

ANTOINE ALLARD

Modélisation mathématique en épidémiologie
par
réseaux de contacts
Introduction de l'hétérogénéité dans la transmissibilité

Mémoire présenté
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en Physique
pour l'obtention du grade de Maître ès Sciences (M. Sc.)

FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL
QUÉBEC

2008

©Antoine Allard, 2008

Résumé

La dernière décennie a vu se développer une approche novatrice et efficace pour modéliser la propagation de maladies infectieuses dans des populations : l'épidémiologie par réseaux de contacts. Utilisant des réseaux complexes pour tenir compte des interactions permettant la transmission de maladies infectieuses entre les individus, la percolation de liens de tels réseaux permet de prédire l'issue d'une épidémie et donc d'aider à tester et à améliorer les politiques de santé publique. Nous présentons un nouveau modèle de percolation de liens sur des réseaux complexes possédant une quelconque distribution de degrés et où l'hétérogénéité de la probabilité d'occupation des liens est explicitement intégrée. Puisque plusieurs maladies infectieuses possèdent une transmissibilité hétérogène, due aux différences physiologiques et comportementales entre les individus, ce modèle permettra des études beaucoup plus précises et détaillées et pourra participer à l'élaboration de politiques de santé publique mieux adaptées à la réalité.

Abstract

In the past decade, contact network epidemiology has revealed itself as a powerful approach to model the spread of infectious diseases in population. Using complex networks to capture the patterns of interactions between individuals that can lead to the transmission of an infectious disease, the bond percolation of such networks allows one to predict the outcome of outbreaks and therefore can help testing and improving public health strategies. We present a new bond percolation model of complex networks with an arbitrary degree distribution that explicitly incorporates heterogeneity in the bond occupation probability. Since several infectious diseases show heterogeneity in their transmissibility, mainly caused by intrinsic physiological and behavioural differences among individuals, this model will permit more realistic investigations and potentially lead to more precise intervention/prevention scenarios.

Avant-propos

Un seul nom figure sur la page couverture de cet ouvrage qui pourtant n'aurait jamais vu le jour sans l'apport de plusieurs personnes. Cette section me permet donc de mentionner et remercier ceux et celles qui ont contribué de près ou de loin (et parfois même à leur insu) à la réussite de ce projet. À votre façon, vous m'avez aidé à évoluer en tant que scientifique mais aussi en tant qu'individu (on espérera pour le mieux...). Grâce à vous, ces années furent riches en émotions, en expériences. Une source intarissable de souvenirs!

J'aimerais d'entrée de jeu remercier mon directeur Louis-J. Dubé qui m'a supporté et fait confiance depuis le tout début de notre collaboration. C'est un directeur qui a à coeur la formation de ses étudiants et qui ne se contente que du meilleur d'eux-mêmes. Il sait transmettre sa passion et donne le goût d'aller voir toujours plus loin. J'aimerais souligner son audace d'oser aborder des sujets de recherche peu orthodoxes, mais ô combien fascinants, qui me permet aujourd'hui de me positionner avantageusement dans un domaine en pleine ébullition.

Je voudrais aussi remercier le Dr. Babak Pourbohloul sans qui ce projet n'aurait pas vu le jour. Il m'a initié à un domaine de recherche peu classique et généralement méconnu des physiciens. Il m'a permis d'explorer la recherche scientifique dans un contexte différent que celui des universités, en plus de redécouvrir le magnifique Ouest Canadien.

Il serait impossible de passer sous le silence la contribution de Pierre-André Noël à ce travail. En plus de patiemment me montrer les rudiments de l'art numérique, il m'a éclairé sur les subtilités des fonctions génératrices et ses nombreux conseils ont permis à ce travail d'avoir l'aboutissement que j'osais à peine espérer.

J'aimerais aussi lever mon chapeau à Josette Lépine avec qui j'ai plongé dans cet univers parfois déroutant que sont les études supérieures. Ses bons mots m'ont plus souvent qu'autrement aidé à me convaincre que la lumière au bout du tunnel n'était

pas celle d'un train arrivant en sens opposé, surtout à 3 heures du matin avec un rapport à remettre le lendemain matin et un algorithme qui ne converge pas sur les bonnes valeurs !

J'aimerais aussi remercier messieurs Pierre Mathieu et Yvan Saint-Aubin qui ont bien voulu s'attaquer à un sujet qui leur était peu familier en premier lieu. J'ai en grande estime leur jugement scientifique et leurs conseils et impressions sur cet ouvrage seront considérés avec attention.

J'aimerais mentionner les autres membres du groupe Dynamica, Julien Poirer et Guillaume Painchaud-April (*le binôme du frigo*), qui ont fait partie de mon quotidien et avec qui, lors des célèbres cafés ou des mémorables soirées chez Habil, nous avons réglé les problèmes du monde plus d'une fois. Finalement, qui est-ce qui a brûlé la bibliothèque d'Alexandrie ?

J'aimerais aussi remercier mon ami Marc Olivier sans qui je ne serais probablement jamais devenu physicien. C'est toujours rassurant d'avoir un ami à ses côtés lorsqu'on plonge tête première dans un tout nouveau monde. Jérôme qui, malgré la distance, a pu continuer à suivre mes *highs* et moins *highs* et vice versa bien sûr ! J'aimerais aussi remercier la gang du band et leur univers rafraîchissant : Jean-Philippe avec ses questions du genre *Mais qu'advient-il du fait en mécanique quantique ?*, Oli et le tourment éternel de l'homme. . .

Mes parents qui m'ont toujours soutenu dans tous mes projets. Vous êtes la graine qui a permis cet ouvrage (et bien sûr tout le reste !) et vous serez toujours une source d'inspiration.

J'aimerais finalement remercier Catherine, mon tourment, pour les folies, les rêves, les caresses, les projets, les douceurs, les chicanes, les réconciliations, les voyages, les sourires. . . hein ?

À la vie, à l'ivresse. . .

*"There is no dark side of the moon,
really. . . As a matter of fact, it's all
dark." — Gerry O'Driscoll, Abbey
Road Studios' Irish Doorman
(1972-1973)*

Table des matières

Résumé	ii
Abstract	iii
Avant-propos	iv
Table des matières	vii
Liste des tableaux	x
Table des figures	xi
Glossaire	xii
Notation	xiii
Prologue	1
1 Concepts élémentaires	4
1.1 Introduction à l'épidémiologie	4
1.1.1 Terminologie	5
1.1.2 Et la modélisation mathématique...	5
1.1.3 Modèles compartimentaux	6
1.2 Introduction à la théorie des réseaux complexes	8
1.2.1 Propriétés des réseaux complexes	9
2 Épidémiologie par réseaux de contacts	14
2.1 Probabilité de transmission	15
2.1.1 L'hypothèse <i>iid</i>	16
2.2 Fonctions génératrices	16
2.3 Formalisme	18
2.3.1 Distribution des degrés occupés	19
2.3.2 Distribution des degrés occupés excédants	20
2.3.3 Distribution du nombre d'infectés	21

2.3.4	Seuil de percolation	23
2.3.5	Composante géante	23
2.4	Adaptation du formalisme aux réseaux semi-dirigés	24
2.4.1	Distribution de degrés occupés	25
2.4.2	Distribution de degrés occupés excédants	28
2.4.3	Distribution du nombre d'infectés	28
2.4.4	Seuil de percolation	29
2.4.5	Composante géante (probabilité et taille)	30
3	Vers une transmissibilité hétérogène (partie 1)	34
3.1	Échec de l'hypothèse <i>iid</i>	35
3.2	Hétérogénéité des noeuds	36
3.2.1	Nouveaux réseaux	36
3.2.2	Nouvelle Transmissibilité	38
3.3	Formalisme (en-dessous du seuil de percolation)	39
3.3.1	Distribution de degrés occupés	39
3.3.2	Distribution de degrés occupés excédants	40
3.3.3	Distribution du nombre d'infectés	41
3.3.4	Seuil de percolation	43
3.3.5	Composante géante	46
4	Vers une transmissibilité hétérogène (partie 2)	47
4.1	Formalisme semi-dirigé	47
4.1.1	Nouveaux réseaux	48
4.1.2	Distribution de degrés	50
4.1.3	Distribution de degrés excédants	50
4.1.4	Distribution des petites composantes	51
4.1.5	Composante géante	53
4.2	Formalisme (au-dessus du seuil de percolation)	54
4.2.1	Ajout de la transmissibilité	54
4.2.2	Composante géante (probabilité et taille)	56
4.2.3	Distribution du nombre d'infectés	57
4.2.4	Composition de la composante géante	59
4.3	Validation	60
4.3.1	Simulations numériques	60
4.3.2	La <i>taille</i> de l'infini	66
4.3.3	Cas particulier	67
5	Conclusions et perspectives	70

Bibliographie	74
A Synthèse du formalisme	79
B Implémentation numérique	82
B.1 Mise en puissances de polynômes par Transformées de Fourier	82
B.2 Simulations numériques de la propagation de maladies sur un réseau . .	85
C Formule intégrale de Cauchy	86
D Preuve de la conjecture	87

Liste des tableaux

1	Glossaire de termes utilisés en théorie des réseaux.	xii
A.1	Quantités de base du formalisme.	79
A.2	Synthèse des différentes quantités présentes dans notre formalisme. . .	81

Table des figures

1.1	Représentation schématique d'un réseau	10
2.1	Représentation schématique des réseaux considérés par le modèle de Newman.	19
2.2	Représentation schématique de la décomposition de $H(x; T)$ en différentes contributions.	22
2.3	Représentation schématique des réseaux considérés par le modèle de Meyers <i>et al.</i>	25
2.4	Représentation schématique des équations (2.29) et (2.30).	28
2.5	Structure simplifiée d'un réseau dirigé possédant une composante géante	30
2.6	Représentation schématique des équations (2.34) et (2.35).	32
3.1	Représentation schématique des réseaux considérés par notre formalisme	37
3.2	Représentation schématique de la décomposition de $H_{ij}(\mathbf{x}; \mathbf{T})$ en différentes contributions	42
4.1	Représentation schématique des réseaux semi-dirigés considérés par notre formalisme généralisé	49
4.2	Distribution de la probabilité jointe du nombre d'individus infectés	63
4.3	Distributions de la probabilité du nombre d'infectés d'un type donné	64
4.4	Comparaison du nombre moyen d'infectés prédit par (3.12) avec les résultats des simulations numériques pour différentes valeurs de γ	65
4.5	Comparaison des valeurs de \mathcal{P} , \mathcal{S} , \mathcal{S}_1 et \mathcal{S}_2 prédites par (4.25), (4.27) et (4.33) avec les résultats des simulations numériques pour différentes valeurs de γ	66
4.6	Comparaison des distributions de la probabilité du nombre d'infectés d'un type donné	67
4.7	Illustration des différences entre les prédictions pour \mathcal{P} et \mathcal{S} et les résultats numériques dues à la taille finie des réseaux.	68

Glossaire

Terme	Signification
Noeud	Unité fondamentale des réseaux complexes correspondant aux éléments en interaction.
Lien	Unité fondamentale des réseaux complexes indiquant une interaction entre les noeuds reliés.
Lien dirigé	Lien ne pouvant être suivi que dans un seul sens (unidirectionnel).
Lien non-dirigé	Lien pouvant être suivi dans les deux sens (bidirectionnel).
Degré	Nombre de liens attachés à un noeud.
Degré excédant	Nombre de liens par lesquels il est possible de quitter un noeud considérant que nous sommes arrivés en suivant un lien (et donc en excluant ce dernier).
Distribution de degrés	Distribution des degrés des noeuds dans un réseau.
<i>Hubs</i>	Noeuds possédant un degré significativement supérieur à la moyenne (aussi appelés <i>grands connecteurs</i>).
Composante	Groupe fermé de noeuds où chaque élément peut être rejoint à partir de n'importe quel autre élément du groupe en suivant des liens.
Petite composante	Composante dont la taille est indépendante de la taille du réseau (quantité intensive) et généralement petite par rapport à la taille du réseau ¹ (<i>i.e.</i> le nombre de noeuds).
Composante géante	Composante occupant une fraction importante (<i>i.e.</i> $\mathcal{O}(1)$) du réseau (quantité extensive). Il n'existe typiquement qu'une seule composante géante dans un réseau, lorsqu'il y en a une.
Réseau dirigé	Réseau dont tous les liens sont unidirectionnels.
Réseau non-dirigé	Réseau dont tous les liens sont bidirectionnels.
Réseau semi-dirigé	Réseau composé de liens dirigés et non-dirigés.
Réseau multipart	Réseau possédant un ou plusieurs types de noeuds et dont les liens ne peuvent exister qu'entre deux types différents. Un réseau possédant un seul type de noeuds est dit <i>uniparti</i> , un réseau possédant deux types de noeuds est dit <i>biparti</i> , etc.
Réseau multitype	Réseau possédant plusieurs types de noeuds dont les réseaux multipartis forment un sous-ensemble.

TAB. 1 – Glossaire de termes utilisés en théorie des réseaux.

¹Par *petite par rapport à la taille du réseau*, nous voulons dire *qui occupe une fraction très inférieure à 1 du réseau*. La distinction entre les petites composantes et la composante géante ne se fait pas typiquement par la taille relative de celles-ci (la frontière n'est pas clairement définie), mais via la propriété d'extensivité - ou inversement d'intensivité - des composantes. Voir la section 1.2.1 pour plus de détails.

Notation

Nous utilisons dans cet ouvrage une notation standard qu'un lecteur initié aux mathématiques avancées devrait être en mesure de bien comprendre. Toutefois, par souci de concision, nous avons utilisé une notation *vectorielle* pour expliciter certaines équations du formalisme présenté aux chapitres 3 et 4. Dans cette notation, la variable inscrite en **caractère gras** correspond à l'ensemble de ces variables avec le dernier indice variant de 1 à M (le nombre de types de noeuds). Voici quelques exemples² :

$$\begin{aligned} \mathbf{k} &: (k_1, k_2, \dots, k_M) & \mathbf{x} &: (x_1, x_2, \dots, x_M) \\ \boldsymbol{\delta}_i &: (\delta_{i1}, \delta_{i2}, \dots, \delta_{iM}) & \mathbf{H}_i(\mathbf{x}; \mathbf{T}) &: (H_{i1}(\mathbf{x}; \mathbf{T}), H_{i2}(\mathbf{x}; \mathbf{T}), \dots, H_{iM}(\mathbf{x}; \mathbf{T})). \end{aligned}$$

Lorsqu'une variable telle que celles-ci est utilisée comme argument d'une fonction, notre notation doit se lire ainsi³ :

$$\begin{aligned} P_i(\mathbf{k}) &= P_i(k_1, k_2, \dots, k_m) \\ G_i(\mathbf{x}; \mathbf{T}) &= G_i(x_1, x_2, \dots, x_m; \mathbf{T}) \\ F_{ij}(\mathbf{H}_i(\mathbf{x}; \mathbf{T}); \mathbf{T}) &= F_{ij}(H_{i1}(\mathbf{x}; \mathbf{T}), H_{i2}(\mathbf{x}; \mathbf{T}), \dots, H_{iM}(\mathbf{x}; \mathbf{T}); \mathbf{T}). \end{aligned}$$

Plusieurs objets du formalisme présenté dans ces mêmes chapitres auront un ou des indices laissés libres (voir ci-dessus). Chacun de ces indices prend des valeurs entières allant de 1 à M . Ainsi, il existe M objets $P_i(\mathbf{k})$ indiquant la distribution de degrés des noeuds de type i (*i.e.* un pour chaque type de noeuds).

Finalement, plusieurs quantités seront définies tout au long de cet ouvrage. Un résumé décrivant sommairement l'ensemble de ces quantités est donné à l'annexe A et le lecteur pourra s'y référer au besoin. Notons aussi que les formalismes de Newman [55] et de Meyers *et al.* [45] ont été réécrits dans notre notation afin de faciliter la lecture et la compréhension.

²La signification de ces symboles sera toujours expliquée à même le texte à chaque nouvelle utilisation.

³Pour la matrice de transmissibilité \mathbf{T} , sa présence en tant que paramètre d'une fonction signifie que certains de ses éléments sont utilisés dans la définition de cette fonction.

Prologue

Qu'ont Leonhard Euler (1707–1783) et Daniel Bernoulli (1700–1782) en commun ? Ils étaient suisses. Ils étaient contemporains. Ils étaient même amis. Tous les deux ont eu un intérêt marqué pour les mathématiques. Ils ont été élèves de Johann Bernoulli (1667–1748). Et bien d'autres choses encore. Toutefois, un autre lien les unissant reste généralement méconnu : ils ont tous les deux été pionniers de disciplines qui ne sont devenues populaires que dans la deuxième moitié du XX^e siècle et dont l'union aujourd'hui révolutionne la modélisation mathématique en épidémiologie.

En 1735, Leonhard Euler, un des plus prolifiques mathématiciens de tous les temps, résolut le problème des 7 ponts de Königsberg en étudiant les propriétés topologiques de structures abstraites appelées aujourd'hui graphes (ou réseaux). La ville de Königsberg possédait alors 7 ponts enjambant la rivière Pregel et reliant deux îles à la terre ferme. Le problème était fort simple : est-il possible de traverser chaque pont une seule fois sans revenir sur ses pas ? Euler réalisa que la solution résidait dans les connexions entre les différents points de terre. De ce point de vue, la configuration géographique de la ville était analogue à un ensemble de 4 points (les deux rives et les deux îles) reliés entre eux par 7 liens (les ponts). L'étude de la topologie de cet objet permettait alors de prouver que la réponse au problème était négative. De cette étude germa une théorie, aujourd'hui appelée *théorie des graphes* (ou des réseaux), qui fut popularisée par Erdős et Rényi au milieu du XX^e siècle et qui occupe un nombre toujours grandissant de chercheurs aujourd'hui.

De son côté, Daniel Bernoulli, grand mathématicien suisse, fut le tout premier en 1760 à utiliser un modèle mathématique pour tirer des conclusions épidémiologiques. À cette époque, la variole était endémique en Angleterre et était la cause d'un décès sur dix. Le seul moyen de prévention disponible, appelé *variolisation*, consistait à causer une infection mineure en mettant des individus sains en contact avec des tissus infectés. Cette infection mineure vaincue, les individus traités se voyaient conférer une immunité permanente face à la variole. Toutefois, la variolisation pouvait entraîner la mort et d'intenses débats faisaient rage à savoir si les gains au niveau de la popula-

tion valaient les risques au niveau individuel. À l'aide d'un modèle mathématique et de données réalistes, Daniel Bernoulli démontra que le déploiement à grande échelle de la variolisation permettait de sauver un plus grand nombre de vies que le nombre de décès potentiels que cela pouvait causer. La variolisation fut éventuellement utilisée à grande échelle, ce qui contribua à éliminer l'endémie de variole en Angleterre. Malgré cet impressionnant triomphe, la modélisation mathématique resta marginale en épidémiologie jusqu'à la fin du XX^e siècle où l'informatique lui donna un nouveau souffle.

Au milieu du XX^e siècle, l'amélioration des conditions sanitaires et le développement de moyens de prévention ou de traitement toujours plus efficaces (*e.g.* antibiotiques, anti-viraux, vaccins) firent croire à plusieurs que les maladies infectieuses seraient bientôt éliminées. Il n'en fut malheureusement pas ainsi. Les trente dernières années ont été témoins de l'émergence de nouvelles maladies infectieuses au rythme effarant de plus d'une par an, un taux jamais vu auparavant. Il y a aujourd'hui près de 40 maladies infectieuses qui étaient inconnues de la génération précédente. De plus, l'explosion démographique et du trafic aérien, le commerce à l'échelle mondiale et la démocratisation du tourisme font aujourd'hui en sorte que ces maladies se propagent géographiquement à une vitesse inégalée dans l'histoire. Avec environ 2.1 milliards de passagers aériens en 2006, une épidémie, aussi petite soit-elle, faisant rage n'importe où dans le monde⁴ est à quelques heures d'être une menace imminente pour n'importe quel autre endroit sur la planète. Les changements climatiques influencent aussi la situation épidémiologique mondiale puisqu'ils provoquent la migration de maladies vers des régions auparavant épargnées. L'ensemble de ces facteurs fait en sorte qu'encore aujourd'hui, les maladies infectieuses sont la cause d'environ 25% des décès dans le monde.

Face à une telle menace, les sociétés doivent se préparer à l'avance pour réagir promptement et efficacement si une telle épidémie est déclarée. Bien qu'il puisse parfois exister des moyens pratiques permettant de combattre l'évolution d'une épidémie (*e.g.* antibiotiques, anti-viraux, vaccins), il n'est pas toujours clair comment ceux-ci devraient être déployés afin de minimiser les coûts humains et économiques, particulièrement en tenant compte de diverses contraintes (*e.g.* quantité limitée de matériel ou d'effectifs). En simulant la propagation de maladies infectieuses dans des populations, la modélisation mathématique permet une évaluation systématique de différents scénarios de prévention et d'intervention avant même qu'une épidémie ne soit déclarée. Ce nouvel outil permet ainsi aux sociétés de se préparer à réagir et est de plus en plus utilisé comme outil supplémentaire dans les processus de prise de décisions en matière de santé publique.

⁴L'Organisation Mondiale de la Santé (OMS) a recensé plus de 1100 épidémies dans le monde entre 2001 et 2006.

Depuis la contribution de Daniel Bernoulli, différentes approches ont été utilisées pour développer des modèles mathématiques en épidémiologie. Chacune possédant ses avantages et ses inconvénients, le choix de la technique à utiliser se fait généralement en fonction de la situation à modéliser. L'essor de la théorie des réseaux (*i.e.* l'utilisation de graphes comme modèles de systèmes réels) à la fin du XX^e siècle a permis le développement d'une toute nouvelle approche analytique : l'*épidémiologie par réseaux de contacts*. L'avantage premier de cette méthode est qu'elle considère explicitement la structure sociale⁵, appelée réseau de contacts, en la modélisant à l'aide d'un réseau complexe. En analysant la topologie de cette structure, il est possible de prévoir et de quantifier l'issue finale d'une épidémie et ainsi, en principe, de tester à l'avance différentes stratégies de prévention ou d'intervention. Toutefois, étant encore à ses premiers balbutiements, cette approche modélise les réseaux de contacts avec des réseaux complexes ironiquement encore trop simplistes. Il est donc nécessaire de poursuivre les efforts faits jusqu'à présent de sorte que les futurs modèles rendent compte de toute la complexité des véritables réseaux de contacts.

Il s'agit d'ailleurs là de la motivation du projet dont les résultats sont présentés dans ce mémoire. En effet, les modèles actuels en épidémiologie par réseaux de contacts permettent à la probabilité de transmission de la maladie de varier d'une paire d'individus à l'autre, sous la condition qu'aucune corrélation n'existe entre les différentes valeurs. Or, pour certaines maladies, cette hypothèse n'est pas réaliste puisque les caractères génétiques ou comportementaux des individus ont une grande influence sur la probabilité de transmission (*e.g.* influenza, M.T.S.). En s'inspirant des formalismes existants, nous avons donc développé un nouveau modèle considérant des réseaux *multi-types* permettant de tenir compte explicitement de ces corrélations. Dorénavant, il sera donc possible d'effectuer des simulations beaucoup plus réalistes et ainsi permettre des analyses de stratégies d'intervention/prévention beaucoup plus précises et détaillées.

La première partie du chapitre 1 introduit une série de concepts élémentaires en épidémiologie et présente un modèle mathématique fréquemment utilisé par les épidémiologistes. La deuxième partie de ce même chapitre introduit la théorie des réseaux et présente quelques propriétés des réseaux réels que les modèles cherchent à reproduire. Le chapitre 2 quant à lui présente deux modèles pionniers de l'épidémiologie par réseaux de contacts qui ont inspiré le formalisme que nous avons développé. Nous les avons adaptés quelque peu de façon à les rendre plus compréhensibles et pour faciliter la transition à notre formalisme. Ainsi, la notation diffère quelque peu de celle utilisée par les auteurs. Les chapitres 3 et 4 se consacrent à la présentation de notre modèle et comparent avantageusement ses prédictions aux résultats de simulations numériques.

⁵Cette structure joue un rôle primordial dans la dynamique de la propagation des maladies infectieuses en étant le support par lequel les maladies se propagent d'un individu à l'autre.

Chapitre 1

Concepts élémentaires

L'épidémiologie par réseaux de contacts est à l'interface de deux disciplines qui ont connu un essor considérable au cours de la deuxième moitié du XX^e siècle : la modélisation mathématique en épidémiologie et la théorie des réseaux. Puisque l'une et l'autre font appel à des notions avec lesquelles les physiciens ne sont généralement pas familiers, nous allons tout d'abord présenter certains concepts fondamentaux en épidémiologie, plus particulièrement en modélisation mathématique, à la §1.1 et en théorie des réseaux à la §1.2. Notre objectif ici, n'est pas de faire une revue exhaustive des sujets, mais d'introduire des notions qui seront utiles à la compréhension des prochains chapitres et qui permettront de mettre en contexte notre travail.

1.1 Introduction à l'épidémiologie

L'épidémiologie consiste en *l'étude des rapports existant entre les maladies ou tout autre phénomène biologique, et divers facteurs (mode de vie, milieu ambiant ou social, particularités individuelles) susceptibles d'exercer une influence sur leur fréquence, leur distribution, leur évolution*¹. Autrement dit, l'épidémiologie est une discipline qui s'intéresse aux facteurs influençant la santé des populations et qui cherche à développer des moyens d'atténuer leur impact lorsque celui-ci est jugé négatif.

Dans le cas des maladies infectieuses, l'épidémiologie cherche à identifier les agents pathogènes et à comprendre leur mode de propagation. Ainsi, les épidémiologistes cherchent à répondre à des questions telles que : *quelles sont les populations à risque ?*

¹Tiré de : Le Petit Robert 2001

par quels moyens et dans quelles circonstances la maladie est-elle transmise ? quels sont les facteurs favorisant cette transmission ? Ces informations permettent une meilleure compréhension de la dynamique de la propagation, ce qui en retour permet d’orienter les moyens de prévention et d’intervention dans le meilleur intérêt de la santé publique.

1.1.1 Terminologie

Nous définissons ici quelques termes qui seront fréquemment utilisés au cours de cet ouvrage.

Maladie infectieuse : maladie provoquée par la transmission d’un agent pathogène : virus, bactérie, parasite, mycose, prion, etc.

Épidémie : augmentation rapide de l’incidence d’une pathologie. Bien que souvent utilisé dans un contexte de maladies infectieuses, ce terme peut être utilisé pour des phénomènes biologiques généraux (obésité, fracture de la hanche chez les personnes âgées, suicide, etc.)².

Seuil épidémique : seuil théorique présent dans les modèles mathématiques au-dessus duquel il y aura (ou pourra y avoir) une épidémie.

Patient zéro : premier cas reconnu d’une pathologie infectieuse qui est à l’origine de tous les autres cas recensés.

Pandémie : épidémie causée par une maladie infectieuse émergente qui prend des proportions continentales voire planétaires.

Endémie : présence habituelle et stable d’une maladie dans une population.

1.1.2 Et la modélisation mathématique...

Tel que dit précédemment, l’épidémiologie cherche, entre autre, à comprendre la dynamique régissant la propagation de maladies infectieuses afin d’établir des stratégies de prévention et d’intervention permettant de diminuer leur impact sur la santé publique. Toutefois, puisque l’introduction volontaire de maladies dans des populations ou la retenue de moyens d’interventions pour les fins d’études scientifiques seraient éthiquement problématiques, ces études sont limitées à des analyses statistiques des données recueillies lors d’épisodes épidémiques antérieurs. Ainsi, lorsque vient le temps de se préparer à une pandémie (*e.g.* la grippe aviaire) ou de planifier l’utilisation d’un

²Bien que la définition présentée pour le terme épidémie ne fasse pas allusion à la taille de celle-ci, nous réserverons son emploi pour des situations où une fraction importante de la population est infectée (*i.e.* où l’infection ne se limite pas à quelques individus).

tout nouveau vaccin (*e.g.* le virus du papillome humain), les études épidémiologiques classiques sont dans l'impossibilité de répondre clairement aux questions soulevées par ce genre de problématiques et se voient dans l'obligation d'extrapoler à partir des connaissances sur des maladies similaires.

Quoique marginale depuis sa première utilisation au XVIII^e siècle par Daniel Bernoulli [17, 27], la modélisation mathématique s'est imposée à la fin du XX^e comme un outil supplémentaire dans le processus de prise de décisions en matière de santé publique. En effet, en permettant la simulation de scénarios épidémiologiques avant même qu'une épidémie ne se produise, elle permet d'évaluer les risques associés à une telle épidémie en plus de quantifier l'impact et l'efficacité de différentes méthodes de prévention et d'intervention. Il existe une multitude d'approches pour modéliser la propagation de maladies. Celles-ci vont de simulations numériques massives [31, 32, 60], à des modèles déterministes [9, 25, 34], en passant par des modèles stochastiques [1, 11, 43, 50]. Chacune de ces approches possédant ses avantages et ses inconvénients, le choix du modèle à préconiser se fait généralement en fonction des questions auxquelles il faut répondre. Nous présentons à la prochaine section une approche déterministe très utilisée par les épidémiologistes et qui, pendant longtemps, a fait consensus dans la communauté.

1.1.3 Modèles compartimentaux

Les modèles compartimentaux sont des modèles déterministes où la population est divisée en un nombre de catégories selon différentes caractéristiques (âge, sexe, caractère génétique particulier) et selon l'état par rapport à la maladie (susceptible à se faire infecter, infecté non-contagieux, infecté contagieux, immunisé, décédé, etc.). Le changement d'état des individus, *i.e.* le changement du nombre d'individus dans chacune de ces *boîtes*, est gouverné par un ensemble d'équations différentielles. Ces dernières sont établies en supposant que le nombre d'individus nouvellement infectés dans un intervalle de temps donné est proportionnel au produit du nombre d'individus contagieux avec le nombre d'individus sains, tous deux dans l'intervalle de temps précédent³. Cette hypothèse implique donc que les contacts entre les individus contagieux et les individus sains se font aléatoirement dans le temps — selon un processus de Poisson — mais aussi au sein de la population puisqu'un individu contagieux pourra infecter n'importe quel individu sain avec la même probabilité (hypothèse *fully-mixed*). De par leur simplicité conceptuelle, les modèles compartimentaux peuvent être aisément adaptés à plusieurs situations épidémiologiques en faisant varier le nombre de catégories dans lesquelles la

³Cette hypothèse est analogue à — et inspirée de — l'hypothèse masse-action utilisée en chimie qui suppose que la vitesse d'une réaction chimique est proportionnelle au produit des concentrations des réactifs.

population est divisée. À titre d'exemple, nous allons expliciter le désormais classique modèle SIR.

Modèle SIR

Le modèle SIR divise la population en trois catégories : les individus susceptibles de se faire infecter, les individus infectés et contagieux et les individus ne pouvant plus transmettre la maladie (guérison, immunité, décès). Nous noterons $s(t)$, $i(t)$ et $r(t)$ la fraction de la population faisant partie de chacune de ces trois catégories. Considérant μ et ν comme étant respectivement le taux avec lequel les individus ont des contacts pouvant transmettre la maladie⁴ et auquel ils cessent d'être contagieux, l'hypothèse masse-action nous permet d'écrire

$$\frac{ds}{dt} = -\mu is \quad (1.1a)$$

$$\frac{di}{dt} = \mu is - \nu i \quad (1.1b)$$

$$\frac{dr}{dt} = \nu i. \quad (1.1c)$$

avec les conditions initiales $s(0) = s_0$, $i(0) = i_0$ et $r(0) = r_0$. Toutefois, puisque $s(t) + i(t) + r(t) = 1$ en tout temps, (1.1c) ne nous est d'aucune utilité et seules (1.1a) et (1.1b) sont nécessaires pour décrire la dynamique du système. Ce modèle est généralement utilisé pour décrire des maladies à propagation rapide⁵ (*e.g.* influenza) auxquelles les malades développent éventuellement une immunité.

En divisant (1.1b) par νi , nous obtenons le nombre d'individus qui seront infectés directement par individu malade pendant la période où ce dernier est contagieux

$$\frac{1}{i} \frac{di}{d(\nu t)} = \frac{\mu s}{\nu} - 1. \quad (1.2)$$

Ainsi, nous voyons que pour $\frac{\mu s(t)}{\nu} > 1$, chaque individu contagieux contaminera en moyenne plus d'un individu susceptible et la maladie se propagera à un nombre toujours grandissant d'individus. Il en sera ainsi jusqu'à ce que le nombre d'individus susceptibles, $s(t)$, soit tel que $\frac{\mu s(t)}{\nu} < 1$ et que la propagation s'arrête éventuellement par elle-même. Le rapport μ/ν s'interprète comme étant le nombre de contacts pouvant transmettre la maladie qu'ont les individus infectés tout au long de leur période de contagion. Ainsi, en multipliant par la fraction des individus susceptibles à un temps

⁴Il y aura transmission si l'individu avec lequel il y a eu contact est susceptible.

⁵L'évolution démographique naturelle de la population n'est pas représentée (*i.e.* naissances, décès).

donné, nous obtenons le nombre de nouveaux infectés causés par un seul individu contagieux. L'équation (1.2) permet donc de mettre en évidence l'importance des conditions initiales dans le modèle SIR puisque pour $\frac{\mu s_0}{\nu} > 1$, il y aura assurément une épidémie tandis que pour $\frac{\mu s_0}{\nu} < 1$, seuls quelques individus seront infectés avant que la propagation ne s'éteigne par elle-même. Le seuil épidémiologique est donc défini par $\frac{\mu s_0}{\nu} = 1$.

Nous n'avons explicité ici que les grands résultats du modèle SIR, à savoir la présence d'un seuil épidémique. Il existe toutefois quelques théorèmes décrivant plus en détails l'évolution de la dynamique du modèle. Le lecteur intéressé est invité à se référer à [34] pour plus de détails. De plus, la modélisation mathématique est un domaine très vaste et nous n'avons offert qu'une description sommaire d'une seule, quoiqu'importante, approche. On pourra se référer à [9, 25, 34, 38, 43] pour plus d'information.

1.2 Introduction à la théorie des réseaux complexes

La théorie des réseaux (théorie des graphes) est présente en mathématiques depuis Euler qui, en 1736, l'utilisa pour résoudre le problème des 7 ponts de Königsberg [16]. Il fallut toutefois attendre les contributions de Erdős et Rényi [29] au courant du XX^e siècle pour que cette théorie prenne vraiment son envol. Ce n'est que tout récemment que les scientifiques (plus particulièrement les physiciens) se sont intéressés aux réseaux complexes⁶ en tant que modèles de systèmes réels. En effet, l'interaction entre les éléments d'une gamme de systèmes biologiques (*e.g.* chaîne alimentaire d'écosystèmes, interactions entre les protéines [35], métabolismes [36]), sociologiques (*e.g.* réseaux de contacts entre les individus [40] et ceux formés par les citations scientifiques [64, 65]) et technologiques (*e.g.* Internet [23, 30, 69], WWW [6, 15, 21], réseaux de distribution électrique [8] ou téléphonique [3, 4]) se laisse aisément décrire par des réseaux complexes. Les études empiriques faites sur ces derniers ont mis en évidence plusieurs propriétés permettant ainsi des descriptions topologiques plus précises. Ces connaissances ont permis à leur tour d'expliquer divers phénomènes associés aux réseaux tels que la résilience (et inversement la défaillance en cascade) et l'efficacité du transfert d'information en plus de permettre le développement de modèles mathématiques beaucoup plus réalistes.

Suite à ces développements, les physiciens se sont intéressés aux processus *dynamiques* ayant cours sur ces structures. Notons au passage la navigation sur les réseaux (*e.g.* moteurs de recherche comme *Google* [20]), l'*attachement préférentiel* menant aux

⁶Nous entendons ici les réseaux complexes comme étant l'ensemble général des réseaux dont un sous-ensemble correspond aux grilles régulières utilisées en physique de l'état solide et un autre sous-ensemble correspond aux réseaux parfaitement aléatoires de Erdős et Rényi.

distributions de degrés en loi de puissance [14] et la percolation de sites et/ou de liens [50, 55]. Ce dernier processus a été utilisé dans une foule d'applications passant de la mesure de la résilience des réseaux à la propagation de rumeurs (via un modèle analogue à celui d'Ising [72]), de virus informatiques ou d'agents pathogènes dans une population [13, 45, 46, 50, 58, 59, 61].

1.2.1 Propriétés des réseaux complexes

Un réseau complexe est un ensemble de noeuds et de liens reliant ces derniers entre eux. Différents types de réseaux sont définis selon la nature des noeuds et des liens [26]. Par exemple, certains types de réseaux attribuent des propriétés aux noeuds pour les différencier les uns des autres. Pour ce qui est des liens les joignant, ceux-ci peuvent être unidirectionnels, bidirectionnels, uniques, multiples, avoir un poids relatif, etc. Il existe aussi des réseaux où les noeuds sont séparés en 2, 3, ... plusieurs groupes (*biparti*, *triparti*, ... *multiparti*) et où les liens ne peuvent joindre que deux noeuds de type différent. Le choix du type de réseau à utiliser est habituellement fait en fonction du système à modéliser. Ainsi, il est commode de modéliser le WWW ou une chaîne alimentaire à l'aide d'un réseau dont les liens sont dirigés alors qu'un réseau biparti modélise relativement bien les contacts sexuels d'une société hétérosexuelle. La figure 1.1 donne un exemple de réseau complexe où les noeuds sont classés en 4 catégories et où les liens sont unidirectionnels ou bidirectionnels. Toutefois, pour des raisons de solvabilité mathématique, la plupart des modèles *dynamiques*, notamment ceux décrivant la percolation, ne considèrent que des réseaux simplifiés où les noeuds sont indiscernables et où les liens peuvent être unidirectionnels et/ou bidirectionnels. Nous n'avons donné jusqu'à présent que des propriétés qualitatives des réseaux complexes. Bien que nous ayons une idée de l'allure de ces réseaux, rien n'a encore été dit sur leurs structures (*e.g.* communautés, petites composantes, composante géante, etc.) ou sur la façon dont les noeuds sont reliés entre eux. Les prochaines sections se consacrent donc à ces quelques propriétés quantitatives.

Distribution des degrés

Le degré d'un noeud est le nombre de liens qu'il a avec d'autres noeuds du réseau. On appelle *voisins* (ou premiers voisins) deux noeuds qui sont directement reliés par un lien⁷, deuxièmes voisins deux noeuds qui peuvent être minimallement joints par

⁷Il existe des réseaux où plusieurs liens peuvent exister entre une même paire de noeuds. On nomme ces réseaux *multigraphes*.

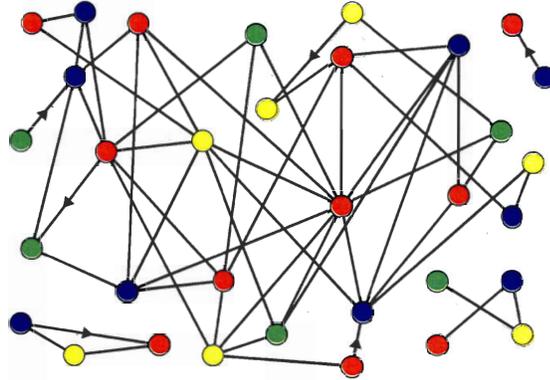


FIG. 1.1 – Représentation schématique d'un réseau composé de 33 noeuds étiquetés en 4 catégories (types). À l'exception des 6 liens unidirectionnels représentés à l'aide d'une flèche, tous les liens du réseau peuvent être parcourus dans l'une ou l'autre des deux directions.

l'intermédiaire d'un autre noeud, troisièmes voisins deux noeuds qui peuvent être minimalement rejoints en passant par deux noeuds intermédiaires, et ainsi de suite. Comme son nom l'indique, la distribution des degrés, typiquement notée $P(k)$ ou p_k , est la distribution du nombre de liens qu'ont les noeuds dans un réseau ou de façon équivalente la probabilité qu'un noeud choisi au hasard soit de degré k . On considère souvent la distribution des degrés comme définissant un ensemble où tous les réseaux respectent cette distribution mais sont à tout autre égard aléatoires. Notons au passage quelques exemples de distributions de degrés.

1) **Régulière** : Les réseaux réguliers sont ceux avec lesquels les physiciens sont les plus familiers. Dans ces réseaux, tous les noeuds ont le même degré d , ce qui mène à une distribution de la forme

$$p_k = \delta_{kd}.$$

Les grilles régulières souvent utilisées en physique statistique tombent dans cette catégorie. Lorsque $d = N - 1$ où N est le nombre de noeuds, on dit que le réseau est *complet*.

2) **Aléatoire** : Les réseaux aléatoires tels que ceux étudiés par Erdős et Rényi sont typiquement construits à partir de N noeuds où un lien existe entre chacune des $\binom{N}{2}$ paires de noeuds possibles avec une probabilité p . Ainsi, leur distribution de degrés est donnée par une distribution binomiale

$$p_k = \binom{N}{k} p^k (1-p)^{N-k} \simeq \frac{(Np)^k e^{-Np}}{k!}$$

qui peut être approximée par une distribution de Poisson (dernière égalité) dans la limite $N \rightarrow \infty$. Bien qu'étant fort intéressants à étudier, les réseaux aléatoires (*i.e.* qui ont une distribution de degrés aléatoire) s'avèrent être de piètres modèles pour la plupart des réseaux réels [14].

3) Exponentielle : Certains réseaux réels tels celui montrant les relations entre les actionnaires siégeant sur les conseils d'administration des 1000 entreprises étatsuniennes les plus riches [55] ont une distribution de degrés suivant approximativement une exponentielle

$$p_k = (1 - e^{-1/\kappa})e^{-k/\kappa} \quad (1.3)$$

où κ est un paramètre libre du système. Certains auteurs ont récemment utilisé une fonction exponentielle comme (1.3) pour décrire la distribution de degrés de plusieurs réseaux sociaux [12]. Toutefois nous croyons qu'il s'agit d'une approximation très grossière et qu'il ne faudrait pas généraliser cette idée.

4) Loi de Puissance : Beaucoup d'attention a été portée depuis quelques années aux réseaux semblant posséder une distribution de degrés suivant une loi de puissance [5, 28, 67]. En effet, une multitude d'études empiriques sur une foule de réseaux réels différents affirment que leur distribution de degrés suit une *loi de puissance pure* [3, 4, 6, 15, 21, 23, 30, 35, 36, 40, 64, 65, 69]

$$p_k \propto k^{-\alpha}$$

ou une *loi de puissance avec coupure exponentielle* [8]

$$p_k = \frac{k^{-\alpha} e^{-k/\kappa}}{\text{Li}_\alpha(e^{-1/\kappa})} \text{ for } k \geq 1 \quad (1.4)$$

où $\text{Li}_n(x)$ est le $n^{\text{ième}}$ polylogarithme de x [2]. Il existe une polémique au sein de la communauté scientifique à savoir si ces distributions de degrés hautement asymétriques sont bel et bien des lois de puissances ou s'il ne s'agit pas plutôt d'un autre type de distribution (*e.g.* une loi *log-normale*) [24, 54]. La distribution des degrés d'un réseau s'obtient aisément (en principe) en traçant un histogramme des degrés des noeuds du réseau. Cependant, dans le cas des lois de puissance, il faut porter ce dernier en échelle logarithmique et rares sont les données empiriques qui évitent d'avoir un bruit important dans la queue de la distribution, élevant ainsi l'identification d'une loi de puissance pure au rang d'acte de foi.

Toutefois, indépendamment de l'issue de cette confrontation, il est important de garder en tête que les distributions de degrés de la plupart des réseaux réels sont hautement asymétriques et possèdent une *queue épaisse*⁸. Ceci a pour conséquence que,

⁸Traduction libre de *heavy tail* ou *fat tail*.

bien que la plupart des noeuds du réseau aient un faible degré, il existe une probabilité non-négligeable qu'il existe des noeuds possédant un degré très élevé. La présence de ces *grands connecteurs (hubs)* a d'importantes répercussions sur la résilience des réseaux au retrait aléatoire des noeuds [7] ou sur la propagation de l'information ou d'une maladie [19, 50]. De plus, quelque soit la forme exacte de ces distributions, l'asymétrie de (1.4) fait que cette distribution est un excellent test pour les modèles et est, de ce fait, de plus en plus utilisée dans les publications scientifiques.

Effet *Small-World*

L'effet *Small-World* [70] pourrait bien s'expliquer par l'adage populaire québécois : *hé que le monde est p'tit!* Imaginé par l'écrivain hongrois Frigyes Karinthy en 1929 [37] puis expérimenté plus rigoureusement par Stanley Milgram dans les années 1960 [47], l'effet *Small-World* consiste au fait que la distance entre les noeuds d'un réseau social est relativement courte par rapport aux dimensions de ce dernier. Ainsi, il est possible à partir d'un noeud donné de rejoindre n'importe quel autre noeud du réseau en un petit nombre de déplacements. Plus formellement, l'effet *Small-World* correspond au fait que la *distance géodésique*, *i.e.* le parcours moyen le plus court pour passer d'un noeud à l'autre, dans un réseau augmente au maximum comme $\mathcal{O}(\log N)$ pour un degré moyen⁹ donné où N est le nombre de noeuds dans le réseau. Bien que cet effet fut mis en évidence dans le contexte des réseaux sociaux, celui-ci est présent dans une foule d'autres catégories de réseaux [70].

Effet de regroupement

L'effet de regroupement (*clustering*) est une mesure de la probabilité que deux noeuds ayant un voisin commun soient eux-mêmes voisins. Encore une fois, ceci peut se traduire par l'adage : *les amis de mes amis sont mes amis*. Il existe plusieurs moyens de quantifier cette tendance dans un réseau. Une définition généralement utilisée consiste à calculer

$$C = \frac{\beta \times [\text{Nombre de triangles dans le réseau}]}{[\text{Nombre de parcours de longueur 2 dans le réseau}]}$$

où un *triangle* est défini comme un groupe de trois noeuds joints les uns aux autres par trois liens. Le coefficient β prend la valeur 3 si les parcours de longueur 2 sont calculés en ne tenant pas compte du sens qu'ils sont parcourus et 6 dans le cas contraire. La

⁹Degré moyen : moyenne de la distribution des degrés.

présence de β assure que $0 \leq C \leq 1$. Ainsi, plus le coefficient C d'un réseau est près de 1, plus les noeuds auront tendance à s'agglomérer. Ceci aura une influence notamment sur la rapidité de propagation d'une épidémie [52].

Composantes

Il n'existe pas toujours une connexion (directe ou indirecte) entre deux noeuds dans un réseau. On dit alors que le réseau n'est plus complètement connecté; qu'il est alors scindé en différentes composantes (voir Glossaire). On distingue les composantes en deux catégories : les *petites composantes* dont la taille ne dépend pas de la taille du réseau (intensive) et la *composante géante* dont la taille est de l'ordre de celle du réseau et varie en fonction de celle-ci (extensive). L'étude de la percolation d'un réseau consiste entre autre à l'étude de l'apparition de la composante géante qui, dans le cas d'un réseau de taille infinie, apparaît brusquement à une valeur critique des paramètres de celui-ci. La figure 1.1 montre un réseau scindé en 3 petites composantes et une que l'on pourrait qualifier de composante géante¹⁰.

Autres propriétés

Il existe une multitude de propriétés des réseaux complexes autres que celles introduites ci-haut. Citons notamment les structures en communautés, les corrélations entre les degrés et le mélange entre les types de noeuds. Pour plus de détails, le lecteur est invité à consulter [5, 18, 28, 33, 49, 51, 53].

Ayant maintenant donné une description de la modélisation mathématique en épidémiologie et de la théorie des réseaux, nous sommes en mesure de présenter comment ces deux disciplines se sont scindées au tournant du dernier millénaire pour former une toute nouvelle approche de modélisation en épidémiologie.

¹⁰Il n'est pas possible de déterminer si cette composante est extensive ou intensive seulement à l'aide de cette figure. Toutefois, considérons-la comme géante pour illustrer notre propos.

Chapitre 2

Épidémiologie par réseaux de contacts

Les modèles compartimentaux présentés au chapitre précédent sont construits en supposant que chaque individu contagieux peut infecter chacun des individus susceptibles d'une population donnée avec une probabilité égale (hypothèse *fully-mixed*). Or, l'expérience de tous les jours indique que cette hypothèse n'est pas réaliste pour une large gamme de maladies infectieuses. En effet, la transmission d'un agent pathogène se fait typiquement entre deux individus dont le comportement (*e.g.* via la cohabitation, le lieu de travail, le transport en commun) facilite des contacts permettant justement cette transmission (*e.g.* baisers, relations sexuelles, éternuements, poignées de main, poignées de porte). Ainsi, certaines paires d'individus sont plus susceptibles d'être vecteur de la propagation que d'autres. Ceci implique donc que seule une fraction de la population pourra potentiellement être directement infectée par un individu contagieux, allant à l'encontre de l'hypothèse *fully-mixed*.

Afin d'offrir une modélisation réaliste de la propagation de maladies et ainsi de permettre l'élaboration de stratégies d'intervention ou de prévention efficaces, la nouvelle génération de modèles épidémiologiques doit considérer la structure sociale par laquelle les maladies infectieuses se propagent. Bien que pouvant sembler aléatoire sur plusieurs aspects, le comportement humain reste répétitif et généralement prévisible (*e.g.* les gens vont à l'école ou au travail à chaque jour, prennent le transport en commun, vont faire les courses). Ceci permet, via l'utilisation de données sociologiques (*e.g.* sondages, recensements, simulations), d'extraire une foule d'information et ainsi de reconstruire¹

¹Il va sans dire que la reconstruction de la structure sociale sera d'autant plus réaliste que les données sociologiques seront précises et détaillées. Ceci est d'ailleurs un des grands défis du modélisateur épidémiologiste puisque cette quête est généralement *périlleuse* et l'issue incertaine.

cette structure que l'on nomme *réseau de contacts*.

La théorie des réseaux complexes permet une description mathématique conceptuellement simple du réseau de contacts où chaque individu est représenté par un noeud et où un lien unissant une paire de noeuds indique une prédisposition à la transmission si l'un ou l'autre est contagieux. Par exemple, les liens d'un réseau simulant la propagation d'une maladie comme le VIH/SIDA correspondraient aux relations sexuelles non-protégées ou aux échanges de seringue lors de l'utilisation de drogues injectables. En permettant à la maladie de se propager d'un individu à un autre en suivant les liens du réseau de contacts avec une probabilité donnée que l'on nomme *transmissibilité*, l'état final de la propagation de la maladie est isomorphe² à la percolation par liens de ce même réseau où l'occupation des liens est déterminée avec cette même probabilité. Ainsi, en développant des modèles de percolation par liens toujours plus complexes, ceux-ci deviendront de plus en plus représentatifs de la situation épidémiologique réelle et donc utiles à la prise de décision en matière de santé publique.

Un des premiers modèles épidémiologiques basés sur un formalisme issu de la théorie des réseaux [55] a été proposé par Mark Newman en 2002 [50]. Ce modèle s'est depuis imposé comme la référence dans le domaine et est le fondement de la plupart des nouveaux développements [41, 45, 48, 57]. Ce projet de maîtrise n'échappant pas à cette tendance, ce chapitre se consacrera à une présentation détaillée de ce modèle. Ainsi, la §2.1 explique quelques considérations en ce qui a trait à la transmissibilité tandis que la §2.2 introduit quelques notions fondamentales sur les fonctions génératrices qui sont à la base du formalisme de Newman. Les §2.3 et §2.4 quant à elles traitent respectivement du formalisme introduit par Newman et d'une extension aux réseaux semi-dirigés proposée par Meyers *et al.* [45].

2.1 Probabilité de transmission

Considérons une paire d'individus reliés par un lien : le premier, qu'on notera i , étant *infectieux* et le second, qu'on notera j , étant *susceptible* de se faire infecter. Supposons que la probabilité que i infecte j pendant un intervalle de temps d'une durée infinitésimale est $r_{ij}dt$ et que i reste infectieux pendant une durée τ_i . La probabilité

²Nous verrons toutefois que cet isomorphisme ne tient que pour des cas particuliers dont le modèle de Newman [50] fait partie. Pour des modèles plus généraux tels que celui développé dans le cadre de cette maîtrise, il faut utiliser des notions de percolation sur réseaux dirigés pour conserver la correspondance.

$1 - T_{ij}$ qu'il n'y ait jamais de transmission entre i et j est donnée par

$$1 - T_{ij} = \lim_{dt \rightarrow 0} (1 - r_{ij} dt)^{\tau_i/dt} = e^{-r_{ij}\tau_i}.$$

La probabilité qu'il y ait éventuellement transmission de la maladie de i vers j est donc

$$T_{ij} = 1 - e^{-r_{ij}\tau_i}. \quad (2.1)$$

Les quantités r_{ij} et τ_i sont dépendantes d'une multitude de paramètres comportementaux et biologiques propres autant à la maladie qu'aux individus et donc varient typiquement d'une paire ij à l'autre. Ceci rend malheureusement difficile, voire impossible, l'élaboration d'un modèle de percolation analytique. Il est toutefois possible, dans certains cas, de faire quelques approximations qui permettront l'élaboration d'un modèle solutionnable.

2.1.1 L'hypothèse *iid*

L'hypothèse *iid* consiste à supposer que r_{ij} et τ_i sont des variables aléatoires *indépendantes identiquement distribuées (iid)* tirées respectivement des distributions $P(r)$ et $P(\tau)$. Ceci implique que T_{ij} est aussi une variable aléatoire *iid* et donc que la probabilité *a priori* qu'il y ait transmission est simplement donnée par la moyenne de T_{ij} sur $P(r)$ et $P(\tau)$:

$$T \equiv \langle T_{ij} \rangle = 1 - \int_0^\infty e^{-r\tau} P(r)P(\tau) dr d\tau \quad (2.2)$$

où T est appelée la *transmissibilité* de la maladie. Ainsi, sous l'hypothèse *iid*, la propagation d'une maladie infectieuse est équivalente à celle pour laquelle toutes les probabilités de transmission sont égales à T . Cette hypothèse s'avère réaliste pour une large gamme de maladies où $P(r)$ et $P(\tau)$ sont concentrées autour d'une valeur donnée r_0 et τ_0 [39, 57].

2.2 Fonctions génératrices

Le formalisme de Newman est basé sur des fonctions génératrices [55, 71] qui possèdent plusieurs propriétés facilitant la manipulation de distributions discrètes de probabilités. Considérant une telle distribution dûment normalisée $\{\psi_k\}$ définie pour

$k \in \mathbb{N}$, on définit sa fonction génératrice $\varphi(x)$ comme étant un polynôme à une variable où les coefficients de x^k sont ψ_k

$$\varphi(x) = \sum_{k=0}^{\infty} \psi_k x^k. \quad (2.3)$$

On voit de (2.3) que $\varphi(1) = 1$ si $\{\psi_k\}$ est normalisée. Tel que dit précédemment, les fonctions génératrices possèdent quelques propriétés qui facilitent la manipulation de distributions discrètes de probabilités :

1. **Réobtention de la distribution** Il est possible de réobtenir les coefficients d'une distribution à partir de sa fonction génératrice par différentiation :

$$\psi_k = \frac{1}{k!} \left. \frac{d^k \varphi(x)}{dx^k} \right|_{x=0}. \quad (2.4)$$

On dit alors que la fonction génératrice $\varphi(x)$ génère la distribution $\{\psi_k\}$.

2. **Moments** Il est aussi possible de calculer les différents moments, s'ils existent, d'une distribution à partir de sa fonction génératrice

$$\langle k^n \rangle = \sum_{k=0}^{\infty} k^n \psi_k = \left[\left(x \frac{d}{dx} \right)^n \varphi(x) \right]_{x=1}.$$

Notons au passage le cas particulier

$$\langle k \rangle = \sum_{k=0}^{\infty} k \psi_k = \varphi'(1) \quad (2.5)$$

où le prime (') correspond à la différentiation de la fonction par rapport à son argument. De plus, toujours à partir de sa fonction génératrice, on peut obtenir la variance de la distribution

$$\begin{aligned} \sigma^2 &= \sum_{k=0}^{\infty} (k - \langle k \rangle)^2 \psi_k \\ &= \sum_{k=0}^{\infty} (k^2 - 2 \langle k \rangle k + \langle k \rangle^2) \psi_k \\ &= \varphi''(1) + \varphi'(1) - [\varphi'(1)]^2. \end{aligned} \quad (2.6)$$

3. **Convolution** La résultante du produit ou de la mise en puissance de fonctions génératrices correspond à la convolution de leur distribution. Par exemple, la

convolution de $\{\psi_k\}$ m fois avec elle-même est générée par sa fonction génératrice élevée à la puissance m :

$$\sum_{k=0}^{\infty} \left[\underbrace{\{\psi_k\} * \{\psi_k\} * \dots * \{\psi_k\}}_{m \text{ fois}} \right] x^k = [\varphi(x)]^m. \quad (2.7)$$

Étant données ces diverses propriétés, il est fréquent qu'une fonction génératrice soit obtenue en premier lieu par une quelconque démarche et que la distribution sous-jacente soit par la suite extraite de la résultante en utilisant (2.4). Toutefois, ces démarches nécessitent souvent la résolution numérique d'équations transcendantes et puisque l'évaluation des dérivées est assujettie à la précision numérique des ordinateurs, il est préférable de les évaluer à l'aide de la formule intégrale de Cauchy

$$\psi_k = \frac{1}{2\pi i} \oint_{\Gamma} \frac{\varphi(z)}{z^{k+1}} dz$$

où Γ correspond généralement au cercle unitaire³.

2.3 Formalisme

Les réseaux complexes considérés par l'approche de Newman (voir figure 2.1) sont définis comme suit :

1. Le réseau complexe est composé de N noeuds (les individus) et est de type *uni-parti* *i.e.* il n'existe qu'un seul type de noeuds. Les noeuds du réseau sont donc indiscernables (il n'existe pas de distinction entre les individus).
2. Les liens sont bidirectionnels. Il ne peut y avoir plus qu'un lien entre une paire de noeuds donnée ou de lien reliant un noeud avec lui-même. Le degré maximal d'un noeud est donc $N - 1$ et il y a un maximum de $\binom{N}{2}$ liens dans le réseau.
3. Le réseau est caractérisé par la distribution de degrés $\{P(k)\}$ et est à tout autre égard parfaitement aléatoire.

Ces propriétés définissent un ensemble de réseaux sur lequel les quantités prédites par le formalisme sont moyennées. Ce dernier est construit en supposant de grands réseaux ($N \gg 1$) et devient exact dans la limite $N \rightarrow \infty$.

³Voir annexe C.

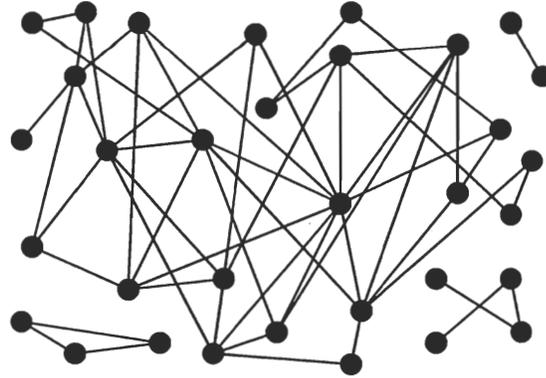


FIG. 2.1 – Représentation schématique des réseaux considérés par le modèle de Newman. Ces réseaux consistent en un seul type de noeuds et seulement un lien pouvant être parcouru dans les deux sens peut exister entre une paire de noeuds donnée. Ce dernier sera occupé (*i.e.* transmettra la maladie) avec une probabilité T que l'on appelle la *transmissibilité*.

2.3.1 Distribution des degrés occupés

Ayant $\{P(k)\}$, on connaît directement la distribution du nombre de voisins qu'ont les noeuds dans le réseau. Toutefois, on s'intéresse aux liens qui transmettent la maladie et donc au *degré efficace* (m) des noeuds. En termes de percolation, on a besoin de la distribution des degrés occupés avec une probabilité T dans le réseau. On supposera que l'état d'occupation d'un lien est indépendant de celui des autres⁴ et donc que la probabilité qu'un noeud ayant un degré k ait m liens occupés est

$$P(m|k) = \binom{k}{m} T^m (1 - T)^{k-m}.$$

La probabilité qu'un noeud choisi aléatoirement dans le réseau ait m liens occupés est

$$\begin{aligned} P(m) &= \sum_{k=m}^{\infty} P(m|k) P(k) \\ &= \sum_{k=m}^{\infty} P(k) \binom{k}{m} T^m (1 - T)^{k-m} \end{aligned} \quad (2.8)$$

⁴Cette hypothèse est justifiée dans un contexte d'application à l'épidémiologie puisque le fait qu'un individu infectieux infecte un de ses voisins n'a aucune influence sur la probabilité qu'il infecte ses autres voisins.

et est générée par la fonction génératrice $G(x; T)$

$$\begin{aligned}
 G(x; T) &= \sum_{m=0}^{\infty} P(m)x^m \\
 &= \sum_{m=0}^{\infty} \sum_{k=m}^{\infty} P(k) \binom{k}{m} T^m (1-T)^{k-m} x^m \\
 &= \sum_{k=0}^{\infty} P(k) \sum_{m=0}^k \binom{k}{m} (xT)^m (1-T)^{k-m} \\
 &= \sum_{k=0}^{\infty} P(k) (1 + (x-1)T)^k.
 \end{aligned} \tag{2.9}$$

On voit que $G(1; T) = 1$ si $\{P(k)\}$ est normalisée. En se référant à (2.5), on a que le degré occupé moyen des noeuds du réseau est

$$\begin{aligned}
 \tilde{z}_1 &\equiv G'(1; T) \\
 &= \left[T \sum_{k=0}^{\infty} k P(k) (1 + (x-1)T)^{k-1} \right]_{x=1} \\
 &= T \sum_{k=0}^{\infty} k P(k) \\
 &= T \langle k \rangle \\
 &= T z_1
 \end{aligned} \tag{2.10}$$

où $z_1 \equiv \langle k \rangle$ est le degré moyen du réseau.

2.3.2 Distribution des degrés occupés excédants

Comme on s'intéresse à la propagation d'une maladie sur un réseau, il s'avère utile de connaître la distribution des liens occupés excédants, $Q(m)$, *i.e.* les liens transmettant la maladie excluant celui par lequel le noeud a été infecté. Cette distribution est proportionnelle à $kP(k)$ puisque les noeuds ayant un plus grand degré ont plus de chances qu'un lien suivi au hasard mène à eux [55]. Ainsi, on a que

$$Q(m) = \frac{1}{z_1} \sum_{k=m}^{\infty} (k+1) P(k+1) \binom{k}{m} T^m (1-T)^{k-m} \tag{2.11}$$

qui est générée par $F(x; T)$

$$F(x; T) = \sum_{m=0}^{\infty} Q(m)x^m = \frac{G'(x; T)}{G'(1; T)}, \tag{2.12}$$

dont le résultat explicite s'obtient en suivant la même procédure qu'en (2.9). À l'aide de la propriété de convolution (2.7) des fonctions génératrices, on peut calculer la distribution de probabilité du nombre de seconds voisins rejoints en suivant des liens occupés

$$\sum_{k=0}^{\infty} P(k) [F(x; T)]^k = G(F(x; T); T).$$

En utilisant (2.5), on obtient que le nombre moyen de seconds voisins pouvant être atteints en ne suivant que les liens occupés est

$$\begin{aligned} \tilde{z}_2 &\equiv \left[\frac{d}{dx} G(F(x; T); T) \right]_{x=1} \\ &= G'(1; T) F'(1; T) \\ &= G''(1; T) \\ &= T^2 \sum_{k=0}^{\infty} k(k-1) P(k) \\ &= T^2 (\langle k^2 \rangle - \langle k \rangle) \\ &= T^2 z_2 \end{aligned} \tag{2.13}$$

où $z_2 \equiv \langle k^2 \rangle - \langle k \rangle$ est le nombre moyen de seconds voisins dans le réseau (en suivant tous les liens) [55].

2.3.3 Distribution du nombre d'infectés

Comme il a été dit en introduction de ce chapitre, l'état final d'une épidémie est isomorphe à la percolation par liens d'un réseau. Ainsi, puisque les composantes sont formées des noeuds liés par des liens occupés (*i.e.* transmettant la maladie), la question : *combien d'individus seront éventuellement infectés suite à l'infection d'un seul individu du réseau de contacts ?* peut se traduire en : *quelle est la taille de la composante⁵ du réseau à laquelle appartient cet individu infecté initialement ?* On définit $H(x; T)$ comme générant la distribution de la taille de la petite composante atteinte en suivant un lien, qu'on suppose occupé, choisi au hasard dans le réseau. Puisque les réseaux considérés par ce formalisme sont aléatoires (sauf en ce qui a trait à la distribution de degrés), la probabilité d'une boucle fermée dans une composante de taille finie est proportionnelle à $\mathcal{O}(N^{-1})$, ce qui est négligeable dans la limite $N \rightarrow \infty$. Ceci implique donc que les petites composantes ont une structure en arbre. Ainsi, il est possible de scinder $H(x; T)$ en un

⁵Dans un contexte d'utilisation en épidémiologie, les composantes sont définies tout comme précédemment à l'exception que les noeuds doivent être rejoints à partir de liens occupés, *i.e.* transmettant la maladie.

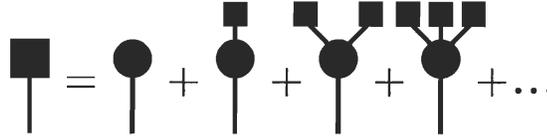


FIG. 2.2 – Représentation schématique de la décomposition de $H(x; T)$ en différentes contributions. Les boîtes correspondent à des composantes dont la distribution en taille est générée par $H(x; T)$ et les cercles aux noeuds rejoints par un lien occupé.

ensemble de contributions (voir figure 2.2) comme suit : la petite composante atteinte en suivant un lien choisi aléatoirement peut consister en 1) un noeud n’ayant pas de liens occupés excédants ou 2) un noeud ayant $m \geq 1$ liens occupés excédants chacun menant à une petite composante dont la distribution de la taille est aussi donnée par $H(x; T)$. Comme la distribution des degrés occupés excédants est générée par $F(x; T)$, on peut résumer cette interdépendance par l’équation d’auto-cohérence

$$H(x; T) = xF(H(x; T); T) \quad (2.14)$$

où on a fait usage de la propriété de convolution des fonctions génératrices (2.7) et où le facteur x est ajouté pour tenir compte du noeud qui a été atteint. Ainsi, dans la limite $N \rightarrow \infty$, on peut dire que $H(x; T)$ est *invariant sous déplacement sur le réseau*; que peu importe où on est rendu sur le réseau, on aura toujours *devant soi* la même distribution des petites composantes. L’annexe B.1 présente une méthode numérique permettant la résolution de (2.14). De façon similaire à $H(x; T)$, on définit $K(x; T)$ comme étant la fonction générant la distribution de la taille de la composante atteinte en choisissant un noeud au hasard dans le réseau

$$K(x; T) = xG(H(x; T); T). \quad (2.15)$$

Ainsi, connaissant les paramètres $\{P(k)\}$ et T , il suffit de résoudre (2.14) pour obtenir la distribution de la taille de la petite composante à laquelle appartient un noeud choisi au hasard. Dans un régime en-dessous du seuil de percolation⁶ (*i.e.* du seuil épidémique), (2.15) donne la distribution du nombre d’infections causées par un noeud infecté au hasard (incluant ce dernier que l’on nomme patient zéro).

⁶En fait, $K(x; T)$ reste valide dans un régime au-dessus du seuil de percolation. Elle nécessite toutefois d’être normalisée puisque la transition de phase implique une perte de la normalisation par un facteur $1 - S$, où S est la taille de la composante géante, comme il sera expliqué à la §2.3.5.

2.3.4 Seuil de percolation

Ayant en main $K(x, T)$, on peut obtenir directement le nombre moyen d'individus infectés $\langle s \rangle$ à l'aide de (2.5) ainsi que la variance associée à l'aide de (2.6). Or, on sait que, dans la limite $N \rightarrow \infty$, le point de percolation d'un réseau est caractérisé par la divergence de la longueur de corrélation, ce qui implique la divergence de la taille moyenne des petites composantes [66]. Ainsi, on obtient à l'aide de (2.14) et (2.15)

$$\langle s \rangle = K'(1; T) = 1 + \frac{G'(1; T)}{1 - F'(1; T)}$$

qui diverge lorsque $F'(1; T) = 1$. À l'aide de (2.12), on peut déterminer la transmissibilité critique

$$T_c = \frac{z_1}{z_2} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \quad (2.16)$$

à laquelle le réseau percole, *i.e.* à laquelle la composante géante apparaît. Ce point marque le moment où la probabilité d'avoir une épidémie à grande échelle cesse d'être nulle. En effet, puisqu'un individu infecté transmettra assurément la maladie à tous les autres membres de la composante à laquelle il appartient, la présence d'une composante géante implique qu'une fraction non-négligeable du réseau sera infectée si le patient zéro en fait partie. Ceci étant dit, il est intéressant de souligner que T_c dépend uniquement de la structure du réseau (*i.e.* distribution de degrés) et qu'il est possible d'évaluer cette quantité globale à l'aide des quantités locales z_1 et z_2 .

2.3.5 Composante géante

Lorsque $T > T_c$, on est en présence d'une composante géante et donc d'un risque non-nul d'épidémie. La composante géante étant une quantité extensive du réseau (formellement, elle a une taille infinie), la probabilité d'y retrouver des boucles fermées n'est pas négligeable. Sa structure n'est donc pas en arbre et elle est alors exclue de (2.14) et (2.15). Ainsi, au-dessus du seuil de percolation, $H(x; T)$ et $K(x; T)$ perdent leur normalisation puisqu'un lien ou un noeud choisi au hasard ne mène plus assurément à une composante de taille finie (*i.e.* petite composante) respectant une structure en arbre. Dans le cas de $K(x; T)$, la valeur manquante correspond donc à la probabilité \mathcal{S} qu'un noeud choisi au hasard mène à la composante géante

$$K(1; T) = 1 - \mathcal{S}. \quad (2.17)$$

En utilisant (2.15), on a que

$$\mathcal{S} = 1 - G(h; T) \quad (2.18)$$

où $h \equiv H(1; T)$ correspond à la probabilité qu'un lien ne mène pas à la composante géante et est la solution de l'équation d'auto-cohérence (2.14)

$$h = F(h; T). \quad (2.19)$$

Puisque le fait qu'un noeud mène à la composante géante implique qu'il en fasse partie, la fraction du réseau occupée par celle-ci est donc simplement \mathcal{S} . Ainsi, le fait que $T > T_c$ n'implique pas nécessairement qu'une infection initiale provoquera une épidémie à grande échelle⁷, mais plutôt qu'il y a une probabilité \mathcal{S} qu'il y ait une épidémie de taille relative \mathcal{S} .

2.4 Adaptation du formalisme aux réseaux semi-dirigés

Quelques années plus tard, Meyers et al. [45] ont présenté une généralisation du modèle de Newman en y ajoutant la possibilité d'avoir des liens dirigés dans le réseau, offrant ainsi un formalisme pour décrire la percolation de réseaux semi-dirigés. Cette généralisation fut motivée par la crise du SRAS qui a eu lieu à Vancouver et à Toronto en 2003 [46]. Cette crise mit en évidence l'importance des centres hospitaliers, et plus particulièrement des travailleurs de la santé, dans la propagation des maladies infectieuses. En effet, alors que les deux villes connurent leur premier cas de SRAS pratiquement au même moment⁸, Toronto dénombra plus de 209 cas probables alors que Vancouver n'en déplora que quatre. Une telle asymétrie s'explique tout d'abord par la différence des réseaux de contacts des deux patients zéros. Celui de Toronto vivait dans une maison multigénérationnelle où cohabitaient 10 personnes, ce qui favorisa un plus grand nombre d'infections secondaires, alors que celui de Vancouver vivait seul en compagnie de son épouse, ce qui eu l'effet inverse⁹ [62]. Par contre, une fois l'épidémie débutée à Toronto, beaucoup des transmissions se sont faites en milieu hospitalier [10, 68] ce qui a sensibilisé les autorités et le public à l'importance d'une hygiène adéquate¹⁰.

⁷On se rappelle que pour $\mathcal{S} < 1$, il existe encore des composantes de taille finie. Cette prédiction contraste avec celle des modèles compartimentaux dans lesquels il y aura assurément une épidémie à grande échelle au-dessus du seuil.

⁸Les deux patients zéros furent exposés au SRAS au même hôtel de Hong Kong le 25 février 2003.

⁹Le fait qu'il soit possible d'expliquer les différentes issues de ces épidémies potentielles à l'aide du réseau de contacts des patients zéros illustre bien la pertinence de ce genre d'approche.

¹⁰Au meilleur de la connaissance de l'auteur, c'est depuis l'épisode du SRAS qu'il y a des distributeurs de désinfectants à l'entrée des hôpitaux.

Les modélisateurs ont eux aussi été sensibilisés à cette problématique et ont voulu incorporer l'effet des milieux hospitaliers dans leurs modèles afin d'étudier de nouvelles stratégies de prévention et d'intervention [13, 61]. Pour tenir compte de la prédisposition

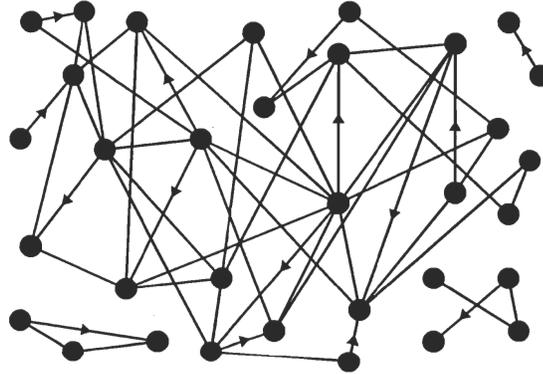


FIG. 2.3 – Représentation schématique des réseaux considérés par le modèle de Meyers *et al.* Tout comme ceux considérés par Newman, ces réseaux consistent en un seul type de noeuds et seulement un lien peut exister entre une paire de noeuds donnée. Toutefois, les liens pourront être bidirectionnels ou unidirectionnels et seront occupés (*i.e.* transmettront la maladie) avec une probabilité T_d ou T_u (selon le type du lien) que l'on appelle la *transmissibilité*.

des travailleurs de la santé à être en contact avec des gens infectieux avec qui ils n'auraient pas eu de contacts si ceux-ci n'avaient pas été infectés, Meyers *et al.* ont simplement ajouté des liens dirigés allant de la population aux travailleurs de la santé au modèle de Newman. Ainsi, ces liens dirigés permettent à un individu tombé malade d'infecter un travailleur de la santé sans toutefois permettre une transmission inverse puisque ces individus, s'ils n'avaient pas été malades, ne se seraient pas présentés à l'hôpital ou à la clinique. La figure 2.3 illustre les réseaux considérés par le modèle de Meyers *et al.*

2.4.1 Distribution de degrés occupés

Étant donné l'ajout des liens dirigés, les noeuds du réseau auront dorénavant 3 types de degrés : entrant (k^{in}), sortant (k^{out}) et non-dirigé (k^{un}). Le premier et le second type correspondent respectivement à l'arrivée et au départ d'un lien dirigé tandis que le dernier correspond au degré tel que considéré par Newman. Dans les scénarios étudiés par Meyer *et al.*, la plupart des noeuds dans la population auront des degrés non-dirigé

et sortant non-nuls et un degré entrant nul :

$$\begin{aligned}k^{\text{in}} &= 0 \\k^{\text{out}} &> 0 \\k^{\text{un}} &> 0\end{aligned}$$

tandis que les noeuds travaillant en milieu hospitalier¹¹ auront en plus des degrés non-dirigé et sortant non-nuls et un degré entrant de l'ordre de la taille du réseau :

$$\begin{aligned}k^{\text{in}} &\sim N \\k^{\text{out}} &> 0 \\k^{\text{un}} &> 0.\end{aligned}$$

Le modèle fut toutefois développé pour une distribution générale des degrés où les réseaux sont définis similairement à ceux de Newman :

1. Le réseau complexe est composé de N noeuds (les individus) et est de type *uni-parti* *i.e.* il n'existe qu'un seul type de noeud. Les noeuds du réseau sont donc indiscernables (il n'existe pas de distinction entre les individus¹¹).
2. Les liens peuvent être bidirectionnels ou unidirectionnels. Il ne peut avoir plus d'un lien (peu importe son type) entre une paire de noeuds donnée ou de lien reliant un noeud avec lui-même. Le degré maximal d'un noeud est donc $N - 1$ pour chaque type et il y a un maximum de $\binom{N}{2}$ liens dans le réseau.
3. Le réseau est caractérisé par la distribution de degrés $\{P(r, t, k)\}$ (r liens entrants, t liens sortants et k liens bidirectionnels¹²) et est à tout autre égard parfaitement aléatoire.

Ces propriétés définissent un ensemble de réseaux sur lequel les quantités prédites par le formalisme sont moyennées. Ce dernier est construit en supposant de grands réseaux ($N \gg 1$) et devient exact dans la limite $N \rightarrow \infty$. L'ajout de liens dirigés dans le réseau ajoute une contrainte à $\{P(r, t, k)\}$ puisque les degrés entrants et sortants correspondent au même objet *réel* : un lien. Ainsi, il est nécessaire qu'il y ait autant de liens sortants dans le réseaux que de liens entrants. Dans la limite $N \rightarrow \infty$, on demande simplement que z_1^{in} (degré moyen entrant) soit égal à z_1^{out} (degré moyen sortant).

De façon analogue au modèle de Newman, il est possible de calculer la distribution des *degrés efficaces*¹³ ($m^{\text{in}}, m^{\text{out}}, m^{\text{un}}$) à partir de $\{P(r, t, k)\}$ en supposant que l'état

¹¹Il est à noter que bien qu'il soit possible de distinguer les individus travaillant en milieu hospitalier du reste de la population à l'aide de la distribution de degrés, il n'y a aucune distinction *explicite* entre les noeuds.

¹²La notation $(k^{\text{in}}, k^{\text{out}}, k^{\text{un}})$ a été remplacée par (r, t, k) pour simplifier la notation.

¹³Ceux-ci seront toutefois notés (a, b, m) pour simplifier la notation.

d'occupation des liens est indépendant

$$P(a, b, m) = \sum_{i=a}^{\infty} \sum_{j=b}^{\infty} \sum_{k=m}^{\infty} P(r, t, k) \binom{i}{a} T_d^a (1 - T_d)^{i-a} \\ \times \binom{j}{b} T_d^b (1 - T_d)^{j-b} \binom{k}{m} T_u^m (1 - T_u)^{k-m} \quad (2.20)$$

La distinction des liens en deux types (*i.e.* bidirectionnels et unidirectionnels) permet l'utilisation de deux transmissibilités, T_u et T_d , ce qui offre un modèle plus général que celui de Newman. Ces deux transmissibilités sont, bien entendu, la moyenne de deux distributions quelconques tel que prescrit par l'hypothèse *iid*. La fonction génératrice associée à $P(a, b, m)$ est alors

$$\bar{G}(x, y, u; T_d, T_u) = \sum_{a,b,m=0}^{\infty} P(a, b, m) x^a y^b u^m \\ = \sum_{r,t,k=0}^{\infty} P(r, t, k) (1 + (x-1)T_d)^r (1 + (y-1)T_d)^t (1 + (u-1)T_u)^k. \quad (2.21)$$

En utilisant la propriété de différentiation des fonctions génératrices, il est possible de calculer les *degrés efficaces* moyens entrants (\bar{z}_1^{in}), sortants (\bar{z}_1^{out}) et non-dirigés (\bar{z}_1^{un})

$$\bar{z}_1^{\text{in}} = \frac{d}{dx} \bar{G}(1, 1, 1; T_d, T_u) = T_d \sum_{i=0}^{\infty} r P(r, t, k) = T_d \langle r \rangle = T_d z_1^{\text{in}} \quad (2.22)$$

$$\bar{z}_1^{\text{out}} = \frac{d}{dy} \bar{G}(1, 1, 1; T_d, T_u) = T_d \sum_{j=0}^{\infty} t P(r, t, k) = T_d \langle t \rangle = T_d z_1^{\text{out}} \quad (2.23)$$

$$\bar{z}_1^{\text{un}} = \frac{d}{du} \bar{G}(1, 1, 1; T_d, T_u) = T_u \sum_{k=0}^{\infty} k P(r, t, k) = T_u \langle k \rangle = T_u z_1^{\text{un}} \quad (2.24)$$

où $z_1^{\text{in}} \equiv \langle i \rangle$, $z_1^{\text{out}} \equiv \langle j \rangle$ et $z_1^{\text{un}} \equiv \langle k \rangle$ sont respectivement les degrés moyens entrants, sortants et non-dirigés du réseau. La contrainte supplémentaire imposée à $\{P(r, t, k)\}$

$$z_1^{\text{in}} = \sum_{i=0}^{\infty} r P(r, t, k) = \sum_{j=0}^{\infty} t P(r, t, k) = z_1^{\text{out}} \quad (2.25)$$

peut alors s'écrire ainsi

$$\frac{d}{dx} \bar{G}(1, 1, 1; 1, T_u) = \frac{d}{dy} \bar{G}(1, 1, 1; 1, T_u).$$

2.4.2 Distribution de degrés occupés excédants

Toujours suivant Newman, on définit trois fonctions génératrices générant la distribution des degrés occupés excédants considérant que l'on ait atteint un noeud via l'une ou l'autre des façons suivantes : en suivant un lien dirigé¹⁴ (2.26), en suivant à rebours un lien dirigé (2.27) ou en suivant un lien non-dirigé (2.28). Les distributions générées sont de forme semblable à (2.8) mais pour trois variables comme pour (2.20) et peuvent être obtenues à partir de $\bar{G}(x, y, u; T_d, T_u)$ par différentiation

$$\bar{F}^{\text{in}}(x, y, u; T_d, T_u) = \frac{1}{T_d z_1^{\text{in}}} \frac{d}{dx} \bar{G}(x, y, u; T_d, T_u) \quad (2.26)$$

$$\bar{F}^{\text{out}}(x, y, u; T_d, T_u) = \frac{1}{T_d z_1^{\text{out}}} \frac{d}{dy} \bar{G}(x, y, u; T_d, T_u) \quad (2.27)$$

$$\bar{F}^{\text{un}}(x, y, u; T_d, T_u) = \frac{1}{T_u z_1^{\text{un}}} \frac{d}{du} \bar{G}(x, y, u; T_d, T_u). \quad (2.28)$$

2.4.3 Distribution du nombre d'infectés

Afin d'obtenir une distribution du nombre d'individus infectés suite à une infection initiale, on définit $\bar{H}^{\text{in}}(v; T_d, T_u)$ et $\bar{H}^{\text{un}}(v; T_d, T_u)$ comme étant les fonctions génératrices générant la distribution du nombre d'individus éventuellement atteints en suivant respectivement un lien dirigé et un lien non-dirigé. Tout comme pour Newman, la probabilité de retrouver une boucle fermée dans une composante de taille finie va comme $\mathcal{O}(N^{-1})$ et est donc négligeable dans la limite $N \rightarrow \infty$. La figure 2.4 illustre comment il

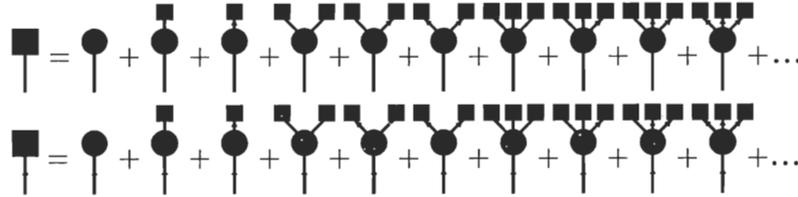


FIG. 2.4 – Représentation schématique des équations (2.29) et (2.30).

Les boîtes correspondent à des composantes dont la distribution en taille est donnée par $\bar{H}^{\text{in}}(x; T_d, T_u)$ ou $\bar{H}^{\text{un}}(x; T_d, T_u)$ selon la nature du lien suivi et les cercles aux noeuds rejoints par un lien occupé.

¹⁴Il est important de réaliser que l'arrivée à un noeud via un lien dirigé correspond à arriver à ce noeud via un de ses liens entrants.

est possible de décomposer $\bar{H}^{\text{in}}(v; T_d, T_u)$ et $\bar{H}^{\text{un}}(v; T_d, T_u)$ afin d'obtenir deux équations de cohérence analogues à (2.14)

$$\bar{H}^{\text{in}}(v; T_d, T_u) = v\bar{F}^{\text{in}}(1, \bar{H}^{\text{in}}(v; T_d, T_u), \bar{H}^{\text{un}}(v; T_d, T_u); T_d, T_u) \quad (2.29)$$

$$\bar{H}^{\text{un}}(v; T_d, T_u) = v\bar{F}^{\text{un}}(1, \bar{H}^{\text{in}}(v; T_d, T_u), \bar{H}^{\text{un}}(v; T_d, T_u); T_d, T_u). \quad (2.30)$$

En solutionnant ces deux équations¹⁵, on obtient directement $\bar{K}(v; T_d, T_u)$ qui génère la distribution de la petite composante formée de noeuds reliés par des liens occupés (dirigés et non-dirigés) à laquelle appartient un noeud choisi aléatoirement

$$\bar{K}(v; T_d, T_u) = v\bar{G}(1, \bar{H}^{\text{in}}(v; T_d, T_u), \bar{H}^{\text{un}}(v; T_d, T_u); T_d, T_u). \quad (2.31)$$

Tout comme pour (2.14) et (2.15), le facteur v dans (2.29)–(2.31) est ajouté pour compter le noeud auquel on vient d'arriver. Ainsi, en spécifiant $\{P(r, t, k)\}$, T_d et T_u , on obtient directement la distribution de la taille de la petite composante à laquelle appartient un noeud choisi au hasard sous le seuil de percolation¹⁶.

2.4.4 Seuil de percolation

Ayant en main $\bar{K}(v; T_d, T_u)$, on peut obtenir directement le nombre moyen d'individus infectés $\langle s \rangle$ à l'aide de (2.5) ainsi que leur variance à l'aide de (2.6). En utilisant (2.5) et (2.21)–(2.31), on obtient¹⁷

$$\langle s \rangle = 1 + \frac{\langle t \rangle T_d \left[1 - T_u \left(\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} - \frac{\langle rk \rangle}{\langle k \rangle} \right) \right] + \langle k \rangle T_u \left[1 - T_d \left(\frac{\langle rt \rangle}{\langle r \rangle} - \frac{\langle tk \rangle}{\langle k \rangle} \right) \right]}{\left(1 - T_d \frac{\langle rt \rangle}{\langle r \rangle} \right) \left(1 - T_u \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right) - T_u T_d \frac{\langle rk \rangle}{\langle r \rangle} \frac{\langle tk \rangle}{\langle k \rangle}} \quad (2.32)$$

qui diverge lorsque

$$\left(1 - T_d \frac{\langle rt \rangle}{\langle r \rangle} \right) \left(1 - T_u \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right) - T_u T_d \frac{\langle rk \rangle}{\langle r \rangle} \frac{\langle tk \rangle}{\langle k \rangle} = 0. \quad (2.33)$$

Cette dernière équation définit une courbe dans l'espace $\{T_d, T_u\}$ au-dessus de laquelle il existe une composante géante et donc où il y a un risque d'épidémie. Dans le cas où $T_d = T_u = T$, on peut calculer une transmissibilité critique en solutionnant

$$\left(1 - T_c \frac{\langle rt \rangle}{\langle r \rangle} \right) \left(1 - T_c \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right) - T_c^2 \frac{\langle rk \rangle}{\langle r \rangle} \frac{\langle tk \rangle}{\langle k \rangle} = 0$$

¹⁵Voir annexe B.1

¹⁶En fait, tout comme pour le formalisme de Newman, $\bar{K}(v; T_d, T_u)$ reste valide dans un régime au-dessus du seuil de percolation. Elle nécessite toutefois d'être normalisée puisque la transition de phase implique une perte de la normalisation par un facteur $1 - \mathcal{P}$, où \mathcal{P} est la probabilité qu'il y ait une épidémie, comme il sera expliqué à la §2.4.5.

¹⁷Les équations de cette section ont été réécrites dans une notation nous semblant plus lisible que celle utilisée par Meyer *et al.* [45] pour faciliter la compréhension.

qui retombe correctement sur l'expression obtenue dans le modèle de Newman (2.16) en prenant la limite où les liens dirigés disparaissent ($P(r, t, k) \rightarrow 0 \forall r, t > 0$).

2.4.5 Composante géante (probabilité et taille)

Dans une situation où le point (T_d, T_u) est à l'extérieur de la courbe de transmissibilité critique (*i.e.* une droite reliant ce point à l'origine est interceptée par la courbe de transmissibilité critique (2.33)), il existe une composante géante dans le réseau et donc un risque non-nul d'épidémie. Tout comme pour le formalisme de Newman, les fonctions $\bar{H}^{\text{in}}(v; T_d, T_u)$, $\bar{H}^{\text{un}}(v; T_d, T_u)$ et $\bar{K}(v; T_d, T_u)$, qui sont obtenues en supposant la structure en arbre des petites composantes, ne représentent plus l'ensemble du réseau — puisqu'elles excluent la composante géante — et perdent donc leur normalisation. Toutefois, là s'arrête l'analogie avec le modèle de Newman puisque la probabilité d'une épidémie et sa taille relative ne se confondent plus en une seule et même quantité. En

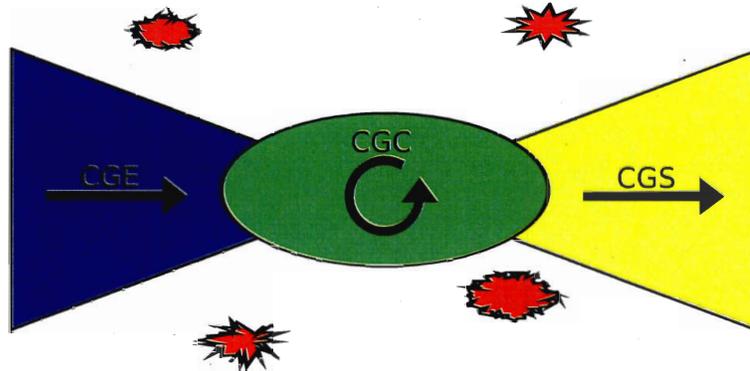


FIG. 2.5 – Structure simplifiée d'un réseau dirigé possédant une composante géante telle que proposée par Broder *et al.* [21]. On y voit la composante géante scindée en trois parties : la CGE en bleu, la CGC en vert et la CGS en jaune. Les petites composantes sont représentées en rouge.

effet, la présence de liens dirigés dans le réseau divise la composante géante en trois parties : la *Composante Géante Entrante* (CGE), la *Composante Géante Connectée* (CGC) et la *Composante Géante Sortante* (CGS). La figure 2.5 illustre cette nouvelle structure que l'on définit ainsi

1. **CGC** : Ensemble analogue à la composante géante telle que connue jusqu'à présent. Il s'agit d'un ensemble extensif de noeuds connectés uniquement par

des liens occupés. Ainsi, si l'un des individus dans cette composante est infecté, tous les autres individus de cet ensemble le seront.

2. **CGE** : Ensemble contenant tous les noeuds qui peuvent rejoindre la CGC mais qui ne peuvent être rejoints à partir de cette dernière (le lien avec la CGC implique un lien dirigé). Ces individus déclencheront donc une épidémie s'ils sont infectés, mais le fait qu'il y en ait une n'implique pas qu'ils seront infectés¹⁸.
3. **CGS** : Ensemble contenant tous les noeuds qui ne peuvent rejoindre la CGC mais qui peuvent être rejoints à partir de cette dernière (le lien de la CGC à la CGS se fait via un lien dirigé). Ces individus ne peuvent pas déclencher une épidémie, mais seront infectés s'il y en a une.

Ainsi, on voit que la probabilité qu'il y ait une épidémie (\mathcal{P}) n'est pas nécessairement égale à sa taille (\mathcal{S}) comme c'était le cas dans le formalisme de Newman où il fallait que le patient zéro fasse partie de la composante géante pour qu'il y ait une épidémie¹⁹. En fait, \mathcal{P} correspond à la fraction du réseau occupée par $\text{CGE} \cup \text{CGC}$ (l'ensemble des noeuds susceptibles de déclencher une épidémie) et \mathcal{S} à celle occupée par $\text{CGC} \cup \text{CGS}$ (l'ensemble des noeuds qui seront infectés lors d'une épidémie).

Calcul de \mathcal{P}

La composante géante formée de $\text{CGE} \cup \text{CGC}$ peut être vue comme la composante extensive des composantes formées en ne considérant que les liens dirigés (en respectant leur sens) et les liens non-dirigés. Ainsi, en connaissant $\bar{K}(v; T_d, T_u)$, la valeur associée à sa dénormalisation²⁰ au-dessus du seuil épidémique indiquera la taille relative de la composante extensive $\text{CGE} \cup \text{CGC}$ et donc \mathcal{P} . Cette distribution a déjà été calculée à la section 2.4.3 et est donnée par (2.31). En définissant

$$\vec{h}^{\text{in}} \equiv \bar{H}^{\text{in}}(1; T_d, T_u), \quad \vec{h}^{\text{un}} \equiv \bar{H}^{\text{un}}(1; T_d, T_u)$$

(2.29) et (2.30), évaluées à $v = 1$, se réduisent à

$$\begin{aligned} \vec{h}^{\text{in}} &= \bar{F}^{\text{in}}(1, \vec{h}^{\text{in}}, \vec{h}^{\text{un}}; T_d, T_u) \\ \vec{h}^{\text{un}} &= \bar{F}^{\text{un}}(1, \vec{h}^{\text{in}}, \vec{h}^{\text{un}}; T_d, T_u). \end{aligned}$$

¹⁸En fait, ils ne seront infectés que s'ils font partie de la chaîne qui a initialement déclenché l'épidémie.

¹⁹Il s'agit ici d'une façon simple et intuitive d'expliquer $\mathcal{P} = \mathcal{S}$. Une explication plus rigoureuse est donnée à la §3.3.5.

²⁰On se rappelle que $\bar{K}(1; T_d, T_u)$ correspond à la probabilité qu'un noeud choisi au hasard mène à une composante de taille finie.

Ces deux équations résolues, on obtient directement la probabilité qu'il y ait une épidémie au-dessus du seuil épidémique (ou de percolation) du réseau à l'aide de (2.31)

$$\mathcal{P} = 1 - \bar{K}(1; T_d, T_u) = 1 - \bar{G}(1, \vec{h}^{\text{in}}, \vec{h}^{\text{un}}; T_d, T_u).$$

Calcul de \mathcal{S}

Le calcul de \mathcal{S} se fait de la même façon que pour \mathcal{P} mais les distributions doivent être calculées en suivant les liens non-dirigés et les liens dirigés à rebours²¹. On définit $\overleftarrow{H}^{\text{out}}(v; T_d, T_u)$ et $\overleftarrow{H}^{\text{un}}(v; T_d, T_u)$ comme étant les fonctions génératrices générant la

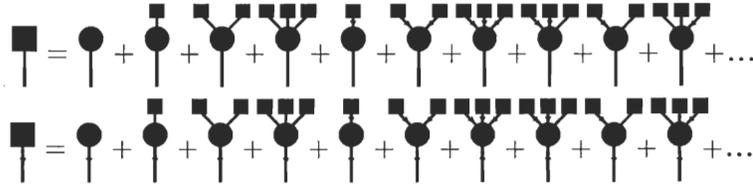


FIG. 2.6 – Représentation schématique des équations (2.34) et (2.35).

Les boîtes correspondent à des composantes dont la distribution en taille est donnée par $\overleftarrow{H}^{\text{out}}(v; T_d, T_u)$ ou $\overleftarrow{H}^{\text{un}}(v; T_d, T_u)$ selon la nature du lien suivi et les cercles aux noeuds rejoints par un lien occupé.

distribution de taille des petites composantes atteintes en suivant respectivement un lien dirigé à rebours et un lien non-dirigé. Ainsi, avec un raisonnement similaire à celui de la section 2.4.3 (voir figure 2.6), on obtient deux équations de cohérence similaires à (2.29) et (2.30)

$$\overleftarrow{H}^{\text{out}}(v; T_d, T_u) = v \bar{F}^{\text{out}}(\overleftarrow{H}^{\text{out}}(v; T_d, T_u), 1, \overleftarrow{H}^{\text{un}}(v; T_d, T_u); T_d, T_u) \quad (2.34)$$

$$\overleftarrow{H}^{\text{un}}(v; T_d, T_u) = v \bar{F}^{\text{un}}(\overleftarrow{H}^{\text{out}}(v; T_d, T_u), 1, \overleftarrow{H}^{\text{un}}(v; T_d, T_u); T_d, T_u). \quad (2.35)$$

En solutionnant ces deux équations²², on obtient directement $\overleftarrow{K}(v; T_d, T_u)$ qui génère la distribution des petites composantes formées de noeuds reliés par des liens occupés (dirigés suivis à rebours et non-dirigés)

$$\overleftarrow{K}(v; T_d, T_u) = v \bar{G}(\overleftarrow{H}^{\text{out}}(v; T_d, T_u), 1, \overleftarrow{H}^{\text{un}}(v; T_d, T_u); T_d, T_u). \quad (2.36)$$

²¹En effet, la probabilité qu'un noeud choisi au hasard mène à une composante extensive en suivant les liens à rebours correspond la fraction des noeuds qui seront infectés si une épidémie se déclare (voir figure 2.5).

²²Voir annexe B.1.

Au-dessus du seuil de percolation, $\overleftarrow{H}^{\text{out}}(v; T_d, T_u)$ et $\overleftarrow{H}^{\text{un}}(v; T_d, T_u)$ perdent leur normalisation. En posant

$$\overleftarrow{h}^{\text{out}} \equiv \overleftarrow{H}^{\text{out}}(1; T_d, T_u), \quad \overleftarrow{h}^{\text{un}} \equiv \overleftarrow{H}^{\text{un}}(1; T_d, T_u)$$

(2.34) et (2.35) deviennent

$$\overleftarrow{h}^{\text{out}} = \overleftarrow{F}^{\text{out}}(\overleftarrow{h}^{\text{out}}, 1, \overleftarrow{h}^{\text{un}}; T_d, T_u) \quad (2.37)$$

$$\overleftarrow{h}^{\text{un}} = \overleftarrow{F}^{\text{un}}(\overleftarrow{h}^{\text{out}}, 1, \overleftarrow{h}^{\text{un}}; T_d, T_u) \quad (2.38)$$

dont les solutions permettent d'obtenir directement \mathcal{S} à partir de (2.36)

$$\mathcal{S} = 1 - \overleftarrow{K}(1; T_d, T_u) = 1 - \overleftarrow{G}(\overleftarrow{h}^{\text{out}}, 1, \overleftarrow{h}^{\text{un}}; T_d, T_u).$$

Ceci met fin à la description de deux importants modèles en épidémiologie par réseaux de contacts. Les concepts introduits par ceux-ci sont à la base du formalisme développé dans le cadre de ce projet de maîtrise. Nous avons donc en main tout le nécessaire pour en offrir une description détaillée.

Chapitre 3

Vers une transmissibilité hétérogène (partie 1)

Le chapitre précédent a introduit les bases de l'épidémiologie par réseaux de contacts et a présenté en détail deux de ses principaux modèles, dont la validité en épidémiologie repose sur le réalisme de l'hypothèse *iid*. Bien que ce soit le cas pour une large gamme de maladies, nous verrons pourtant que cette hypothèse ne permet pas toujours une modélisation adéquate des problèmes épidémiologiques. C'est cette problématique qui a motivé ce projet de maîtrise où on a voulu développer un nouveau formalisme qui pourrait y palier. Les prochains chapitres présentent donc les résultats des efforts qui ont été déployés au cours des deux dernières années. Conservant la majorité des concepts utilisés dans les modèles de Newman et Meyers, nous avons développé un formalisme qui pourra combler certaines des lacunes qu'ont ces modèles lorsque la situation épidémiologique à modéliser ne permet pas l'hypothèse *iid*.

Nous discuterons tout d'abord des causes de l'invalidité de l'hypothèse *iid* à la §3.1 et proposerons ensuite une approche pour palier à cette problématique à la §3.2. Nous jetterons enfin les bases de notre formalisme décrivant la topologie des réseaux de contacts dans un régime en-dessous du seuil de percolation à la §3.3. Leur description au-dessus du seuil épidémique nécessite une approche semi-dirigée et fera l'objet du chapitre 4. Nous appuyerons finalement nos propos en y comparant les prédictions de notre modèle avec les résultats de simulations numériques et en y montrant qu'il retombe bel et bien sur les limites théoriques appropriées déjà connues.

3.1 Échec de l'hypothèse *iid*

Comme il a été expliqué à la section 2.1, la probabilité de transmission d'une maladie infectieuse dépend d'une foule de paramètres comportementaux et biologiques et varie donc typiquement d'une paire d'individus à une autre. Ceci complique grandement l'élaboration de modèles de percolation analytiques et il est nécessaire de recourir à des approximations telle l'hypothèse *iid* pour obtenir des modèles résolubles.

Cette hypothèse est présente dans la plupart des modèles analytiques actuels et son utilisation est justifiée lors de la modélisation de maladies pour lesquelles il n'y a pas de corrélation entre la probabilité de transmission et un caractère quelconque des individus¹. Or, cette hypothèse n'est malheureusement pas réaliste pour plusieurs maladies. Par exemple, les femmes sont jusqu'à deux fois plus vulnérables à être infectées par un partenaire porteur du VIH lors de relations sexuelles que les hommes dans la situation inverse [56]. Citons aussi le cas de l'influenza dont la probabilité de transmission varie, tout dépendant des souches, selon si l'infection se fait d'un adulte à un enfant, d'un enfant à un adulte, d'un enfant à un enfant ou d'un adulte à un adulte [22].

De plus, il a été soulevé par Kenah et Robins (K&R) [39] que l'hypothèse *iid* n'est pas applicable lorsque la distribution du temps d'infectiosité (τ_i) n'est plus concentrée autour d'une valeur particulière (*e.g.* distribution bimodale). En effet, cette hypothèse suppose que la probabilité d'occupation des liens dans un réseau est *iid* et donc varie indépendamment d'une paire de noeuds à l'autre. Or, pour tous les liens quittant un certain noeud i vers ses voisins j , T_{ij} telle que définie à l'équation (2.1)

$$T_{ij} = 1 - e^{-r_{ij}\tau_i}$$

varie certes d'une paire à l'autre, mais doit être calculée avec la même valeur de τ_i pour chacune d'entres elles. Or, en moyennant sur $P(\tau)$ tel que prescrit dans l'hypothèse *iid*

$$T = \langle T_{ij} \rangle = 1 - \int_0^\infty e^{-r\tau} P(r) P(\tau) dr d\tau,$$

on néglige cette corrélation et dans les cas où $P(\tau)$ n'est pas concentrée autour d'une certaine valeur, les formalismes basés sur l'hypothèse *iid* feront défaut. Dans ce même article, K&R proposent un formalisme basé sur une approche semi-dirigée comme à la section 2.4 permettant de palier à ce problème. De plus, Noël *et al.* [57], dans un article à paraître, mettent en perspective le problème soulevé par K&R en montrant les limites où l'hypothèse *iid* reste valide.

¹*i.e.* la transmission de l'agent pathogène est indépendante de la *nature* (*e.g.* homme/femme, adulte/enfant) des individus.

Le formalisme que nous présentons fut motivé en premier lieu par la première problématique énoncée ici. Toutefois, lorsque les différences entre les temps d’infectiosité pourront être associés à des différences de caractère chez les individus permettant de les distinguer entre différents types, notre formalisme offrira une alternative pour palier au problème soulevé par K&R.

3.2 Hétérogénéité des noeuds

Plusieurs approches sont possibles pour contourner et/ou régler les problématiques mentionnées ci-haut [39, 48, 57], mais encore peu de modélisateurs s’y sont attaqués. L’approche que nous avons préconisée consiste à considérer la solution de la façon suivante :

Le formalisme qui solutionnera la problématique mentionnée à la section 3.1 devra être en mesure d’appliquer la bonne probabilité d’occupation (i.e. la transmissibilité) aux liens selon le sens de la transmission.

En langage épidémiologique, ceci nécessite de connaître *qui infecte qui* afin d’utiliser la bonne transmissibilité. Nous avons donc développé un formalisme mathématique décrivant l’état final de la percolation d’un réseau où les noeuds sont étiquetés avec un certain type (*e.g.* adulte, enfant, homme, femme, etc.). Ainsi, il nous est possible de connaître *qui est qui* et donc d’utiliser la transmissibilité appropriée pour chaque paire de liens selon le sens de l’infection.

3.2.1 Nouveaux réseaux

Comme il a été dit en introduction de ce chapitre, le formalisme que nous avons développé utilise plusieurs des concepts présents chez Newman². Ainsi, les réseaux que nous considérons sont très semblables à ceux introduits à la section 2.3 (voir figure 3.1) :

1. Le réseau complexe est composé de N noeuds (les individus) où chacun d’entre eux est étiqueté d’un des M types (*e.g.* 1, 2, 3, ..., M) possibles. Ceci permet donc de faire une distinction entre les individus. On définit w_i comme étant la fraction du réseau occupée par les noeuds de type i .

²Nous pouvons le voir comme une extension à plusieurs types de noeuds de son formalisme.

2. Les liens sont bidirectionnels. Il ne peut y avoir plus d'un lien entre une paire de noeuds donnée ou de lien reliant un noeud avec lui-même. Le degré maximal d'un noeud est donc $N - 1$ et il y a un maximum de $\binom{N}{2}$ liens dans le réseau.
3. Le réseau est caractérisé par la distribution de degrés des noeuds de type i $\{P_i(\mathbf{k})\}$ où \mathbf{k} correspond au vecteur (k_1, k_2, \dots, k_M) (k_1 liens allant vers des noeuds de type 1, k_2 liens allant vers des noeuds de type 2, etc.) et est à tout autre égard parfaitement aléatoire.

Ces propriétés définissent un ensemble de réseaux sur lequel les quantités prédites par le formalisme sont moyennées. Ce dernier est construit en supposant de grands réseaux ($N \gg 1$) et devient exact dans la limite $N \rightarrow \infty$. Bien que les liens n'aient aucune directionnalité, l'ajout de types de noeuds dans les réseaux implique une directionnalité *artificielle* des liens. En effet, l'infection peut désormais se faire d'un noeud de type i à un type j (nous noterons $i \rightarrow j$) ou dans le sens inverse ($j \rightarrow i$). Puisque la distribution de degrés est définie telle qu'elle indique les degrés d'un noeud de type i vers les autres types de noeuds, un même lien joignant un type i et un type j sera *regardé de deux points de vue différents* (i.e. $i \rightarrow j$ et $j \rightarrow i$). Ainsi, tout comme le formalisme de Meyers, il est

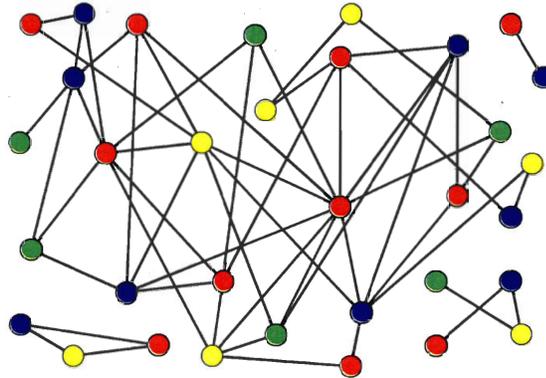


FIG. 3.1 – Représentation schématique des réseaux considérés par notre formalisme avec $M = 4$, $N = 33$. $w_1 = \frac{3}{11}$, $w_2 = \frac{1}{3}$, $w_3 = \frac{2}{11}$ and $w_4 = \frac{7}{33}$ et où les couleurs bleu, rouge, vert et jaune correspondent respectivement aux noeuds de type 1, 2, 3 et 4. Tous les liens sont bidirectionnels et peuvent donc être parcourus dans l'un ou l'autre des sens.

nécessaire d'imposer une contrainte à la distribution de degrés afin que l'ensemble des réseaux défini par $P_i(\mathbf{k})$ soit cohérent (i.e. que les réseaux soient *fermés*³). Définissons tout d'abord z_{ij} comme étant le nombre de premiers voisins de type j qu'a un noeud

³Nous entendons ici un réseau dont les prescriptions topologiques (i.e. $P_i(\mathbf{k})$ et w_i) sont telles que chaque lien quittant un noeud correspond à un lien arrivant à un autre noeud.

de type i . Cette quantité s'obtient directement de $P_i(\mathbf{k})$

$$z_{ij} = \sum_{\mathbf{k}=0}^{\infty} k_j P_i(\mathbf{k}) \quad (3.1)$$

où la somme sur \mathbf{k} correspond à une somme sur tous ses éléments (k_1, k_2, \dots, k_M) de 0 à l'infini comme les bornes l'indiquent. La contrainte à imposer à $P_i(\mathbf{k})$ consiste à demander qu'il y ait autant de liens $i \rightarrow j$ que de liens $j \rightarrow i$, ce qui, dans la limite $N \rightarrow \infty$, peut s'écrire comme

$$\mathbf{wz} = (\mathbf{wz})^\dagger \quad (3.2)$$

où

$$\mathbf{w} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_M \end{bmatrix} ; \quad \mathbf{z} = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1M} \\ z_{21} & z_{22} & \dots & z_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ z_{M1} & z_{M2} & \dots & z_{MM} \end{bmatrix}.$$

Notons que (3.2) implique $\binom{M}{2}$ équations non-triviales et que ce système d'équations linéaires est généralement surdéterminé. Ainsi, il n'est typiquement pas possible de fixer w_i ou $P_i(\mathbf{k})$ et d'obtenir l'autre à l'aide de (3.2); il faut plutôt utiliser une paire w_i et $P_i(\mathbf{k})$ respectant déjà (3.2) dès le départ⁴. Ajoutons par contre que le cas $M = 2$ est le seul où il est possible de définir $P_i(\mathbf{k})$ et d'en déduire w_i par la suite à l'aide de (3.2). En effet, puisque $\text{Tr}(\mathbf{w}) = 1$, nous obtenons directement que

$$w_1 = \frac{z_{21}}{z_{12} + z_{21}} \quad w_2 = \frac{z_{12}}{z_{12} + z_{21}}. \quad (3.3)$$

3.2.2 Nouvelle Transmissibilité

Le nouveau type de réseaux présenté ci-haut nous permet de connaître la nature de chaque noeud du réseau et ainsi de répondre à la question *qui infecte qui*? Définissons la matrice de transmissibilité

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} & \dots & T_{1M} \\ T_{21} & T_{22} & \dots & T_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ T_{M1} & T_{M2} & \dots & T_{MM} \end{bmatrix}$$

⁴Notons toutefois que ces quantités seront habituellement prises de scénarios épidémiologiques réels et respecteront nécessairement (3.2). Cette condition pourra aussi aider les modélisateurs tentant de reproduire un véritable réseau de contacts à s'assurer de la cohérence de celui-ci.

où chacun des éléments T_{ij} est la probabilité de transmission⁵ de la maladie d'un noeud de type i à un de type j . Bien entendu, la valeur de T_{ij} varie d'une paire d'individus à l'autre pour les mêmes raisons que celles énumérées à la §2.1. Nous suposerons donc qu'à l'intérieur d'un même type de paire $i \rightarrow j$, les probabilités de transmission sont *iid* et donc que nous pouvons utiliser la valeur moyenne des distributions associées. Ainsi, pour tout problème de modélisation où l'on veut utiliser ce formalisme, il suffit de séparer les noeuds du réseau en différentes catégories (types) tel que l'hypothèse *iid* soit valide et d'utiliser les valeurs de transmissibilité correspondantes. Ainsi, ce formalisme permet d'éviter les problèmes rencontrés par les formalismes précédents et énumérés à la §3.1. Tout comme pour ceux de Newman et de Meyers *et al.*, il est possible de déduire une foule d'informations sur la percolation du réseau (*e.g.* l'état final de l'épidémie, s'il y en a une) en ne connaissant que $P_i(\mathbf{k})$, w_i et \mathbf{T} .

3.3 Formalisme (en-dessous du seuil de percolation)

Tel que mentionné en introduction de ce chapitre, notre formalisme, à l'instar de ceux présentés précédemment, est basé sur des fonctions génératrices telles que celles présentées à la section 2.2. Les prochaines sections expliquent comment nous avons adapté ce type de formalisme pour lui permettre de décrire les structures introduites à la section 3.2.

3.3.1 Distribution de degrés occupés

Connaissant $P_i(\mathbf{k})$ et supposant l'indépendance entre les états d'occupation des liens, la probabilité qu'un noeud de type i de degrés \mathbf{k} ait \mathbf{m} (m_1, m_2, \dots, m_M) liens occupés est simplement

$$P_i(\mathbf{m}|\mathbf{k}) = \prod_{l=1}^M \binom{k_l}{m_l} T_{il}^{m_l} (1 - T_{il})^{k_l - m_l}.$$

La probabilité qu'un noeud de type i choisi aléatoirement ait \mathbf{m} liens occupés est alors

$$P_i(\mathbf{m}) = \sum_{\mathbf{k}=\mathbf{m}}^{\infty} P_i(\mathbf{k}) \prod_{l=1}^M \binom{k_l}{m_l} T_{il}^{m_l} (1 - T_{il})^{k_l - m_l} \quad (3.4)$$

⁵Ne pas confondre l'actuel T_{ij} où les indices i et j vont de 1 à M (chaque type du réseau) avec le T_{ij} de la section 2.1 où les indices vont de 1 à N (chaque noeud du réseau).

et est générée par la fonction $G_i(\mathbf{x}; \mathbf{T})$

$$\begin{aligned}
G_i(\mathbf{x}; \mathbf{T}) &= \sum_{m=0}^{\infty} P_i(\mathbf{m}) \prod_{l=1}^M x_l^{m_l} \\
&= \sum_{m=0}^{\infty} \sum_{\mathbf{k}=m}^{\infty} P_i(\mathbf{k}) \prod_{l=1}^M \binom{k_l}{m_l} T_{il}^{m_l} (1 - T_{il})^{k_l - m_l} x_l^{m_l} \\
&= \sum_{\mathbf{k}=0}^{\infty} P_i(\mathbf{k}) \prod_{l=1}^M \sum_{m_l=0}^{k_l} \binom{k_l}{m_l} (x_l T_{il})^{m_l} (1 - T_{il})^{k_l - m_l} \\
&= \sum_{\mathbf{k}=0}^{\infty} P_i(\mathbf{k}) \prod_{l=1}^M [1 + (x_l - 1)T_{il}]^{k_l} \tag{3.5}
\end{aligned}$$

où \mathbf{x} correspond au vecteur (x_1, x_2, \dots, x_M) . Il est possible d'obtenir le degré occupé moyen \tilde{z}_{ij} (*i.e.* nombre moyen de liens quittant un noeud de type i vers un noeud de type j) à l'aide de $G_i(\mathbf{x}; \mathbf{T})$

$$\begin{aligned}
\tilde{z}_{ij} &= \frac{d}{dx_j} G_i(\mathbf{1}; \mathbf{T}) \\
&= T_{ij} \left[\sum_{\mathbf{k}=0}^{\infty} k_j P_i(\mathbf{k}) \prod_{l=1}^M [1 + (x_l - 1)T_{il}]^{k_l - 1} \right]_{\mathbf{x}=\mathbf{1}} \tag{3.6} \\
&= T_{ij} \sum_{\mathbf{k}=0}^{\infty} k_j P_i(\mathbf{k}) \\
&= T_{ij} z_{ij}
\end{aligned}$$

où z_{ij} est le degré moyen du réseau tel que défini à l'équation (3.1) et $\mathbf{1}$ correspond au vecteur $\underbrace{(1, 1, \dots, 1)}_{M \text{ fois}}$.

3.3.2 Distribution de degrés occupés excédants

Ainsi que pour les formalismes de Newman et Meyers *et al.*, il s'avère fort utile, voire nécessaire, de connaître la distribution des degrés excédants. Le degré excédant correspond au nombre de liens par lesquels il est possible de quitter un noeud auquel nous sommes arrivés en suivant un quelconque lien (donc en excluant celui-ci). Alors que dans le modèle de Newman cette quantité correspond simplement au degré des noeuds moins un, le fait d'avoir plusieurs types de noeuds requiert de connaître *d'où* nous sommes arrivés afin de savoir *vers où* nous pouvons aller. Ainsi, il faut introduire une *mémoire* au système. Définissons $Q_{ij}(\mathbf{n})$, où \mathbf{n} est aux degrés occupés excédants ce que \mathbf{k} et \mathbf{m} sont aux degrés et aux degrés occupés, comme étant la distribution de

degrés excédants d'un noeud de type j auquel nous sommes arrivés d'un noeud de type i (*i.e.* en suivant un lien $i \rightarrow j$). En suivant un raisonnement similaire à celui pour obtenir (2.11), nous obtenons

$$Q_{ij}(\mathbf{n}) = \frac{1}{z_{ji}} \sum_{\mathbf{k}=\mathbf{n}}^{\infty} (k_i + 1) P_j(\mathbf{k} + \boldsymbol{\delta}_i) \prod_{l=1}^M \binom{k_l}{n_l} T_{jl}^{n_l} (1 - T_{jl})^{k_l - n_l} \quad (3.7)$$

où $\boldsymbol{\delta}_i$ correspond au vecteur de Kronecker $(\delta_{i1}, \delta_{i2}, \dots, \delta_{iM})$. Cette distribution est générée par $F_{ij}(\mathbf{x}; \mathbf{T})$

$$\begin{aligned} F_{ij}(\mathbf{x}; \mathbf{T}) &= \sum_{\mathbf{n}=\mathbf{0}}^{\infty} Q_{ij}(\mathbf{n}) \prod_{l=1}^M x_l^{n_l} \\ &= \frac{1}{z_{ji}} \sum_{\mathbf{n}=\mathbf{0}}^{\infty} \sum_{\mathbf{k}=\mathbf{n}}^{\infty} (k_i + 1) P_j(\mathbf{k} + \boldsymbol{\delta}_i) \prod_{l=1}^M \binom{k_l}{n_l} T_{jl}^{n_l} (1 - T_{jl})^{k_l - n_l} x_l^{n_l} \\ &= \frac{1}{z_{ji}} \sum_{\mathbf{k}=\mathbf{0}}^{\infty} (k_i + 1) P_j(\mathbf{k} + \boldsymbol{\delta}_i) \prod_{l=1}^M \sum_{n_l=0}^{k_l} \binom{k_l}{n_l} (x_l T_{jl})^{n_l} (1 - T_{jl})^{k_l - n_l} \\ &= \frac{1}{z_{ji}} \sum_{\mathbf{k}=\mathbf{0}}^{\infty} k_i P_j(\mathbf{k}) \prod_{l=1}^M [1 + (x_l - 1) T_{jl}]^{k_l - \delta_{il}}. \end{aligned} \quad (3.8)$$

En comparant (3.6) et (3.8), nous constatons qu'il est possible de relier directement $F_{ij}(\mathbf{x}; \mathbf{T})$ à $G_i(\mathbf{x}; \mathbf{T})$ de la façon suivante

$$F_{ij}(\mathbf{x}; \mathbf{T}) = \frac{1}{z_{ji}} \frac{dG_j(\mathbf{x}; \mathbf{T})}{dx_i}.$$

3.3.3 Distribution du nombre d'infectés

Ayant maintenant en main les fonctions génératrices associées aux degrés occupés et aux degrés occupés excédants, il nous est possible de construire une équation de récurrence analogue à (2.14). La définition de nos nouveaux réseaux nous permet encore de supposer une structure en arbre des petites composantes puisque la probabilité des boucles fermées est d'ordre $\mathcal{O}(N^{-1})$ et est donc négligeable dans la limite $N \rightarrow \infty$. Définissons $H_{ij}(\mathbf{x}; \mathbf{T})$ comme étant la fonction génératrice de la distribution de la taille de la petite composante atteinte en suivant un lien $i \rightarrow j$ (donc en comptant le noeud de type j dans la composante atteinte). La structure en arbre des petites composantes nous permet une fois de plus de décomposer $H_{ij}(\mathbf{x}; \mathbf{T})$ en un ensemble de contributions. Ainsi, en arrivant par un lien $i \rightarrow j$, le noeud atteint pourra 1) n'avoir aucun autre lien par lesquels nous pourrions continuer notre chemin (*i.e.* degré occupé excédant nul) ou 2) un ou plusieurs liens occupés menant chacun à une petite composante (*i.e.* degré

occupé excédant non-nul) dont la distribution en taille est aussi générée par $H_{ij}(\mathbf{x}; \mathbf{T})$ (invariance sous déplacement dans le réseau). La figure 3.2 illustre la décomposition

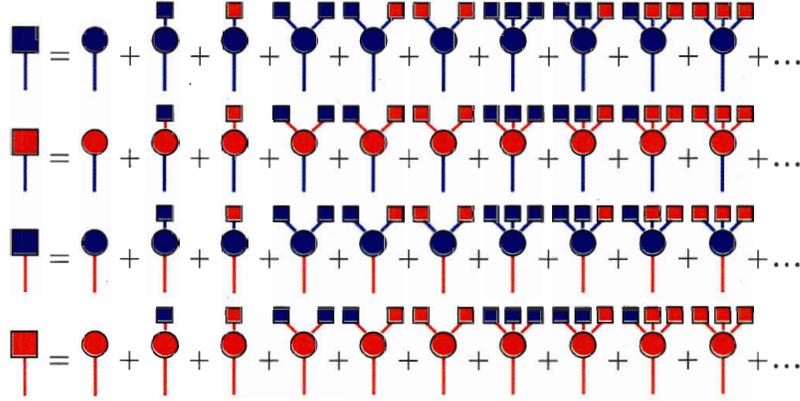


FIG. 3.2 – Représentation schématique de la décomposition de $H_{ij}(\mathbf{x}; \mathbf{T})$ en différentes contributions pour $M = 2$. Les boîtes correspondent à des composantes dont la distribution en taille est donnée par $H_{ij}(\mathbf{x}; \mathbf{T})$ et les cercles aux noeuds rejoints par un lien occupé. La couleur des boîtes correspond à la nature du premier noeud atteint (j) et la couleur du lien à celle du noeud d'où nous sommes partis (i).

de $H_{ij}(\mathbf{x}; \mathbf{T})$ pour $M = 2$. Puisque la distribution des degrés excédants est donnée par $F_{ij}(\mathbf{x}; \mathbf{T})$ et en utilisant la propriété de convolution des fonctions génératrices, nous pouvons résumer l'interdépendance de $H_{ij}(\mathbf{x}; \mathbf{T})$ par l'équation transcendante suivante :

$$H_{ij}(\mathbf{x}; \mathbf{T}) = x_j F_{ij}(\mathbf{H}_j(\mathbf{x}; \mathbf{T}); \mathbf{T}) \quad (3.9)$$

où le facteur x_j a été ajouté pour compter le noeud auquel nous sommes arrivés et où $\mathbf{H}_j(\mathbf{x}; \mathbf{T})$ correspond au vecteur $(H_{j1}(\mathbf{x}; \mathbf{T}), H_{j2}(\mathbf{x}; \mathbf{T}), \dots, H_{jM}(\mathbf{x}; \mathbf{T}))$. Cette équation se résout en cherchant l'état stationnaire du système dynamique itératif suivant

$$H_{ij}^{(n)}(\mathbf{x}; \mathbf{T}) = x_j F_{ij}(\mathbf{H}_j^{(n-1)}(\mathbf{x}; \mathbf{T}); \mathbf{T})$$

pour $n \geq 1$ avec des conditions initiales normalisées. L'existence d'une solution stationnaire est assurée par la présence du facteur x_j (une explication intuitive et une méthode de résolution numérique sont présentées à l'annexe B.1). Nous définissons maintenant $K(\mathbf{x}; \mathbf{T})$ comme étant la fonction génératrice de la distribution de la taille de la petite composante atteinte à partir d'un noeud quelconque du réseau. De façon analogue à Newman, $x_i G_i(\mathbf{H}_i(\mathbf{x}; \mathbf{T}); \mathbf{T})$ génère la distribution du nombre de noeuds rejoints par

des liens occupés à partir d'un noeud de type i . En pondérant par la fraction du réseau occupée par les noeuds de type i , nous obtenons

$$K(\mathbf{x}; \mathbf{T}) = \sum_{i=1}^M w_i x_i G_i(\mathbf{H}_i(\mathbf{x}; \mathbf{T}); \mathbf{T}). \quad (3.10)$$

Ainsi, en ne connaissant que $P_i(\mathbf{k})$, w_i et \mathbf{T} , il suffit de résoudre (3.9) et (3.10) pour obtenir la distribution de la taille de la petite composante à laquelle appartient un noeud choisi au hasard. De plus, puisque $H_{ij}(\mathbf{x}; \mathbf{T})$ et $K(\mathbf{x}; \mathbf{T})$ possèdent M variables, la distribution ainsi obtenue aura M dimensions et il sera possible de connaître la composition de ces composantes *i.e.* la distribution du nombre d'individus infectés de chaque type.

3.3.4 Seuil de percolation

À l'aide de $K(\mathbf{x}; \mathbf{T})$, il est possible de calculer directement le nombre moyen d'individus de type i infectés $\langle s_i \rangle$ à l'aide de (2.5) et la variance associée à l'aide de (2.6). En dérivant (3.10) par rapport à x_i , nous obtenons

$$\begin{aligned} \frac{dK(\mathbf{x}; \mathbf{T})}{dx_i} &= \sum_{l=1}^M w_l \frac{d}{dx_i} x_l G_l(\mathbf{H}_l(\mathbf{x}; \mathbf{T}); \mathbf{T}) \\ &= \sum_{l=1}^M w_l G_l(\mathbf{H}_l(\mathbf{x}; \mathbf{T}); \mathbf{T}) \delta_{il} + \sum_{l=1}^M w_l x_l \frac{d}{dx_i} G_l(\mathbf{H}_l(\mathbf{x}; \mathbf{T}); \mathbf{T}) \\ &= w_i G_i(\mathbf{H}_i(\mathbf{x}; \mathbf{T}); \mathbf{T}) + \sum_{l=1}^M w_l x_l \sum_{j=1}^M \frac{\partial G_l(\mathbf{H}_l(\mathbf{x}; \mathbf{T}); \mathbf{T})}{\partial H_{lj}(\mathbf{x}; \mathbf{T})} \frac{dH_{lj}(\mathbf{x}; \mathbf{T})}{dx_i} \end{aligned} \quad (3.11)$$

où nous avons explicité les dérivées en chaîne. En évaluant (3.11) à $\mathbf{x} = \mathbf{1}$, nous obtenons⁶ le nombre moyen de noeuds de type i dans les petites composantes

$$\langle s_i \rangle = w_i + \sum_{l=1}^M w_l \sum_{j=1}^M \tilde{z}_{lj} \alpha_{lj}^{(i)} \quad (3.12)$$

où $\alpha_{ij}^{(l)} \equiv \left. \frac{dH_{ij}(\mathbf{x}; \mathbf{T})}{dx_i} \right|_{\mathbf{x}=\mathbf{1}}$ sont les solutions de

$$\alpha_{ij}^{(l)} = \delta_{jl} + \sum_{m=1}^M T_{jm} \beta_{ij}^{(m)} \alpha_{jm}^{(l)} \quad (3.13)$$

⁶Rappelons-nous que $H_{ij}(\mathbf{1}; \mathbf{T}) = 1 \forall i, j = 1, \dots, M$ en-dessous du seuil de percolation puisque tous les liens occupés mènent à une composante de taille finie possédant une structure en arbre.

obtenue en dérivant (3.9) par x_l et en évaluant à $\mathbf{x} = \mathbf{1}$. Afin d'alléger la notation et extraire la dépendance par rapport à la transmissibilité, nous avons noté dans (3.13)

$$\left. \frac{\partial F_{ij}(\mathbf{x}; \mathbf{T})}{\partial x_l} \right|_{\mathbf{x}=\mathbf{1}} = \frac{1}{z_{ji}} \left. \frac{d^2 G_j(\mathbf{x}; \mathbf{T})}{dx_i dx_l} \right|_{\mathbf{x}=\mathbf{1}} = T_{jl} \beta_{ij}^{(l)}$$

où $\beta_{ij}^{(l)}$ n'est fonction que de la structure du réseau

$$\beta_{ij}^{(l)} = \frac{1}{z_{ji}} \sum_{\mathbf{k}=0}^{\infty} k_i (k_l - \delta_{il}) P_j(\mathbf{k}) \quad (3.14)$$

et est donc connu. Ainsi il suffit, pour obtenir $\langle s_i \rangle$, de résoudre (3.13) qui correspond en fait à M systèmes de M^2 équations linéaires et de M^2 inconnues qui peut se résoudre aisément à la main ou à l'aide d'un logiciel tel *Maple* ou *Mathematica*.

Cas $M = 2$

Afin de donner une idée de la forme des solutions, nous avons résolu le système (3.13) pour le cas $M = 2$:

$$\begin{aligned} \alpha_{11}^{(1)} &= \frac{1}{\zeta_2} \left(T_{12} T_{21} [\beta_{11}^{(2)} - \beta_{21}^{(2)}] [T_{22} (\beta_{12}^{(1)} \beta_{22}^{(2)} - \beta_{12}^{(2)} \beta_{22}^{(1)}) - \beta_{12}^{(1)}] + T_{22} \beta_{22}^{(2)} - 1 \right) \\ \alpha_{12}^{(1)} &= \frac{T_{21}}{\zeta_2} \left([T_{22} (\beta_{12}^{(1)} \beta_{22}^{(2)} - \beta_{12}^{(2)} \beta_{22}^{(1)}) - \beta_{12}^{(1)}] [1 - T_{11} (\beta_{11}^{(1)} - \beta_{21}^{(1)})] \right) \\ \alpha_{21}^{(1)} &= \frac{1}{\zeta_2} \left([T_{22} \beta_{22}^{(2)} - 1] [1 - T_{11} (\beta_{11}^{(1)} - \beta_{21}^{(1)})] \right) \\ \alpha_{22}^{(1)} &= \frac{T_{21} \beta_{22}^{(1)}}{\zeta_2} \left(T_{11} [\beta_{11}^{(1)} - \beta_{21}^{(1)}] - 1 \right) \\ \alpha_{11}^{(2)} &= \frac{T_{12} \beta_{11}^{(2)}}{\zeta_2} \left(T_{22} [\beta_{22}^{(2)} - \beta_{12}^{(2)}] - 1 \right) \\ \alpha_{12}^{(2)} &= \frac{1}{\zeta_2} \left([T_{11} \beta_{11}^{(1)} - 1] [1 - T_{22} (\beta_{22}^{(2)} - \beta_{12}^{(2)})] \right) \\ \alpha_{21}^{(2)} &= \frac{T_{12}}{\zeta_2} \left([T_{11} (\beta_{21}^{(2)} \beta_{11}^{(1)} - \beta_{21}^{(1)} \beta_{11}^{(2)}) - \beta_{21}^{(2)}] [1 - T_{22} (\beta_{22}^{(2)} - \beta_{12}^{(2)})] \right) \\ \alpha_{22}^{(2)} &= \frac{1}{\zeta_2} \left(T_{12} T_{21} [\beta_{22}^{(1)} - \beta_{12}^{(1)}] [T_{11} (\beta_{21}^{(2)} \beta_{11}^{(1)} - \beta_{21}^{(1)} \beta_{11}^{(2)}) - \beta_{21}^{(2)}] + T_{11} \beta_{11}^{(1)} - 1 \right) \end{aligned}$$

où

$$\begin{aligned} \zeta_2 &= T_{12} T_{21} [T_{11} (\beta_{21}^{(2)} \beta_{11}^{(1)} - \beta_{21}^{(1)} \beta_{11}^{(2)}) - \beta_{21}^{(2)}] [T_{22} (\beta_{12}^{(1)} \beta_{22}^{(2)} - \beta_{12}^{(2)} \beta_{22}^{(1)}) - \beta_{12}^{(1)}] \\ &\quad - [T_{11} \beta_{11}^{(1)} - 1] [T_{22} \beta_{22}^{(2)} - 1] \quad (3.15) \end{aligned}$$

est un polynôme en T_{ij} dont les coefficients ne dépendent que de la topologie du réseau. Nous voyons donc que tous les $\alpha_{ij}^{(l)}$ sont inversement proportionnels à la même quantité ζ_2 et donc que le nombre moyen d'infectés diverge selon

$$\langle s \rangle = \langle s_1 \rangle + \langle s_2 \rangle \sim \zeta_2^{-1}$$

lorsque $\zeta_2 \rightarrow 0$. Il y aura donc une transition de phase à $\zeta_2 = 0$ qui correspond à l'apparition de la composante géante. Ainsi, $\zeta_2 = 0$ définit une *hypersurface* dans un espace à 4 dimensions⁷, dont la base est formée des éléments de \mathbf{T} (T_{11} , T_{12} , T_{21} et T_{22}), qui définit le seuil de percolation du réseau. Ce résultat pour le cas où $M = 2$ et la forme particulière de (3.13) nous amène à conjecturer un comportement similaire pour un M général.

Conjecture

Nous conjecturons⁸ donc que la forme particulière de (3.13) (plus particulièrement la présence du terme δ_{jl}) assure une solution non-triviale où la forme fermée des $\alpha_{ij}^{(l)}$ est rationnelle avec un même dénominateur commun ζ_M . Celui-ci est un polynôme formé des éléments de \mathbf{T} ayant des coefficients qui dépendent uniquement de la structure du réseau — $P_i(\mathbf{k})$ — et s'annule à l'apparition de la composante géante causant ainsi la divergence simultanée de l'ensemble des $\langle s_i \rangle$ — et donc la divergence de $\langle s \rangle$ — selon

$$\langle s \rangle = \sum_{i=1}^M \langle s_i \rangle \sim \zeta_M^{-1}$$

lorsque $\zeta_M \rightarrow 0$. Ainsi, $\zeta_M = 0$ correspond à une *hypersurface* dans un espace M^2 dimensionnel dont la base est formée par les éléments⁹ de la matrice de transmissibilité \mathbf{T} . Cette *hypersurface* définit le seuil de percolation du réseau. En langage épidémiologique, $\zeta_M = 0$ définit, pour une structure sociale donnée (*i.e.* $P_i(\mathbf{k})$ et w_i), l'ensemble des *transmissibilités critiques* marquant le début d'un risque d'épidémie à grande échelle puisque tout scénario se situant à l'*extérieur* de cette hypersurface possédera une composante géante.

⁷En réalité, le fait que T_{12} et T_{21} n'apparaissent que sous la forme $T_{12}T_{21}$ réduit la dimensionnalité effective de l'espace à 3.

⁸La preuve de cette conjecture est présentée à l'annexe D.

⁹Chaque T_{ij} correspond à une dimension. Il se peut toutefois que ζ_M présente des symétries (comme c'est le cas pour $M = 2$), ce qui réduira la dimensionnalité effective de l'espace.

3.3.5 Composante géante

Vue la ressemblance frappante de notre formalisme avec celui développé par Newman, on pourrait être tenté de continuer sur l'élan et généraliser les équations (2.17)–(2.19) à plusieurs types de noeuds pour obtenir une description du réseau au-dessus du seuil de percolation. Les équations ainsi obtenues ne correspondraient toutefois pas à la taille d'une éventuelle épidémie puisque le fait que typiquement $\mathbf{T} \neq \mathbf{T}^\dagger$ (il s'agit là d'ailleurs de la motivation à développer ce genre de formalisme) induit une asymétrie dans le réseau.

L'approche introduite par Newman consiste en fait à calculer une distribution de degrés efficaces (la distribution de degrés occupés) où chacun des liens de chaque noeud est conservé avec la probabilité d'occupation appropriée. Une analyse topologique [55] du réseau résultant permet ensuite d'obtenir les quantités d'intérêt (*e.g.* la taille de la composante géante). Comme un lien non-dirigé peut transmettre la maladie dans l'une ou l'autre des directions, celui-ci restera non-dirigé avec une probabilité T^2 , deviendra dirigé (dans l'une ou l'autre des directions) avec une probabilité $T(1 - T)$ ou sera retiré avec une probabilité $(1 - T)^2$. Ceci illustre que tout réseau non-dirigé — ou du moins tout réseau possédant des liens non-dirigés — où nous appliquons une probabilité d'occupation des liens est *intrinsèquement* semi-dirigé et donc que sa composante géante se divise en trois ensembles (CGE, CGC et CGS) comme à la figure 2.5. Toutefois, puisque que dans ce cas-ci la probabilité qu'un lien devienne entrant ou sortant est la même, la CGE a la même taille que la CGS et la probabilité d'une épidémie \mathcal{P} et sa taille \mathcal{S} se confondent en une seule et même quantité. L'isomorphisme avec la percolation de liens est alors exact.

Dans l'approche présentée dans ce chapitre, un lien non-dirigé entre un noeud de type i et un noeud de type j restera non-dirigé avec une probabilité $T_{ij}T_{ji}$, deviendra dirigé du type i au type j avec une probabilité $T_{ij}(1 - T_{ji})$, deviendra dirigé du type j au type i avec une probabilité $T_{ji}(1 - T_{ij})$ ou sera retiré avec une probabilité $(1 - T_{ij})(1 - T_{ji})$. Ainsi, nous voyons que le fait que $\mathbf{T} \neq \mathbf{T}^\dagger$ implique qu'il y aura plus de liens entre les noeuds de type i et les noeuds de type j courant dans une direction que dans l'autre. Ceci fait en sorte que la CGE n'aura pas la même taille que la CGS et donc que $\mathcal{P} \neq \mathcal{S}$. Il faut donc avoir recours à une approche semi-dirigée inspirée de celle proposée par Meyers *et al.* pour expliciter cette asymétrie et ainsi décrire correctement la topologie du réseau au-dessus du seuil de percolation¹⁰. Une version semi-dirigée de notre formalisme est donc présentée au chapitre suivant et est ensuite utilisée pour correctement décrire la topologie du réseau en présence d'une composante géante.

¹⁰Tel que mentionné précédemment, il faut faire appel à des concepts de percolation de liens sur réseaux dirigés pour préserver l'isomorphisme.

Chapitre 4

Vers une transmissibilité hétérogène (partie 2)

Au chapitre 3, nous avons présenté une solution à l'incapacité des modèles épidémiologiques par réseaux de contacts actuels à décrire correctement la propagation de maladies dont la probabilité de transmission est corrélée avec un caractère des individus (*e.g.* âge, sexe). Nous avons ensuite développé un formalisme mathématique inspiré des modèles présentés au chapitre 2 afin de décrire la topologie des nouveaux réseaux de contacts que notre solution propose sous leur seuil de percolation. Tel que mentionné à la fin du chapitre précédent, la description de tels réseaux au-dessus de leur seuil de percolation nécessite une approche semi-dirigée analogue à celle développée par Meyers *et al.* [45]. Ainsi, une version généralisée à des réseaux semi-dirigés de notre formalisme est présentée à la §4.1 et est par la suite utilisée à la §4.2 pour décrire la topologie de nos réseaux dans un régime *au-dessus* de leur seuil de percolation. Finalement, la §4.3 compare les prédictions de notre formalisme aux résultats de simulations numériques et montre que celui-ci retombe bel et bien sur des limites théoriques déjà connues.

4.1 Formalisme semi-dirigé

Nous présentons d'abord une généralisation de notre formalisme permettant de décrire la percolation de réseaux semi-dirigés. Cette version ne tient pas compte d'une probabilité d'occupation des liens (*i.e.* la transmissibilité) comme pour les autres versions présentées jusqu'à maintenant. Celle-ci sera introduite d'une nouvelle façon à la section 4.2. Ce formalisme décrit donc simplement la topologie du réseau en considérant

tous les liens comme étant occupés (*i.e.* $\mathbf{T} = \mathbf{1}$). Il serait toutefois possible de calculer une distribution des degrés occupés en introduisant un terme binomial comme aux équations (2.8), (2.20) et (3.4) et ainsi d'obtenir un formalisme décrivant la percolation des liens d'un réseau semi-dirigé à plusieurs types de noeuds. Puisque cette généralisation supplémentaire ne nous sera d'aucune utilité pour la suite des choses, l'obtention d'un tel formalisme, dont la dérivation est directe, est laissée à la discrétion du lecteur intéressé.

4.1.1 Nouveaux réseaux

Les réseaux décrits par ce formalisme sont définis de façon similaire à ceux de la section 3.2.1 à l'exception qu'ils permettent aux liens d'avoir un sens ou non :

1. Le réseau complexe est composé de N noeuds (les individus) où chacun d'entre eux est étiqueté d'un des M types (*e.g.* $1, 2, 3, \dots, M$) possibles. Ceci permet donc de faire une distinction entre les individus. On définit w_i comme étant la fraction du réseau occupée par les noeuds de type i .
2. Les liens peuvent être bidirectionnels ou unidirectionnels. Il ne peut avoir plus d'un lien (peu importe son type) entre une paire de noeuds donnée ou de lien reliant un noeud avec lui-même. Le degré maximal d'un noeud est donc $N - 1$ et il y a un maximum de $\binom{N}{2}$ liens dans le réseau.
3. Le réseau est uniquement défini par la distribution de degrés $\{P_i(\mathbf{r}, \mathbf{t}, \mathbf{k})\}$ où \mathbf{r} , \mathbf{t} et \mathbf{k} correspondent¹ respectivement aux liens entrants, sortants et non-dirigés et est à tout autre égard parfaitement aléatoire.

Tout comme pour les autres formalismes, ces propriétés définissent un ensemble de réseaux sur lequel toutes les quantités prédites par le formalisme sont moyennées. Celui-ci est construit en supposant de grands réseaux ($N \gg 1$) et devient exact dans la limite $N \rightarrow \infty$. La figure 4.1 illustre une réalisation possible de ces réseaux. Définissons z_{ij}^{in} , z_{ij}^{out} et z_{ij}^{un} comme étant le nombre de premiers voisins de type j qu'a un noeud de type i en suivant respectivement des liens entrants, sortants et non-dirigés. Ces quantités

¹La notation utilisée ici est la même que celle de la section 3.2.1.

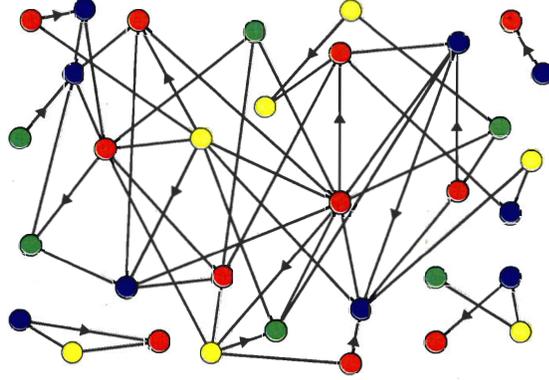


FIG. 4.1 – Représentation schématique des réseaux semi-dirigés considérés par notre formalisme généralisé avec $M = 4$, $N = 33$, $w_1 = \frac{3}{11}$, $w_2 = \frac{1}{3}$, $w_3 = \frac{2}{11}$ and $w_4 = \frac{7}{33}$ et où les couleurs bleu, rouge, vert et jaune correspondent respectivement aux noeuds de type 1, 2, 3 et 4. Les liens peuvent être dirigés ou non.

s'obtiennent directement de $P_i(\mathbf{r}, \mathbf{t}, \mathbf{k})$

$$z_{ij}^{\text{in}} = \sum_{\mathbf{r}, \mathbf{t}, \mathbf{k}=0}^{\infty} r_j P_i(\mathbf{r}, \mathbf{t}, \mathbf{k}) \quad (4.1)$$

$$z_{ij}^{\text{out}} = \sum_{\mathbf{r}, \mathbf{t}, \mathbf{k}=0}^{\infty} t_j P_i(\mathbf{r}, \mathbf{t}, \mathbf{k}) \quad (4.2)$$

$$z_{ij}^{\text{un}} = \sum_{\mathbf{r}, \mathbf{t}, \mathbf{k}=0}^{\infty} k_j P_i(\mathbf{r}, \mathbf{t}, \mathbf{k}). \quad (4.3)$$

De plus, à l'instar des réseaux de Meyers *et al.* et de ceux décrits à la section 3.2.1, les réseaux définis ci-haut sont assujettis à des contraintes de *fermeture* similaires à (2.25) et (3.2). Celles-ci demandent d'une part qu'il y ait autant de liens non-dirigés de type $i \rightarrow j$ que de type $j \rightarrow i$

$$\mathbf{wz}^{\text{un}} = (\mathbf{wz}^{\text{un}})^{\dagger} \quad (4.4)$$

et d'autre part que les noeuds de type i aient autant de liens entrants provenant de noeuds de type j qu'il y a de liens sortants de noeuds de type j allant vers des noeuds de type i

$$\mathbf{wz}^{\text{out}} = (\mathbf{wz}^{\text{in}})^{\dagger} \quad (4.5)$$

où nous avons noté

$$\mathbf{w} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_M \end{bmatrix} ; \quad \mathbf{z}^{\text{un}} = \begin{bmatrix} z_{11}^{\text{un}} & z_{12}^{\text{un}} & \dots & z_{1M}^{\text{un}} \\ z_{21}^{\text{un}} & z_{22}^{\text{un}} & \dots & z_{2M}^{\text{un}} \\ \vdots & \vdots & \ddots & \vdots \\ z_{M1}^{\text{un}} & z_{M2}^{\text{un}} & \dots & z_{MM}^{\text{un}} \end{bmatrix}$$

$$\mathbf{z}^{\text{out}} = \begin{bmatrix} z_{11}^{\text{out}} & z_{12}^{\text{out}} & \dots & z_{1M}^{\text{out}} \\ z_{21}^{\text{out}} & z_{22}^{\text{out}} & \dots & z_{2M}^{\text{out}} \\ \vdots & \vdots & \ddots & \vdots \\ z_{M1}^{\text{out}} & z_{M2}^{\text{out}} & \dots & z_{MM}^{\text{out}} \end{bmatrix} ; \quad \mathbf{z}^{\text{in}} = \begin{bmatrix} z_{11}^{\text{in}} & z_{12}^{\text{in}} & \dots & z_{1M}^{\text{in}} \\ z_{21}^{\text{in}} & z_{22}^{\text{in}} & \dots & z_{2M}^{\text{in}} \\ \vdots & \vdots & \ddots & \vdots \\ z_{M1}^{\text{in}} & z_{M2}^{\text{in}} & \dots & z_{MM}^{\text{in}} \end{bmatrix}$$

Ces contraintes impliquent respectivement $\binom{M}{2}$ et M^2 équations non-triviales ce qui a pour conséquence que le système linéaire correspondant sera surdéterminé tout comme à la section 3.2.1. Il faudra donc choisir une distribution de degrés $(P_i(\mathbf{r}, \mathbf{t}, \mathbf{k}))$ de même qu'un vecteur de population (w_i) respectant *a priori* les contraintes (4.4) et (4.5). Toutefois, comme il a été expliqué à la section 3.2.1, ces quantités seront habituellement prises de scénarios épidémiologiques réels qui respecteront nécessairement (4.4) et (4.5).

4.1.2 Distribution de degrés

De façon analogue au formalisme de Meyers *et al.*, nous définissons $\tilde{G}_i(\mathbf{x}, \mathbf{y}, \mathbf{u})$ comme étant la fonction génératrice associée à la distribution de degrés $P_i(\mathbf{r}, \mathbf{t}, \mathbf{k})$

$$\tilde{G}_i(\mathbf{x}, \mathbf{y}, \mathbf{u}) = \sum_{\mathbf{r}=0}^{\infty} \sum_{\mathbf{t}=0}^{\infty} \sum_{\mathbf{k}=0}^{\infty} P_i(\mathbf{r}, \mathbf{t}, \mathbf{k}) \prod_{l=1}^M x_l^{r_l} y_l^{t_l} u_l^{k_l}$$

Cette fonction permet de calculer directement les premiers moments

$$z_{ij}^{\text{in}} = \frac{d}{dx_j} \tilde{G}_i(\mathbf{1}, \mathbf{1}, \mathbf{1})$$

$$z_{ij}^{\text{out}} = \frac{d}{dy_j} \tilde{G}_i(\mathbf{1}, \mathbf{1}, \mathbf{1})$$

$$z_{ij}^{\text{un}} = \frac{d}{du_j} \tilde{G}_i(\mathbf{1}, \mathbf{1}, \mathbf{1}).$$

tels que définis aux équations (4.1)–(4.3).

4.1.3 Distribution de degrés excédants

Définissons maintenant $\tilde{F}_{ij}^{\text{in}}(\mathbf{x}, \mathbf{y}, \mathbf{u})$, $\tilde{F}_{ij}^{\text{out}}(\mathbf{x}, \mathbf{y}, \mathbf{u})$ et $\tilde{F}_{ij}^{\text{un}}(\mathbf{x}, \mathbf{y}, \mathbf{u})$ comme étant les fonctions génératrices associées à la distribution de degrés excédants obtenue en suivant

respectivement un lien dirigé $i \rightarrow j$ en respectant son sens, un lien dirigé $i \rightarrow j$ à rebours et un lien non-dirigé $i \rightarrow j$. Ces fonctions peuvent être obtenues directement de $\tilde{G}_i(\mathbf{x}, \mathbf{y}, \mathbf{u})$ par différentiation

$$\tilde{F}_{ij}^{\text{in}}(\mathbf{x}, \mathbf{y}, \mathbf{u}) = \frac{1}{z_{ji}^{\text{in}}} \frac{d}{dx_i} \tilde{G}_j(\mathbf{x}, \mathbf{y}, \mathbf{u}) \quad (4.6)$$

$$\tilde{F}_{ij}^{\text{out}}(\mathbf{x}, \mathbf{y}, \mathbf{u}) = \frac{1}{z_{ji}^{\text{out}}} \frac{d}{dy_i} \tilde{G}_j(\mathbf{x}, \mathbf{y}, \mathbf{u}) \quad (4.7)$$

$$\tilde{F}_{ij}^{\text{un}}(\mathbf{x}, \mathbf{y}, \mathbf{u}) = \frac{1}{z_{ji}^{\text{un}}} \frac{d}{du_i} \tilde{G}_j(\mathbf{x}, \mathbf{y}, \mathbf{u}) \quad (4.8)$$

et leur distribution associée peut être extraite par différentiation (voir (2.4)).

4.1.4 Distribution des petites composantes

Nous nous intéressons maintenant à la distribution de la taille de la petite composante à laquelle appartient un noeud choisi au hasard dans les réseaux en-dessous du seuil de percolation. Comme il a été montré par Meyers *et al.*, la présence de liens dirigés dans les réseaux semi-dirigés implique que la probabilité qu'un noeud choisi au hasard mène à la composante géante (\mathcal{P}) n'est pas nécessairement égale à celle qu'un quelconque noeud puisse être rejoint à partir de celle-ci (\mathcal{S}). Il est donc nécessaire de connaître la distribution associée aux composantes naturelles (formées en considérant les liens dirigés et les liens non-dirigés), mais aussi celle pour les composantes à rebours (formées en ne suivant que les liens dirigés à rebours et les liens non-dirigés).

Composantes naturelles

Encore une fois, la taille *infinie* ($N \rightarrow \infty$) des réseaux considérés nous permet de supposer que les petites composantes respectent une structure en arbre. Définissons $\vec{H}_{ij}^{\text{in}}(\mathbf{v})$ et $\vec{H}_{ij}^{\text{un}}(\mathbf{v})$ comme étant les fonctions génératrices générant la taille des petites composantes atteintes en suivant respectivement un lien dirigé $i \rightarrow j$ (*i.e.* en respectant son sens) et un lien non-dirigé $i \rightarrow j$. En décomposant ces distributions en différentes contributions comme à la figure 2.4 (en imaginant plus d'un type de noeuds), nous obtenons que $\vec{H}_{ij}^{\text{in}}(\mathbf{v})$ et $\vec{H}_{ij}^{\text{un}}(\mathbf{v})$ doivent respecter les équations de cohérence suivantes

$$\vec{H}_{ij}^{\text{in}}(\mathbf{v}) = v_j \tilde{F}_{ij}^{\text{in}}(1, \vec{H}_j^{\text{in}}(\mathbf{v}), \vec{H}_j^{\text{un}}(\mathbf{v})) \quad (4.9)$$

$$\vec{H}_{ij}^{\text{un}}(\mathbf{v}) = v_j \tilde{F}_{ij}^{\text{un}}(1, \vec{H}_j^{\text{in}}(\mathbf{v}), \vec{H}_j^{\text{un}}(\mathbf{v})) \quad (4.10)$$

où $\vec{H}_j^{\text{in}}(\mathbf{v})$ correspond au vecteur

$$(\vec{H}_{j1}^{\text{in}}(\mathbf{v}), \vec{H}_{j2}^{\text{in}}(\mathbf{v}), \dots, \vec{H}_{jM}^{\text{in}}(\mathbf{v}))$$

et $\vec{H}_j^{\text{un}}(\mathbf{v})$ au vecteur

$$(\vec{H}_{j1}^{\text{un}}(\mathbf{v}), \vec{H}_{j2}^{\text{un}}(\mathbf{v}), \dots, \vec{H}_{jM}^{\text{un}}(\mathbf{v})).$$

Ainsi, en solutionnant² (4.9) et (4.10), nous obtenons la distribution de la taille de la petite composante, formée en ne considérant que les liens dirigés (*i.e.* en respectant leur sens) et les liens non-dirigés, et rejointe à partir d'un noeud quelconque du réseau

$$\vec{K}(\mathbf{v}) = \sum_{l=1}^M w_l v_l \tilde{G}_l(\mathbf{1}, \vec{H}_l^{\text{in}}(\mathbf{v}), \vec{H}_l^{\text{un}}(\mathbf{v})). \quad (4.11)$$

En prenant la dérivée de (4.11) par rapport à chacune des variables v_j , nous pourrions obtenir un critère pour décrire le seuil de percolation des réseaux semi-dirigés. Cet exercice est, ici aussi, laissé à la discrétion du lecteur intéressé.

Composantes à rebours

Définissons maintenant $\overleftarrow{H}_{ij}^{\text{out}}(\mathbf{v})$ et $\overleftarrow{H}_{ij}^{\text{un}}(\mathbf{v})$ comme étant les fonctions génératrices générant la taille des petites composantes atteintes en suivant respectivement un lien dirigé $i \rightarrow j$ à rebours et un lien non-dirigé $i \rightarrow j$. Par un raisonnement analogue à celui utilisé précédemment, il nous est possible de décomposer $\overleftarrow{H}_{ij}^{\text{out}}(\mathbf{v})$ et $\overleftarrow{H}_{ij}^{\text{un}}(\mathbf{v})$ en une somme de contributions (voir la figure 2.6 mais avec plus d'un type de noeuds) et ainsi d'obtenir les équations de cohérence suivantes

$$\overleftarrow{H}_{ij}^{\text{out}}(\mathbf{v}) = v_j \tilde{F}_{ij}^{\text{out}}(\overleftarrow{H}_j^{\text{out}}(\mathbf{v}), \mathbf{1}, \overleftarrow{H}_j^{\text{un}}(\mathbf{v})) \quad (4.12)$$

$$\overleftarrow{H}_{ij}^{\text{un}}(\mathbf{v}) = v_j \tilde{F}_{ij}^{\text{un}}(\overleftarrow{H}_j^{\text{out}}(\mathbf{v}), \mathbf{1}, \overleftarrow{H}_j^{\text{un}}(\mathbf{v})). \quad (4.13)$$

En solutionnant³ (4.12) et (4.13), nous obtenons la distribution de la taille de la petite composante, formée en ne considérant que les liens dirigés à rebours et les liens non-dirigés, et atteinte à partir d'un noeud quelconque du réseau

$$\overleftarrow{K}(\mathbf{v}) = \sum_{l=1}^M w_l v_l \tilde{G}_l(\overleftarrow{H}_l^{\text{out}}(\mathbf{v}), \mathbf{1}, \overleftarrow{H}_l^{\text{un}}(\mathbf{v})). \quad (4.14)$$

²Voir l'annexe B.1.

³Voir l'annexe B.1.

4.1.5 Composante géante

Au-dessus du seuil de percolation, il existe une composante géante dans le réseau. Ainsi, $\vec{K}(\mathbf{v})$ et $\overleftarrow{K}(\mathbf{v})$ perdent leur normalisation puisqu'un noeud choisi au hasard ne mène plus assurément vers une composante de taille finie.

Calcul de \mathcal{P}

Comme nous l'avons vu à la section 2.4.5, la probabilité qu'un noeud choisi au hasard nous mène à la composante géante (\mathcal{P}) correspond à la taille relative de $\text{CGE} \cup \text{CGC}$. Cette quantité correspond à la valeur perdue de $\vec{K}(\mathbf{1})$. Ainsi, posons

$$\vec{h}_{ij}^{\text{in}} \equiv \vec{H}_{ij}^{\text{in}}(\mathbf{1}), \quad \vec{h}_{ij}^{\text{un}} \equiv \vec{H}_{ij}^{\text{un}}(\mathbf{1})$$

correspondant respectivement à la probabilité qu'un lien $i \rightarrow j$ dirigé (en respectant son sens) ou non-dirigé mène à une composante naturelle de taille finie. Ces derniers objets sont les solutions de (4.9) et (4.10) évaluées à $\mathbf{v} = \mathbf{1}$:

$$\vec{h}_{ij}^{\text{in}} = \tilde{F}_{ij}^{\text{in}}(\mathbf{1}, \vec{h}_j^{\text{in}}, \vec{h}_j^{\text{un}}) \quad (4.15)$$

$$\vec{h}_{ij}^{\text{un}} = \tilde{F}_{ij}^{\text{un}}(\mathbf{1}, \vec{h}_j^{\text{in}}, \vec{h}_j^{\text{un}}). \quad (4.16)$$

À l'aide de ces solutions, nous obtenons directement de (4.11) la probabilité qu'un noeud quelconque mène à la composante géante

$$\begin{aligned} \mathcal{P} &= 1 - \vec{K}(\mathbf{1}) \\ &= 1 - \sum_{l=1}^M w_l \tilde{G}_l(\mathbf{1}, \vec{h}_l^{\text{in}}, \vec{h}_l^{\text{un}}). \end{aligned} \quad (4.17)$$

Calcul de \mathcal{S}

La taille de la composante géante est obtenue en calculant la fraction des noeuds qui peuvent être rejoints à partir de la CGC *i.e.* en calculant la taille de $\text{CGC} \cup \text{CGS}$. Cette quantité correspond donc à la valeur perdue de $\overleftarrow{K}(\mathbf{1})$. Posons

$$\overleftarrow{h}_{ij}^{\text{out}} \equiv \overleftarrow{H}_{ij}^{\text{out}}(\mathbf{1}), \quad \overleftarrow{h}_{ij}^{\text{un}} \equiv \overleftarrow{H}_{ij}^{\text{un}}(\mathbf{1})$$

correspondant respectivement à la probabilité qu'un lien $i \rightarrow j$ dirigé suivi à rebours ou non-dirigé mène à une composante à rebours de taille finie. Ces quantités peuvent

être obtenues en solutionnant les équations

$$\overleftarrow{h}_{ij}^{\text{out}} = \tilde{F}_{ij}^{\text{out}}(\overleftarrow{h}_j^{\text{out}}, \mathbf{1}, \overleftarrow{h}_j^{\text{un}}) \quad (4.18)$$

$$\overleftarrow{h}_{ij}^{\text{un}} = \tilde{F}_{ij}^{\text{un}}(\overleftarrow{h}_j^{\text{out}}, \mathbf{1}, \overleftarrow{h}_j^{\text{un}}) \quad (4.19)$$

obtenues en posant $\mathbf{v} = \mathbf{1}$ dans (4.12) et (4.13). À l'aide de (4.14), nous obtenons directement la taille relative de la composante géante

$$\begin{aligned} S &= 1 - \overleftarrow{K}(\mathbf{1}) \\ &= 1 - \sum_{l=1}^M w_l \tilde{G}_l(\overleftarrow{h}_l^{\text{out}}, \mathbf{1}, \overleftarrow{h}_l^{\text{un}}). \end{aligned} \quad (4.20)$$

4.2 Formalisme (au-dessus du seuil de percolation)

Grâce à ce formalisme permettant de décrire la percolation de réseaux semi-dirigés, nous sommes en mesure de décrire la topologie des réseaux introduits à la section 3.2.1 (*i.e.* des réseaux non-dirigés avec plusieurs types de noeuds) au-dessus du seuil de percolation.

4.2.1 Ajout de la transmissibilité

Pour utiliser le formalisme semi-dirigé présenté à la section 4.1, nous devons transformer la distribution de degrés $P_i(\mathbf{k})$ en une distribution semi-dirigée. Cette transformation se fait de façon similaire aux équations (2.8), (2.20) et (3.4), *i.e.* en introduisant un terme multinomial, ce qui suppose l'indépendance de l'état d'occupation des liens. Ainsi, la probabilité qu'un noeud de type i ayant \mathbf{k} liens ait \mathbf{a} liens occupés entrants, \mathbf{b} liens occupés sortants et \mathbf{m} liens occupés non-dirigés est donnée par

$$\begin{aligned} P_i(\mathbf{a}, \mathbf{b}, \mathbf{m} | \mathbf{k}) &= \prod_{l=1}^M \binom{k_l}{a_l b_l m_l} [T_{li}(1 - T_{il})]^{a_l} [T_{il}(1 - T_{li})]^{b_l} \\ &\quad \times [T_{il} T_{li}]^{m_l} [(1 - T_{li})(1 - T_{il})]^{k_l - a_l - b_l - m_l}. \end{aligned}$$

La probabilité qu'un noeud choisi au hasard ait \mathbf{a} liens occupés entrants, \mathbf{b} liens occupés sortants et \mathbf{m} liens occupés non-dirigés est obtenue en sommant sur \mathbf{k} , ce qui

donne

$$P_i(\mathbf{a}, \mathbf{b}, \mathbf{m}) = \sum_{\mathbf{k}=\mathbf{a}+\mathbf{b}+\mathbf{m}}^{\infty} P_i(\mathbf{k}) \prod_{l=1}^M \binom{k_l}{a_l b_l m_l} [T_{li}(1-T_{il})]^{a_l} [T_{il}(1-T_{li})]^{b_l} \\ \times [T_{il}T_{li}]^{m_l} [(1-T_{li})(1-T_{il})]^{k_l-a_l-b_l-m_l}.$$

La fonction génératrice associée à $P_i(\mathbf{a}, \mathbf{b}, \mathbf{m})$

$$\tilde{G}_i(\mathbf{x}, \mathbf{y}, \mathbf{u}) = \sum_{\mathbf{a}, \mathbf{b}, \mathbf{m}=0}^{\infty} P_i(\mathbf{a}, \mathbf{b}, \mathbf{m}) \prod_{l=1}^M x_l^{a_l} y_l^{b_l} u_l^{m_l}.$$

se simplifie, après quelques manipulations algébriques similaires à celles explicitées précédemment dans de pareils cas, à

$$\tilde{G}_i(\mathbf{x}, \mathbf{y}, \mathbf{u}) = \sum_{\mathbf{k}=0}^{\infty} P_i(\mathbf{k}) \prod_{l=1}^M [(1-T_{il})(1-T_{li}) \\ + T_{li}(1-T_{il})x_l + T_{il}(1-T_{li})y_l + T_{il}T_{li}u_l]^{k_l}. \quad (4.21)$$

Cette expression nous permet de calculer directement les premiers moments pour les différents types de liens

$$\tilde{z}_{ij}^{\text{in}} = \sum_{\mathbf{a}, \mathbf{b}, \mathbf{m}=0}^{\infty} a_j P_i(\mathbf{a}, \mathbf{b}, \mathbf{m}) = \frac{d}{dx_j} G_i(\mathbf{1}, \mathbf{1}, \mathbf{1}; \mathbf{T}) = T_{ji}(1-T_{ij})z_{ij} \\ \tilde{z}_{ij}^{\text{out}} = \sum_{\mathbf{a}, \mathbf{b}, \mathbf{m}=0}^{\infty} b_j P_i(\mathbf{a}, \mathbf{b}, \mathbf{m}) = \frac{d}{dy_j} G_i(\mathbf{1}, \mathbf{1}, \mathbf{1}; \mathbf{T}) = T_{ij}(1-T_{ji})z_{ij} \\ \tilde{z}_{ij}^{\text{un}} = \sum_{\mathbf{a}, \mathbf{b}, \mathbf{m}=0}^{\infty} m_j P_i(\mathbf{a}, \mathbf{b}, \mathbf{m}) = \frac{d}{du_j} G_i(\mathbf{1}, \mathbf{1}, \mathbf{1}; \mathbf{T}) = T_{ji}T_{ij}z_{ij}$$

où z_{ij} est le nombre moyen de voisins tel que défini à l'équation (3.1). Étant donné la forme particulière de (4.21), nous voyons que les fonctions $\tilde{F}_{ij}^{\text{in}}(\mathbf{x}, \mathbf{y}, \mathbf{u})$, $\tilde{F}_{ij}^{\text{out}}(\mathbf{x}, \mathbf{y}, \mathbf{u})$ et $\tilde{F}_{ij}^{\text{un}}(\mathbf{x}, \mathbf{y}, \mathbf{u})$ définies par (4.6), (4.7) et (4.8) se fondent en une seule et même fonction

$$\tilde{F}_{ij}^{\text{in}}(\mathbf{x}, \mathbf{y}, \mathbf{u}) = \tilde{F}_{ij}^{\text{out}}(\mathbf{x}, \mathbf{y}, \mathbf{u}) = \tilde{F}_{ij}^{\text{un}}(\mathbf{x}, \mathbf{y}, \mathbf{u}) \equiv \tilde{F}_{ij}(\mathbf{x}, \mathbf{y}, \mathbf{u}) \quad (4.22)$$

où

$$\tilde{F}_{ij}(\mathbf{x}, \mathbf{y}, \mathbf{u}) = \frac{1}{z_{ji}} \sum_{\mathbf{k}=0}^{\infty} k_i P_j(\mathbf{k}) \prod_{l=1}^M [(1-T_{jl})(1-T_{lj}) \\ + T_{jl}(1-T_{lj})x_l + T_{jl}T_{lj}u_l]^{k_l-\delta_{il}}. \quad (4.23)$$

Comme nous le verrons à la section suivante, cette façon d'introduire la transmissibilité dans le formalisme semi-dirigé a d'importantes répercussions sur la suite des choses, plus particulièrement sur les équations permettant de quantifier la composante géante.

4.2.2 Composante géante (probabilité et taille)

Le fait que $\tilde{F}_{ij}^{\text{in}}(\mathbf{x}, \mathbf{y}, \mathbf{u})$, $\tilde{F}_{ij}^{\text{out}}(\mathbf{x}, \mathbf{y}, \mathbf{u})$ et $\tilde{F}_{ij}^{\text{un}}(\mathbf{x}, \mathbf{y}, \mathbf{u})$ soient équivalentes lorsque la transmissibilité est insérée dans le formalisme comme à la section précédente simplifie grandement les équations (4.15)–(4.20). Nous verrons notamment que certaines quantités associées aux composantes naturelles, que nous notons à l'aide d'une flèche de gauche à droite (\rightarrow), sont analogues aux quantités notées de façon similaire à la §3.3 et sont équivalentes *en-dessous* du seuil de percolation. Toutefois, puisqu'au-dessus du seuil de percolation il est nécessaire de considérer à la fois les composantes naturelles et celles *à rebours*, nous continuerons à les distinguer l'une de l'autre à l'aide de flèches.

Calcul de \mathcal{P}

Considérant (4.22), (4.15) et (4.16) deviennent

$$\vec{h}_{ij}^{\text{in}} = \tilde{F}_{ij}(\mathbf{1}, \vec{\mathbf{h}}_j^{\text{in}}, \vec{\mathbf{h}}_j^{\text{un}}; \mathbf{T}), \quad \vec{h}_{ij}^{\text{un}} = \tilde{F}_{ij}(\mathbf{1}, \vec{\mathbf{h}}_j^{\text{in}}, \vec{\mathbf{h}}_j^{\text{un}}; \mathbf{T})$$

i.e. une seule et même équation impliquant l'égalité $\vec{h}_{ij}^{\text{in}} = \vec{h}_{ij}^{\text{un}} \equiv \vec{h}_{ij}$ où \vec{h}_{ij} est la probabilité qu'un lien $i \rightarrow j$ occupé ne mène pas à la composante géante. En utilisant (4.23), nous voyons que \vec{h}_{ij} s'obtient en solutionnant

$$\vec{h}_{ij} = \frac{1}{z_{ji}} \sum_{\mathbf{k}=0}^{\infty} k_i P_j(\mathbf{k}) \prod_{l=1}^M [1 + (\vec{h}_{jl} - 1) T_{jl}]^{k_l - \delta_{il}}. \quad (4.24)$$

Une fois \vec{h}_{ij} obtenue, la probabilité d'une épidémie s'écrit directement

$$\mathcal{P} = 1 - \sum_{j=1}^M w_j \sum_{\mathbf{k}=0}^{\infty} P_j(\mathbf{k}) \prod_{l=1}^M [1 + (\vec{h}_{jl} - 1) T_{jl}]^{k_l} \quad (4.25)$$

à l'aide de (4.17). Comparant (4.24) à (3.8) et (4.25) à (3.5), nous voyons qu'il est possible de réécrire les 2 équations précédentes en termes des quantités de la §3.3

$$\begin{aligned} \vec{h}_{ij} &= F_{ij}(\vec{\mathbf{h}}_j; \mathbf{T}) \\ \mathcal{P} &= 1 - \sum_{i=1}^M w_i G_i(\vec{\mathbf{h}}_i; \mathbf{T}). \end{aligned}$$

Calcul de \mathcal{S}

De la même façon, (4.22) implique que (4.18) et (4.19) deviennent

$$\overleftarrow{h}_{ij}^{\text{out}} = \tilde{F}_{ij}(\overleftarrow{\mathbf{h}}_j^{\text{out}}, \mathbf{1}, \overleftarrow{\mathbf{h}}_j^{\text{un}}; \mathbf{T}), \quad \overleftarrow{h}_{ij}^{\text{un}} = \tilde{F}_{ij}(\overleftarrow{\mathbf{h}}_j^{\text{out}}, \mathbf{1}, \overleftarrow{\mathbf{h}}_j^{\text{un}}; \mathbf{T})$$

et donc que $\overleftarrow{h}_{ij}^{\text{out}} = \overleftarrow{h}_{ij}^{\text{un}} \equiv \overleftarrow{h}_{ij}$, où \overleftarrow{h}_{ij} est obtenu de

$$\overleftarrow{h}_{ij} = \frac{1}{z_{ji}} \sum_{\mathbf{k}=0}^{\infty} k_i P_j(\mathbf{k}) \prod_{l=1}^M [1 + (\overleftarrow{h}_{jl} - 1) T_{lj}]^{k_l - \delta_{il}}. \quad (4.26)$$

Avec \overleftarrow{h}_{ij} , nous obtenons directement la taille relative d'une épidémie

$$\mathcal{S} = 1 - \sum_{j=1}^M w_j \sum_{\mathbf{k}=0}^{\infty} P_j(\mathbf{k}) \prod_{l=1}^M [1 + (\overleftarrow{h}_{jl} - 1) T_{lj}]^{k_l} \quad (4.27)$$

à l'aide de (4.20). Ici aussi, il est possible de réécrire (4.26) et (4.27) en fonction des quantités de la §3.3, à savoir

$$\begin{aligned} \overleftarrow{h}_{ij} &= F_{ij}(\overleftarrow{\mathbf{h}}_j; \mathbf{T}^\dagger) \\ \mathcal{S} &= 1 - \sum_{i=1}^M w_i G_i(\overleftarrow{\mathbf{h}}_i; \mathbf{T}^\dagger). \end{aligned} \quad (4.28)$$

On remarque deux choses en observant attentivement (4.25) et (4.27). D'une part, (4.25) correspond simplement à la perte de normalisation de (3.10). Ceci nous montre que nous aurions fait fausse route en généralisant directement le modèle de Newman à plusieurs types de noeuds puisque que nous aurions obtenu une expression pour \mathcal{P} sans toutefois en avoir une pour \mathcal{S} puisque typiquement $\mathcal{P} \neq \mathcal{S}$.

D'autre part, (4.25) et (4.27) ne diffèrent en fait que par l'ordre des indices du terme de la transmissibilité⁴. Ceci montre que la différence quantitative entre \mathcal{P} et \mathcal{S} réside strictement dans le fait que \mathbf{T} n'est pas symétrique confirmant explicitement notre hypothèse de départ. Ainsi, lorsque $\mathbf{T} = \mathbf{T}^\dagger$, la taille de la composante géante correspond aussi à la probabilité d'avoir une épidémie à grande échelle comme c'est le cas pour le modèle de Newman où \mathbf{T} est manifestement symétrique.

4.2.3 Distribution du nombre d'infectés

En présence d'une composante géante, il est encore possible de déterminer la distribution du nombre d'infectés. Celle-ci est toujours générée par $K(\mathbf{x}; \mathbf{T})$, qui s'obtient en solutionnant⁵ (3.9) et (3.10), mais qui doit être renormalisée puisque $\mathcal{P} \neq 0$

$$\frac{K(\mathbf{x}; \mathbf{T})}{1 - \mathcal{P}}. \quad (4.29)$$

⁴Il en va de même pour (4.24) et (4.26).

⁵Notons que (4.22) implique que $\overrightarrow{H}_{ij}^{\text{in}}(\mathbf{x}) = \overrightarrow{H}_{ij}^{\text{un}}(\mathbf{x}) = H_{ij}(\mathbf{x})$ et donc que $\overrightarrow{K}(\mathbf{x}) = K(\mathbf{x})$. Ainsi, résoudre (4.9)–(4.11) est équivalent à résoudre (3.9) et (3.10).

Il est toutefois possible que, pour un problème donné, seuls les premiers moments ne soient requis, rendant ainsi la résolution de (3.9) et (3.10) — ce qui pour de grandes valeurs de M et/ou de larges distributions de degrés peut nécessiter de longs calculs numériques — tout à fait superflue. En suivant une démarche similaire à celle de la section 3.3.4, il est possible d'obtenir des équations de récurrence simples permettant d'obtenir aisément le nombre moyen d'infectés. À l'aide de (4.29), nous avons directement que

$$\langle \vec{s}_i \rangle = \frac{1}{K(\mathbf{1}; \mathbf{T})} \left. \frac{dK(\mathbf{x}; \mathbf{T})}{dx_i} \right|_{\mathbf{x}=\mathbf{1}} = \frac{1}{1 - \mathcal{P}} \left. \frac{dK(\mathbf{x}; \mathbf{T})}{dx_i} \right|_{\mathbf{x}=\mathbf{1}}$$

En évaluant (3.11) à $\mathbf{x} = \mathbf{1}$, nous obtenons en posant $\vec{\alpha}_{ij}^{(l)} \equiv \left. \frac{dH_{ij}(\mathbf{x})}{dx_l} \right|_{\mathbf{x}=\mathbf{1}}$

$$\left. \frac{dK(\mathbf{x}; \mathbf{T})}{dx_i} \right|_{\mathbf{x}=\mathbf{1}} = w_i G_i(\vec{\mathbf{h}}_i; \mathbf{T}) + \sum_{l=1}^M w_l \sum_{j=1}^M T_{lj} z_{lj} \vec{h}_{jl} \vec{\alpha}_{lj}^{(i)}$$

puisque

$$\begin{aligned} \left. \frac{\partial G_l(\mathbf{H}_l(\mathbf{x}; \mathbf{T}); \mathbf{T})}{\partial H_{lj}(\mathbf{x}; \mathbf{T})} \right|_{\mathbf{x}=\mathbf{1}} &= \left. \frac{dG_l(\mathbf{x}; \mathbf{T})}{dx_j} \right|_{\mathbf{x}=\vec{\mathbf{h}}_l} \\ &= T_{lj} \sum_{k=0}^{\infty} k_j P_l(\mathbf{k}) \prod_{n=1}^M [1 + (\vec{h}_{ln} - 1) T_{ln}]^{k_n - \delta_{nj}} \\ &= T_{lj} z_{lj} \vec{h}_{jl} \end{aligned}$$

en utilisant (4.24). Ainsi, nous avons l'expression généralisée pour le nombre moyen d'infectés

$$\langle \vec{s}_i \rangle = \frac{w_i G_i(\vec{\mathbf{h}}_i; \mathbf{T})}{1 - \mathcal{P}} + \frac{1}{1 - \mathcal{P}} \sum_{l=1}^M w_l \sum_{j=1}^M T_{lj} z_{lj} \vec{h}_{jl} \vec{\alpha}_{lj}^{(i)}. \quad (4.30)$$

Nous interprétons $\vec{\alpha}_{ij}^{(l)}$ comme étant simplement le nombre moyen de noeuds de type l présents dans la composante atteinte par un lien de type $i \rightarrow j$ et $w_i G_i(\vec{\mathbf{h}}_i; \mathbf{T})$ comme étant la probabilité que le patient zéro soit de type i et ne provoque aucune infection secondaire. Tout comme à la §3.3.4, l'équation de récurrence associée aux $\vec{\alpha}_{ij}^{(l)}$ est obtenue en dérivant (3.9) par rapport à x_l

$$\frac{dH_{ij}(\mathbf{x}; \mathbf{T})}{dx_l} = F_{ij}(\mathbf{H}_j(\mathbf{x}; \mathbf{T}); \mathbf{T}) \delta_{jl} + x_j \sum_{n=1}^M \frac{\partial F_{ij}(\mathbf{H}_j(\mathbf{x}; \mathbf{T}); \mathbf{T})}{\partial H_{jn}(\mathbf{x}; \mathbf{T})} \frac{dH_{jn}(\mathbf{x}; \mathbf{T})}{dx_l}$$

et en évaluant à $\mathbf{x} = \mathbf{1}$

$$\vec{\alpha}_{ij}^{(l)} = F_{ij}(\vec{\mathbf{h}}_j; \mathbf{T}) \delta_{jl} + \sum_{n=1}^M T_{jn} \vec{\beta}_{ij}^{(n)} \vec{\alpha}_{jn}^{(l)}. \quad (4.31)$$

Afin d'alléger la notation, nous avons noté dans (4.31)

$$\begin{aligned} \left. \frac{\partial F_{ij}(\mathbf{H}_j(\mathbf{x}; \mathbf{T}); \mathbf{T})}{\partial H_{jn}(\mathbf{x}; \mathbf{T})} \right|_{\mathbf{x}=\mathbf{1}} &= \left. \frac{1}{T_{ji}z_{ji}} \frac{d^2 G_j(\mathbf{x}; \mathbf{T})}{dx_i dx_n} \right|_{\mathbf{x}=\vec{h}_j} \\ &= T_{jn} \frac{1}{z_{ji}} \sum_{\mathbf{k}=0}^{\infty} k_i (k_n - \delta_{in}) P_j(\mathbf{k}) \prod_{l=1}^M [1 + (\vec{h}_{jl} - 1) T_{jl}]^{k_l - \delta_{il} - \delta_{ln}} \\ &= T_{jn} \vec{\beta}_{ij}^{(n)} \end{aligned}$$

où $\vec{\beta}_{ij}^{(n)}$ n'est fonction que de $P(\mathbf{k})$ et de \vec{h}_{ij}

$$\vec{\beta}_{ij}^{(n)} = \frac{1}{z_{ji}} \sum_{\mathbf{k}=0}^{\infty} k_i (k_n - \delta_{in}) P_j(\mathbf{k}) \prod_{l=1}^M [1 + (\vec{h}_{jl} - 1) T_{jl}]^{k_l - \delta_{il} - \delta_{ln}} \quad (4.32)$$

et est donc connu. Cette quantité correspond simplement au nombre moyen de noeuds de type n qui seront infectés directement par un noeud de type j , lui-même infecté par un noeud de type i (*i.e.* degré occupé excédant moyen).

Il est donc possible d'obtenir le nombre moyen d'infectés en présence d'une composante géante sans devoir obtenir préalablement la distribution de probabilité générée par $K(\mathbf{x}; \mathbf{T})$, ce qui peut correspondre à une importante économie de temps de calcul⁶. De plus, soulignons que (4.30)–(4.32) sont les versions généralisées à toutes valeurs de \mathbf{T} de (3.12)–(3.14). Ainsi, $\langle \vec{s}_i \rangle$, $\vec{\alpha}_{ij}^{(l)}$ et $\vec{\beta}_{ij}^{(n)}$ retombent sur leur quantité associée — $\langle s_i \rangle$, $\alpha_{ij}^{(l)}$ et $\beta_{ij}^{(n)}$ — sous le seuil de percolation puisqu'alors $\vec{h}_{ij} = 1$ et $\mathcal{P} = 0$.

4.2.4 Composition de la composante géante

À la §4.2.2, nous avons interprété \overleftarrow{h}_{ij} comme étant la probabilité qu'un lien de type $i \rightarrow j$ ne mène pas à la composante géante. De façon analogue, nous pouvons interpréter \overleftarrow{h}_{ij} comme étant la probabilité qu'un noeud de type i ne soit pas infecté par un noeud de type j lors d'une épidémie⁷. Ainsi, $G_i(\overleftarrow{\mathbf{h}}_i; \mathbf{T}^\dagger)$ nous donne la probabilité qu'un noeud choisi au hasard parmi tous les noeuds de type i ne fasse pas partie de la composante géante (*i.e.* de $\text{CGC} \cup \text{CGS}$). En pondérant par w_i , nous obtenons directement la partie de la composante géante occupée par des noeuds de type i

$$\mathcal{S}_i = w_i [1 - G_i(\overleftarrow{\mathbf{h}}_i; \mathbf{T}^\dagger)]. \quad (4.33)$$

⁶Il serait également possible d'obtenir une équation de récurrence pour les deuxièmes moments de $K(\mathbf{x}; \mathbf{T})$ évalués à $\mathbf{x} = \mathbf{1}$ et ainsi, à l'aide de (2.6), obtenir la variance associée.

⁷Autrement dit, \overleftarrow{h}_{ij} est la probabilité qu'un lien de type $j \rightarrow i$ ne lie pas le noeud de type i à la composante géante. Ce lien pourrait transmettre la maladie mais uniquement lorsque celle-ci est limitée à une composante de taille finie.

Ceci est cohérent avec l'expression (4.28), considérant que $\sum_{i=1}^M \mathcal{S}_i = \mathcal{S}$.

4.3 Validation

Il est coutume en théorie des réseaux de confronter les prédictions d'un formalisme nouvellement introduit aux résultats obtenus à l'aide de simulations numériques. Typiquement, ces simulations consistent en la génération de réseaux répondant aux critères prescrits par le modèle en question et d'en mesurer les différentes propriétés pour fins de comparaison. Nous présentons donc en détail les paramètres des réseaux qui ont été utilisés et confirmons les prédictions théoriques de notre modèle en les comparant aux résultats des simulations numériques.

4.3.1 Simulations numériques

Dans la §1.2.1, il a été mentionné que la plupart des réseaux réels possèdent une distribution de degrés très asymétrique que plusieurs chercheurs ont associée à une loi de puissance. Nous avons donc choisi pour illustrer et valider notre formalisme un ensemble de réseaux dont la distribution de degrés entre les individus est prescrite par une loi de puissance que nous avons atténuée à l'aide d'une exponentielle⁸. Comme nous ne connaissions pas de tendance générale concernant la distribution des liens entre les différentes catégories de noeuds, nous avons opté pour une simple distribution aléatoire indépendante du degré total du noeud ; la probabilité qu'un lien d'un noeud de type i soit relié à un noeud de type j étant donnée par p_{ij} . Ainsi, nous avons utilisé⁹

$$P_i(k_1, k_2) = \frac{(k_1 + k_2)^{-\alpha_i} e^{-(k_1+k_2)/\kappa_i}}{\text{Li}_{\alpha_i}(e^{-1/\kappa_i})} \cdot \binom{k_1 + k_2}{k_1} p_{i1}^{k_1} p_{i2}^{k_2} \quad (4.34)$$

avec les paramètres

$$\boldsymbol{\alpha} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}; \quad \boldsymbol{\kappa} = \begin{bmatrix} 8 \\ 10 \end{bmatrix}; \quad \boldsymbol{p} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}.$$

⁸La loi de puissance amortie d'une exponentielle est une distribution théorique fréquemment utilisée pour valider les modèles, notamment puisque la présence de cette coupure assure l'existence d'un seuil épidémique (ce qui n'est pas nécessairement le cas pour la loi de puissance pure).

⁹Afin de préserver la normalisation, nous avons imposé la contrainte supplémentaire $\sum_{j=1}^M p_{ij} = 1 \forall i$.

Le facteur de normalisation de (4.34), $\text{Li}_\alpha(z)$, correspond au α ème polylogarithme de z [2] et est défini comme suit

$$\text{Li}_\alpha(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^\alpha}.$$

qui possède la propriété suivante

$$\frac{\partial \text{Li}_n(z)}{\partial z} = \frac{1}{z} \text{Li}_{n-1}(z). \quad (4.35)$$

Puisque l'introduction d'une transmissibilité hétérogène induit généralement une asymétrie entre la probabilité d'une épidémie et sa taille relative (*i.e.* $\mathcal{P} \neq \mathcal{S}$), nous avons utilisé la matrice de transmissibilité suivante

$$\mathbf{T} = \gamma \begin{bmatrix} 0.95 & 0.99 \\ 0.61 & 1.00 \end{bmatrix}$$

où γ est un paramètre permettant de varier la transmissibilité de la maladie tout en préservant cette asymétrie. L'annexe B.2 donne le détail de l'algorithme de simulation.

Expressions analytiques

Nous calculons maintenant les expressions analytiques pour différentes quantités de notre formalisme. Pour $G_i(x_1, x_2; \mathbf{T})$, nous avons

$$\begin{aligned} G_i(x_1, x_2; \mathbf{T}) &= \sum_{k_1+k_2>0}^{\infty} \frac{(k_1+k_2)^{-\alpha_i} e^{-(k_1+k_2)/\kappa_i}}{\text{Li}_{\alpha_i}(e^{-1/\kappa_i})} \cdot \binom{k_1+k_2}{k_1} p_{i1}^{k_1} p_{i2}^{k_2} \prod_{l=1}^2 [1 + (x_l - 1)T_{il}]^{k_l} \\ &= \sum_{k=1}^{\infty} \frac{k^{-\alpha_i} e^{-k/\kappa_i}}{\text{Li}_{\alpha_i}(e^{-1/\kappa_i})} \sum_{k_1=0}^k \binom{k}{k_1} [p_{i1}(1 + (x_1 - 1)T_{i1})]^{k_1} [p_{i2}(1 + (x_2 - 1)T_{i2})]^{k-k_1} \\ &= \frac{\text{Li}_{\alpha_i}(e^{-1/\kappa_i} [p_{i1}(1 + (x_1 - 1)T_{i1}) + p_{i2}(1 + (x_2 - 1)T_{i2})])}{\text{Li}_{\alpha_i}(e^{-1/\kappa_i})}. \end{aligned}$$

En utilisant (4.35), nous obtenons directement

$$\tilde{z}_{ij} = T_{ij} z_{ij} = T_{ij} \frac{p_{ij} \text{Li}_{\alpha_i-1}(e^{-1/\kappa_i})}{\text{Li}_{\alpha_i}(e^{-1/\kappa_i})} \quad (4.36)$$

et

$$F_{ij}(x_1, x_2; \mathbf{T}) = \frac{\text{Li}_{\alpha_j-1}(e^{-1/\kappa_j} [p_{j1}(1 + (x_1 - 1)T_{j1}) + p_{j2}(1 + (x_2 - 1)T_{j2})])}{[p_{j1}(1 + (x_1 - 1)T_{j1}) + p_{j2}(1 + (x_2 - 1)T_{j2})] \text{Li}_{\alpha_j-1}(e^{-1/\kappa_j})}.$$

À l'aide de (3.3), (4.36) et des paramètres numériques donnés ci-haut, nous obtenons

$$\mathbf{z} \simeq \begin{bmatrix} 2.4552 & 1.0522 \\ 0.7170 & 1.0755 \end{bmatrix}; \quad \mathbf{w} \simeq \begin{bmatrix} 0.4053 & 0 \\ 0 & 0.5947 \end{bmatrix}.$$

Finalement, en utilisant (3.14) et (4.36) nous avons que

$$\beta_{ij}^{(l)} = p_{jl} \left(\frac{\text{Li}_{\alpha_j-2}(e^{-1/\kappa_j})}{\text{Li}_{\alpha_j-1}(e^{-1/\kappa_j})} - 1 \right)$$

ce qui permet de résoudre (3.15) et d'obtenir la valeur critique du paramètre γ

$$\gamma_c \simeq 0.1834$$

marquant le seuil de percolation du réseau et donc l'apparition de la composante géante.

Résultats

Lors des simulations numériques, nous avons généré un ensemble de 2000 réseaux de taille relativement grande ($N = 10^5$) respectant la distribution de degrés (4.34) sur lesquels nous avons simulé la propagation d'une maladie. La figure 4.2 compare la distribution du nombre d'infectés pour $\gamma = 0.1$ prédites par (3.10) avec les résultats obtenus des simulations numériques. Nous y constatons une très bonne correspondance entre les deux. Notons que les résultats numériques sont le résultat de processus stochastiques non-corrélés et convergent donc comme l'inverse de la racine carrée du nombre de réalisations effectuées. Pour $\gamma = 0.1$, un total de 3.1×10^8 simulations ont été effectuées. Ainsi, nous nous attendons donc à une convergence jusqu'à une précision grossièrement d'environ $\mathcal{O}(10^{-5})$, ce qui est bel et bien observé. Afin de mieux montrer cet accord entre les prédictions théoriques et les résultats numériques, la figure 4.3 présente les distributions du nombre d'infectés d'un type indépendamment du nombre d'infectés de l'autre type¹⁰ pour $\gamma = 0.1$ et $\gamma = 0.5$. Pour $\gamma = 0.5$, nous avons effectué un total de 1.68×10^7 simulations et nous attendons donc à ce que l'accord soit grossièrement de l'ordre de 10^{-4} . Cette figure permet de constater que notre modèle prédit très précisément le comportement obtenu par simulations pour ce qui est des petites composantes autant en-dessous qu'au-dessus du seuil de percolation. La figure 4.4 compare les valeurs de $\langle s_1 \rangle$, $\langle s_2 \rangle$ et $\langle s \rangle \equiv \langle s_1 \rangle + \langle s_2 \rangle$ pour des valeurs de γ entre 0 et 1 obtenues par simulations à celles prédites par (4.30). Nous y voyons d'une part que le nombre moyen d'infectés diverge bel et bien à la valeur prédite par $\zeta_2 = 0$ et d'autre part que (4.30) prédit

¹⁰Il s'agit simplement de la projection des matrices telles que celles montrées à la figure 4.2 sur l'un ou l'autre d'un de leurs axes, *i.e.* $K(1, x_2; \mathbf{T})$ ou $K(x_1, 1; \mathbf{T})$.

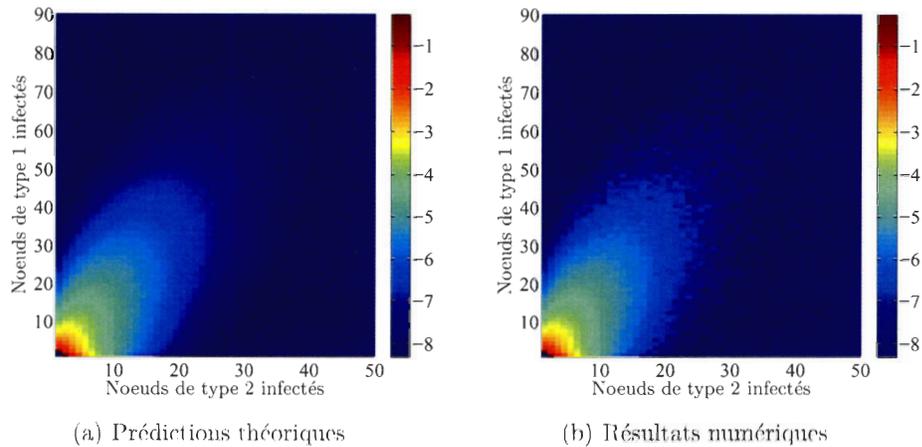


FIG. 4.2 – Distribution de la probabilité jointe du nombre d'individus infectés suite à l'infection d'un individu choisi au hasard dans le réseau. La distribution prédite par (3.10) est montrée à gauche tandis que les résultats des simulations numériques sont montrés à droite. La carte des couleurs correspond à une échelle logarithmique et est identique pour chacune des deux distributions.

correctement le comportement de $\langle s_1 \rangle$, $\langle s_2 \rangle$ et $\langle s \rangle$ sauf à l'approche de γ_c où le formalisme surestime le nombre moyen d'individus infectés. Il s'agit d'un effet dû à la taille finie des réseaux utilisés qui sera discuté à la §4.3.2. Finalement, nous avons mesuré numériquement \mathcal{P} , \mathcal{S}_1 , \mathcal{S}_2 et \mathcal{S} pour plusieurs valeurs de γ entre 0 et 1 afin de vérifier que celles-ci coïncident bel et bien avec les valeurs prédites par (4.25), (4.27) et (4.33). La figure 4.5 montre ici aussi un accord très précis entre les prédictions théoriques et l'expérience sauf à l'approche de γ_c (voir encadré). Il s'agit là aussi d'un effet de la taille finie du système qui sera traité à la §4.3.2.

Autre exemple

Le lecteur attentif aura peut-être remarqué quelque chose de particulier à propos des réseaux utilisés précédemment : $F_{ij}(\mathbf{x}; \mathbf{T})$ est indépendant de i ! En effet, notre choix de distribution de degrés fait en sorte que la distribution de degrés occupés excédants est indépendante du type du noeud par lequel nous sommes arrivés. Ceci simplifie significativement plusieurs équations de notre modèle lorsqu'appliquées à ce cas particulier. Toutefois, cela ne sabote en rien la démonstration de la validité de notre formalisme faite précédemment puisqu'en aucun moment lors de son élaboration nous avons sup-

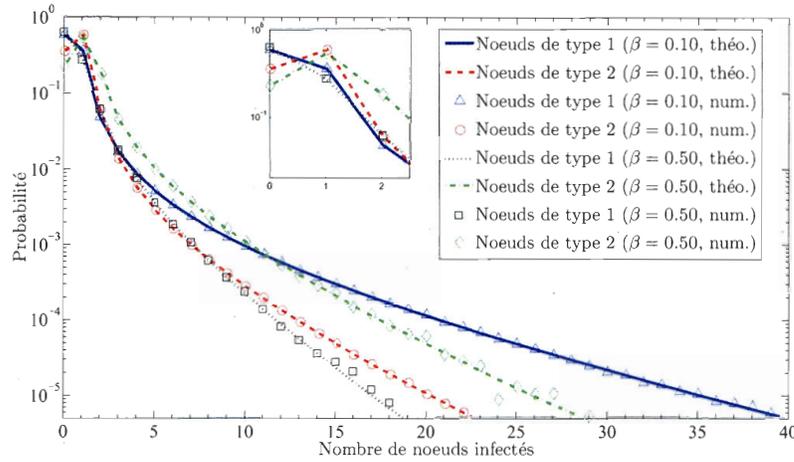


FIG. 4.3 – Distributions de la probabilité du nombre d'infectés d'un type donné indépendamment du nombre d'infectés de l'autre type. Les lignes indiquent les prédictions obtenues en projetant le résultat de (3.10) sur l'un ou l'autre des axes (*i.e.* $K(x_1, 1; \mathbf{T})$ pour les types 1 et $K(1, x_2; \mathbf{T})$ pour les types 2). Les symboles correspondent à ces mêmes quantités mais obtenues à l'aide des simulations numériques. Un agrandissement du début des distributions est montré dans l'encadré.

posé une telle indépendance. Afin d'en persuader le lecteur, nous avons refait certaines de nos simulations mais en utilisant une distribution de degrés qui n'aura pas cette particularité :

$$P_i(k_1, k_2) = \prod_{l=1}^2 \frac{k_l^{-\alpha_{il}} e^{-k_l/\kappa_{il}}}{\text{Li}_{\alpha_{il}}(e^{-1/\kappa_{il}})}$$

avec les paramètres suivant

$$\boldsymbol{\alpha} = \begin{bmatrix} 3 & 2.5 \\ 2 & 1.5 \end{bmatrix}; \quad \boldsymbol{\kappa} = \begin{bmatrix} 0.8 & 0.6 \\ 0.4 & 0.2 \end{bmatrix}; \quad \mathbf{T} = \begin{bmatrix} 0.1 & 0.55 \\ 0.3 & 0.45 \end{bmatrix}.$$

La fonction génératrice associée à la distribution de degrés occupés devient alors

$$G_i(\mathbf{x}; \mathbf{T}) = \prod_{l=1}^2 \frac{\text{Li}_{\alpha_{il}}(e^{-1/\kappa_{il}} [1 + (x_l - 1)T_{il}])}{\text{Li}_{\alpha_{il}}(e^{-1/\kappa_{il}})}$$

d'où on calcule aisément

$$\tilde{z}_{ij} = T_{ij} \prod_{l=1}^2 \frac{\text{Li}_{\alpha_{il} - \delta_{jl}}(e^{-1/\kappa_{il}})}{\text{Li}_{\alpha_{il}}(e^{-1/\kappa_{il}})}.$$

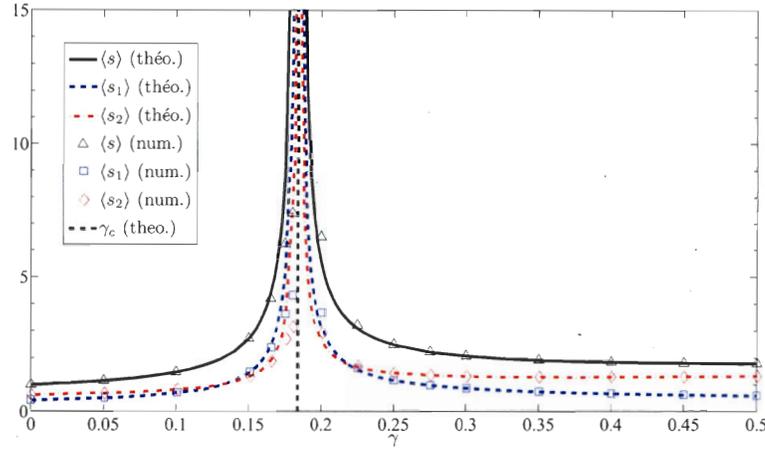


FIG. 4.4 – Comparaison du nombre moyen d’infectés prédit par (3.12) avec les résultats des simulations numériques pour différentes valeurs de γ . Seules les valeurs de γ entre 0 et 0.5 sont montrées pour mettre en évidence le comportement autour du point de percolation mais l’accord avec les prédictions est vérifié pour tout l’intervalle $[0,1]$.

De même, $F_{ij}(\mathbf{x}; \mathbf{T})$ prend alors la forme

$$F_{ij}(\mathbf{x}; \mathbf{T}) = \frac{1}{1 + (x_i - 1)T_{ji}} \prod_{l=1}^2 \frac{\text{Li}_{\alpha_{jl} - \delta_{il}}(e^{-1/\kappa_{jl}} [1 + (x_l - 1)T_{jl}])}{\text{Li}_{\alpha_{jl} - \delta_{il}}(e^{-1/\kappa_{jl}})}$$

ce qui nous permet d’obtenir

$$\beta_{ij}^{(n)} = \prod_{l=1}^2 \frac{\text{Li}_{\alpha_{jl} - \delta_{il} - \delta_{in}}(e^{-1/\kappa_{jl}})}{\text{Li}_{\alpha_{jl} - \delta_{il}}(e^{-1/\kappa_{jl}})} - \delta_{in}$$

d’où il est possible de déterminer à l’aide de (3.15) que les paramètres de cet exemple se situent sous le seuil épidémique. Les simulations ont été faites sur des réseaux de plus petites tailles ($N = 1000$) afin de diminuer le temps de calcul et aussi de montrer que certaines des prédictions de notre formalisme restent valides pour de petits réseaux (voir §4.3.2). La figure 4.6 montre les distributions du nombre d’infectés de chacun des types (indépendamment du nombre d’infectés de l’autre type) obtenues par simulations et les compare à celles prédites par (3.10). Près de 7×10^8 simulations ont été effectuées : les résultats numériques devraient donc être précis jusqu’à grossièrement environ $\mathcal{O}(10^{-5})$. Nous y constatons un excellent accord avec les prédictions de notre modèle.

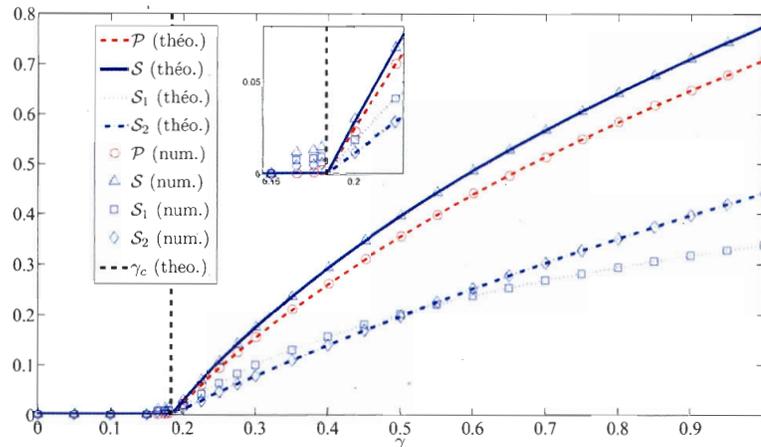


FIG. 4.5 – Comparaison des valeurs de \mathcal{P} , \mathcal{S} , \mathcal{S}_1 et \mathcal{S}_2 prédites par (4.25), (4.27) et (4.33) avec les résultats des simulations numériques pour différentes valeurs de γ . L'encadré est un agrandissement du graphique autour du point de percolation.

4.3.2 La *taille* de l'infini

Nous avons vu à maintes reprises qu'il fut fort utile de considérer des réseaux de taille infinie pour développer notre formalisme. Il est toutefois clair qu'aucun réseau réel n'aura une telle taille. Ainsi, nous pouvons nous interroger sur le régime d'application de notre formalisme, *i.e.* à partir de quelle taille les effets de la taille finie ne seront-ils plus négligeables? La réponse à cette question varie selon les quantités considérées.

Seuil de percolation : Noël *et al.* [57] ont démontré que la taille finie fait disparaître la singularité au point de percolation. Ainsi, cette valeur correspond encore au point où une composante commence à occuper une fraction importante du réseau, mais cette transition n'est plus soudaine et se fait de façon continue.

Petites composantes : dans un régime suffisamment loin du seuil de percolation, les petites composantes ont une taille finie généralement petite par rapport à la taille du réseau et notre formalisme prédit correctement leur taille (voir la figure 4.7 où $N = 10^3$). Toutefois, autour du seuil, les petites composantes dont la taille moyenne devrait alors diverger seront limitées par la taille finie du réseau (voir figure 4.4 où $N = 10^5$) et notre formalisme surestime alors leur taille.

Composante géante : le fait que le réseau ait une taille finie implique que la composante géante, formellement de taille infinie, est aussi de taille finie. Ainsi, autour du seuil de percolation, certaines petites composantes auront une taille

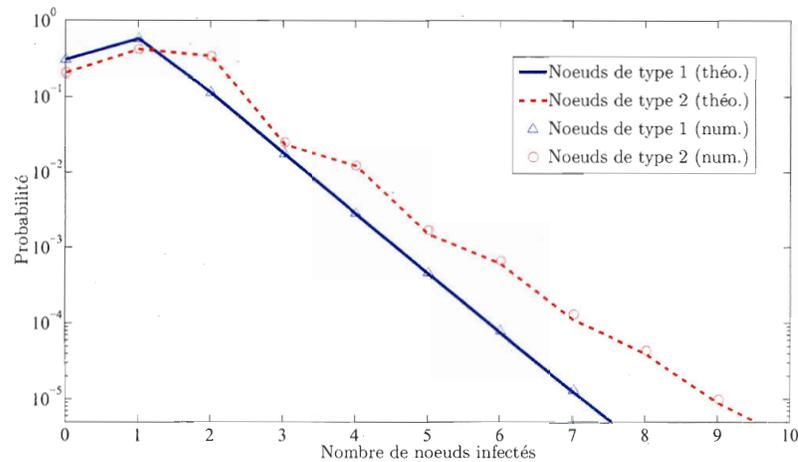


FIG. 4.6 – Comparaison des distributions de la probabilité du nombre d'infectés d'un type donné indépendamment du nombre d'infectés de l'autre type aux résultats de simulations numériques effectuées sur des réseaux de 1000 noeuds.

similaire à celle prédite pour la composante géante et il sera difficile de distinguer ces deux quantités l'une de l'autre. Ceci aura pour effet que l'analyse des résultats numériques surestimera les véritables valeurs comme à la figure 4.5. De plus, au fur et à mesure que la transmissibilité augmente, la progression de la maladie est limitée par la taille finie puisque des liens qui devraient infecter de nouveaux noeuds dans un réseau infini mèneront, dans un réseau fini, vers des noeuds déjà infectés qui ne peuvent être infectés plus d'une fois. Ainsi, notre formalisme surestimera les quantités associées à la composante géante. La figure 4.7 illustre cet effet pour un réseau similaire à celui utilisé pour les figures 4.2–4.5 mais où les simulations ont été réalisées sur des réseaux de 1000 noeuds. Malgré tout, l'accord avec le formalisme infini demeure tout à fait acceptable.

4.3.3 Cas particulier

Finalement, nous montrons que notre formalisme retombe sur les expressions théoriques connues pour les réseaux bipartis [44, 50]. Rappelons que ces réseaux possèdent deux types de noeuds mais que les liens ne peuvent exister qu'entre deux types différents. Ainsi, nous imposons la contrainte suivante

$$P_i(k_1, k_2) = 0 \quad \forall k_i \neq 0 \quad (4.37)$$

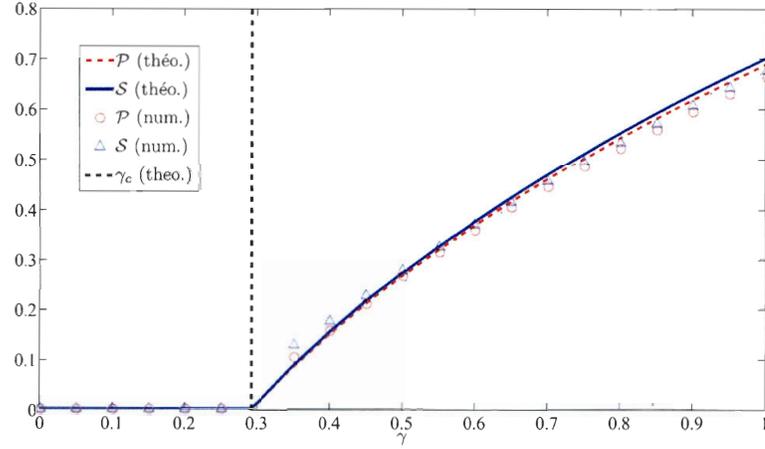


FIG. 4.7 – Illustration des différences entre les prédictions pour \mathcal{P} et \mathcal{S} et les résultats numériques dues à la taille finie des réseaux.

qui implique que

$$z_{ii} = 0; \quad T_{ii} = 0; \quad \beta_{ii}^{(j)} = 0; \quad \beta_{ji}^{(i)} = 0; \quad \alpha_{ii}^{(j)} = 0.$$

et que $F_{ii}(x_1, x_2; \mathbf{T})$ et $H_{ii}(x_1, x_2; \mathbf{T})$ ne sont pas définies puisque ces fonctions considèrent une situation (*i.e.* un lien $i \rightarrow i$) impossible dans les réseaux bipartis¹¹. À l'aide de (3.15), nous obtenons directement le critère de percolation des réseaux bipartis

$$T_{12}T_{21}\beta_{12}^{(1)}\beta_{21}^{(2)} = 1$$

ce qui correspond à l'équation (2) de [44] et l'équation (54) de [50]. Le formalisme utilisé par ces deux auteurs s'intéressent au nombre moyen d'infectés considérant que le patient zéro est d'un type donné ce qui nécessite donc de modifier légèrement (3.10)

$$K_i(x_1, x_2; \mathbf{T}) = x_i G_i(H_{i1}(x_1, x_2; \mathbf{T}), H_{i2}(x_1, x_2; \mathbf{T}); \mathbf{T}) \quad (4.38)$$

pour obtenir la fonction générant la distribution du nombre d'infectés causés par un patient zéro de type i . Le nombre moyen de noeuds de type 1 ainsi infectés est donné par

$$\langle s_1 \rangle = 1 + T_{12}z_{12}\alpha_{12}^{(1)}$$

¹¹Notre formalisme reste toutefois solide face à cette indétermination puisque le coefficient devant l'un ou l'autre de ces objets sera toujours nul lorsqu'utilisé dans $G_i(x_1, x_2; \mathbf{T})$ ou $F_{ij}(x_1, x_2; \mathbf{T})$ étant donné (4.37).

où $\alpha_{12}^{(1)}$ est solution de

$$\begin{aligned}\alpha_{12}^{(1)} &= T_{21}\beta_{12}^{(1)}\alpha_{21}^{(1)} \\ \alpha_{21}^{(1)} &= 1 + T_{12}\beta_{21}^{(2)}\alpha_{12}^{(1)}\end{aligned}$$

obtenues de (3.13). Ainsi, nous obtenons

$$\langle s_1 \rangle = 1 + \frac{T_{12}T_{21}z_{12}\beta_{12}^{(1)}}{1 - T_{12}T_{21}\beta_{12}^{(1)}\beta_{21}^{(2)}}$$

ce qui correspond à l'équation (1) de [44] et à l'équation (53) de [50]. Il est également possible de retrouver les équations pour la composante géante obtenues dans [44] en posant $x_2 = 1$ dans (3.9) et (4.38) et en considérant les contraintes citées ci-haut.

Chapitre 5

Conclusions et perspectives

La modélisation mathématique est de plus en plus utilisée en épidémiologie comme un outil supplémentaire d'aide à la décision en matière de santé publique. En permettant la simulation de la propagation de maladies infectieuses dans une population, elle permet l'analyse systématique de différentes stratégies pour contrer ou prévenir des épidémies de maladies infectieuses en fonction de différentes contraintes. En retour, ceci permet aux autorités concernées d'optimiser leurs plans d'urgence ou leurs programmes de prévention de manière à minimiser les coûts humains et économiques entraînés par de telles épidémies. Parmi les différentes approches, l'épidémiologie par réseaux de contacts est un modèle analytique très puissant qui, en plus de prédire et quantifier l'issue finale d'une épidémie, offre une meilleure compréhension de la dynamique sous-jacente à la propagation de maladies infectieuses via la perspective des réseaux de contacts. Étant à ses premiers balbutiements, cette approche est encore basée sur des hypothèses simplificatrices qui limitent les catégories de maladies infectieuses pouvant être adéquatement modélisées. Il est donc nécessaire que des efforts soient déployés pour complexifier les modèles afin de les rendre toujours plus réalistes.

Pour permettre l'élaboration des premiers formalismes, il a été supposé que la probabilité de transmission de la maladie puisse varier d'une paire d'individus à l'autre selon une distribution donnée, en autant qu'il n'y ait aucune corrélation entre ces différentes valeurs (hypothèse *iid*). Cette hypothèse a pour conséquences qu'au niveau de la population, l'état final de l'épidémie est équivalent à celui où la probabilité de transmission est, pour chaque paire d'individus, égale à la moyenne de cette même distribution. Cette équivalence permet l'élaboration des modèles de percolation actuels où la probabilité de transmission n'est prescrite que par cette seule valeur moyenne nommée la *transmissibilité*. Or, cette hypothèse restreint le nombre de maladies infectieuses pouvant être adéquatement modélisées. En effet, les caractères génétiques ou comportementaux

des individus influencent significativement la probabilité de transmission de plusieurs maladies infectieuses. Il est donc nécessaire que les futurs modèles prennent en compte ces corrélations pour être davantage réalistes.

En réponse à cette problématique, nous avons développé un nouveau formalisme¹ qui permet de distinguer les noeuds du réseau en différentes catégories. Cet ajout qualitatif permet de connaître *qui est qui* et donc *qui infecte qui*. Ainsi, en séparant la population en catégories selon les divers caractères influençant la probabilité de transmission, il nous est possible d'utiliser la transmissibilité appropriée et donc de tenir compte des corrélations présentées ci-haut. Notons que la probabilité de transmission peut varier entre les différentes paires d'individus d'un même type de transmission (*i.e.* d'un noeud de type i à un noeud d'un type j) mais que celle-ci doit varier selon les restrictions de l'hypothèse *iid*. En d'autres termes, nous avons élaboré un formalisme permettant de *factoriser* la population en différentes catégories de sorte que l'hypothèse *iid*, nécessaire à l'élaboration des modèles de percolation, puisse être utilisée.

L'incorporation de l'hétérogénéité de la transmissibilité dans les modèles épidémiologiques est une quête relativement récente. Quelques modèles ont été suggérés mais aucun n'a encore su faire consensus au sein de la communauté scientifique. L'approche que nous présentons permet d'obtenir plusieurs quantités d'intérêt épidémiologique courantes : la distribution du nombre d'individus infectés (générée par $K(\mathbf{x}; \mathbf{T})$) et les moments associés, le seuil épidémique (donné par $\zeta_M = 0$), la probabilité qu'il y ait une épidémie (\mathcal{P}) de même que sa composition (\mathcal{S}_i). De plus, notre modèle possède l'avantage que l'hétérogénéité de la transmissibilité y est explicite. Cela permet d'une part de voir directement son effet sur la dynamique de la propagation², mais aussi d'offrir des prédictions plus détaillées. En effet, le fait de distinguer les noeuds en différentes catégories permet de connaître le nombre d'individus infectés pour chacun des types. Il en résulte la possibilité d'effectuer des analyses épidémiologiques beaucoup plus précises et adaptées aux besoins réels en termes de santé publique³.

Par contre, bien que notre formalisme ait été développé pour un nombre arbitraire de

¹Rappelons que bien que l'épidémiologie ait été la motivation première de ce travail, le formalisme que nous avons développé est un modèle de percolation où la probabilité d'occupation des liens est hétérogène. À cet égard, notre modèle peut être utilisé dans une foule d'autres contextes.

²En effet, l'explicitation de l'hétérogénéité de la transmissibilité permet une interprétation en termes de réseaux semi-dirigés de la percolation de liens de réseaux complexes initialement non-dirigés. Ceci offre une vision plus générale du concept de composante géante qui permet, entre autres, de mettre en évidence et d'expliquer le fait que $\mathcal{S} \neq \mathcal{P}$.

³Par exemple, notre approche pourrait permettre d'étudier l'effet d'une méthode de prévention déployée dans l'ensemble de la population sur les travailleurs de la santé. Il serait également possible d'étudier l'effet de la vaccination contre l'influenza des étudiants d'âge primaire sur l'incidence de la maladie chez les personnes âgées.

catégories (M), le temps de calcul nécessaire à l'obtention de la distribution du nombre d'individus infectés ($K(\mathbf{x}; \mathbf{T})$) peut devenir considérable. Celui-ci croît très rapidement avec le nombre de type de noeuds, ce qui peut rendre notre formalisme peu pratique lorsqu'il y a plusieurs types de noeuds ou lorsque les distributions de degrés sont larges⁴. Pour contourner ce problème, nous avons développé des équations de récurrence qui permettent de calculer rapidement les valeurs moyennes⁵ de cette distribution sans l'avoir préalablement obtenue. Il est donc possible d'obtenir en quelques minutes suffisamment d'information (*i.e.* $\langle s_i \rangle$, \mathcal{P} et \mathcal{S}_i) pour sonder l'issue d'une situation épidémiologique donnée. Lorsque les paramètres optimaux sont ainsi établis, rien n'empêche alors de lancer le calcul pour obtenir $K(\mathbf{x}; \mathbf{T})$ et ainsi avoir encore plus d'information. Bien entendu, celle-ci est obtenue dans une fraction du temps qu'aurait nécessité son obtention à l'aide de simulations numériques massives.

Les réseaux de contacts réels sont des structures très complexes dont les connexions évoluent dans le temps et sont établies selon un ensemble de règles arbitraires. Ainsi, bien que nous ayons démontré l'exactitude *théorique* de notre formalisme en comparant avantageusement ses prédictions aux résultats de simulations numériques, il est important de rappeler que notre modèle ne reste qu'une *approximation* de la réalité. Les réseaux de contacts réels possèdent une foule de propriétés topologiques dont certaines influencent significativement la dynamique de la propagation des maladies infectieuses. Tel que mentionné précédemment, il est important que des efforts soient déployés pour que les modèles reproduisent le plus fidèlement possible la complexité de cette dynamique ; leur utilité en épidémiologie s'en verrait alors maximisée. Les modèles précédents présentés au chapitre 2 tiennent compte de l'effet *Small World*⁶ et du fait que les distributions de degrés sont hautement asymétriques. Suite à notre contribution, il est désormais possible d'étiqueter les noeuds du réseau et d'utiliser une transmissibilité hétérogène. Bien qu'il s'agisse là d'une avancée qualitative significative, les prochains modèles devraient être en mesure d'incorporer des propriétés structurelles telles que l'effet de regroupement (*i.e.* *les amis de mes amis sont mes amis*) et la présence de communautés à caractère social (*e.g.* race, religion, âge, sexe) ou circonstanciel (*e.g.* lieu de travail, voisinage, école, hôpitaux). Il serait également intéressant que les modèles puissent non seulement décrire l'évolution temporelle de la propagation des maladies, mais aussi l'évolution temporelle de la structure des réseaux eux-mêmes. Une évolution temporelle conjointe de la propagation de la maladie et du réseau permettrait une

⁴En effet, le calcul de $K(\mathbf{x}; \mathbf{T})$ pour la première distribution de degrés utilisée à la §4.3 a nécessité un calcul d'environ 24 heures puisque cette distribution est relativement large et nécessite l'utilisation de matrices de 120 x 100 éléments.

⁵Il est également possible d'obtenir une équation de récurrence pour la deuxième dérivée de $K(\mathbf{x}; \mathbf{T})$ et ainsi d'obtenir la variance de la distribution.

⁶L'effet *Small World* est présent dans tout réseau dont les connexions sont établies aléatoirement, ce qui est le cas ici.

implémentation en *temps réel* des stratégies d'intervention lors d'une épidémie⁷. De plus, mentionnons que les modèles actuels sont développés en considérant des réseaux de contacts infinis. Il va de soi qu'aucun réseau de contacts réel ne possède une telle taille et, malgré que leurs prédictions restent relativement valides pour des réseaux aussi petits que 1000 noeuds, il serait pertinent que les modèles futurs puissent tenir compte de la taille finie des systèmes réels.

Finalement, il est important de souligner que la validité de la modélisation mathématique en épidémiologie repose sur la capacité à obtenir des données représentatives des situations réelles. En effet, quelle valeur peut-on donner à des conclusions obtenues à partir de données qui ne représentent pas la réalité? Même les modèles les plus réalistes se verront relégués au titre de jouets mathématiques⁸ s'il est impossible de fournir à ces modèles les paramètres nécessaires. Par contre, le besoin de compiler de telles données ne s'étant jamais manifesté, celles-ci sont généralement difficiles à trouver, incomplètes voire inexistantes. L'importance d'une collaboration entre les modélisateurs (*e.g.* mathématiciens, physiciens) et les gens de la santé publique (*e.g.* épidémiologistes, microbiologistes, statisticiens) est alors manifeste. D'une part, elle permet une éducation mutuelle entre ces deux groupes. Les modélisateurs peuvent ainsi faire la promotion des différents modèles, démontrer leur utilité, de sorte qu'ils puissent être utilisés à leur plein potentiel. Ils ont aussi l'opportunité de sensibiliser les gens de la santé publique à la nécessité de rechercher des données réalistes et fiables pour nourrir les modèles. Inversement, les gens de la santé publique peuvent informer les modélisateurs des besoins en santé publique de sorte que les futurs modèles puissent y répondre. D'autre part, au fil de sa progression et de ses accomplissements, cette collaboration permet de démontrer la pertinence de la modélisation mathématique en épidémiologie aux autorités de la santé publique et aux organismes subventionnaires qui, jusqu'à tout récemment, regardaient cette discipline d'un oeil sceptique. Une telle synergie assurera la continuation de ce domaine de recherche et ultimement contribuera à l'amélioration de la santé publique.

⁷Par exemple, alors que la maladie progresse au sein de la population, il serait possible de mesurer l'impact qu'auraient des mises en quarantaine ou la fermeture de lieux publics.

⁸Notons toutefois que l'intérêt premier du théoricien réside habituellement dans l'étude théorique de ces jouets. Le fait qu'ils contribuent à l'amélioration de la santé publique n'est qu'une douceur supplémentaire.

Bibliographie

- [1] H. ABBEY, *An examination of the Reed-Frost theory of epidemics*, Human Biology, 24 (1952), pp. 201–233.
- [2] M. ABRAMOWITZ AND I. A. STEGUN, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, Ninth Dover Printing, Tenth GPO Printing ed., 1964.
- [3] W. AIELLO, F. CHUNG, AND L. LU, *A random graph model for massive graphs*, in Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, New York, 2000, Association of Computing Machinery, pp. 171–180.
- [4] —, *Random evolution of massive graphs*, in Handbook of Massive Data Sets, J. Abello, P. M. Pardalos, and M. G. C. Resende, eds., Kluwer Academic, Dordrecht, 2002, pp. 97–122.
- [5] R. ALBERT AND A.-L. BARABÁSI, *Statistical mechanics of complex networks*, Rev. Mod. Phys., 74 (2002), pp. 47–97.
- [6] R. ALBERT, H. JEONG, AND A.-L. BARABÁSI, *Diameter of the World-Wide Web*, Nature, 401 (1999), pp. 130–131.
- [7] —, *Attack and error tolerance of complex networks*, Nature, 406 (2000), pp. 378–382.
- [8] L. A. N. AMARAL, A. SCALA, M. BARTHÉLÉMY, AND H. E. STANLEY, *Classes of small-world networks*, Proc. Natl. Acad. Sci. U.S.A., 97 (2000), pp. 11149–11152.
- [9] R. M. ANDERSON AND R. M. MAY, *Infectious Diseases of Humans : Dynamics and Control*, Oxford University Press, Oxford, 1991.
- [10] M. AVENDANO, P. DERKACH, AND S. SWAN, *Clinical course and management of SARS in health care workers in Toronto : A case series*, CMAJ, 168 (2003), p. 1649–60.
- [11] N. T. J. BAILEY, *The Mathematical Theory of Infectious Diseases and its Applications*, Hafner Press, New York, 1975.
- [12] S. BANSAL, B. T. GRENFELL, AND L. A. MEYERS, *When individual behaviour matters : homogeneous and network models in epidemiology*, J. R. Soc. Interface, 4 (2007), pp. 879–891.

- [13] S. BANSAL, B. POURBOHLOUL, AND L. A. MEYERS, *A comparative analysis of influenza vaccination programs*, PLoS Medicine, 3 (2006), pp. 1–10.
- [14] A.-L. BARABÁSI AND R. ALBERT, *Emergence of scaling in random networks*, Science, 286 (1999), pp. 509–512.
- [15] A.-L. BARABÁSI, R. ALBERT, AND H. JEONG, *Scale-free characteristics of random networks : The topology of the World Wide Web*, Physica A, 281 (2000), pp. 69–77.
- [16] N. BIGGS, E. LLOYD, AND R. WILSON, *Graph Theory, 1736–1936*, Oxford University Press, 1986.
- [17] S. BLOWER AND D. BERNOULLI, *An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it*, Rev. Mod. Virol., 14 (2004), pp. 275–288.
- [18] S. BOCCALETTI, V. LATORA, Y. MORENO, M. CHAVEZ, AND D.-U. HWANG, *Complex networks : Structure and dynamics*, Phys. Rep., 424 (2006), pp. 175–308.
- [19] M. BOGUÑÁ, R. PASTOR-SATORRAS, AND A. VESPIGNANI, *Absence of epidemics threshold in scale-free networks with degree correlations*, Phys. Rev. Lett., 90 (2003), pp. 028701(1–4).
- [20] S. BRIN AND L. PAGE, *The anatomy of a large-scale hypertextual web search engine*, Computer Networks, 30 (1998), p. 107–117.
- [21] A. BRODER, R. KUMAR, F. MAGHOUL, P. RAGHAVAN, S. RAJAGOPALAN, R. STATA, A. TOMKINS, , AND J. WIENER, *Graph structure in the web*, Computer Networks, 33 (2000), pp. 309–320.
- [22] S. CAUCHEMEZ, F. CARRAT, C. VIBOUD, A. J. VALLERON, AND P. Y. BOËLLE, *A Bayesian MCMC approach to study transmission of influenza : application to household longitudinal data*, Statist. Med., 23 (2004), pp. 3469–3487.
- [23] Q. CHEN, H. CHANG, R. GOVINDAN, S. JAMIN, S. J. SHENKER, AND W. WILLINGER, *The origin of power laws in internet topologies revisited*, in Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies, Los Alamitos, 2002, IEEE Computer Society.
- [24] A. CLAUSET, C. R. SHALIZI, AND M. E. J. NEWMAN, *Power-law distributions in empirical data*, submitted to SIAM Review, (2008).
- [25] O. DIEKMANN AND J. A. P. HEESTERBEEK, *Mathematical Epidemiology of Infectious Diseases : Model Building, Analysis and Interpretation*, John Wiley & Sons Ltd, 2000.
- [26] R. DIESTEL, *Graph Theory*, Springer-Verlag, Heidelberg, third ed., 2005. Free online version available.
- [27] K. DIETZ AND J. A. P. HEESTERBEEK, *Bernoulli was ahead of modern epidemiology*, Nature, 408 (2000), pp. 513–514.

- [28] S. N. DOROGOVTSSEV AND J. F. F. MENDES, *Evolution of networks*, Adv. Phys., 51 (2002), pp. 1079–1187.
- [29] P. ERDŐS AND A. RÉNYI, *On random graphs*, Publ. Math. Debrecen, 6 (1959), pp. 290–297.
- [30] M. FALOUTSOS, P. FALOUTSOS, AND C. FALOUTSOS, *On power-law relationships of the internet topology*, Computer Communications Rev., 29 (1999), pp. 251–262.
- [31] N. M. FERGUSON, D. A. T. CUMMINGS, S. CAUCHEMEZ, C. F. ANS S. RILEY, A. MEEYAIL, S. IAMSIRITHAWORN, AND D. S. BURKE, *Strategies for containing an emerging influenza pandemic in Southeast Asia*, Nature, 437 (2005), pp. 209–214.
- [32] N. M. FERGUSON, D. A. T. CUMMINGS, C. FRASER, J. C. CAJKA, P. C. COOLEY, AND D. S. BURKE, *Strategies for mitigating an influenza pandemic*, Nature, 442 (2006), pp. 448–452.
- [33] M. GIRVAN AND M. E. J. NEWMAN, *Community structure in social and biological networks*, Proc. Natl. Acad. Sci. U.S.A., 99 (2002), pp. 7821–7826.
- [34] H. S. HETHCOTE, *The mathematics of infectious diseases*, SIAM Review, 42 (2000), pp. 599–653.
- [35] H. JEONG, S. MASON, A.-L. BARABASI, AND Z. N. OLTVAI, *Lethality and centrality in protein networks*, Nature, 411 (2001), pp. 41–42.
- [36] H. JEONG, B. TOMBOR, R. ALBERT, Z. N. OLTVAI, AND A.-L. BARABASI, *The large-scale organization of metabolic networks*, Nature, 407 (2000), pp. 651–654.
- [37] F. KARINTHY, *Chains*, in Everything is Different, Budapest, 1929.
- [38] M. J. KEELING AND K. T. D. EAMES, *Networks and epidemic models*, J. R. Soc. Interface, 2 (2005), pp. 295–307.
- [39] E. KENAH AND J. M. ROBINS, *Second look at the spread of epidemics on networks*, Phys. Rev. E, 76 (2007), p. 036113.
- [40] F. LILJEROS, C. R. EDLING, L. A. N. AMARAL, H. E. STANLEY, AND Y. ABERG, *The web of a human sexual contacts*, Nature, 411 (2001), pp. 907–908.
- [41] M. MARDER, *Dynamics of epidemics on random networks*, Phys. Rev. E, 75 (2007), p. 066103.
- [42] S. MASLOV, K. SNEPPEN, AND A. ZALIZNYAK, *Detection of topological patterns in complex networks : correlation profile of the internet*, Physica A, 333 (2004), pp. 529–540.
- [43] L. A. MEYERS, *Contact network epidemiology : Bond percolation applied to infectious disease prediction and control*, Bulletin (New Series) of the American Mathematical Society, 44 (2007), pp. 63–86.

- [44] L. A. MEYERS, M. E. J. NEWMAN, M. MARTIN, AND S. SCHRAG, *Applying network theory to epidemics : Control measures for mycoplasma pneumonia outbreaks*, *Emerg. Infect. Dis.*, 9 (2003), pp. 204–210.
- [45] L. A. MEYERS, M. E. J. NEWMAN, AND B. POURBOHLOUL, *Predicting epidemics on directed contact networks*, *J. Theor. Biol.*, 240 (2006), pp. 400–418.
- [46] L. A. MEYERS, B. POURBOHLOUL, M. E. J. NEWMAN, D. M. SKOWRONSKI, AND R. C. BRUNHAM, *Network theory and SARS : predicting outbreak diversity*, *J. Theor. Biol.*, 232 (2005), pp. 71–81.
- [47] S. MILGRAM, *The small world problem*, *Psych. Today*, 2 (1967), pp. 60–67.
- [48] J. C. MILLER, *Epidemic size and probability in populations with heterogeneous infectivity and susceptibility*, *Phys. Rev. E*, 76 (2007), p. 010101.
- [49] Y. MORENO, J. B. GOMEZ, AND A. F. PACHECO, *Epidemic incidence in correlated complex networks*, *Phys. Rev. E*, 68 (2003), p. 035103.
- [50] M. E. J. NEWMAN, *Spread of epidemic disease on networks*, *Phys. Rev. E*, 66 (2002), p. 016128.
- [51] ———, *Mixing patterns in networks*, *Phys. Rev. E*, 67 (2003), pp. 026126(1–13).
- [52] ———, *Properties of highly clustered networks*, *Phys. Rev. E*, 68 (2003), pp. 026121(1–6).
- [53] ———, *The structure and function of complex networks*, *SIAM Review*, 45 (2003), pp. 167–256.
- [54] ———, *Power laws, Pareto distributions and Zipf’s law*, *Contemporary Physics*, 46 (2005), pp. 323–351.
- [55] M. E. J. NEWMAN, S. H. STROGATZ, AND D. J. WATTS, *Random graphs with arbitrary degree distributions and their applications*, *Phys. Rev. E*, 64 (2001), p. 026118.
- [56] A. NICOLosi, M. C. LEITE, M. MUSICCO, C. ARICI, G. GAVAZZENI, AND A. LAZZARIN, *The efficiency of male-to-female and female-to-male sexual transmission of the human immunodeficiency virus : a study of 730 stable couples. italian study group on hiv heterosexual transmission.*, *Epidemiology*, 5 (1994), pp. 570–575.
- [57] P.-A. NOËL, B. DAVOUDI, R. C. BRUNHAM, L. J. DUBÉ, AND B. POURBOHLOUL, *Time evolution of epidemic disease on finite and infinite networks*, *Phys. Rev. E*, (2008, to be published).
- [58] R. PASTOR-SATORRAS AND A. VESPIGNANI, *Epidemic spreading in scale-free networks*, *Phys. Rev. Lett.*, 86 (2001), pp. 3200–3203.
- [59] ———, *Immunization of complex networks*, *Phys. Rev. E*, 65 (2002), pp. 036104(1–8).
- [60] P. PATLOLLA, V. GUNUPUDI, A. R. MIKLER, AND R. T. JACOB, *Agent-Based simulation tools in computational epidemiology*, *Lecture Notes in Computer Science*, 3473 (2006), pp. 212–223.

- [61] B. POURBOHLOUL, L. A. MEYERS, D. M. SKOWRONSKI, M. KRAJDEN, D. M. PATRICK, AND R. C. BRUNHAM, *Modeling control strategies of respiratory pathogens*, *Emerg. Infect. Dis.*, 11 (2005), pp. 1249–1256.
- [62] S. M. POUTANEN, D. E. LOW, B. HENRY, S. FINKELSTEIN, D. ROSE, K. GREEN, R. TELLIER, R. DRAKER, D. ADACHI, M. AYERS, A. K. CHAN, D. M. SKOWRONSKI, I. SALIT, A. E. SIMOR, A. S. SLUTSKY, P. W. DOYLE, M. KRAJDEN, M. PETRIC, R. C. BRUNHAM, AND A. J. MCGEER, *Identification of severe acute respiratory syndrome in Canada*, *N. Engl. J. Med.*, 348 (2003), pp. 1995–2005.
- [63] W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING, AND B. P. FLANNERY, *Numerical Recipes : The Art of Scientific Computing*, Cambridge University Press, third ed., 2007.
- [64] S. REDNER, *How popular is your paper? An empirical study of the citation distribution*, *Eur. Phys. J. B*, 4 (1998), pp. 131–134.
- [65] P. O. SEGLEN, *The skewness of science*, *J. Amer. Soc. Inform. Sci.*, 43 (1992), pp. 628–638.
- [66] D. STAUFFER AND A. AHARONY, *Introduction to Percolation Theory*, Taylor & Francis, revised second ed., 1994.
- [67] S. H. STROGATZ, *Exploring complex networks*, *Nature*, 410 (2001), pp. 268–276.
- [68] M. VARIA, S. WILSON, S. SARWAL, A. MCGEER, E. GOURNIS, E. GALANIS, B. HENRY, AND H. TEAM, *Investigation of a nosocomial outbreak of severe acute respiratory syndrome (SARS) in Toronto, Canada*, *CMAJ*, 169 (2003), p. 285–292.
- [69] A. VAZQUEZ, R. PASTOR-SATORRAS, AND A. VESPIGNANI, *Large-scale topological and dynamical properties of the Internet*, *Phys. Rev. E*, 65 (2002), p. 066130.
- [70] D. J. WATTS AND S. H. STROGATZ, *Collective dynamics of 'small-world' networks*, *Nature*, 393 (1998), pp. 440–442.
- [71] H. S. WILF, *generatingfunctionology*, Academic Press inc., second ed., 1994.
- [72] H. P. YOUNG, *The diffusion of innovations in social networks*, in *The Economy as an Evolving Complex System*, Vol. 3, L. E. Blume and S. N. Durlauf, eds., Oxford University Press, Oxford, 2003.

Annexe A

Synthèse du formalisme

Cette annexe résume les concepts de base ainsi que les résultats principaux de notre formalisme.

Quantités de base

Le tableau A.1 présente les quantités de base de notre formalisme. Ces quantités sont typiquement imposées par le problème à modéliser.

Quantité	Définition
N	Nombre de noeuds dans le réseau. Notre formalisme est développé dans la limite $N \rightarrow \infty$, mais comme il a été expliqué à la §4.3, plusieurs prédictions de notre formalisme restent suffisamment précises pour des réseaux où $N = 1000$.
M	Nombre de types de noeuds dans le réseau.
w_i	Fraction du réseau occupée par les noeuds de type i .
$P_i(\mathbf{k})$	Distribution de degrés des noeuds de type i (k_1 liens vers les noeuds de type 1, k_2 liens vers les noeuds de type 2, etc.).
\mathbf{T}	Matrice de transmissibilité. Les éléments T_{ij} de cette matrice indique la probabilité de transmission d'un noeud infecté de type i à un noeud susceptible de type j . Ces éléments peuvent aussi être vus comme la probabilité d'occupation des liens dans la direction $i \rightarrow j$ dans un langage de percolation de liens.

TAB. A.1 – Quantités de base du formalisme.

Contrainte

Les quantités $P_i(\mathbf{k})$ et w_i sont reliées l'une à l'autre par la relation (voir le tableau A.2 pour la signification de z_{ij})

$$\mathbf{wz} = (\mathbf{wz})^\dagger$$

où

$$\mathbf{w} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_M \end{bmatrix} ; \quad \mathbf{z} = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1M} \\ z_{21} & z_{22} & \dots & z_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ z_{M1} & z_{M2} & \dots & z_{MM} \end{bmatrix}.$$

Cette contrainte demande en fait qu'il y ait autant de liens non-dirigés allant d'un noeud de type i à un noeud de type j que de liens non-dirigés allant d'un noeud de type j à un noeud de type i . Ceci assure la *fermeture* (*i.e.* la cohérence) du réseau. Cette contrainte correspond généralement à un système linéaire d'équations surdéterminé, ce qui implique que $P_i(\mathbf{k})$ et w_i doivent être fixés *a priori* afin de la respecter.

Seuil de percolation

Nous conjecturons que la taille moyenne de la petite composante atteinte en choisissant un noeud au hasard diverge selon

$$\langle s \rangle = \sum_{i=1}^M \langle s_i \rangle \sim \zeta_M^{-1}$$

où ζ_M est un polynôme formé des éléments de \mathbf{T} ayant des coefficients qui dépendent uniquement de la structure du réseau. $\zeta_M = 0$ correspond à une *hypersurface* dans un espace M^2 dimensionnel dont la base est formée par les éléments de la matrice de transmissibilité \mathbf{T} . Cette *hypersurface* définit le seuil de percolation du réseau (point où la composante géante apparaît). En langage épidémiologique, $\zeta_M = 0$ définit, pour une structure sociale $P_i(\mathbf{k})$ donnée, l'ensemble des *transmissibilités critiques* marquant le début d'un risque d'épidémie à grande échelle puisque tout scénario se situant à l'*extérieur* de cette hypersurface possédera une composante géante.

Formalisme

Le tableau A.2 résume les quantités importantes de notre formalisme.

Quantité	Définition	Définition mathématique	# éq.
$G_i(\mathbf{x}; \mathbf{T})$	Fonction générant la distribution de degrés occupés.	$G_i(\mathbf{x}; \mathbf{T}) = \sum_{\mathbf{k}=0}^{\infty} P_i(\mathbf{k}) \prod_{l=1}^M [1 + (x_l - 1)T_{il}]^{k_l}$	(3.5)
z_{ij}	Nombre moyen de liens partant d'un noeud de type i à un noeud de type j .	$z_{ij} = \frac{1}{T_{ij}} \frac{dG_i(\mathbf{1}; \mathbf{T})}{dx_j}$	(3.1)
\tilde{z}_{ij}	Nombre moyen de liens occupés partant d'un noeud de type i à un noeud de type j .	$\tilde{z}_{ij} = \frac{dG_i(\mathbf{1}; \mathbf{T})}{dx_j}$	(3.6)
$F_{ij}(\mathbf{x}; \mathbf{T})$	Fonction générant la distribution de degrés excédants occupés.	$F_{ij}(\mathbf{x}; \mathbf{T}) = \frac{1}{\tilde{z}_{ji}} \frac{dG_j(\mathbf{x}; \mathbf{T})}{dx_i}$	(3.8)
$H_{ij}(\mathbf{x}; \mathbf{T})$	Fonction générant la distribution de la taille de la composante atteinte en suivant un lien de type $i \rightarrow j$.	$H_{ij}(\mathbf{x}; \mathbf{T}) = x_j F_{ij}(\mathbf{H}_j(\mathbf{x}; \mathbf{T}); \mathbf{T})$	(3.9)
$K(\mathbf{x}; \mathbf{T})$	Fonction générant la distribution de la taille de la composante atteinte à partir d'un noeud choisi au hasard.	$K(\mathbf{x}; \mathbf{T}) = \frac{1}{1-\mathcal{P}} \sum_{i=1}^M w_i x_i G_i(\mathbf{H}_i(\mathbf{x}; \mathbf{T}); \mathbf{T})$	(3.10)
\overrightarrow{h}_{ij}	Probabilité qu'un lien $i \rightarrow j$ occupé <i>ne mène pas</i> à la composante géante.	$\overrightarrow{h}_{ij} = F_{ij}(\overrightarrow{\mathbf{h}}_j; \mathbf{T})$	(4.24)
\mathcal{P}	Probabilité qu'un noeud choisi au hasard mène à la composante géante.	$\mathcal{P} = 1 - \sum_{i=1}^M w_i G_i(\overrightarrow{\mathbf{h}}_i; \mathbf{T})$	(4.25)
\overleftarrow{h}_{ij}	Probabilité qu'un lien de type $j \rightarrow i$ <i>ne lie pas</i> le noeud de type i à la composante géante.	$\overleftarrow{h}_{ij} = F_{ij}(\overleftarrow{\mathbf{h}}_j; \mathbf{T}^\dagger)$	(4.26)
\mathcal{S}_i	Taille relative du réseau occupée par des noeuds de type i faisant partie de la composante géante.	$\mathcal{S}_i = w_i [1 - G_i(\overleftarrow{\mathbf{h}}_i; \mathbf{T}^\dagger)]$	(4.33)
$\overrightarrow{\alpha}_{ij}^{(l)}$	Nombre moyen de noeuds de type l présents dans la composante atteinte en suivant un lien de type $i \rightarrow j$.	$\overrightarrow{\alpha}_{ij}^{(l)} = F_{ij}(\overrightarrow{\mathbf{h}}_j; \mathbf{T}) \delta_{jl} + \sum_{n=1}^M T_{jn} \overrightarrow{\beta}_{ij}^{(n)} \overrightarrow{\alpha}_{jn}^{(l)}$	(4.31)
$\langle \overrightarrow{s}_i \rangle$	Nombre moyen de noeuds de type i présents dans la composante atteinte à partir d'un noeud choisi au hasard.	$\langle \overrightarrow{s}_i \rangle = \frac{w_i G_i(\overrightarrow{\mathbf{h}}_i; \mathbf{T})}{1-\mathcal{P}} + \frac{1}{1-\mathcal{P}} \sum_{l=1}^M w_l \sum_{j=1}^M T_{lj} z_{lj} \overrightarrow{h}_{jl} \overrightarrow{\alpha}_{lj}^{(i)}$	(4.30)
$\overrightarrow{\beta}_{ij}^{(n)}$	Degré excédant occupé moyen.	$\overrightarrow{\beta}_{ij}^{(n)} = \frac{1}{T_{jn}} \frac{dF_{ij}(\mathbf{H}_j(\mathbf{1}; \mathbf{T}); \mathbf{T})}{dH_{jn}(\mathbf{1}; \mathbf{T})}$	(4.32)

TAB. A.2 – Synthèse des différentes quantités présentes dans notre formalisme.

Annexe B

Implémentation numérique

B.1 Mise en puissances de polynômes par Transformées de Fourier

Les modèles de percolation utilisant des fonctions génératrices présentés dans cet ouvrage possèdent des équations, transcendantes ou non, qui impliquent la multiplication ou la mise en puissances de polynômes à une ou plusieurs variables et de taille arbitraire. Du point de vue numérique, la résolution directe de ces équations (*i.e.* effectuer les multiplications un terme à la fois) ne permet pas un algorithme général et facilement réutilisable, en plus de nécessiter des temps de calcul d'ordre géologique. Nous présentons dans cette section une méthode basée sur l'utilisation de la *Fast Fourier Transform* (FFT) [63] qui permet de sauver plusieurs ordres de grandeur dans le temps de calcul (*i.e.* le ramener à une échelle humaine) et ainsi le rendre tout à fait acceptable dans un contexte d'utilisation en épidémiologie.

Rappelons tout d'abord que les équations que nous avons à résoudre sont de la forme générale suivante

$$f(x_1, \dots, x_M) = \sum_{k_1, \dots, k_M=0}^{\infty} p_{k_1, \dots, k_M} [h_1(x_1, \dots, x_M)]^{k_1} \dots [h_M(x_1, \dots, x_M)]^{k_M} \quad (\text{B.1})$$

où p_{k_1, \dots, k_M} est un simple coefficient. Cette équation illustre, numériquement les bornes supérieures sont bien entendu limitées à une valeur finie, que la résolution directe de ce genre d'équations peut nécessiter un temps de calcul excessivement long. Il faut donc avoir recours à une autre approche.

Pour illustrer la méthode que nous avons utilisée, considérons le cas $M = 1$ (*i.e.* une seule variable). Le produit de deux polynômes de même variable, pour deux séries absolument convergentes¹, donne le résultat suivant par la règle du produit de Cauchy

$$\begin{aligned} A(x)B(x) &= \left(\sum_{i=0}^{\infty} a_i x^i \right) \left(\sum_{j=0}^{\infty} b_j x^j \right) \\ &= \sum_{i=0}^{\infty} \left[\sum_{j=0}^i a_j b_{i-j} \right] x^i \\ &= \sum_{i=0}^{\infty} \left[\{a_j\} * \{b_j\} \right] x^i \end{aligned} \quad (\text{B.2})$$

où les coefficients du polynôme résultant sont la convolution discrète (notée par $*$) des coefficients des polynômes d'origine. Par récurrence, on peut aisément montrer que ce résultat se généralise au produit d'un nombre arbitraire de polynômes. Autrement dit, ce résultat nous indique que la distribution générée par un produit de fonctions génératrices s'obtient par la convolution des distributions générées par ces fonctions. En utilisant le théorème de la convolution, nous voyons qu'il est possible d'obtenir le polynôme résultant du produit $A(x)B(x)$ à l'aide de Transformées de Fourier Discrètes (notée par \mathcal{F} et son inverse par \mathcal{F}^{-1})

$$A(x)B(x) = \sum_{i=0}^{\infty} \left[\mathcal{F}^{-1} \left(\mathcal{F}(\{a_j\}) \cdot \mathcal{F}(\{b_j\}) \right) \right] x^i \quad (\text{B.3})$$

où \cdot dénote la multiplication *élément par élément*². Nous voyons que ce résultat peut se généraliser au produit d'un nombre arbitraire de fonctions qui peuvent être élevées à une quelconque puissance entière. De plus, considérant M variables linéairement indépendantes, il est possible de généraliser ce résultat à des valeurs arbitraires de M pour résoudre des équations telles que (B.1).

Il existe plusieurs algorithmes fiables permettant d'effectuer des convolutions discrètes telles dans (B.2). Toutefois, ceux-ci nécessitent généralement un temps de calcul proportionnel à $\mathcal{O}(N^2)$ (où N est le nombre d'opérations à effectuer) ce qui, dans les cas qui nous concernent, implique des temps de calcul encore trop longs. Toutefois, en utilisant un algorithme FFT dans (B.3), le temps de calcul sera proportionnel à $\mathcal{O}(N \log_2 N)$, ce qui peut correspondre à une économie de temps de plusieurs ordres de grandeur. C'est d'ailleurs cette approche que nous avons utilisée pour obtenir les prédictions théoriques de la §4.3. Nous allons maintenant faire quelques remarques sur la résolution numérique

¹Toutes les séries de puissances utilisées dans cet ouvrage convergent absolument.

²Dans ce contexte, la multiplication *élément par élément* correspond à la multiplication des éléments associés à la même *fréquence* discrète. Numériquement, nous obtenons un vecteur pour chaque Transformée de Fourier et multiplions les éléments possédant le même indice dans leur vecteur respectif.

des équations transcendantes présentes dans le modèle de Newman, celui de Meyers *et al.* et le nôtre.

Équations à une variable (Newman et Meyers)

Les équations transcendantes des modèles de Newman et Meyers *et al.* ne possèdent qu'une seule variable et se solutionnent en itérant respectivement les applications

$$H_n(x; T) = xF(H_{n-1}(x; T); T)$$

et

$$\begin{aligned}\bar{H}_n^{\text{in}}(v; T_d, T_u) &= v\bar{F}^{\text{in}}(1, \bar{H}_{n-1}^{\text{in}}(v; T_d, T_u), \bar{H}_{n-1}^{\text{un}}(v; T_d, T_u); T_d, T_u) \\ \bar{H}_n^{\text{un}}(v; T_d, T_u) &= v\bar{F}^{\text{un}}(1, \bar{H}_{n-1}^{\text{in}}(v; T_d, T_u), \bar{H}_{n-1}^{\text{un}}(v; T_d, T_u); T_d, T_u)\end{aligned}$$

pour $n \geq 1$ jusqu'à la convergence vers un point fixe en utilisant les conditions initiales

$$H_0(x; T) = x; \quad \bar{H}_0^{\text{in}}(v; T_d, T_u) = \bar{H}_0^{\text{un}}(v; T_d, T_u) = v.$$

La présence des variables x et v devant les fonctions $F(x; T)$, $\bar{F}^{\text{in}}(x, y, u; T_d, T_u)$ et $\bar{F}^{\text{un}}(x, y, u; T_d, T_u)$ assurent la convergence de ces applications et donc l'existence d'un point fixe. En effet, leur présence implique que la n ème itération ne modifiera que les coefficients d'indice supérieur à n . Ainsi, il est possible de connaître la valeur exacte des n premiers coefficients en itérant l'application $n + 1$ fois et puisque les distributions générées par les fonctions génératrices utilisées dans cet ouvrage sont normalisées, il sera possible en principe d'obtenir exactement (*i.e.* jusqu'à la précision numérique de l'ordinateur utilisé) la distribution en un nombre fini d'itérations. De plus, puisque tous les termes impliqués sont positifs, que les distributions sont toutes normalisées et que les opérations ne sont que des additions et des multiplications, nous sommes assurés qu'avec les conditions initiales explicitées ci-haut, les distributions correspondant au point fixe de ces applications seront aussi normalisées et que tous leurs termes seront positifs.

Équations à plusieurs variables

Les mêmes remarques s'appliquent aussi aux équations transcendantes présentes dans notre modèle et le détail de l'application à itérer est donné à la §3.3.3. La convergence de l'application est ici aussi assurée par la présence de la variable x_j devant la fonction $F_{ij}(\mathbf{x}; \mathbf{T})$ et il faudra itérer l'application $n + 1$ fois pour connaître exactement

tous les coefficients dont la somme des indices est inférieure ou égale à n . Toutefois, une petite fraction des $n + 1$ itérations est souvent suffisante pour obtenir les distributions à une précision tout à fait acceptable et il n'est pas nécessaire d'effectuer chacune des $n + 1$ itérations, une tâche qui peut s'avérer fort longue.

B.2 Simulations numériques de la propagation de maladies sur un réseau

À la §4.3, nous avons validé notre formalisme en comparant ses prédictions aux résultats obtenus lors de simulations numériques. Nous présentons maintenant la procédure que nous avons suivie pour obtenir ces résultats.

Considérant que nous voulons simuler la propagation d'une maladie, dont la matrice de transmissibilité \mathbf{T} est connue, sur des réseaux de N noeuds divisés en M types dont la distribution de degrés $P_i(\mathbf{k})$ et le vecteur de population w_i sont aussi connus.

1. Générer une séquence de Nw_i nombres aléatoires respectant la distribution de degrés $P_i(\mathbf{k})$ prescrite pour chaque type de lien ($i \rightarrow j$, il y en a M^2 au total). Ces nombres peuvent être générés en utilisant des méthodes de transformation et/ou de rejet [63]. Ces séquences prescrivent le nombre de liens qui quitteront les noeuds pour aller vers des noeuds des M types différents. Chaque élément de ces séquences correspond donc au nombre de *demi-liens* émergeant d'un noeud. Afin de respecter la contrainte (3.2), ces séquences doivent respecter les conditions suivantes :
 - la somme des degrés intra-type (*i.e.* $i \rightarrow i$) doit être paire ;
 - la somme des degrés inter-type de type $i \rightarrow j$ doit être égale à la somme des degrés de type $j \rightarrow i$.

Si l'une ou l'autre de ces conditions n'est pas respectée, recommencer l'étape 1.

2. Assembler aléatoirement par paire (*i.e.* les demi-liens de type $i \rightarrow j$ avec ceux de type $j \rightarrow i$) tous les demi-liens disponibles. Les liens reliant un noeud avec lui-même ou les liens supplémentaires entre deux noeuds peuvent être retirés en utilisant un algorithme de reconnexion [42].
3. Choisir un noeud au hasard dans le réseau et le considérer comme étant infecté. Laisser la maladie se propager avec une probabilité prescrite par \mathbf{T} jusqu'à ce qu'il n'y ait plus de noeuds à infecter. Noter le nombre d'infectés de chaque type.
4. Recommencer l'étape 3 autant de fois que nécessaire.
5. Recommencer les étapes 1 à 3 jusqu'à convergence satisfaisante des résultats.

Annexe C

Formule intégrale de Cauchy

Soit une fonction complexe $f(z)$ analytique sur la courbe fermée Γ et dans la région R bornée par Γ . Le théorème intégral de Cauchy prescrit que

$$\oint_{\Gamma} \frac{f(z)}{z - z_0} dz = 2\pi i f(z_0) \quad \forall z_0 \in R$$

et a comme corollaire que

$$\frac{d^n f(z_0)}{dz^n} = \frac{n!}{2\pi i} \oint_{\Gamma} \frac{f(z)}{(z - z_0)^{n+1}} dz \quad \forall z_0 \in R. \quad (\text{C.1})$$

Ayant obtenu $\varphi(x)$ par un quelconque raisonnement, l'évaluation des coefficients ψ_k se fait à l'aide (2.4). En utilisant (C.1), nous obtenons directement que

$$\psi_k = \frac{1}{k!} \left. \frac{d^k \varphi(z)}{dz^k} \right|_{z=0} = \frac{1}{2\pi i} \oint_{\Gamma} \frac{\varphi(z)}{z^{k+1}} dz$$

où Γ doit être choisi tel que $\varphi(z)$ soit analytique sur Γ et dans la région délimitée par Γ . Pour les cas étudiés dans ce mémoire, $\varphi(z)$ est de façon générale une série infinie de puissances

$$\varphi(z) = \sum_{k=0}^{\infty} \psi_k z^k$$

où $\{\psi_k\}$ correspond à une distribution de probabilités (*i.e.* $\sum_{k=0}^{\infty} \psi_k = \sum_{k=0}^{\infty} |\psi_k| = 1$). Ainsi $\varphi(z)$ convergera uniformément et absolument dans une région certainement aussi grande que $|z| \leq 1$. Puisque la précision numérique sera d'autant plus grande que la région délimitée par Γ le sera (en n'incluant toutefois aucun pôle de $\varphi(z)$), le plus grand contour qui respectera de façon générale ces conditions sera le cercle unitaire $|z| = 1$.

Annexe D

Preuve de la conjecture

Après le dépôt initial à la Faculté des Études Supérieures de ce mémoire, la preuve de la conjecture de la section 3.3.4 a été trouvée. L'équation (3.13) peut s'écrire sous la forme matricielle

$$\boldsymbol{\alpha}^{(i)} = \mathbf{A}\boldsymbol{\alpha}^{(i)} + \mathbf{B}_i$$

où

$$\boldsymbol{\alpha}^{(i)} = [\alpha_{11}^{(i)}, \alpha_{12}^{(i)}, \dots, \alpha_{1M}^{(i)}, \alpha_{21}^{(i)}, \dots, \alpha_{MM}^{(i)}]^\dagger, \quad \mathbf{B}_i = [\delta_{1i}, \delta_{2i}, \dots, \delta_{Mi}, \delta_{1i}, \delta_{2i}, \dots, \delta_{Mi}]^\dagger$$

et

$$\mathbf{A} = \begin{bmatrix} T_{11}\beta_{11}^{(1)} & T_{12}\beta_{11}^{(2)} & \cdots & T_{1M}\beta_{11}^{(M)} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & T_{21}\beta_{12}^{(1)} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & T_{MM}\beta_{1M}^{(M)} \\ T_{11}\beta_{21}^{(1)} & T_{12}\beta_{21}^{(2)} & \cdots & T_{1M}\beta_{21}^{(M)} & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & T_{MM}\beta_{MM}^{(M)} \end{bmatrix}.$$

Nous avons alors directement que

$$\boldsymbol{\alpha}^{(i)} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{B}_i = \frac{\mathbf{C}\mathbf{B}_i}{\det(\mathbf{I} - \mathbf{A})}$$

où \mathbf{I} est la matrice identité et \mathbf{C} est la matrice adjointe de $\mathbf{I} - \mathbf{A}$. Ainsi, à l'aide de l'équation (3.12), nous voyons que $\zeta_M = \det(\mathbf{I} - \mathbf{A})$ et que tous les $\alpha_{ij}^{(i)}$, et donc $\langle s \rangle$, divergent simultanément. De plus, ceci permet de montrer que la présence du terme inhomogène \mathbf{B}_i permet effectivement l'existence d'une solution non-triviale tel que présumé dans la conjecture.