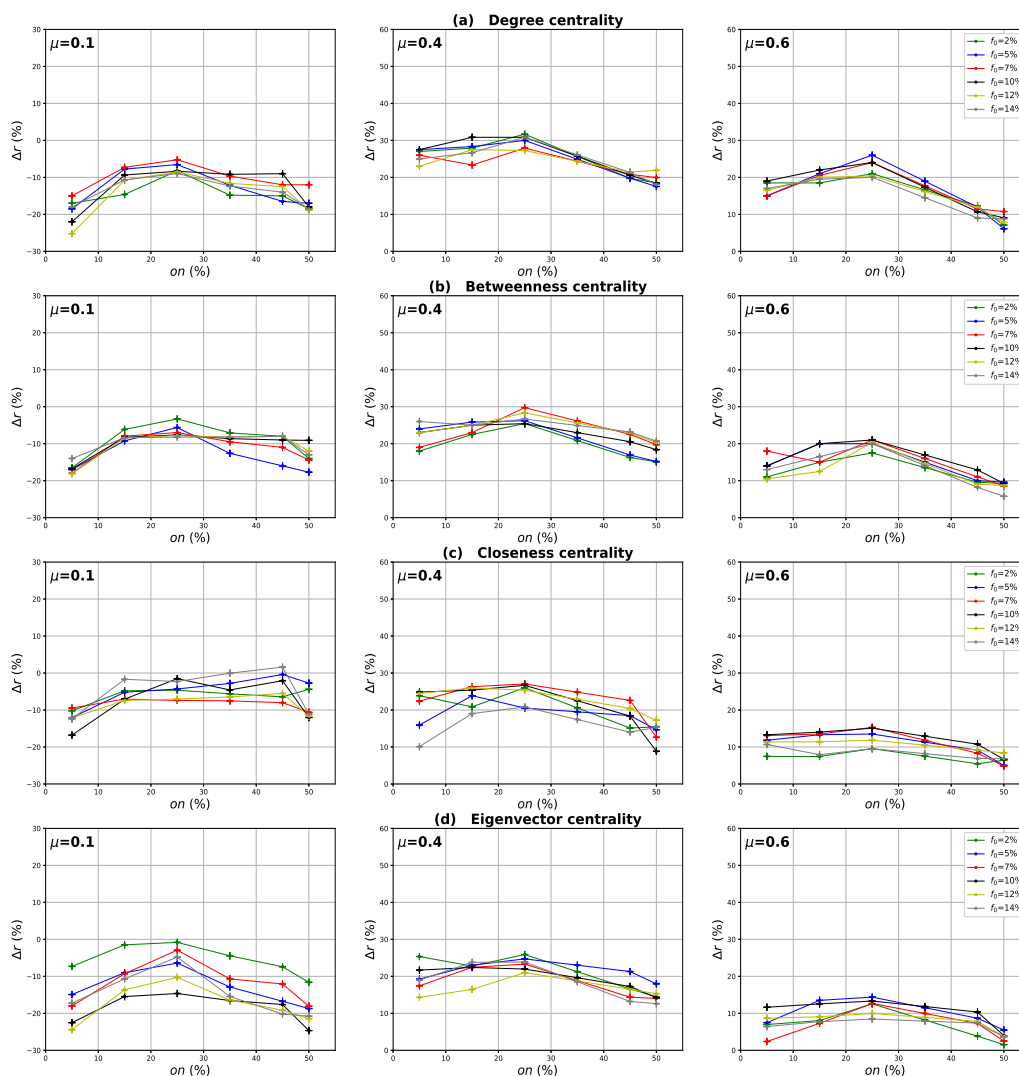


FIGURE 8 – Influence du **nombre de noeuds qui se chevauchent** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure globale** en fonction de la fraction des noeuds qui se chevauchent on . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés.

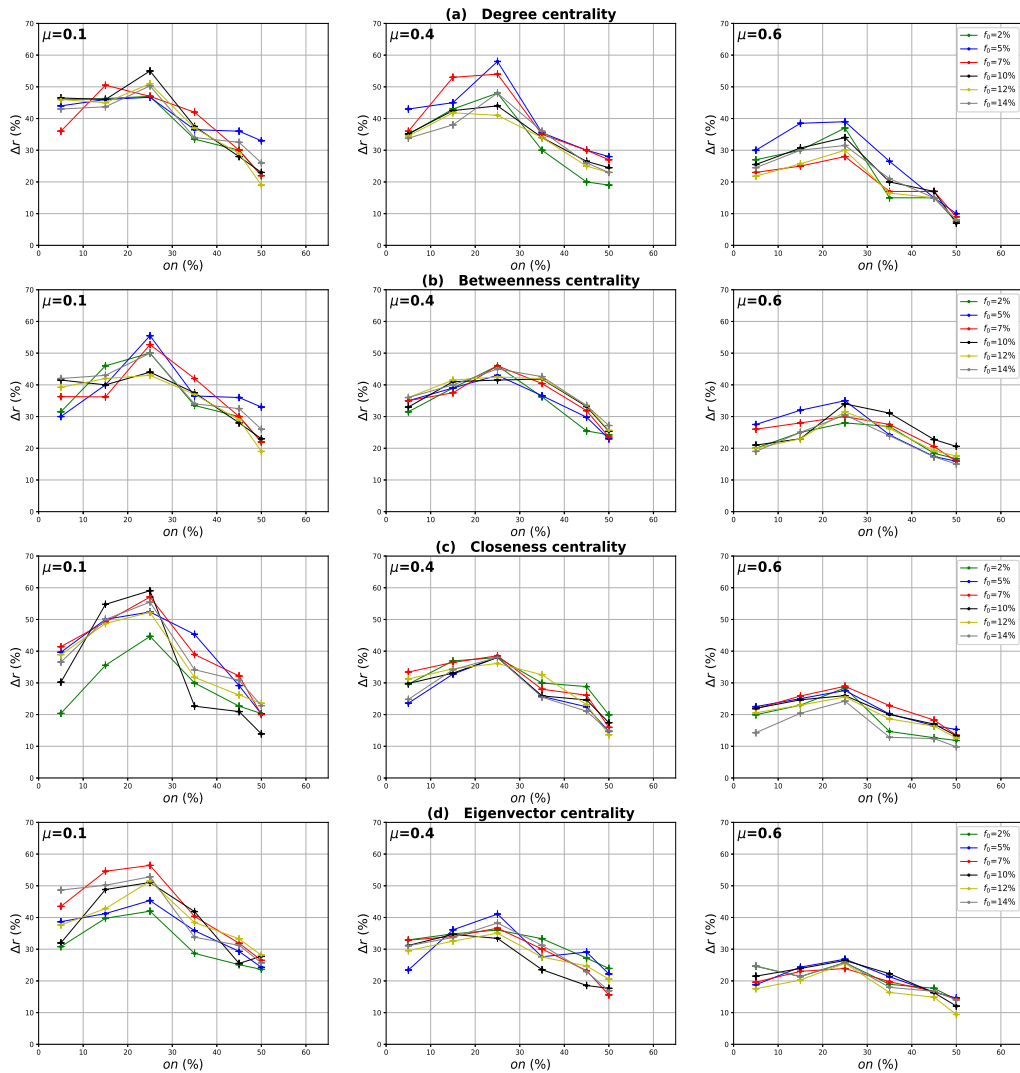


FIGURE 9 – Influence du **nombre de noeuds qui se chevauchent** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure de module** en fonction de la fraction des noeuds qui se chevauchent on . Nous avons fixé le degré d'appartenance om à 10% du nombre total de communautés.

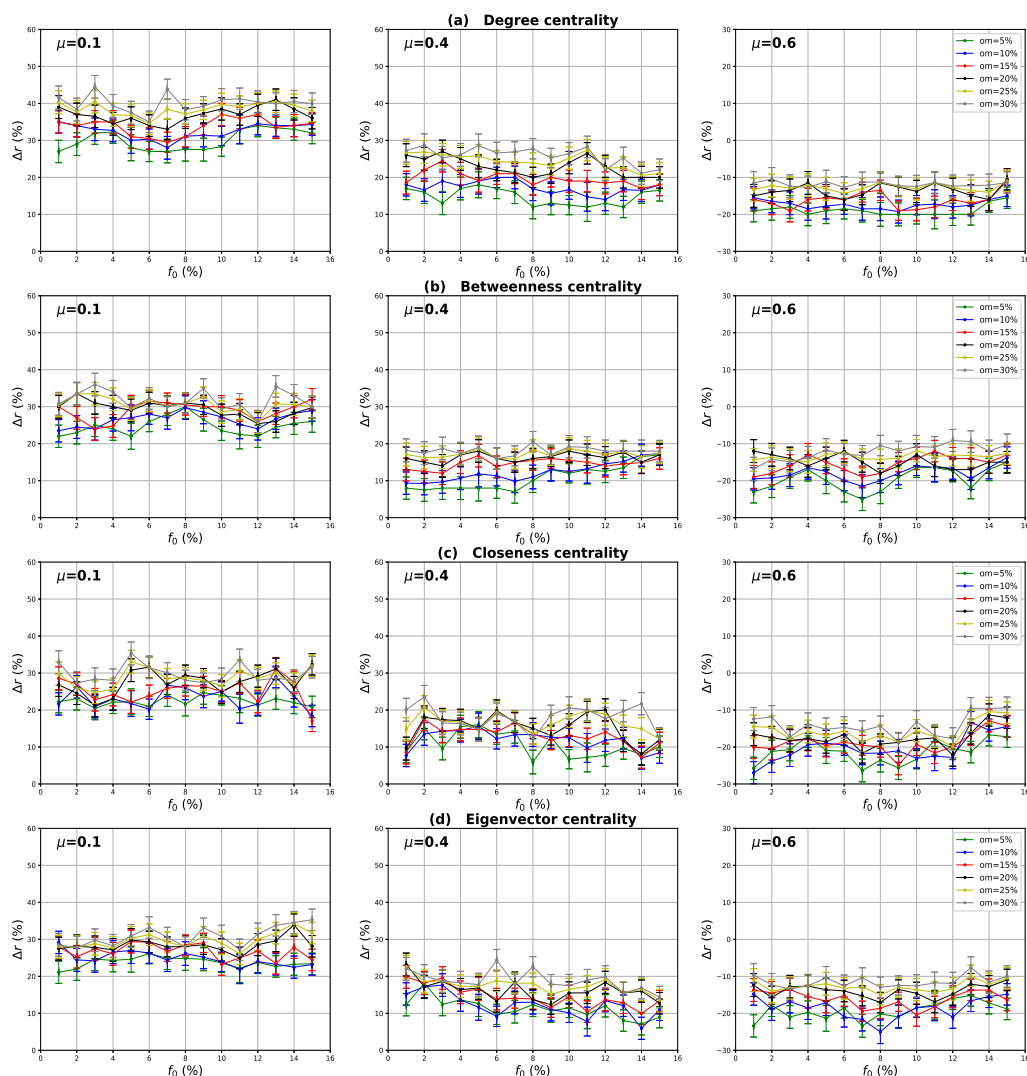


FIGURE 10 – Influence du **degré d'appartenance aux communautés** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure locale** en fonction de la fraction des épandeurs initiaux f_0 . Nous définissons le nombre de noeuds qui se chevauchent *on* à 10% du nombre total de noeuds.

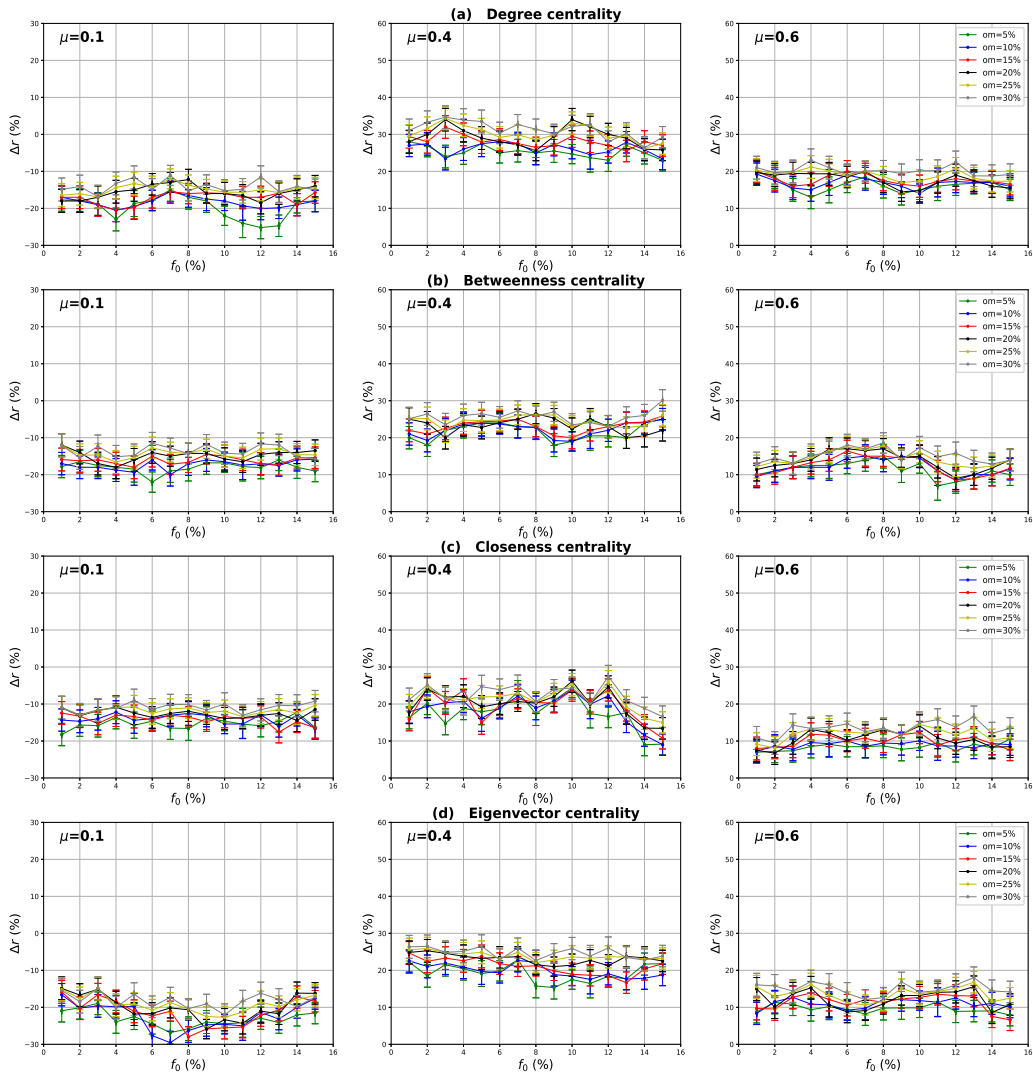


FIGURE 11 – Influence du **degré d'appartenance aux communautés** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure globale** en fonction de la fraction des épandeurs initiaux f_0 . Nous définissons le nombre de noeuds qui se chevauchent om à 10% du nombre total de noeuds.

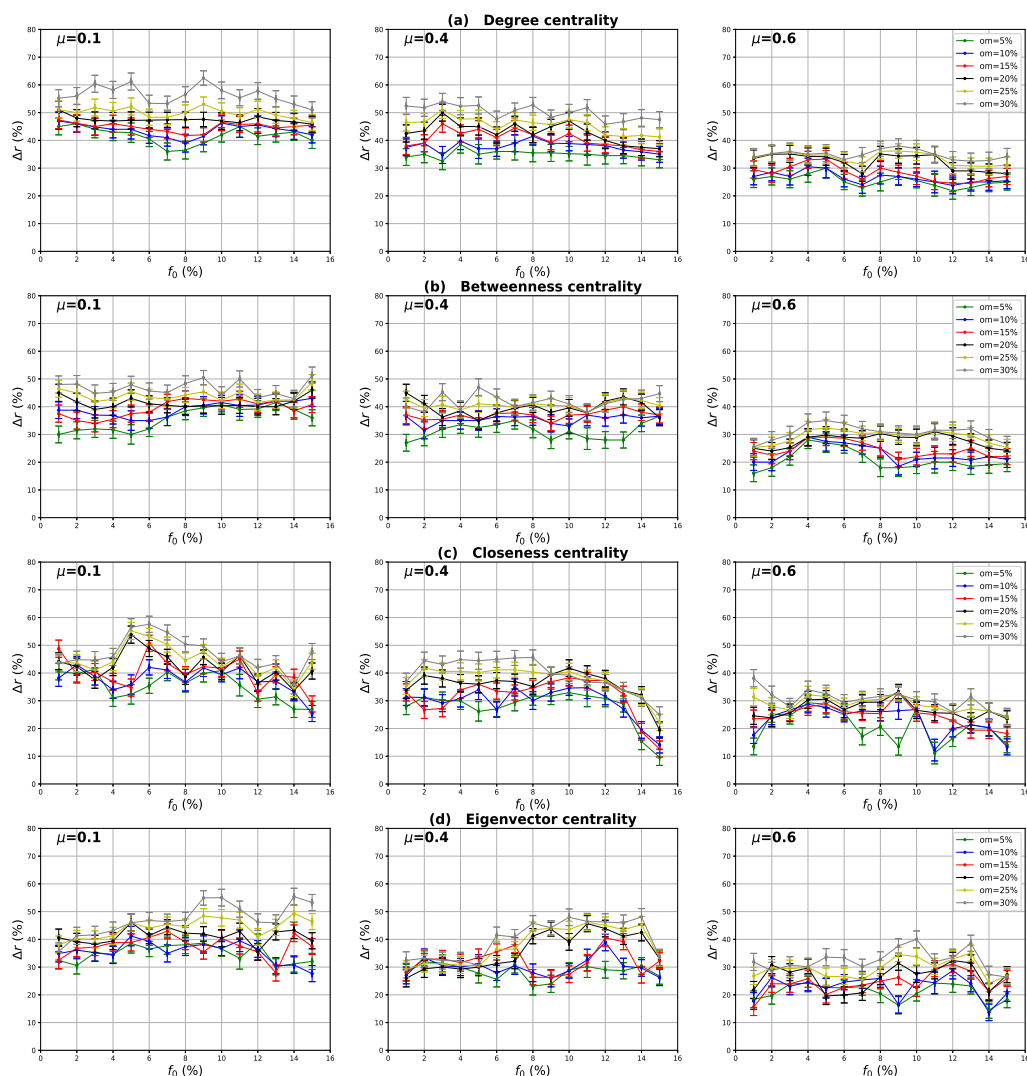


FIGURE 12 – Influence du **degré d'appartenance aux communautés** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure de module** en fonction de la fraction des épandeurs initiaux f_0 . Nous définissons le nombre de noeuds qui se chevauchent *on* à 10% du nombre total de noeuds.

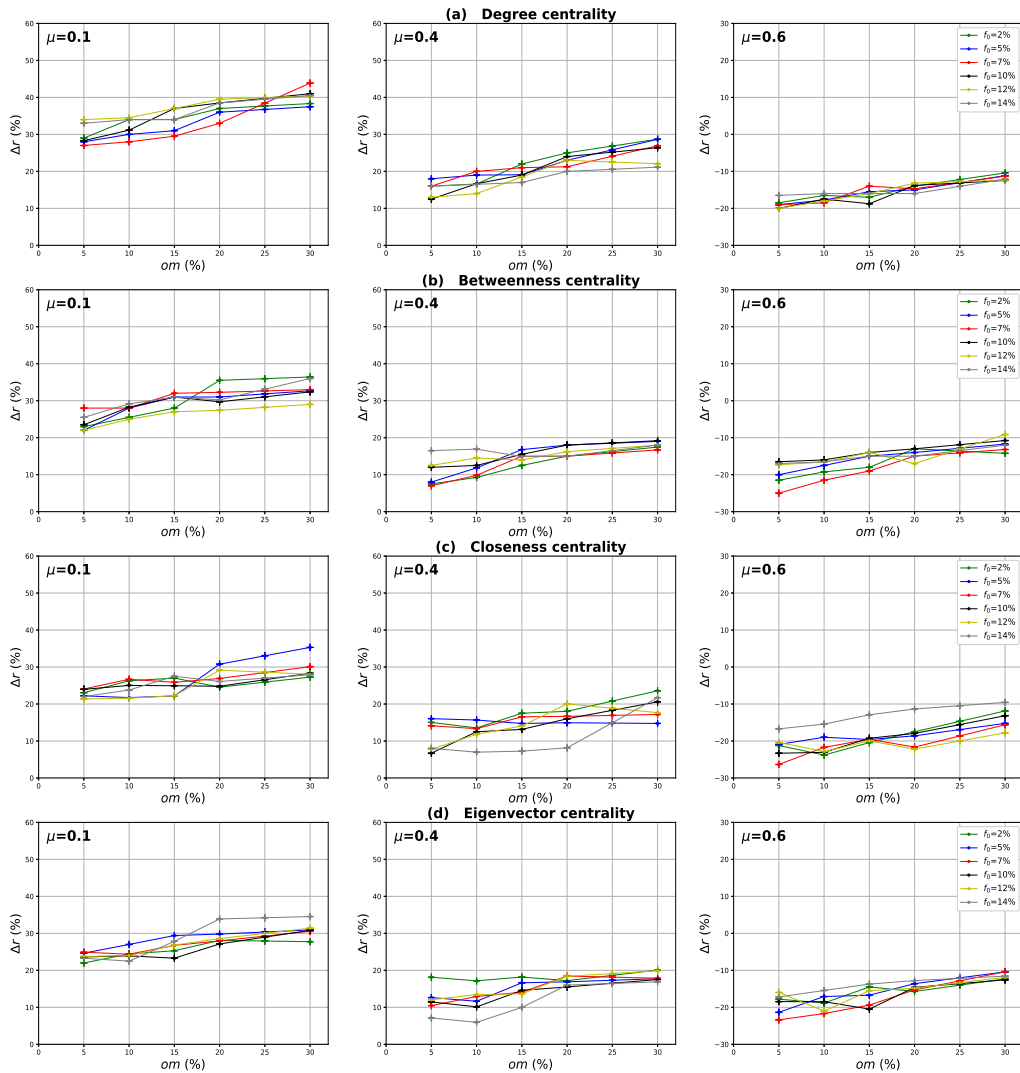


FIGURE 13 – Influence du **degré d'appartenance aux communautés** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure locale** en fonction du degré d'appartenance om . Nous avons défini le nombre de noeuds qui se chevauchent om à 10% du nombre total de communautés.

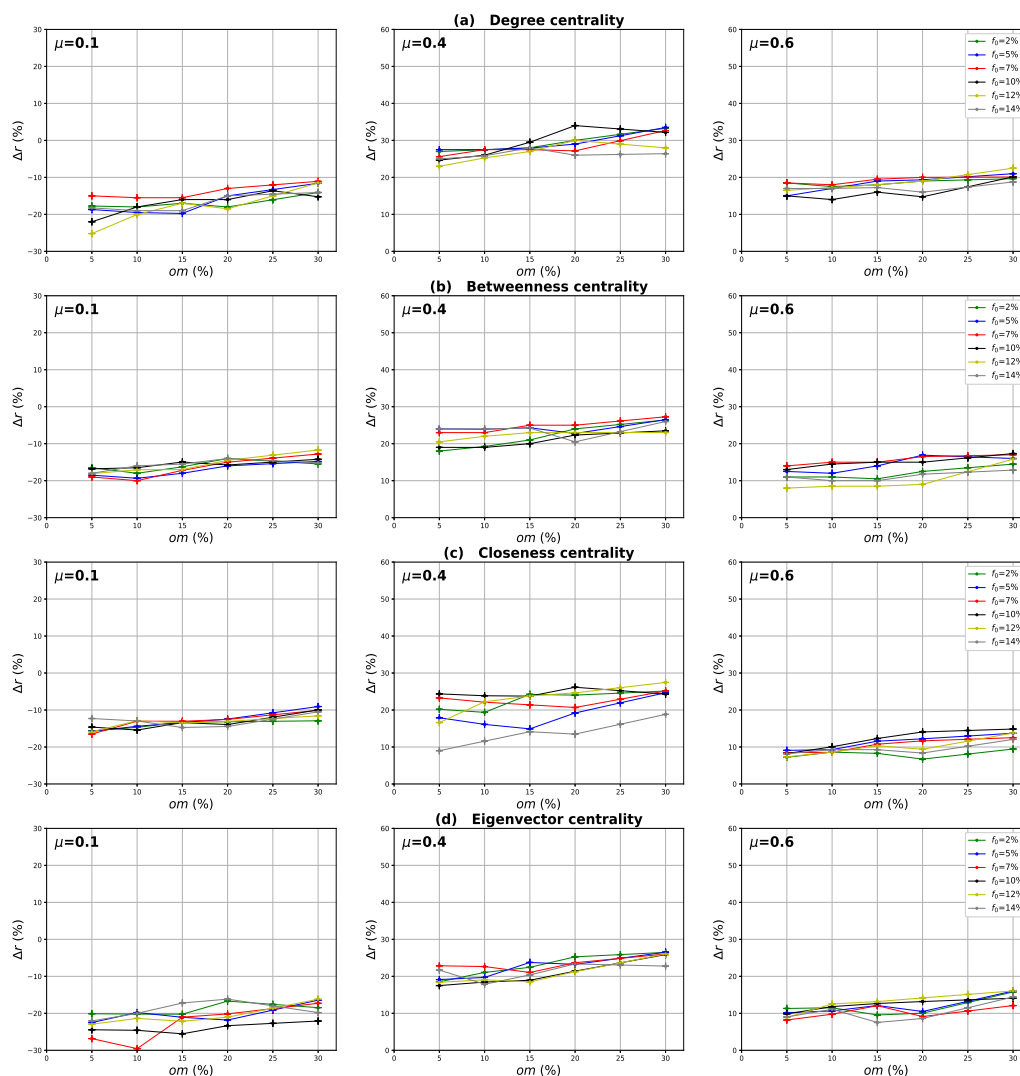


FIGURE 14 – Influence du **degré d'appartenance aux communautés** en terme de taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure globale** en fonction du degré d'appartenance om . Nous avons défini le nombre de noeuds qui se chevauchent on à 10% du nombre total de communautés.

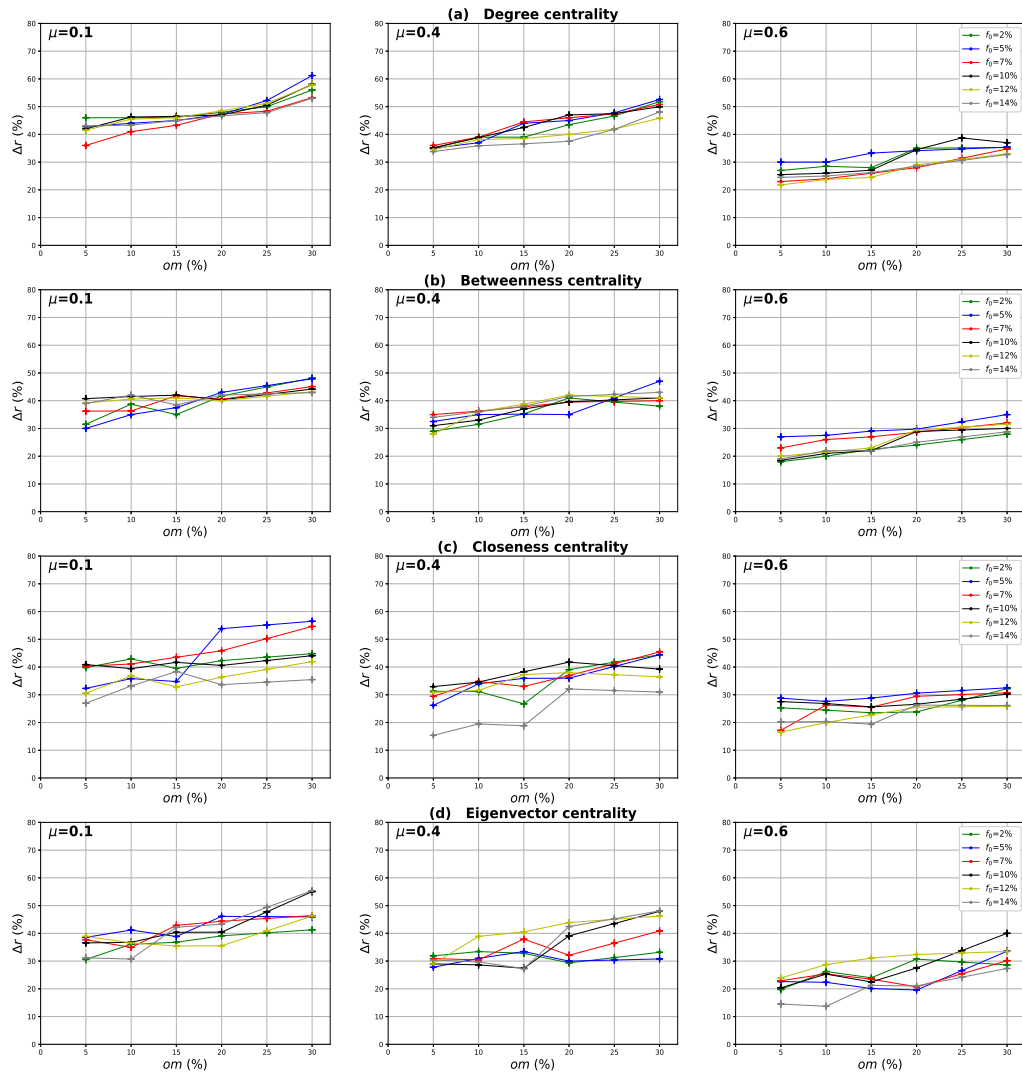


FIGURE 15 – Influence du **degré d'appartenance aux communautés** en terme de la taille épidémique. La différence relative de la taille de l'épidémie Δr entre **la mesure standard et la mesure de module** en fonction du degré d'appartenance om . Nous avons défini le nombre de noeuds qui se chevauchent on à 10% du nombre total de communautés.

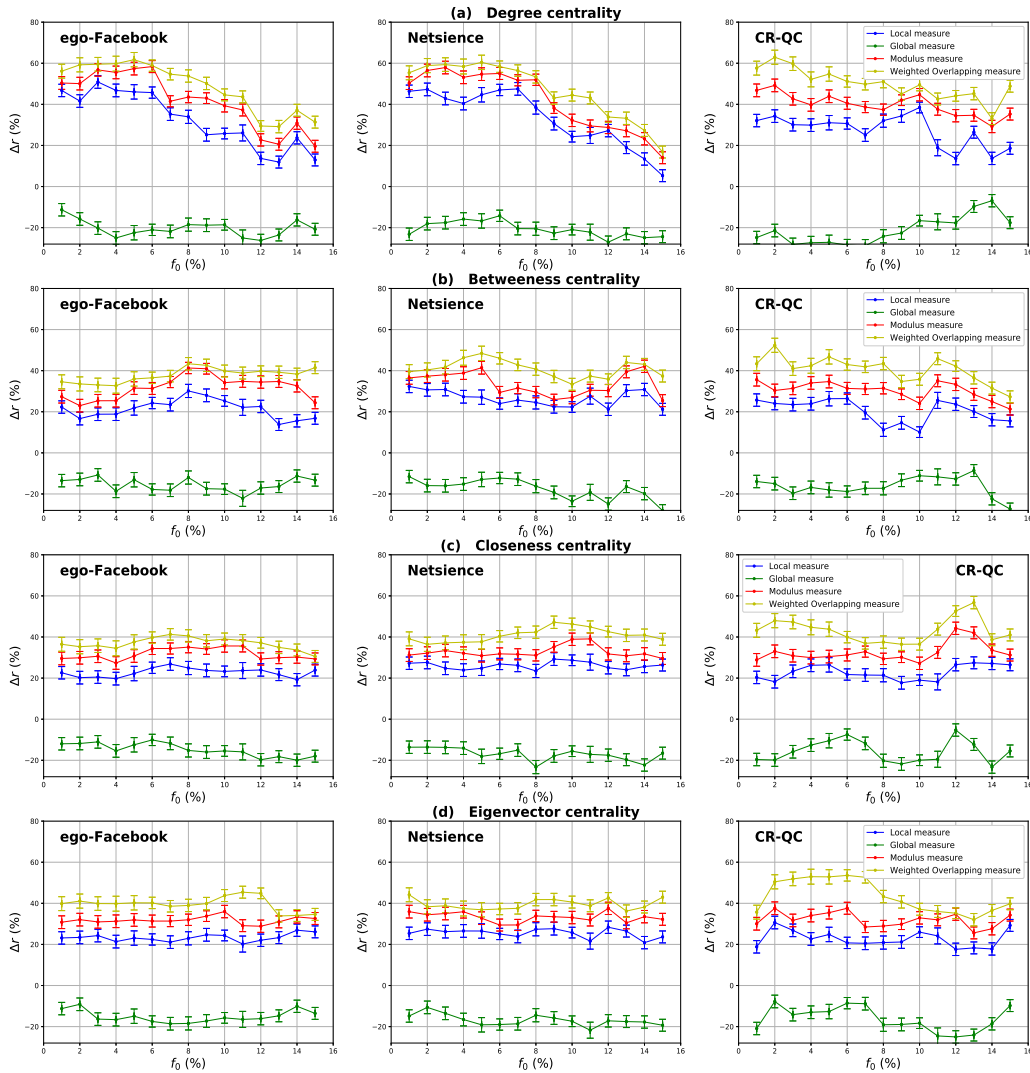


FIGURE 16 – La différence relative de la taille de l'épidémie Δr par rapport à la fraction des épandeurs initiaux f_0 . Les mesures de centralité du degré (a), d'intermédiarité (b), de la proximité (c) et du vecteur propre (d) sont comparées à leurs extensions du centralité modulaire chevauchante. Des réseaux du monde réel avec une structure communautaire bien définie (réseaux ego-Facebook, Netscience et ca-GrQc) sont utilisés. Les valeurs estimées de leur coefficient de mélange sont respectivement égales à 0.02, 0.03 et 0.11.

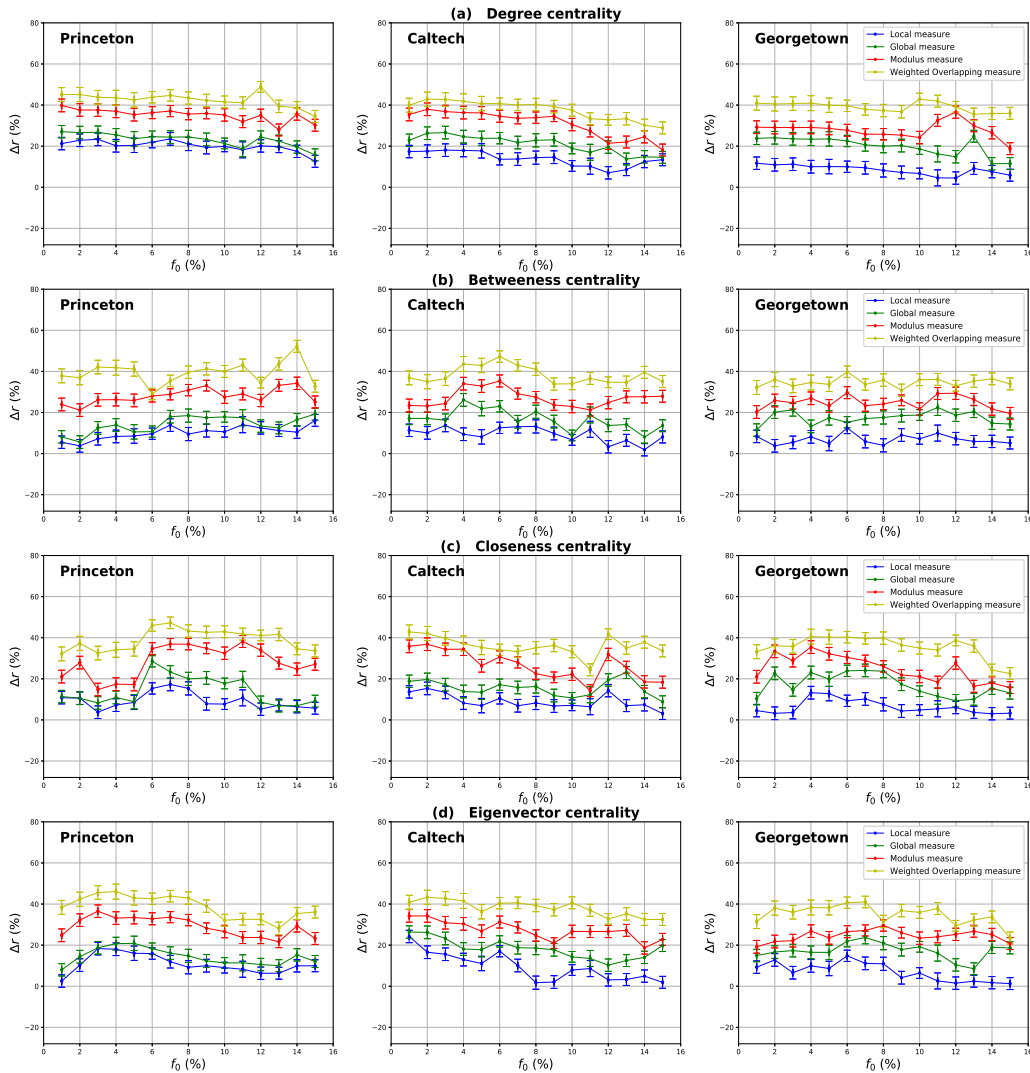


FIGURE 17 – La différence relative de la taille de l'épidémie Δr en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité du degré (a), d'intermédiarité (b), de proximité (c) et du vecteur propre (d) dérivées de la centralité modulaire chevauchante sont comparées à leurs homologues standard conçu pour les réseaux sans structure communautaire. Des réseaux du monde réel avec une structure communautaire de force moyenne (réseaux de Princeton, Caltech et Georgetown) sont utilisés. Les valeurs estimées de leur coefficient de mélange sont respectivement égales à 0.22, 0.26 et 0.3.

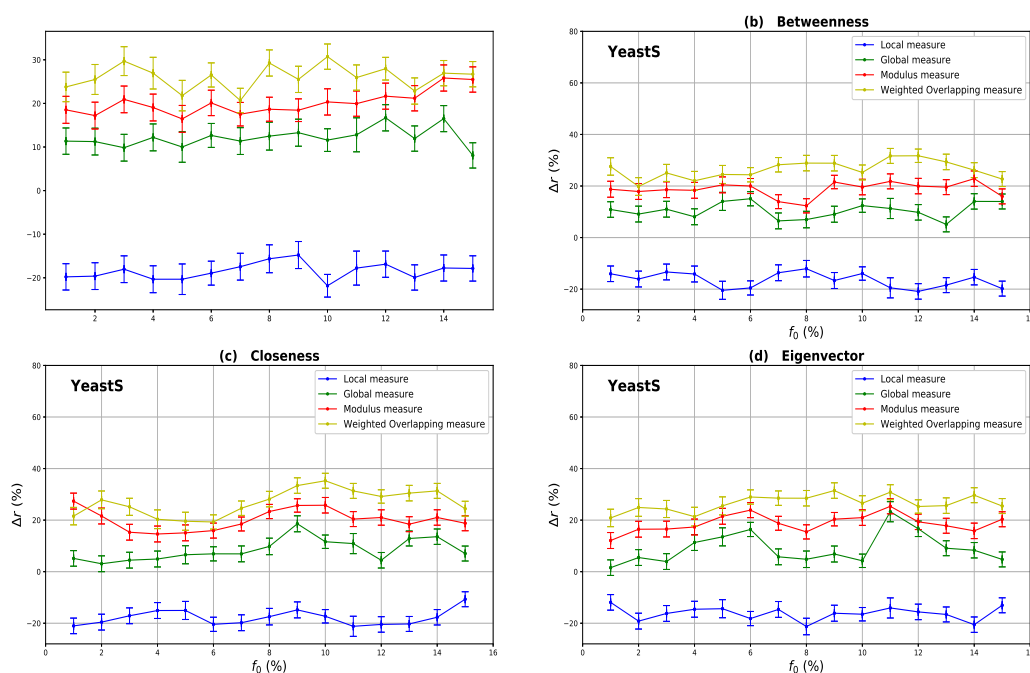


FIGURE 18 – La différence relative de la taille de l'épidémie Δr en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité du degré (a), d'intermédiarité (b), de proximité (c) et du vecteur propre (d) dérivées de la centralité modulaire chevauchante sont comparées à leurs homologues standard conçu pour les réseaux sans structure communautaire. Un réseau du monde réel (YeastS) avec une structure communautaire lâche est utilisé. La valeur estimée de son coefficient de mélange est égale à 0.47.

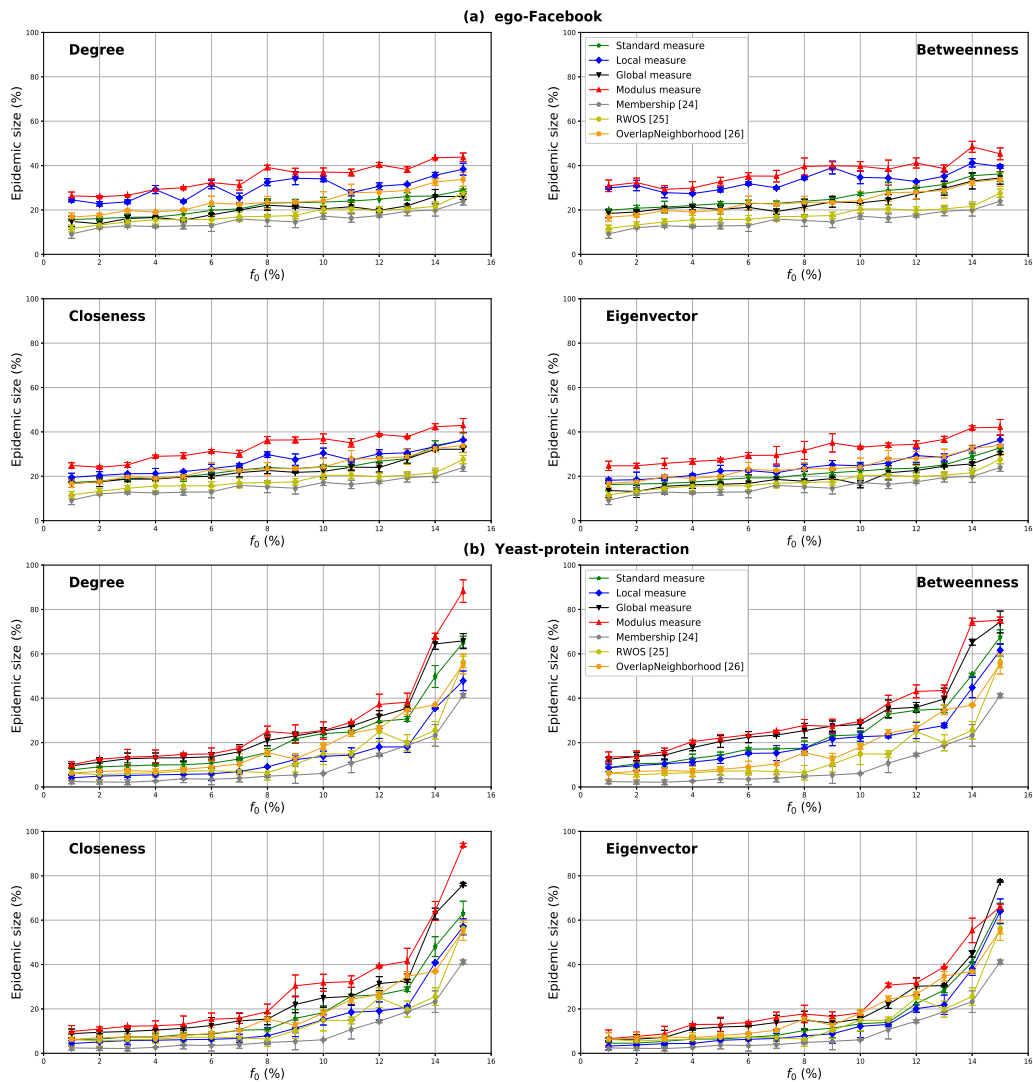


FIGURE 19 – La taille de l'épidémie en fonction de la fraction d'épandeurs initiaux f_0 . La mesure standard et les variations modulaires de centralité du degré, d'intermédiarité, de proximité et du vecteur propre ainsi que certaines mesures alternatives (Membership, RWOS et OverlapNeighborhood) sont testées sur deux réseaux du monde réel de force de structure communautaire différente.



FIGURE 20 – La différence relative de la taille de l'épidémie Δr en fonction de la fraction des épandeurs initiaux f_0 . Les mesures de centralité de degré, d'intermédiarité dérivées de la centralité modulaire chevauchante sont comparées à leurs homologues standard. Les mesures sont effectuées sur le réseau ego-Facebook en (a) et le réseau Yeast- protein interaction en (b) pour les algorithmes SLPA, LINKC et Louvain.



BIBLIOGRAPHIE

- AHN, Y.-Y., BAGROW, J. P. et LEHMANN, S. (2010). Link communities reveal multiscale complexity in networks. *nature*, 466(7307):761.
- ALBERT, R. et BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47.
- ALTMANN, M. (1993). Reinterpreting network measures for models of disease transmission. *Social networks*, 15(1):1–17.
- ARASU, A., NOVAK, J., TOMKINS, A. et TOMLIN, J. (2002). Pagerank computation and the structure of the web : Experiments and algorithms. *In Proceedings of the Eleventh International World Wide Web Conference, Poster Track*, pages 107–117.
- BARABÁSI, A.-L. (2009). Scale-free networks : a decade and beyond. *science*, 325(5939):412–413.
- BARABÁSI, A.-L. et ALBERT, R. (1999). Emergence of scaling in random networks. *science*, 286(5439):509–512.
- BARTHÉLEMY, M., BARRAT, A., PASTOR-SATORRAS, R. et VESPIGNANI, A. (2004). Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Physical review letters*, 92(17):178701.
- BAVELAS, A. (1948). A mathematical model for group structures. *Applied anthropology*, 7(3):16–30.

- BERAHMAND, K., BOUYER, A. et SAMADI, N. (2018). A new local and multidimensional ranking measure to detect spreaders in social networks. *In Computing*, pages 1–23.
- BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R. et LEFEBVRE, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics : theory and experiment*, 2008(10):P10008.
- BOCCALETTI, S., LATORA, V., MORENO, Y., CHAVEZ, M. et HWANG, D.-U. (2006). Complex networks : Structure and dynamics. *Physics reports*, 424(4-5):175–308.
- BOLLEN, J., RODRIQUEZ, M. A. et Van de SOMPEL, H. (2006). Journal status. *Scientometrics*, 69(3):669–687.
- BOMZE, I. M., BUDINICH, M., PARDALOS, P. M. et PELILLO, M. (1999). The maximum clique problem. *In Handbook of combinatorial optimization*, pages 1–74. Springer.
- BONACICH, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of mathematical sociology*, 2(1):113–120.
- BONACICH, P. (1987). Power and centrality : A family of measures. *American journal of sociology*, 92(5):1170–1182.
- BONACICH, P. (2007). Some unique properties of eigenvector centrality. *Social networks*, 29(4):555–564.
- BONACICH, P. et LLOYD, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social networks*, 23(3):191–201.
- BORGATTI, S. P. (2005). Centrality and network flow. *Social networks*, 27(1):55–71.
- BORGATTI, S. P. et EVERETT, M. G. (2006). A graph-theoretic perspective on centrality. *Social networks*, 28(4):466–484.
- BRIN, S. et PAGE, L. (1998). The anatomy of a large-scale hypertextual web search engine.
- BUTLER, D. (2008). Free journal-ranking tool enters citation market. *Nature*, 451(7174):6–7.

- CHAKRABORTY, D., SINGH, A. et CHERIFI, H. (2016). Immunization strategies based on the overlapping nodes in networks with community structure. *In International Conference on Computational Social Networks*, pages 62–73. Springer.
- CHAN, S. Y., LEUNG, I. X. et LIÒ, P. (2009). Fast centrality approximation in modular networks. *In Proceedings of the 1st ACM international workshop on Complex networks meet information & knowledge management*, pages 31–38. ACM.
- CHEN, D., LÜ, L., SHANG, M.-S., ZHANG, Y.-C. et ZHOU, T. (2012). Identifying influential nodes in complex networks. *Physica a : Statistical mechanics and its applications*, 391(4):1777–1787.
- CHEN, P., XIE, H., MASLOV, S. et REDNER, S. (2007). Finding scientific gems with google’s pagerank algorithm. *Journal of Informetrics*, 1(1):8–15.
- CHEPELIANSKII, A. D. (2010). Towards physical laws for software architecture. *arXiv preprint arXiv :1003.5455*.
- CHERIFI, H. (2014). Chapter eleven non-overlapping community detection. *Complex Networks and Their Applications*, page 320.
- CHU, X., GUAN, J., ZHANG, Z. et ZHOU, S. (2009). Epidemic spreading in weighted scale-free networks with community structure. *Journal of Statistical Mechanics : Theory and Experiment*, 2009(07):P07043.
- COHEN, R., HAVLIN, S. et BEN-AVRAHAM, D. (2003). Efficient immunization strategies for computer networks and populations. *Physical review letters*, 91(24):247901.
- COSCIA, M., GIANNOTTI, F. et PEDRESCHI, D. (2011). A classification for community discovery methods in complex networks. *Statistical Analysis and Data Mining : The ASA Data Science Journal*, 4(5):512–546.
- COSCIA, M., ROSSETTI, G., GIANNOTTI, F. et PEDRESCHI, D. (2012). Demon : a local-first discovery method for overlapping communities. *In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 615–623. ACM.

- CROFTS, J. J. et HIGHAM, D. J. (2011). Googling the brain : Discovering hierarchical and asymmetric network structures, with applications in neuroscience. *Internet Mathematics*, 7(4):233–254.
- DANON, L., DIAZ-GUILERA, A., DUCH, J. et ARENAS, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics : Theory and Experiment*, 2005(09):P09008.
- DING, Y., YAN, E., FRAZHO, A. et CAVERLEE, J. (2009). Pagerank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60(11):2229–2243.
- DOROGOVTSSEV, S. N., GOLTSEV, A. V. et MENDES, J. F. F. (2006). K-core organization of complex networks. *Physical review letters*, 96(4):040601.
- FERGUSON, N. M., CUMMINGS, D. A., CAUCHEMEZ, S., FRASER, C., RILEY, S., MEEYAL, A., IAMSIRITHAWORN, S. et BURKE, D. S. (2005). Strategies for containing an emerging influenza pandemic in southeast asia. *Nature*, 437(7056):209–214.
- FERGUSON, N. M., CUMMINGS, D. A., FRASER, C., CAJKA, J. C., COOLEY, P. C. et BURKE, D. S. (2006). Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452.
- FERRARA, E. (2012). Community structure discovery in facebook. *International Journal of Social Network Mining*, 1(1):67–90.
- FERRARA, E., DE MEO, P., FIUMARA, G. et PROVETTI, A. (2012). The role of strong and weak ties in facebook : a community structure perspective. *arXiv preprint arXiv :1203.0535*.
- FORTUNATO, S. (2010). Community detection in graphs. *Physics reports*, 486(3-5):75–174.
- FORTUNATO, S. et HRIC, D. (2016). Community detection in networks : A user guide. *Physics Reports*, 659:1–44.

- FREEMAN, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41.
- FREEMAN, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239.
- FRESCHI, V. (2007). Protein function prediction from interaction networks using a random walk ranking algorithm. *In 2007 IEEE 7th International Symposium on BioInformatics and BioEngineering*, pages 42–48. IEEE.
- GAO, C., WEI, D., HU, Y., MAHADEVAN, S. et DENG, Y. (2013). A modified evidential methodology of identifying influential nodes in weighted networks. *Physica A : Statistical Mechanics and its Applications*, 392(21):5490–5500.
- GHALMANE, Z., EL HASSOUNI, M. et CHERIFI, H. (2018). Betweenness centrality for networks with non-overlapping community structure. *In 2018 IEEE Workshop on Complexity in Engineering (COMPENG)*, pages 1–5. IEEE.
- GHOSHAL, G. et BARABÁSI, A.-L. (2011). Ranking stability and super-stable nodes in complex networks. *Nature communications*, 2(1):1–7.
- GIABBANELLI, P. J. et PETERS, J. G. (2011). Réseaux complexes et epidemies. *TSI-Technique et Science Informatiques*, 30(2):181.
- GLEICH, D. F. (2015). Pagerank beyond the web. *SIAM Review*, 57(3):321–363.
- GONG, K. (2014). Local immunization based on degree-community-bridge-find in heterogeneity community networks. *In International Conference on Logistics Engineering, Management and Computer Science (LEMCS 2014)*. Atlantis Press.
- GONG, K., TANG, M., HUI, P. M., ZHANG, H. F., YOUNGHAЕ, D. et LAI, Y.-C. (2013). An efficient immunization strategy for community networks. *PloS one*, 8(12):e83489.
- GOVAN, A. Y., MEYER, C. D. et ALBRIGHT, R. (2008). Generalizing google’s pagerank to rank national football league teams. *In Proceedings of the SAS Global Forum*, volume 2008.

- GRANOVETTER, M. S. (1977). The strength of weak ties. pages 347–367.
- GRENFELL, B. et BJØRNSTAD, O. (2005). Epidemic cycling and immunity. *Nature*, 433(7024):366–367.
- GUAN-RONG, C., XIAO-FAN, W. et XIANG, L. (2012). Introduction to complex networks : Models, structures and dynamics.
- GUPTA, N., SINGH, A. et CHERIFI, H. (2015). Community-based immunization strategies for epidemic control. *In Communication Systems and Networks (COMSNETS), 2015 7th International Conference on*, pages 1–6. IEEE.
- GUPTA, N., SINGH, A. et CHERIFI, H. (2016). Centrality measures for networks with community structure. *Physica A : Statistical Mechanics and its Applications*, 452:46–59.
- HAGE, P. et HARARY, F. (1995). Eccentricity and centrality in networks. *Social networks*, 17(1):57–63.
- HALLORAN, M. E., FERGUSON, N. M., EUBANK, S., LONGINI, I. M., CUMMINGS, D. A., LEWIS, B., XU, S., FRASER, C., VULLIKANTI, A., GERMANN, T. C. *et al.* (2008). Modeling targeted layered containment of an influenza pandemic in the united states. *Proceedings of the National Academy of Sciences*, 105(12):4639–4644.
- HAVLIN, S., KENETT, D. Y., BEN-JACOB, E., BUNDE, A., COHEN, R., HERMANN, H., KANTELHARDT, J., KERTÉSZ, J., KIRKPATRICK, S., KURTHS, J. *et al.* (2012). Challenges in network science : Applications to infrastructures, climate, social systems and economics. *The European Physical Journal Special Topics*, 214(1):273–293.
- HE, J.-L., FU, Y. et CHEN, D.-B. (2015). A novel top-k strategy for influence maximization in complex networks with community structure. *PloS one*, 10(12):e0145283.
- HÉBERT-DUFRESNE, L., ALLARD, A., YOUNG, J.-G. et DUBÉ, L. J. (2013). Global efficiency of local immunization on complex networks. *Scientific reports*, 3:2171.

- HOTELLING, H. (1936). Simplified calculation of principal components. *Psychometrika*, 1(1):27–35.
- HUANG, L., PARK, K. et LAI, Y.-C. (2006). Information propagation on modular networks. *Physical Review E*, 73(3):035103.
- HUANG, W. et LI, C. (2007). Epidemic spreading in scale-free networks with community structure. *Journal of Statistical Mechanics : Theory and Experiment*, 2007(01):P01014.
- HUBERT, L. et ARABIE, P. (1985). Comparing partitions. *Journal of classification*, 2(1): 193–218.
- HURLEY, M., JACOBS, G. et GILBERT, M. (2006). The basic si model. *New Directions for Teaching and Learning*, 2006(106):11–22.
- IYER, S., KILLINGBACK, T., SUNDARAM, B. et WANG, Z. (2013). Attack robustness and centrality of complex networks. *PloS one*, 8(4).
- JAVA, A., KOLARI, P., FININ, T., OATES, T. *et al.* (2006). Modeling the spread of influence on the blogosphere. *UMBC TR-CS-06-03*.
- JEBABLI, M., CHERIFI, H., CHERIFI, C. et HAMMOUDA, A. (2014). Overlapping community structure in co-authorship networks : A case study. *In u-and e-Service, Science and Technology (UNESST), 2014 7th International Conference on*, pages 26–29. IEEE.
- JEBABLI, M., CHERIFI, H., CHERIFI, C. et HAMMOUDA, A. (2015). User and group networks on youtube : A comparative analysis. *In Computer Systems and Applications (AICCSA), 2015 IEEE/ACS 12th International Conference of*, pages 1–8. IEEE.
- JEBABLI, M., CHERIFI, H., CHERIFI, C. et HAMMOUDA, A. (2018). Community detection algorithm evaluation with ground-truth data. *Physica A : Statistical Mechanics and its Applications*, 492:651–706.
- JENSEN, P., MORINI, M., KARSAI, M., VENTURINI, T., VESPIGNANI, A., JACOMY, M., COINTET, J.-P., MERCKLÉ, P. et FLEURY, E. (2015). Detecting global bridges in networks. *Journal of Complex Networks*, 4(3):319–329.

- JEUB, L. G., BALACHANDRAN, P., PORTER, M. A., MUCHA, P. J. et MAHONEY, M. W. (2015). Think locally, act locally : Detection of small, medium-sized, and large communities in large networks. *Physical Review E*, 91(1):012821.
- JIA-SHENG, W., XIAO-PING, W., BO, Y. et JIANG-WEI, G. (2011). Improved method of node importance evaluation based on node contraction in complex networks. *Procedia Engineering*, 15:1600–1604.
- JING, Y. et BALUJA, S. (2008). Pagerank for product image search. In *Proceedings of the 17th international conference on World Wide Web*, pages 307–316.
- JONES, J. H. et HANDCOCK, M. S. (2003). Sexual contacts and epidemic thresholds. *Nature*, 423(6940):605–606.
- KATZ, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43.
- KEELING, M. J. et ROHANI, P. (2011). *Modeling infectious diseases in humans and animals*. Princeton University Press.
- KIM, M., SUMBALY, R. et SHAH, S. (2013). Root cause detection in a service-oriented architecture. *ACM SIGMETRICS Performance Evaluation Review*, 41(1):93–104.
- KITROMILIDIS, M. et EVANS, T. S. (2018). Community detection with metadata in a network of biographies of western art painters. *arXiv preprint arXiv :1802.07985*.
- KITSAK, M., GALLOS, L. K., HAVLIN, S., LILJEROS, F., MUCHNIK, L., STANLEY, H. E. et MAKSE, H. A. (2010). Identification of influential spreaders in complex networks. *Nature physics*, 6(11):888–893.
- KUMAR, M., SINGH, A. et CHERIFI, H. (2018). An efficient immunization strategy using overlapping nodes and its neighborhoods. In *Companion of the The Web Conference 2018 on The Web Conference 2018*, pages 1269–1275. International World Wide Web Conferences Steering Committee.

- LANCICHINETTI, A., FORTUNATO, S. et KERTESZ, J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. *New journal of physics*, 11(3):033015.
- LANCICHINETTI, A., FORTUNATO, S. et RADICCHI, F. (2008). Benchmark graphs for testing community detection algorithms. *Physical review E*, 78(4):046110.
- LANGVILLE, A. N. et MEYER, C. D. (2011). *Google's PageRank and beyond : The science of search engine rankings*. Princeton university press.
- LESKOVEC, J., LANG, K. J., DASGUPTA, A. et MAHONEY, M. W. (2009). Community structure in large networks : Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123.
- LESKOVEC, J. et MCAULEY, J. J. (2012). Learning to discover social circles in ego networks. *In Advances in neural information processing systems*, pages 539–547.
- LIU, X., BOLLEN, J., NELSON, M. L. et Van de SOMPEL, H. (2005). Co-authorship networks in the digital library research community. *Information processing & management*, 41(6):1462–1480.
- LIU, Z. et HU, B. (2005). Epidemic spreading in community networks. *EPL (Europhysics Letters)*, 72(2):315.
- LÜ, L., CHEN, D., REN, X.-L., ZHANG, Q.-M., ZHANG, Y.-C. et ZHOU, T. (2016). Vital nodes identification in complex networks. *Physics Reports*, 650:1–63.
- LÜ, L., ZHANG, Y.-C., YEUNG, C. H. et ZHOU, T. (2011). Leaders in social networks, the delicious case. *PloS one*, 6(6).
- LUO, S.-L., GONG, K. et KANG, L. (2016). Identifying influential spreaders of epidemics on community networks. *CoRR*, abs/1601.07700.
- MA, N., GUAN, J. et ZHAO, Y. (2008). Bringing pagerank to the citation analysis. *Information Processing & Management*, 44(2):800–810.

- MA, Y.-F., LU, L., ZHANG, H.-J. et LI, M. (2002). A user attention model for video summarization. pages 533–542.
- MADAR, N., KALISKY, T., COHEN, R., BEN-AVRAHAM, D. et HAVLIN, S. (2004). Immunization and epidemic dynamics in complex networks. *The European Physical Journal B*, 38(2):269–276.
- MANTZARIS, A. V. (2014). Uncovering nodes that spread information between communities in social networks. *EPJ Data Science*, 3(1):26.
- MASUDA, N. (2009). Immunization of networks with community structure. *New Journal of Physics*, 11(12):123018.
- MEDO, M., ZHANG, Y.-C. et ZHOU, T. (2009). Adaptive model for recommendation of news. *EPL (Europhysics Letters)*, 88(3):38005.
- MENG, X. (2009). Computing bookrank via social cataloging. *In Web slides for CADS 2010 conference., February*, volume 22.
- MOONEY, B. L., CORRALES, L. R. et CLARK, A. E. (2012). Molecular networks : An integrated graph theoretic and data mining tool to explore solvent organization in molecular simulation. *Journal of computational chemistry*, 33(8):853–860.
- MORENO, Y., PASTOR-SATORRAS, R. et VESPIGNANI, A. (2002). Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 26(4):521–529.
- MORRISON, J. L., BREITLING, R., HIGHAM, D. J. et GILBERT, D. R. (2005). Gene-rank : using search engine technology for the analysis of microarray experiments. *BMC bioinformatics*, 6(1):233.
- NEWMAN, M. E. (2002). Spread of epidemic disease on networks. *Physical review E*, 66(1):016128.
- NEWMAN, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2):167–256.

- NEWMAN, M. E. (2004). Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6):066133.
- NEWMAN, M. E. (2006a). Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104.
- NEWMAN, M. E. (2006b). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582.
- ORMAN, G. K., LABATUT, V. et CHERIFI, H. (2011). On accuracy of community structure discovery algorithms. *Journal of Convergence Information Technology*, Vol. 6 Issue 11:p283.
- ORMAN, G. K., LABATUT, V. et CHERIFI, H. (2012). Comparative evaluation of community detection algorithms : a topological approach. *Journal of Statistical Mechanics : Theory and Experiment*, 2012(08):P08001.
- ORMAN, G. K., LABATUT, V. et CHERIFI, H. (2013). Towards realistic artificial benchmark for community detection algorithms evaluation. *International Journal of Web Based Communities*, 9(3):349–370.
- PALLA, G., DERÉNYI, I., FARKAS, I. et VICSEK, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *nature*, 435(7043):814.
- PASTOR-SATORRAS, R. et VESPIGNANI, A. (2001). Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200.
- PEI, S., MUCHNIK, L., ANDRADE JR, J. S., ZHENG, Z. et MAKSE, H. A. (2014). Searching for superspreaders of information in real-world social media. *Scientific reports*, 4:5547.
- PONS, P. et LATAPY, M. (2006). Computing communities in large networks using random walks. *J. Graph Algorithms Appl.*, 10(2):191–218.
- PORTER, M. A., ONNELA, J.-P. et MUCHA, P. J. (2009). Communities in networks. *Notices of the AMS*, 56(9):1082–1097.

- POULIN, R., BOILY, M.-C. et MÂSSE, B. R. (2000). Dynamical systems to define centrality in social networks. *Social networks*, 22(3):187–220.
- RADICCHI, F. (2011). Who is the best player ever? a complex network analysis of the history of professional tennis. *PloS one*, 6(2).
- RAGHAVAN, U. N., ALBERT, R. et KUMARA, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, 76(3):036106.
- RESTREPO, J. G., OTT, E. et HUNT, B. R. (2008). Weighted percolation on directed networks. *Physical review letters*, 100(5):058701.
- ROSVALL, M. et BERGSTROM, C. T. (2007). An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences*, 104(18):7327–7331.
- ROSVALL, M. et BERGSTROM, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123.
- SABIDUSSI, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4):581–603.
- SALATHÉ, M. et JONES, J. H. (2010). Dynamics and control of diseases in networks with community structure. *PLoS computational biology*, 6(4):e1000736.
- SALAVATI, C., ABDOLLAHPOURI, A. et MANBARI, Z. (2019). Ranking nodes in complex networks based on local structure and improving closeness centrality. *Neurocomputing*, 336:36–45.
- SAXENA, R., KAUR, S. et BHATNAGAR, V. (2018). Social centrality using network hierarchy and community structure. *arXiv preprint arXiv :1806.08964*.
- SAYYADI, H. et GETOOR, L. (2009). Futurerank : Ranking scientific articles by predicting their future pagerank. In *Proceedings of the 2009 SIAM International Conference on Data Mining*, pages 533–544. SIAM.

- SHAW, M. E. (1954). Group structure and the behavior of individuals in small groups. *The Journal of psychology*, 38(1):139–149.
- SHIMBEL, A. (1953). Structural parameters of communication networks. *The bulletin of mathematical biophysics*, 15(4):501–507.
- SINGH, A. et SINGH, Y. N. (2012a). Nonlinear spread of rumor and inoculation strategies in the nodes with degree dependent tie strength in complex networks. *arXiv preprint arXiv :1208.6063*.
- SINGH, A. et SINGH, Y. N. (2012b). Rumor spreading and inoculation of nodes in complex networks. *In Proceedings of the 21st International Conference on World Wide Web*, pages 675–678.
- SINGH, A. et SINGH, Y. N. (2013). Rumor dynamics with inoculations for correlated scale free networks. *In 2013 National Conference on Communications (NCC)*, pages 1–5. IEEE.
- STEPHENSON, K. et ZELEN, M. (1989). Rethinking centrality : Methods and examples. *Social networks*, 11(1):1–37.
- SUN, H. et GAO, Z. (2007). Dynamical behaviors of epidemics on scale-free networks with community structure. *Physica A : Statistical Mechanics and its Applications*, 381:491–496.
- TAGHAVIAN, F., SALEHI, M. et TEIMOURI, M. (2017). A local immunization strategy for networks with overlapping community structure. *Physica A : Statistical Mechanics and its Applications*, 467:148–156.
- TRAUD, A. L., MUCHA, P. J. et PORTER, M. A. (2012). Social structure of facebook networks. *Physica A : Statistical Mechanics and its Applications*, 391(16):4165–4180.
- TULU, M. M., HOU, R. et YOUNAS, T. (2018). Identifying influential nodes based on community structure to speed up the dissemination of information in complex network. *IEEE ACCESS*, 6:7390–7401.

- WANG, L., ZHANG, Y., HUANG, T. et LI, X. (2012). Estimating the value of containment strategies in delaying the arrival time of an influenza pandemic : A case study of travel restriction and patient isolation. *Physical Review E*, 86(3):032901.
- WANG, W., LIU, Q.-H., ZHONG, L.-F., TANG, M., GAO, H. et STANLEY, H. E. (2016). Predicting the epidemic threshold of the susceptible-infected-recovered model. *Scientific reports*, 6:24676.
- WANG, Z., MORENO, Y., BOCCALETTI, S. et PERC, M. (2017). Vaccination and epidemics in networked populations, an introduction.
- WASSERMAN, S., FAUST, K. *et al.* (1994). *Social network analysis : Methods and applications*, volume 8. Cambridge university press.
- WATTS, D. J. et STROGATZ, S. H. (1998). Collective dynamics of small-world networks. *nature*, 393(6684):440.
- WEI, H., PAN, Z., HU, G., ZHANG, L., YANG, H., LI, X. et ZHOU, X. (2018). Identifying influential nodes based on network representation learning in complex networks. *PloS one*, 13(7):e0200091.
- WENG, J., LIM, E.-P., JIANG, J. et HE, Q. (2010). Twitterank : finding topic-sensitive influential twitterers. *In Proceedings of the third ACM international conference on Web search and data mining*, pages 261–270.
- WEST, J. D., BERGSTROM, T. C. et BERGSTROM, C. T. (2010). The eigenfactor metricstm : A network approach to assessing scholarly journals. *College & Research Libraries*, 71(3):236–244.
- WITTENBAUM, G. M., HUBBELL, A. P. et ZUCKERMAN, C. (1999). Mutual enhancement : Toward an understanding of the collective preference for shared information. *Journal of personality and social psychology*, 77(5):967.
- WU, X. et LIU, Z. (2008). How community structure influences epidemic spread in social networks. *Physica A : Statistical Mechanics and its Applications*, 387(2-3):623–630.

- XIE, J., KELLEY, S. et SZYMANSKI, B. K. (2013). Overlapping community detection in networks : The state-of-the-art and comparative study. *Acm computing surveys (csur)*, 45(4):43.
- XIE, J., SZYMANSKI, B. K. et LIU, X. (2011). Slpa : Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. *In 2011 IEEE 11th International Conference on Data Mining Workshops*, pages 344–349. IEEE.
- YANG, J. et LESKOVEC, J. (2012). Structure and overlaps of communities in networks. *arXiv preprint arXiv :1205.6228*.
- YANG, J. et LESKOVEC, J. (2014). Structure and overlaps of ground-truth communities in networks. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(2):26.
- YOSHIDA, T. et YAMADA, Y. (2017). A community structure-based approach for network immunization. *Computational Intelligence*, 33(1):77–98.
- ZHANG, X.-S., WANG, R.-S., WANG, Y., WANG, J., QIU, Y., WANG, L. et CHEN, L. (2009). Modularity optimization in community detection of complex networks. *EPL (Europhysics Letters)*, 87(3):38002.
- ZHANG, Z.-K., LIU, C., ZHAN, X.-X., LU, X., ZHANG, C.-X. et ZHANG, Y.-C. (2016). Dynamics of information diffusion and its applications on complex networks. *Physics Reports*, 651:1–34.
- ZHAO, Y., LI, S. et JIN, F. (2016). Identification of influential nodes in social networks with community structure based on label propagation. *Neurocomputing*, 210:34–44.
- ZHAO, Z., WANG, X., ZHANG, W. et ZHU, Z. (2015). A community-based approach to identifying influential spreaders. *Entropy*, 17(4):2228–2252.
- ZUO, X.-N., EHMKE, R., MENNES, M., IMPERATI, D., CASTELLANOS, F. X., SPORNS, O. et MILHAM, M. P. (2012). Network centrality in the human functional connectome. *Cerebral cortex*, 22(8):1862–1875.

Résumé

L'immunisation de réseaux complexes vise à les fragmenter en petites parties avec un petit nombre de nœuds supprimés (immunisés). Ses applications incluent la prévention de la propagation des épidémies, la protection contre les attaques intentionnelles sur les réseaux, etc. Elle utilise les propriétés dynamiques de l'infection ainsi que les propriétés topologiques de la structure du réseau à travers les mesures de centralité. Ceci afin d'immuniser les nœuds les plus influents. Cependant, bien que la plupart des réseaux du monde réel aient une structure communautaire, peu de travaux ont été consacrés à l'étude de l'impact de cette propriété sur la propagation de l'épidémie. Dans cette thèse, nous nous concentrons sur le développement de nouvelles stratégies d'immunisation en utilisant plus de connaissances et de propriétés sur la structure de la communauté afin de mieux identifier et immuniser les nœuds influents. Dans un premier temps, nous visons à comprendre comment la structure communautaire affecte la dynamique épidémique. Ensuite, trois stratégies d'immunisation sont proposées en utilisant différents niveaux d'informations sur la structure de la communauté. Elles sont conçues pour des réseaux avec différentes forces de structure communautaire. Dans un deuxième temps, nous avons introduit un framework pour adapter toutes les mesures de centralité classiques proposées pour les réseaux sans structure communautaire aux réseaux modulaires. Dans ces types de réseaux, un nœud a deux types d'influences: une influence locale sur les nœuds appartenant à sa propre communauté à travers les liens intra-communautaires, et une influence globale sur les nœuds des autres communautés à travers les liens inter-communautaires. Ainsi, au lieu de représenter la centralité par une simple valeur scalaire, nous avons proposé de la représenter par un vecteur bidimensionnel, où chaque dimension représente un type d'influence différent que les nœuds peuvent exercer dans le réseau. Sa première composante est mesurée en calculant la centralité classique sur le réseau local, formé uniquement de liaisons intra-communautaires du réseau d'origine. Sa deuxième composante est quantifiée en calculant la centralité classique sur le réseau global, formé uniquement des liens inter-communautaires du réseau d'origine. Selon la force de la structure communautaire, ces deux composantes sont plus ou moins influentes. Dans notre troisième contribution, nous avons étendu ce framework aux réseaux avec des modules qui se chevauchent. Nous prenons en compte l'influence locale et globale des nœuds qui se chevauchent. Les performances des méthodes proposées ont été évaluées à l'aide de réseaux synthétiques et empiriques en utilisant le modèle épidémiologique Susceptible-Infecté-Rétabli.

Mots-clefs (6) : Réseaux complexes, stratégie d'immunisation, mesure de centralité, structure communautaire, modules qui se chevauchent, modèle Susceptible-Infecté-Rétabli.

Abstract

The immunization of complex networks aims at fragmenting them into small parts with a small number of removed (immunized) nodes. Its applications include prevention of epidemic spreading, protection against intentional attacks on networks, etc. It uses the dynamic properties of the infection as well as the topological properties of the network structure through the centrality measures in order to immunize the most influential nodes. However, although most real-world networks have a community structure, little work has been devoted to studying the impact of this property on the epidemic spread. In this thesis, we focus on developing new immunization strategies using more knowledge and features about the community structure in order to better identify and immunize the influential nodes. Initially, we aim to understand how community structure affects epidemic dynamics. Then, three immunization strategies are proposed using different levels of information about the community structure. They are designed for networks with different community structure strengths. In a second step, we have introduced a framework to adapt all the classic centrality measures proposed for networks without a community structure to modular networks. In these type of networks, a node has two types of influences: A local influence on the nodes belonging to its own community through the intra-community links, and a global influence on the nodes of the other communities through the inter-community links. Thus, instead of representing the centrality by a simple scalar value, we have proposed to represent it by a two-dimensional vector, where each dimension stands for a different type of influence that the nodes can exert in the network. Its first component is measured by calculating classical centrality on the local network. This network is formed only from intra-community links of the original network. In addition, its second component is quantified by calculating classical centrality on the global network. This network is formed solely from the inter-community links of the original network. Depending on the strength of the community structure, these two components are more or less influential. In our third contribution, we extended this framework to networks with overlapping modules. We take into account the local and global influence of overlapping nodes. The performances of the proposed methods have been evaluated using synthetic and empirical networks using the Susceptible-Infected-Recovered epidemiological model.

Key Words (6) : Complex networks, immunization strategy, centrality measure, community structure, overlapping modules, Susceptible-Infected-Recovered epidemiological model.

