



UNIVERSITÉ
DE NAMUR

University of Namur

Institutional Repository - Research Portal Dépôt Institutionnel - Portail de la Recherche

researchportal.unamur.be

THESIS / THÈSE

MASTER EN SCIENCES MATHÉMATIQUES

Processus stochastiques sur réseaux : au-delà des hypothèses de Markov

GUEUNING, Martin

Award date:
2013

Awarding institution:
Universite de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

MASTER EN MATHÉMATIQUES

Processus stochastiques sur réseaux : au-delà des hypothèses de Markov

Martin Gueuning

2013

Facultés universitaires Notre-Dame de la Paix



FACULTÉS UNIVERSITAIRES NOTRE-DAME DE LA PAIX

*Dynamique sur réseaux temporels :
au-delà des processus de Markov*

Gueuning Martin

Promoteur : R. Lambiotte

ANNÉE ACADÉMIQUE 2012-2013

Résumé

La théorie des graphes est une branche des mathématiques en plein essor dont les applications sont multiples, notamment l'étude de propagation de maladies ou de l'information. Cependant, l'approche classique se concentre principalement sur les propriétés intrinsèques des réseaux étudiés et se base sur des hypothèses de Markov qui négligent la composante dynamique du problème. Or, ces hypothèses sont mises en défaut dans les études sur données réelles. Ainsi est-il important de prendre en considération la mémoire du réseau et les caractéristiques liées aux réseaux ou aux propagations impliquant des humains, comme, par exemple, le fait que les hommes ont de courtes périodes d'activité intense suivies de longues périodes d'inactivité, la présence de communautés ou encore la temporalité des liens sociaux. Le rejet de ces hypothèses de Markov implique des différences fondamentales dans les résultats observés, telles que le passage d'une équation différentielle à une équation intégro-différentielle pour décrire la dynamique d'infection dans un modèle compartimental.

Deux pistes impliquant des processus différents sont développées dans ce mémoire afin d'aller au-delà des hypothèses de Markov.

D'une part, nous nous intéressons à une généralisation du processus de percolation, appelée processus de branchement de Bellman-Harris, que nous appliquons sur une nouvelle matrice qui permet de prendre en compte la force dynamique des liens lors d'une propagation sur un réseau.

D'autre part, nous étudions les résultats que l'on peut obtenir via l'étude de l'agrégation de marches aléatoires sur un réseau grâce à des manipulations dans l'espace de Laplace associé.

Il existe toutefois une différence fondamentale entre l'étude d'une propagation compartimentale et l'agrégation de marches aléatoires. Celle-ci réside dans la corrélation de l'état de tous les noeuds possédant un voisin commun qui existe dans le cas d'une propagation globale unique telle que simulée via percolation mais qui disparaît lorsqu'on l'approxime par agrégation de processus indépendants.

Enfin, des simulations numériques sur un réseau connu et contrôlé mettent en évidence l'impact du rôle de la variance de la distribution du temps d'attente avant infection lors d'une propagation dans un réseau. Les résultats indiquent que, en un temps donné, tant la profondeur moyenne maximale que la taille moyenne de l'épidémie augmentent lorsque la variance augmente.

Abstract

Graph theory is a booming branch of mathematics that has a wide range of application, from neuronal connections to epidemiological propagations or information diffusion.

However, the classical approach mainly focuses on underlying properties of the network of interest and relies on Markovian hypotheses that neglect the dynamical patterns of the problem. Unfortunately, these hypotheses are usually not verified in the studies with real data. This is why it is important to take into account the memory of the network as well as the underlying characteristics of networks or propagations involving human beings, such as the bursty nature of human interactions, communities or non-persistent social ties. Rejecting these Markovian hypotheses implies fundamental differences in the found results, such as moving from a differential to an integro-differential equation.

Two different ways, involving different processes, are developed in this dissertation in order to go beyond those Markovian hypotheses.

On the one hand, we focus on a generalization of the percolation process, namely Bellman-Harris branching model, that we apply on a new matrix taking into account the dynamical strength of social ties in spreading over a network.

On the other hand, we look at the results we can obtain by aggregating random walks using Laplace associated space.

There is a fundamental distinction to make between the study of compartmental propagation and its approximation by aggregation of random walks. This difference lies in the correlations between the compartmental states of nodes sharing a common neighbor. While these correlations exist in the case of a unique and global propagation such as performed in the percolation process, they are destroyed when independent random walks are aggregated to approximate the propagation.

Finally, numerical simulations on a controlled network enlighten the impact of the variance of the waiting time distribution leading the propagation on a network. Results show that the average maximal depth and the average size of the epidemic in a given time increase when the variance increases.

Avant-propos

La réalisation de ce mémoire a été permise grâce au soutien de plusieurs personnes à qui je voudrais témoigner toute ma reconnaissance.

Je tiens avant tout à remercier mon promoteur, Renaud Lambiotte, pour avoir accepté de m'encadrer pour ce mémoire, pour les lectures qu'il m'a conseillées, ses conseils, sa disponibilité et ses relectures précieuses.

Je remercie aussi ma maman qui a courageusement relu plusieurs versions provisoires de ce mémoire, apportant à chaque fois des corrections pertinentes.

Enfin, je remercie mes proches, mes cokoteurs et mes condisciples pour le support moral qu'ils m'ont prodigués, parfois consciemment, parfois à leur insu.

Table des matières

Résumé	i
Abstract	ii
Avant-propos	iii
Introduction	1
1 Propagation dans un graphe statique, théorie de la percolation et nécessité d'aller au-delà	3
1.1 Théorie de la percolation	4
1.2 Propagation d'épidémie dans des graphes statiques	7
1.2.1 Graphe complet	7
1.2.2 Graphe quelconque	10
1.3 Analogie entre les modèles de propagation et la théorie de la percolation dans le modèle <i>small-world</i>	13
1.3.1 Percolation par noeuds	14
1.3.2 Percolation par arêtes	17
1.3.3 Percolation mixte	18
1.4 Au-delà des processus de Markov	19
1.5 Caractéristiques des réseaux sociaux	21
1.6 Destruction des corrélations	23
2 Propagation dans des graphes dynamiques	27
2.1 Propagations	28
2.1.1 Force dynamique des liens	28
2.1.2 Processus de Bellman-Harris	30
2.2 Marche aléatoire	36
2.2.1 Détermination de la densité de marcheurs	37
2.2.2 Equation intégrro-différentielle	39
2.3 Différences entre marche aléatoire et propagation <i>SI</i>	41

2.3.1	Variation sur les résultats obtenus avec la marche aléatoire . .	44
3	Application numérique	45
3.1	Motivations	45
3.2	Description du processus	46
3.3	Justification des choix	48
3.4	Résultats : description	50
3.4.1	Processus Markovien	50
3.4.2	Etude en fonction de la moyenne	52
3.4.3	Etude en fonction de la variance	55
3.4.4	Etude en fonction de la moyenne et de la variance	67
3.5	Résultats : conclusions	72
	Conclusion	73
	Bibliographie	75
	Annexes	77

Introduction

La théorie des graphes connaît depuis une dizaine d'années un intérêt croissant de la part de domaines aussi variés que les sciences sociales, la biologie, l'informatique ou les mathématiques appliquées, pour n'en citer que quelques-uns. En effet, que l'on s'intéresse aux connexions cérébrales, à la communication humaine sous l'ère Internet ou à la propagation de maladies ou de virus informatiques, le problème peut être modélisé grâce à la théorie des graphes. Son approche consiste à représenter chaque élément du système complexe qui interagit par un noeud, et chaque interaction entre deux de ces éléments par une arête reliant les noeuds correspondants.

Cependant, l'approche classique se concentre principalement sur les propriétés intrinsèques des réseaux étudiés et néglige la composante dynamique du problème. Or la plupart des réseaux sont sujets à l'apparition et à la disparition d'arêtes au cours du temps. Négliger cette composante nous restreint alors à nous concentrer sur des instantanés du réseau, comme si nous nous focalisions sur une photo au lieu de considérer le film dans son entièreté. C'est pour cette raison que se développe une théorie qui prend en compte la composante dynamique des réseaux.

Dans cette optique, nous nous intéressons dans ce mémoire à des processus stochastiques sur graphes tels que la propagation d'information, de maladie ou encore les marches aléatoires. Nous proposons ainsi dans un premier temps une revue de la littérature actuelle sur un sujet de recherche moderne en plein essor, ainsi qu'une mise en perspective de différents travaux à ce sujet.

Le premier chapitre commence par une présentation de la théorie des propagations dans le cas de graphes statiques ainsi que l'analogie qui existe entre celle-ci et la théorie de la percolation. Nous y justifions ensuite la nécessité de prendre en compte l'aspect dynamique des graphes et de rejeter ipso facto les hypothèses de Markov. Après avoir présenté quelques caractéristiques sous-jacentes des réseaux sociaux humains, nous cloturons ce chapitre par la présentation de différentes méthodes de

manipulations de données qui permettent de supprimer différentes corrélations dans des processus stochastiques sur réseaux afin de mettre en évidence l'impact de celles-ci.

Dans le deuxième chapitre, nous nous intéressons à deux types de processus stochastiques sur réseaux utilisés afin de modéliser les propagations. Le premier consiste à d'abord développer une nouvelle matrice, différente de la matrice d'adjacence, caractérisant la force dynamique des liens, pour ensuite y simuler une propagation d'épidémie via un processus de branchement de Bellman-Harris, généralisation du phénomène de percolation. Le second processus étudié est celui de marche aléatoire. Nous concluons ce chapitre en exposant des différences fondamentales entre modèles épidémiques et marches aléatoires.

Le troisième et dernier chapitre présente des simulations numériques dont l'objectif est d'identifier le rôle de la variance de la distribution de probabilités du temps d'infection sur la taille et la profondeur maximale moyenne d'une épidémie dans un environnement connu et contrôlé.

Chapitre 1

Propagation dans un graphe statique, théorie de la percolation et nécessité d'aller au-delà

Le phénomène de percolation par arêtes ou par noeuds possède une analogie très intéressante avec les modèles de propagation de maladie dans des réseaux. Nous présentons d'abord la notion de seuil de transition pour le phénomène de percolation. Ensuite, nous introduisons la théorie de propagation¹ dans des graphes, d'abord complets puis quelconques. Par après, nous illustrons comment la théorie de la percolation permet de mener à bien une analyse de propagation d'une épidémie sur des réseaux de type *small-world*. Dans un second temps, nous introduisons la notion d'hypothèse de Markov, l'importance qu'il y a de rejeter ce type d'hypothèse et mettons en avant quelques caractéristiques des réseaux sociaux humaines. Enfin, nous décrivons brièvement quelques méthodes de manipulation de données afin de mesurer l'impact de différentes corrélations topologiques ou temporelles sur un réseau donné.

1. Nous parlerons de modèles de propagation de maladie, mais ce concept doit s'entendre au sens large car il s'applique également à la propagation de l'information ou de la peur par exemple.

1.1 Théorie de la percolation

Les développements présentés dans cette section ainsi que dans la section suivante sont inspirés du livre [1].

Une manière intuitive de se représenter le phénomène de percolation par arêtes (respectivement par noeuds) est la suivante : considérons un graphe quelconque et supposons qu'à chaque seconde soit coloriée aléatoirement une arête (resp. un noeud). La question que l'on se pose alors est de savoir sous quelle condition on peut s'attendre à voir l'émergence d'un sous-graphe connexe, ou *cluster*, géant coloré, i.e. plus grand que tous les autres et du même ordre de grandeur que le système entier. Ce passage d'un grand nombre de petits groupes à un grand sous-groupe est appelé transition de percolation. L'objectif est alors de déterminer le seuil de percolation, i.e. sous quel(s) critère(s) se produit cette transition.

Illustrons ce concept avec un graphe appelé graphe aléatoire d'Erdős-Rényi. Il est constitué d'un ensemble de n noeuds dont chaque paire est reliée avec une probabilité p . La probabilité qu'un noeud soit de degré k , p_k , est donnée par :

$$\begin{aligned} p_k &= C_n^k p^k (1-p)^{n-k} \\ &\approx \frac{(np)^k e^{-np}}{k!} \text{ pour } n \text{ grand} \end{aligned}$$

Pour un grand nombre de noeuds, la distribution de degrés suit donc une loi de Poisson de paramètre $\lambda = np$. C'est pourquoi ce graphe est également appelé parfois graphe aléatoire de Poisson.

La proportion S du cluster géant par rapport au réseau entier peut être calculée en fonction de p . Nommons u la proportion de noeuds n'appartenant pas au sous-graphe géant. La probabilité qu'un noeud n'appartienne pas au cluster géant est égale à celle qu'aucun de ses voisins n'appartienne au graphe géant, soit u^k pour un noeud de degré k .

D'où

$$\begin{aligned}u &= \sum_{k=0}^{\infty} p_k u^k \\&= \sum_{k=0}^{\infty} (np)^k e^{-np} \frac{u^k}{k!} \\&= e^{-np} \sum_{k=0}^{\infty} \frac{(np u)^k}{k!} \\&= e^{-np} e^{np u} \\&= e^{-np(1-u)}\end{aligned}$$

Puisque $S = 1 - u$, on en déduit que S est solution de

$$S = 1 - e^{-npS}.$$

- Si $np < 1$, alors la seule solution positive correspond à $S = 0$.
- Si $np > 1$, alors il existe une solution strictement positive qui correspond à la taille du cluster géant. Celle-ci est donnée en fonction du produit np sur la figure 1.1.

Ainsi, le seuil de percolation se situe à $np = 1$.

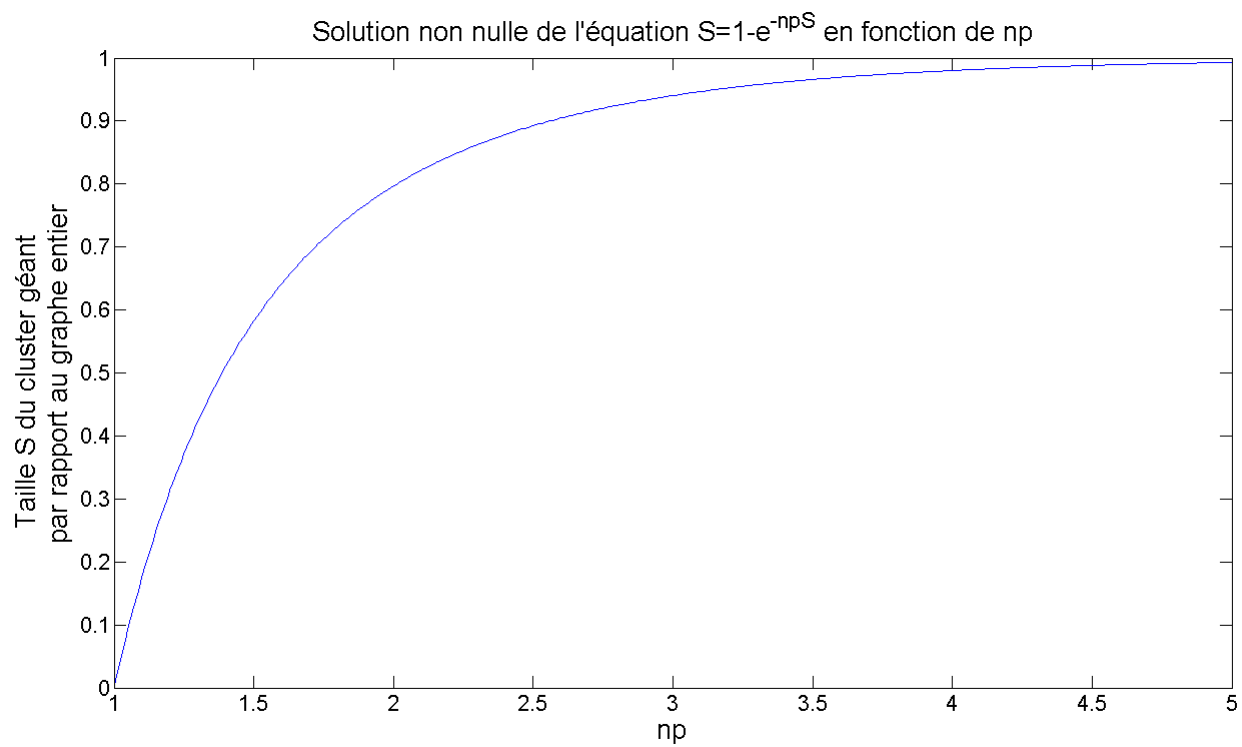


FIGURE 1.1 – Taille du cluster géant lorsque $np > 1$.

1.2 Propagation d'épidémie dans des graphes statiques

Nous nous intéressons maintenant aux modèles de propagation dans les réseaux. Nous envisageons dans ce chapitre la propagation dans des graphes statiques. L'aspect dynamique des graphes sera pris en compte dans les chapitres suivants.

1.2.1 Graphe complet

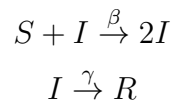
Dans un modèle de propagation sur réseau, chaque individu est représenté par un noeud et chaque contact entre deux individus par une arête. Nous nous plaçons ici sous l'hypothèse où la population est distribuée de manière homogène, ce qui signifie que tous les individus interagissent aléatoirement (selon un processus de Poisson) entre eux avec la même probabilité (hypothèse *fully-mixed*). Cette société est représentée via un graphe complet, graphe dont tous les sommets sont reliés deux à deux par une arête.

La modélisation se fait alors via un modèle déterministe dans lequel les individus sont séparés dans différentes classes selon leur état par rapport à la maladie. Les classes peuvent également dépendre des caractéristiques comportementales ou intrinsèques des individus ([2]). Les changements d'états sont alors gouvernés par un ensemble d'équations différentielles. Nous n'expliquons ici que le modèle compartimental SIR qui est le plus courant.

Dans le modèle SIR, la population est divisée en trois catégories : susceptible (S), infectée (I) et rétablie (R)². Ce modèle fonctionne sous les hypothèses suivantes :

- une fois rétabli, un individu acquiert une immunité face à la maladie ;
- l'échelle de temps d'une épidémie est négligeable par rapport à la durée de vie moyenne d'un individu, ce qui nous permet de négliger les processus démographiques tels que la naissance ou la mort ;
- personne ne possède initialement d'immunité face à la maladie.

Le modèle SIR est alors régi par 2 réactions. La première a lieu lorsqu'un individu infecté contamine un susceptible, avec un taux de transmission β qui représente le taux de contamination par individu. La seconde consiste en la guérison spontanée d'un individu infecté, à un taux moyen de rémission γ :



2. S , I et R représentent des pourcentages de la population.

On en tire le système d'équations différentielles suivant :

$$\frac{dS}{dt} = -\beta SI \quad (1.1)$$

$$\frac{dI}{dt} = \beta SI - \gamma I \quad (1.2)$$

$$\frac{dR}{dt} = \gamma I \quad (1.3)$$

Ce système obéit en outre à la contrainte de conservation de la population, que l'on exprime par : $S + I + R = 1$.

On peut intégrer numériquement ce système, comme le montre la figure 1.2.

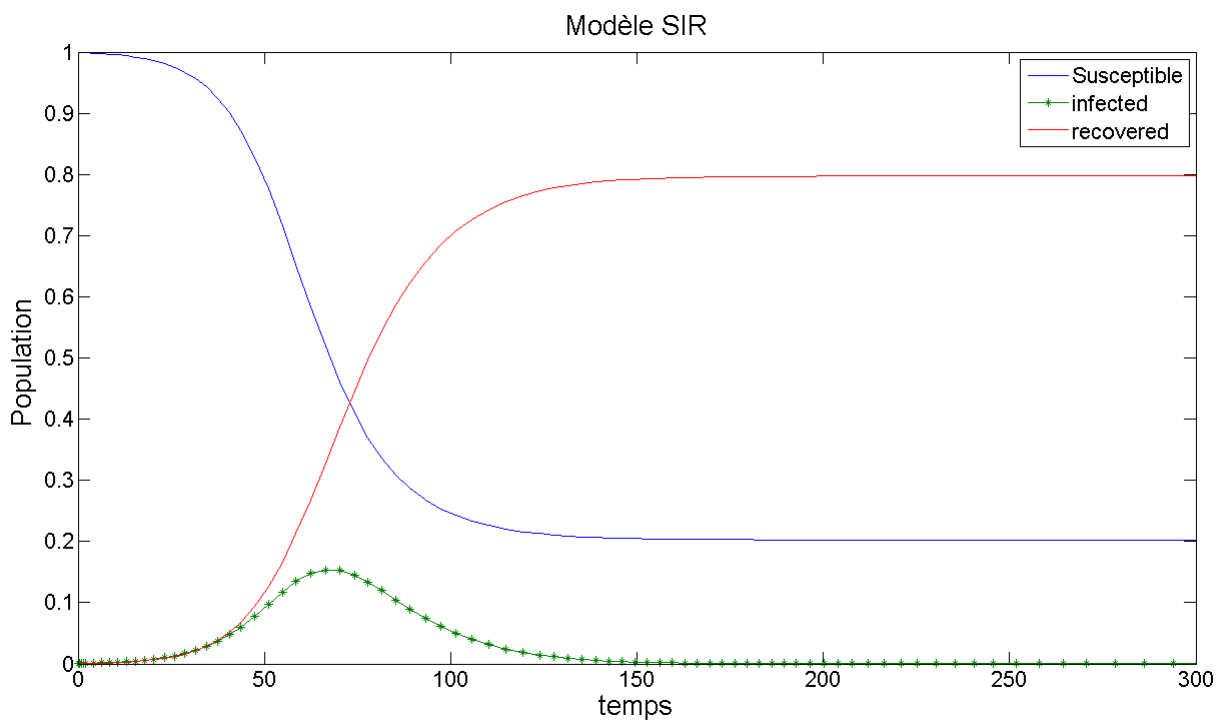


FIGURE 1.2 – Modèle SIR avec $I(0) = 0.01, \gamma = 0.1$ et $\beta = 0.2$

Après un certain temps, le nombre de personnes infectées devient nul. Nous pouvons alors calculer le nombre de personnes touchées par la maladie lors de l'épidémie, noté R_∞ .

En combinant les équations (1.1) et (1.3), on obtient

$$\frac{1}{S} \frac{dS}{dt} = \frac{-\beta}{\gamma} \frac{dR}{dt}$$

Nous pouvons intégrer de part et d'autre, et obtenir l'expression de S en fonction d'une exponentielle de R :

$$S = S(0)e^{\frac{-\beta}{\gamma}R}.$$

En réinjectant l'expression calculée pour S , nous obtenons une équation différentielle pour R :

$$\frac{dR}{dt} = \gamma(1 - R - S(0)e^{\frac{-\beta}{\gamma}R}).$$

Nous pouvons faire l'approximation $S(0) \simeq 1$.

A la fin de l'épidémie, nous nous trouvons à un point d'équilibre, ce qui implique $\frac{dR}{dt} = 0$.

On en tire R_∞ comme étant la solution de :

$$R_\infty = 1 - e^{\frac{-\beta}{\gamma}R_\infty}$$

Cette équation est la même que celle qui détermine la taille du cluster géant d'un graphe aléatoire de Poisson avec $\frac{\beta}{\gamma} = np$. Le seuil épidémique de la maladie est donc l'équivalent du seuil de percolation.

Le coefficient $R_0 := \frac{\beta}{\gamma}$ est appelé coefficient de reproduction. Il représente le nombre de personnes qu'un nouvel infecté va, en moyenne, contaminer avant sa guérison.

La figure 1.2.1 illustre nos conclusions précédentes :

- Si $R_0 > 1$, on a une épidémie ;
- Si $R_0 < 1$, la maladie ne se propage pas assez rapidement par rapport au temps de guérison pour provoquer une épidémie ;
- $R_0 = 1$ est la valeur critique, ou seuil épidémique.

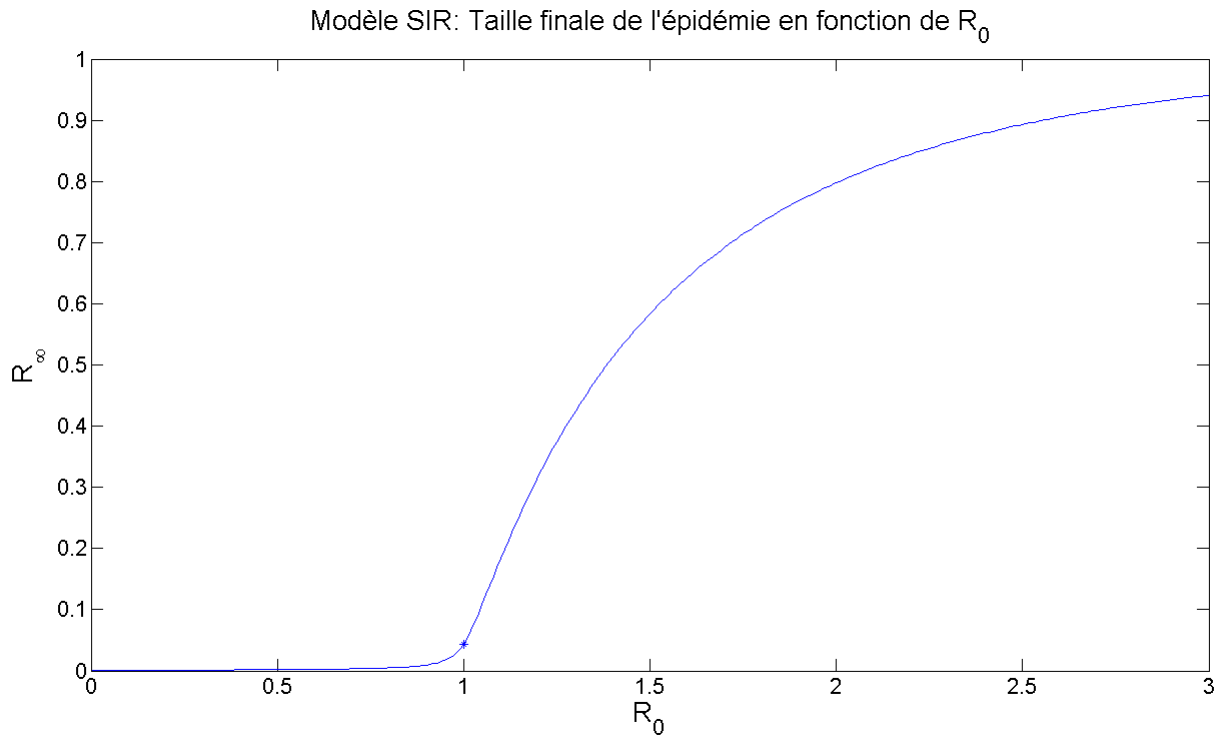


FIGURE 1.3 – Etude du nombre total d’infectés en fonction de R_0 . Le seuil épidémique apparaît pour une valeur de $R_0 = 1$.

1.2.2 Graphe quelconque

Le cadre de travail de la section précédente se place sous l’hypothèse d’une population homogène. Cependant, cette hypothèse n’est pas réaliste car la structure de notre société ne nous amène pas à cotoyer chacun de nos semblables avec une même probabilité. Nous pouvons représenter les contacts potentiels de chacun via un graphe et la structure de celui-ci peut avoir des conséquences sur la propagation d’une maladie et l’émergence ou non d’une épidémie. Dans ce cadre, le taux de transmission β représente le taux de contact entre un individu infecté et un autre, contrairement au cas précédent où il représentait le taux de contact entre un individu infecté et tous les autres. En d’autres mots, β représente le taux d’infection par arête et plus le taux d’infection total du réseau.

Dans le cas où la propagation part d’un foyer unique, la taille maximale de l’épidémie est celle du cluster du foyer de départ. Celle-ci est négligeable si le foyer n’appartient pas à l’éventuelle composante géante. Nous nous intéressons donc aux cas où le foyer

est un élément du cluster géant.

Etant donné β , nous pouvons calculer la durée τ durant laquelle un individu reste sain alors que son voisin est infecté. La probabilité que ce dernier transmette la maladie sur un intervalle de temps donné $\delta\tau$ est $\beta\delta\tau$. Donc, la probabilité que l'individu ne soit pas contaminé après un temps τ est donnée par

$$\lim_{\delta\tau \rightarrow 0} (1 - \beta\delta\tau)^{\frac{\tau}{\delta\tau}} = e^{-\beta\tau}.$$

Dès lors la probabilité ϕ que le maladie soit transmise est

$$\phi = 1 - e^{-\beta\tau}.$$

Une manière astucieuse d'aborder le problème est d'utiliser la percolation par liens. Colorions en effet chaque arête avec une probabilité ϕ . Les arêtes coloriées sont alors celles à travers lesquelles la maladie sera transmise : si l'une de deux extrémités de l'arête est infectée, alors la seconde le sera également. Le graphe colorié représente dès lors la transmission potentielle de la maladie.

Le seuil de percolation du réseau correspond donc au seuil épidémique de la maladie sur ce réseau, où le taux d'occupation des arêtes ϕ est donné en fonction du taux de contamination β et du temps de rémission τ de la maladie. Cependant, il est important de noter que même si ϕ est au-dessus du seuil épidémique, cela ne garantit pas qu'il y aura une épidémie. En effet, si le seuil de percolation est atteint, alors une composante géante occupe une fraction S du réseau entier. Le foyer de la maladie a alors une probabilité S d'apparaître dans le cluster géant et se répand une épidémie touchant une même fraction S de la population. Plus ϕ augmente, plus à la fois la probabilité d'apparition d'une épidémie et sa taille augmentent.

Par ailleurs, il est important de noter que le processus de percolation est un processus stochastique et non déterministe : partant de mêmes conditions initiales, le bilan final de l'épidémie ne sera pas nécessairement le même, car le coloriage des noeuds est réalisé de manière aléatoire, afin de traduire la nature aléatoire des contacts qui entraînent une transmission de la maladie. Nous ne pouvons dès lors plus calculer de manière exacte la taille d'une épidémie mais seulement son espérance ou la probabilité qu'un noeud soit infecté.

Notons A la matrice d'adjacence du graphe, $s_i(t)$, $x_i(t)$ et $r_i(t)$ les probabilités qu'un noeud soit respectivement susceptible, infecté et rétabli au temps t .

Alors le système d'équations différentielles devient :

$$\frac{ds_i}{dt} = -\beta s_i \sum_j A_{ij} x_j \quad (1.4)$$

$$\frac{dx_i}{dt} = \beta s_i \sum_j A_{ij} - \gamma x_i \quad (1.5)$$

$$\frac{dr_i}{dt} = \gamma x_i \quad (1.6)$$

Comportement aux premiers instants

Aux premiers instants ($t \approx 0$), nous pouvons approximer la probabilité d'être infecté à 0 et celle d'être susceptible à 1.

L'équation (1.5) donne donc

$$\frac{dx_i}{dt} = \sum_j (\beta A_{ij} - \gamma \delta_{ij}) x_i$$

Nous pouvons réécrire cette équation d'un point de vue matriciel :

$$\frac{dx}{dt} = \beta M x,$$

où $M = A - \frac{\gamma}{\beta} I$.

La matrice M est une perturbation de l'identité de la matrice A . Elle possède donc les mêmes vecteurs propres que A et ses valeurs propres sont celles de A translatées de $\frac{\gamma}{\beta}$. En effet, si v_r est un vecteur propre de A associé à la valeur propre κ_r , alors

$$\begin{aligned} M v_r &= A v_r - \frac{\gamma}{\beta} I v_r \\ &= (\kappa_r - \frac{\gamma}{\beta}) v_r \end{aligned}$$

En décomposant x dans la base des vecteurs propres de A , nous obtenons :

$$x(t) = \sum_{r=1}^n a_r(0) v_r e^{(\beta \kappa_r - \gamma)t}$$

La croissance sera alors dominée par la plus grande valeur propre positive κ_1 . Cependant, si la valeur de κ_1 est suffisamment petite, il est possible qu'il n'y ait pas d'épidémie. Le seuil épidémique se situe en effet à

$$\frac{\beta}{\gamma} = \frac{1}{\kappa_1}.$$

Nous voyons donc que le seuil épidémique ne dépend plus uniquement des taux β et γ mais aussi de la matrice d'adjacence A via sa plus grande valeur propre κ_1 . Plus celle-ci est grande, et plus la probabilité que l'épidémie se répande augmente. Ceci est tout à fait intuitif, car les grandes valeurs de κ_1 correspondent à des matrices denses, alors que de faibles valeurs de κ_1 correspondent à des matrices creuses.

1.3 Analogie entre les modèles de propagation et la théorie de la percolation dans le modèle *small-world*

Afin d'illustrer encore l'analogie entre la théorie de la percolation et celle des épidémies, nous présentons maintenant un développement effectué dans [3]. Considérons un réseau de type *small-world*, *i.e* soumis à l'effet *small-world*, qui est le fait que le degré de séparation entre deux individus quelconques du réseau est très petit comparé à la taille du réseau. Ces réseaux sont typiquement composés de plusieurs clusters fortement connectés et reliés entre eux. Dans ce type de réseau, les épidémies peuvent se propager très vite. L'effet *small-world* est réaliste par rapport à notre univers social, les expériences de Milgram suggérant que deux individus sont reliés en moyenne par une chaîne sociale de 6 individus ([4]).

Considérons un réseau où les individus sont disposés sur une sphère (souvent de dimension 1) où chaque noeud est relié à un petit nombre de ses voisins. Une petite densité de raccourcis sont ajoutés aléatoirement sur le réseau. Nous avons montré qu'une épidémie peut survenir si le système a atteint ou dépassé son seuil de percolation. Nous allons montrer une solution exacte de la fraction p_c d'individus infectés par la maladie en fonction de la densité des raccourcis sur le réseau. Nous allons également traiter le cas où la probabilité d'avoir une transmission entre deux individus est inférieure à 100%, ce qui peut être modélisé par un processus de percolation par arêtes.

Définissons deux paramètres qui nous intéresseront pour la suite : la transmissibilité et la susceptibilité.

- La transmissibilité est la probabilité qu'un individu susceptible contracte la maladie au contact d'un individu infecté.
- La susceptibilité est la probabilité qu'un individu exposé à la maladie la contracte.

Un noeud est colorié si l'individu est susceptible et un lien l'est s'il laisse passer la maladie. Dès lors, déterminer s'il y aura une épidémie revient à résoudre, sur un graphe *small-world*, le problème standard de percolation par noeuds et/ou arêtes, à savoir déterminer la proportion de noeuds et/ou arêtes à colorier pour qu'émerge une composante géante.

1.3.1 Percolation par noeuds

Considérons dans un premier temps que la transmissibilité vaut 100%, mais que moins de 100% des individus sont susceptibles. La proportion p de noeuds coloriés correspond à la susceptibilité.

Considérons un graphe *small-world* de N noeuds dans lequel tous les noeuds de distance inférieure à k sont reliés. Ajoutons-y aléatoirement des liens entre des paires de noeuds pris aléatoirement que nous appellerons des raccourcis. Nommons ϕ le nombre moyen de raccourcis par noeuds, et ψ la probabilité que deux noeuds pris aléatoirement soient reliés.

Nous pouvons exprimer ψ en fonction de k , ϕ et N .

En effet, pour chaque arête, la probabilité qu'elle ne relie pas deux noeuds donnés vaut $1 - \frac{2}{N^2}$. Comme il y a au total $k\phi N$ arêtes, la probabilité ψ est donnée par

$$\psi = 1 - \left(1 - \frac{2}{N^2}\right)^{k\phi N}.$$

Nous pouvons approximer cette égalité pour un grand nombre de noeuds en utilisant l'expansion de Taylor en $x := \frac{1}{L}$ autour de 0 :

$$\psi(x) = \lim_{x \rightarrow 0} \psi(x) + \left(\lim_{x \rightarrow 0} \frac{d\psi}{dx}(x) \right) x + O(x^2)$$

$$\begin{aligned} \lim_{x \rightarrow 0} \psi(x) &= 1 - e^{k\phi \lim_{x \rightarrow 0} \frac{\ln(1-2x^2)}{x}} \\ &= 1 - e^{k\phi \lim_{x \rightarrow 0} \frac{-4x}{1-2x^2}} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{d\psi}{dx}(x) &= \lim_{x \rightarrow 0} k\phi \left[\frac{\ln(1-2x^2)}{x^2} (1-2x^2)^{\frac{k\phi}{x}} + 4(1-2x^2)^{\frac{k\phi}{x}-1} \right] \\ &= k\phi \left[e \lim_{x \rightarrow 0} k\phi \frac{\ln(1-2x^2)}{x} \lim_{x \rightarrow 0} \frac{\ln(1-2x^2)}{x^2} + 4e \lim_{x \rightarrow 0} \left(\frac{k\phi-x}{x} \right) \ln(1-2x^2) \right] \\ &= k\phi \left[e \lim_{x \rightarrow 0} \frac{-4k\phi x}{1-2x^2} \lim_{x \rightarrow 0} \frac{-2}{1-2x^2} + 4e \lim_{x \rightarrow 0} \left(-\ln(1-2x^2) - \frac{(k\phi-x)4x}{1-2x^2} \right) \right] \\ &= k\phi(-2+4) \\ &= 2k\phi \end{aligned}$$

Donc $\psi(x) = 0 + 2k\phi x + o(x^2)$.

Ce développement nous permet donc de faire l'approximation suivante pour un N grand :

$$\psi(N) \approx \frac{2k\phi}{N}.$$

La situation est donc typique de celle de clusters présentant le phénomène *small-world* : le réseau est constitué de composantes locales de personnes susceptibles reliées par des raccourcis.

Intéressons-nous à présent au nombre moyen de clusters de longueur i , N_i . Pour qu'un noeud soit dans un cluster de longueur i , il faut qu'un noeud soit susceptible, que ses $i-1$ voisins soient susceptibles, et que les $2k$ noeuds reliés au bord du cluster

ne soient pas susceptibles. Donc N_i est donné par

$$\begin{aligned} N_i &= (1-p)^{2k} p (1 - (1-p)^k)^{i-1} N \\ &= (1-q)^2 p q^{i-1} N, \end{aligned}$$

avec $q = 1 - (1-p)^k$.

Considérons l'agrégation de ces petits clusters en un cluster géant. Partant d'un cluster, on le relie à tous ceux qui sont accessibles via un raccourci. A l'étape suivante, les clusters qui sont devenus accessibles sont ajoutés, et ainsi de suite. A chaque étape est défini le vecteur v dont chaque composante v_i correspond à la probabilité qu'un cluster de taille i vienne d'être ajouté au cluster géant en formation. Déterminons la valeur de v à une étape, v' , en fonction de sa valeur à l'étape précédente. Sous le seuil de percolation, les composantes de v sont faibles, donc la dépendance des composantes de v' est linéaire, suivant une matrice de transition M . Chaque composante de taille i a une probabilité d'avoir un raccourci avec un cluster de taille j égale à $1 - (1-\psi)^{ij}$ car il y a ij paires possibles pour relier les clusters.

D'où

$$M_{ij} = N_i (1 - (1-\psi)^{ij}).$$

Une nouvelle fois, nous considérons la plus grande valeur propre de cette matrice M , κ_1 . Il y a un seuil de percolation pour $\kappa_1 = 1$.

Si $\kappa_1 < 1$, alors v tend exponentiellement vers 0. Le taux de nouveaux clusters ajoutés décroît exponentiellement et les composantes connectées sont de tailles finies et distribuées exponentiellement.

Si $\kappa_1 > 1$, alors v croît exponentiellement jusqu'à la taille du système complet. En général, déterminer la plus grande valeur propre de manière exacte n'est pas évident. Cependant, pour un N grand, ψ tend vers 0. Donc, utilisant une nouvelle fois l'approximation de Taylor du premier ordre, nous pouvons approximer M via

$$M_{ij} = ij\psi N_i.$$

Nous pouvons alors déterminer les composantes du vecteur propre de M associé à κ_1 :

$$\begin{aligned} \kappa_1 v_i &= \sum_j M_{ij} v_j \\ &= \sum_j ij\psi N_i v_j \\ &= i\psi N_i \sum_j j v_j \end{aligned}$$

En sommant sur les composantes, on en tire la valeur de κ_1 :

$$\begin{aligned}
\kappa_1 &= \psi \sum_i i^2 N_j & (1.7) \\
&= \frac{2k\phi}{N} Np \frac{(1-q)^2}{q} \sum_i i^2 q^i \\
&= 2k\phi p \frac{(1-q)^2}{q} \frac{q(q+1)}{(1-q)^3} \\
&= 2k\phi p \frac{q+1}{1-q} \\
&= 2k\phi p \frac{2 - (1-p)^k}{(1-p)^k}
\end{aligned}$$

Nous savons que le seuil se situe pour $\kappa = 1$. Dès lors la valeur seuil de ϕ est donnée par

$$\phi = \frac{(1-p)^k}{2kp(2 - (1-p)^k)}.$$

La proportion de noeuds coloriés au seuil épidémique p_c est donc donnée par une racine du polynôme Pol de degré $k + 1$:

$$Pol(p) = -4k\phi p - (2kp\phi + 1)(1-p)^k$$

1.3.2 Percolation par arêtes

Considérons dans un second temps que ce soit cette fois la susceptibilité qui vaille 100%, mais que la transmission entre un individu infecté et un sain ne soit pas garantie à 100%. La proportion d'arêtes coloriées corespond alors à la transmissibilité. Nous ne traiterons que le cas où k vaut 1, car la détermination de N_i est beaucoup plus compliquée.

D'une part, un cluster composé de i noeuds est composé de $i - 1$ arêtes coloriées et de deux arêtes non-coloriées sur sa frontière. Donc, le nombre total de clusters de taille i est augmenté d'un facteur p^{-1} par rapport au cas de la percolation par noeuds et vaut

$$N_i = (1-p)^2 p^{i-1}.$$

D'autre part, la probabilité d'avoir un raccourci entre deux noeuds possède un facteur supplémentaire p , car pour qu'il soit pris en compte, il faut qu'il soit colorié. Ce qui signifie qu'il faut considérer le terme $p\psi$ au lieu de ψ .

Dès lors, nous obtenons la même expression pour le calcul de κ_1 , et donc la même proportion d'arêtes colorées p_c au seuil, que l'on peut aisément calculer pour $k = 1$ (puisque'il s'agit d'une racine d'une équation du second degré) :

$$p_c = \frac{\sqrt{4\phi^2 + 12\phi + 1} - 1 - 2\phi}{4\phi}.$$

1.3.3 Percolation mixte

Le type de propagation le plus courant est celui où à la fois la transmissibilité et la susceptibilité sont inférieures à 100%.

Notons p_{noeud} (respectivement p_{lien}) la densité de coloriage des noeuds (resp. arêtes), et plaçons-nous à nouveau dans le cadre $k = 1$.

Un cluster de taille i est alors composé de i noeuds coloriés reliés par $i - 1$ arêtes colorières. Dès lors, le nombre total de clusters de taille i est donné par

$$N_i = (1 - p_{noeud}p_{lien})^2 p_{noeud}^i p_{lien}^{i-1}.$$

Nous pouvons alors réutiliser l'équation (1.7) en remplaçant ψ par ψp_{lien} pour recalculer la valeur de la plus grande valeur propre de la matrice de transition M et obtenir

$$\begin{aligned} \kappa_1 &= \psi p_{lien} \sum_j j^2 N_j \\ &= 2\phi p \frac{1+p}{1-p} \end{aligned}$$

où $p = p_{noeud}p_{lien}$.

Donc, en d'autres termes, la position du seuil de percolation est la même que pour les cas précédents, mis à part qu'elle représente le produit des densités de coloriage des arêtes et celle des noeuds.

1.4 Au-delà des processus de Markov

Dans le modèle SIR que nous avons traité dans ce chapitre, nous supposons que les taux de transmission et de rémission de la maladie étaient constants. Sous cette hypothèse, nous avons pu écrire des équations différentielles pour l'évolution du SIR. Cependant, dans la réalité, les durées de transmission et guérison sont relativement concentrées autour d'une moyenne et suivent dès lors une distribution qui est loin d'être exponentielle, ce qui indique que les hypothèse que nous avons faites, dites de Markov, ne sont pas vérifiées.

Dans un processus de Markov, la probabilité qu'un évènement se produise ne dépend que de l'état présent, et pas du passé : il n'y a aucune mémoire de l'état du système avant le temps observé ("d'où" l'état actuel provient) ni du délai écoulé depuis le dernier changement ("depuis quand" le système est dans cet état). Pour des processus en temps continu tels que ceux que nous étudions, si on suppose que les probabilités de transmission et de guérison sont constants, alors la probabilité qu'un évènement se produise durant un certain intervalle suit une distribution exponentielle.

Comme montré dans [5], nous pouvons montrer que les résultats obtenus sont fondamentalement différents lorsque nous rejetons les hypothèses de Markov. En effet, si nous définissons, dans ce cas, d'une part $s(\tau)d\tau$ la probabilité qu'un individu infecté crée un contact suffisant pour transmettre la maladie à un voisin en particulier en un temps compris entre τ et $\tau + d\tau$, et d'autre part $r(\tau)d\tau$ la probabilité qu'un individu infecté guérisse dans un délai compris entre τ et $\tau + d\tau$ après infection, nous pouvons écrire la probabilité totale de transmission sur l'intervalle $[\tau, \tau + d\tau]$, notée $f(\tau)d\tau$. En effet, pour qu'il y ait transmission entre les temps τ et $\tau + d\tau$, il faut que l'individu infecté le soit toujours au moment où le contact a été suffisant pour qu'il y ait contamination. Cette probabilité est alors égale à $s(\tau)d\tau$ multipliée par la probabilité que l'individu infecté ne soit pas encore rétabli :

$$\begin{aligned} f(\tau)d\tau &= s(\tau)d\tau \left(1 - \int_0^\tau r(\tau') d\tau'\right) \\ &= s(\tau)d\tau \int_\tau^{+\infty} r(\tau') d\tau' \end{aligned}$$

La transmissibilité, probabilité que le noeud transmette la maladie à ses voisins avant de guérir, est donnée par l'intégrale de $f(\tau)d\tau$.

Nous souhaitons maintenant pouvoir déterminer des données épidémiologiquement intéressantes telles que la proportion de personnes présentes dans chaque compartiment ou la probabilité qu'a un individu d'être dans un état donné.

Pour cela, intéressons-nous d'abord à la probabilité qu'un noeud j n'ait **pas** transmis

la maladie à un voisin i en un temps t , que nous notons $H^{ij}(t)$. Nous nous restreignons ici au cas où le graphe est simplement un arbre.

Il y a deux cas pour que cette transmission échoue :

- d’une part, il se peut que lorsque j contracte la maladie, il ne la transmette pas à i dans un intervalle de temps t après l’infection. Dans ce cas, i n’a pas eu le temps d’être contaminé. Cet événement arrive avec une probabilité de $1 - \int_0^t f(\tau) d\tau$;
- d’autre part, il se peut que j soit capable de contaminer i dans un délai inférieur à t , mais que j soit contaminé par l’un de ses autres voisins trop tard pour que la transmission puisse avoir lieu, voire jamais. Donc, si j peut contaminer i dans un délai τ , mais qu’il n’attrape pas la maladie avant le temps $t' = t - \tau$, alors i n’est pas contaminé avant un temps t . La probabilité que j ne soit pas contaminé avant un temps t' est donnée par $z \prod_{l \in N(j) \setminus i} H^{jl}(t')$, où z représente la probabilité que j n’était pas un noeud initialement infecté, et $N(j) \setminus i$ représente l’ensemble des voisins de j duquel est exclu i . Intégrer selon t' entre 0 et t fournit alors la probabilité de cette seconde possibilité.

Réunissant ces deux conditions, nous obtenons l’équation de transmission du message :

$$H^{ij}(t) = 1 - \int_0^t f(\tau) [1 - z \prod_{l \in N(j) \setminus i} H^{jl}(t - \tau)] d\tau$$

Dans notre cas particulier où le graphe est simplement un arbre, cette équation nous fournit une solution pour $H^{ij}(t)$ pour tous t et $f(\tau)$ arbitraires. La probabilité qu’un individu soit susceptible en un temps t , notée $P(S_i)$, est donnée par la probabilité que chacun de ses voisins ait échoué à lui transmettre la maladie, multipliée par la probabilité qu’il n’était pas initialement contaminé :

$$P(S_i) = z \prod_{j \in N(i)} H^{ij}(t).$$

Une fois connue cette probabilité, nous pouvons en déduire les probabilités qu’un individu soit infecté (resp. rétabli) en un temps t , notée $P(I_i)$ (resp. $P(R_i)$).

En effet, la variation de $P(I_i)$ est celle de $P(S_i)$ diminuée du taux de guérison des individus. Ce taux de guérison possède deux composantes : la possibilité que i était initialement infecté et soit guéri un temps t plus tard, et la possibilité que i ait été infecté en un temps $t' < t$ puis rétabli dans un délai $t - t'$. Cela nous fournit alors le taux de variation de $P(I_i)$:

$$\frac{dP(I_i)}{dt} = -\frac{dP(S_i)}{dt} - (1 - z)r(t) + \int_0^t r(t - t') \frac{dP(S_i)}{dt'} dt'.$$

Il suffit alors d'intégrer cette équation pour obtenir $P(I_i)$.
Finalement, nous obtenons $P(R_i)$ grâce au principe de conservation de la population :

$$P(R_i) = 1 - P(S_i) - P(I_i).$$

Les équations que nous avons obtenues pour décrire les probabilités d'appartenir à un compartiment donné diffèrent fondamentalement de celles obtenues sous l'hypothèse de Markov. En effet, d'une équation différentielle nous sommes passés à une équation intégro-différentielle dans laquelle apparaît une intégrale de 0 à t , ce qui signifie que l'évolution du système dépend d'événements du passé et plus uniquement du présent.

1.5 Caractéristiques des réseaux sociaux

Nous avons développé dans la section précédente l'importance de considérer l'historique du système pour développer des modèles plus réalistes dans le cadre de propagations. C'est l'une des causes du rejet des hypothèses de Markov lorsque l'on étudie des réseaux dirigés par des interactions humaines, mais ce n'est pas la seule. Nous allons ici brièvement citer une série non-exhaustive de caractéristiques des propagations dues au caractère humain sous-jacent du réseau, présentées notamment dans [6].

Burstiness et rythme circadien

L'activité humaine est décrite en anglais par le terme *burstiness*, difficilement traduisible en français. Ce terme exprime le fait que les humains ont tendance à agir de manière "explosive", sur un intervalle de temps très court avant de redevenir inactifs. Les différentes activités d'un individu ne sont donc pas espacées de manière constante dans le temps mais regroupées en "paquets" d'activités de durées assez courtes et très espacés (comme l'illustre la figure 1.4). Par exemple, une fois par jour un individu répond à tous ses e-mails. Cette haute volatilité de l'activité humaine a pour impact de ralentir à la fois la vitesse mais également la taille de la propagation de l'information ([7],[8]). Le rythme circadien signifie simplement que l'être humain vit selon des journées de 24h, ce qui a forcément un impact sur la distribution des temps d'activité et implique une certaine périodicité dans celle-ci (jour-nuit-jour-nuit-...).



FIGURE 1.4 – L’activité humaine est caractérisée par des périodes de grande activité séparées par de longues périodes d’inactivité.

Conversations de groupes

Les réseaux sociaux sont composés de communautés. Celles-ci se caractérisent par des noeuds fortement reliés avec ceux de leur communauté et beaucoup moins avec ceux extérieurs à leur communauté. Cela génère des conversations de groupes qui peuvent engendrer localement des cascades de contamination mais cette structure en communautés a pour effet de ralentir la propagation globale de l’information ou de la maladie ([8]).

Temporalité des liens

Les réseaux sociaux ne sont pas figés, nous tissons de nouveaux liens, en perdons d’autres, et en réactivons parfois certains. Cette temporalité dans les relations humaines peut être modélisée par une temporalité de l’activation des arêtes. Cette modification dynamique de la topologie du graphe a pour effet de ralentir la propagation([6]).

L’information n’est pas une maladie

Classiquement, les propagations d’information et de virus (biologiques ou informatiques) étaient traitées de manière similaire. Cependant, depuis quelques années les différences entre ces deux types de propagation sont mises en avant ([6], [9]). En effet, alors qu’un virus se transmet par interaction passive entre deux noeuds, la propagation d’information est décidée consciemment. La valeur de l’information est estimée par les acteurs, et sa transmission possède un certain coût perçu. Ces deux facteurs influencent le propogateur dans la décision de transférer ou non l’information, et à qui l’envoyer. Ses cibles peuvent à nouveau l’évaluer en fonction de la confiance qu’ils ont envers le propogateur et choisir de l’accepter ou non. Le mécanisme de décision de transmission de l’information est complexe car il peut être justifié par différents concepts selon les individus : altruisme, besoin d’identification ou de reconnaissance, évangélisation,...

1.6 Destruction des corrélations

Pour clôturer ce chapitre, nous proposons de présenter rapidement quelques méthodes utilisées pour manipuler un échantillon de données (tous les temps d'interactions entre tous les noeuds d'un réseau) afin de détruire diverses corrélations, et ainsi pouvoir étudier l'impact de celles-ci. Ces méthodes ont été utilisées dans les articles [10] et [11], qui sont à la base de la motivation de la simulation numérique étudiée dans ce mémoire.

Considérons connus les temps d'activation de chaque arête du réseau étudié (figure 1.5).

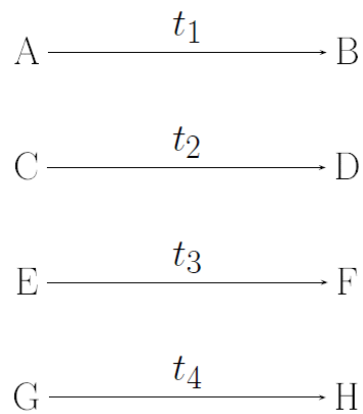


FIGURE 1.5 – Les relations d'origine reliant les différents noeuds.

Réseau statique

Afin d'annihiler l'effet de la *burstiness* du graphe, on considère que les arêtes peuvent s'activer de manière équiprobable selon une certaine distribution discrète déterminée par le rapport du nombre total d'arêtes sur le produit du nombre total de noeuds et de pas discrets sur l'intervalle étudié (figure 1.6).

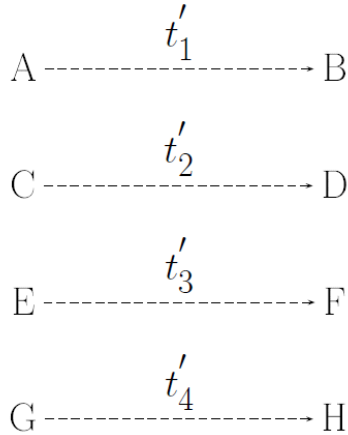


FIGURE 1.6 – On tire pour chaque arête un nouveau temps d'attente t'_i issu d'une distribution unique pour tout i .

Destruction de l'impact de la topologie (*Random topological*)

Une autre manipulation consiste à tirer aléatoirement des couples d'arêtes et à permuter l'une de leurs extrémités (figure 1.7). Cette manipulation a pour effet de détruire l'influence de la topologie du réseau en supprimant les communautés présentes mais conserve la dynamique du processus étudié.

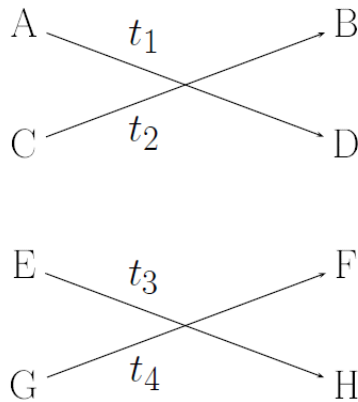


FIGURE 1.7 – On permute les extrémités de couples d'arêtes tirées aléatoirement.

Destruction de la dynamique (*Random dynamic* et *DCW*)

Pour supprimer l'impact de la dynamique du graphe et la *burstiness* des noeuds, les arêtes sont préservées mais leurs temps d'activation respectifs sont soit permutés deux à deux (figure 1.8), soit complètement mélangés(figure 1.9). Ainsi les corrélations entre les différents temps d'activation des arêtes relatives à un noeud donné sont supprimées mais la topologie du réseau est conservée.

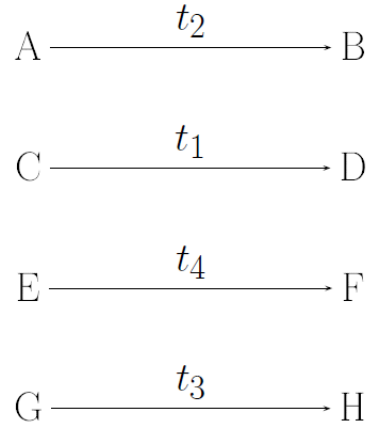


FIGURE 1.8 – RD : On permute deux à deux les temps correspondant à des arêtes tirées aléatoirement.

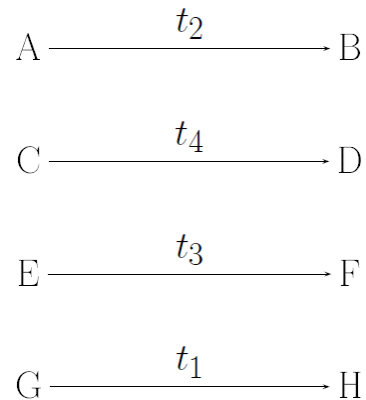


FIGURE 1.9 – DCW : On redistribue aléatoirement les temps d'activation.

Mélange total

On peut effectuer simultanément les destructions de l'impact de la topologie et de la dynamique pour supprimer les deux corrélations (figure 1.10).

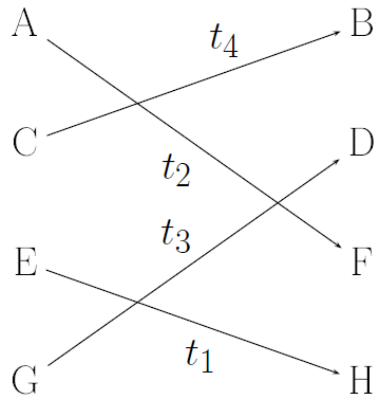


FIGURE 1.10 – On redistribue aléatoirement les temps d'activation et les extrémités des arêtes.

Chapitre 2

Propagation dans des graphes dynamiques

Dans un graphe statique, on peut connaître facilement l'importance d'une arête ou d'un noeud à partir de la matrice d'adjacence du graphe A . Pour une arête reliant les noeuds i et j , la valeur, notée s_{ij} , est donnée directement par l'élément A_{ij} ; pour un noeud k , il suffit de sommer tous les éléments de la ligne (ou la colonne) correspondant à ce noeud : $s_k = \sum_j A_{kj}$. Il existe d'autres mesures de l'importance d'un noeud qui sont toutes fonctions de la matrice d'adjacence, comme par exemple le *pageranking* ou l'*eigenvector centrality*.

Pour étudier la propagation dans un réseau dynamique, nous considérons que la propagation d'un noeud i à un noeud j adjacent s'effectue, et cela de manière instantanée, seulement lorsque l'arête reliant ces deux noeuds s'active. Nous supposons qu'une arête ne reste activée que durant un intervalle de temps infinitésimal.

Nous pourrions être tentés de considérer à nouveau simplement la matrice d'adjacence dans le cas dynamique mais cela n'est pas suffisant. En effet, la probabilité d'activation d'une arête sur un intervalle dt est alors indépendante du temps qui s'est écoulé depuis le début de l'épidémie. Cette hypothèse markovienne facilite certes l'analyse mathématique, mais n'est uniquement valide que pour des systèmes dans lesquels les taux auxquels les événements se déroulent sont indépendants de l'historique du réseau.

Nous présentons dans ce chapitre deux propositions de solution à ce problème, dans le traitement de deux processus stochastiques sur réseaux distincts.

D'abord, nous verrons qu'il est possible de considérer une nouvelle matrice, différente de la matrice d'adjacence, qui permet de prendre en compte les aspects dynamiques

du processus. Cette matrice permet de se ramener à la propagation sur un graphe statique, mais déterminé par la dynamique du réseau initial et de la propagation. Ensuite, nous nous intéresserons au réseau initial dont les arêtes ne sont pas toujours actives. Nous verrons qu'il est alors possible d'approximer une propagation par une agrégation de processus stochastiques appelés marche aléatoire. Enfin, nous mettrons en avant les différences qui existent entre ces deux démarches.

2.1 Propagations

2.1.1 Force dynamique des liens

Nous avons dit que la matrice d'adjacence n'était pas suffisante, dans le cas dynamique, pour décrire l'impact des composantes d'un graphe. Une solution pour comprendre l'importance d'une arête donnée est de définir une nouvelle mesure de l'importance d'une arête : la force dynamique (*dynamical strength*), qui correspond à la probabilité qu'une transmission ait lieu par un lien fixé étant donnée la séquence d'activations des arêtes. Cela permet de décrire le réseau par un graphe statique et de prendre en compte les aspects dynamiques des transmissions. Cette approche est exposée dans les articles [7] et [8], dont nous allons présenter quelques-uns des résultats.

Considérons une propagation suivant le modèle SIR avec taux de contamination β et de guérison T . L'approche proposée est à nouveau de passer par la théorie de la percolation, en considérant la matrice de transmissibilité \mathcal{T} , dont chaque élément \mathcal{T}_{ij} est fonction de β et T . La clé du problème est de déterminer cette matrice.

Posons t_α le moment auquel le noeud i est infecté et $n_{ij}(t_\alpha)$ le nombre de contacts entre i et j sur un intervalle $[t_\alpha, t_\alpha + T]$. La probabilité qu'il n'y ait pas contamination lors d'un contact est donnée par $(1 - \beta)$. Les $n_{ij}(t_\alpha)$ événements étant indépendants, la probabilité qu'aucune contamination n'ait lieu sur un intervalle $[t_\alpha, t_\alpha + T]$ vaut $(1 - \beta)^{n_{ij}(t_\alpha)}$. Donc \mathcal{T}_{ij} vaut $1 - (1 - \beta)^{n_{ij}(t_\alpha)}$. On peut alors moyenniser cette valeur sur toutes les valeurs possibles de t_α pour obtenir la valeur de $\mathcal{T}_{ij}(\beta, T)$.

Au lieu de moyenniser sur les temps d'infection de i , on peut plus simplement moyenniser sur le nombre d'interactions entre i et j sur un intervalle de temps T , en notant $P(n_{ij} = n; T)$ la probabilité que ce nombre soit n .

Nous obtenons alors

$$\mathcal{T}_{ij}(\beta, T) = \sum_{n=0}^{+\infty} P(n_{ij} = n; T) (1 - (1 - \beta)^n),$$

C'est donc cette quantité qui représente la force dynamique des liens. Deux états sont faciles à traiter : lorsque β est très petit et lorsqu'il est proche de 1. Premièrement, considérons le développement de Taylor de $f(\beta) = 1 - (1 - \beta)^n$ autour de 0 :

$$\begin{aligned} f(\beta) &= 1 - (1 - 0)^n + \beta n + O(n^2) \\ &\cong \beta n. \end{aligned}$$

On en tire, pour des valeurs de β très petites la valeur de $\mathcal{T}_{ij}(\beta, T)$:

$$\begin{aligned} \mathcal{T}_{ij}(\beta, T) &\cong \sum_{n=0}^{+\infty} P(n_{ij} = n; T) \beta n \\ &\cong \beta \sum_{n=0}^{+\infty} n P(n_{ij} = n; T) \\ &\cong \beta \langle n_{ij} \rangle_{t_\alpha}, \end{aligned}$$

où la mise entre chevrons représente la moyenne.

Considérons maintenant le cas où β est proche de 1. Le développement de Taylor est alors simplement égal à 1. Nous obtenons alors

$$\begin{aligned} \mathcal{T}_{ij}(\beta, T) &= \sum_{n=1}^{+\infty} P(n_{ij} = n; T) (1 - (1 - \beta)^n) \\ &\cong \sum_{n=1}^{+\infty} P(n_{ij} = n; T) \\ &\cong 1 - P_{ij}^0, \end{aligned}$$

où $P_{ij}^0 = P(n_{ij} = 0; T)$ est la probabilité qu'aucune interaction n'ait lieu sur un intervalle de temps T .

Cette estimation de la force dynamique est très pratique car elle ne dépend que de valeurs qu'on peut tirer relativement facilement de données sur un graphe.

Notons déjà qu'un comportement intensif suivi d'une longue période de repos (typiquement humain) a pour effet d'augmenter la valeur de P_{ij}^0 et donc réduire la force dynamique d'une arête, comparativement à un comportement régulier de type poissonien. Par ailleurs, l'appartenance à une communauté a pour effet d'augmenter la valeur de $\langle n_{ij} \rangle_{t_\alpha}$, augmentant ainsi la force dynamique des liens intra-communautaire.

Une valeur qui est souvent intéressante et étudiée est le nombre moyen de secondes contaminations R_1 , parfois appelé également nombre de reproduction efficace, qui consiste en le nombre moyen de secondes infections provoquées par un individu, c'est-à-dire le nombre d'individus qu'infecteront directement tous ceux que le premier individu a contaminé. Ce nombre R_1 donne de nombreuses informations ([12]),([13]), comme le seuil de percolation d'un processus SIR (qui survient à $R_1 = 1$), la vitesse de diffusion (proportionnelle à R_1) et la taille des cascades (fonction croissante de R_1). Si on suppose que le réseau est complètement aléatoire, R_1 peut être exprimé comme ceci :

$$\begin{aligned}
 R_1 &= \frac{\langle \text{Poids des noeuds}^2 \rangle - \langle \text{Poids}^2 \text{ des arêtes} \rangle}{\langle \text{Poids des noeuds} \rangle} \\
 &= \frac{\left\langle \left(\sum_j \mathcal{T}_{ij} \right)^2 \right\rangle_i - \left\langle \sum_j \mathcal{T}_{ij}^2 \right\rangle_i}{\left\langle \sum_j \mathcal{T}_{ij} \right\rangle_i}
 \end{aligned}$$

Ce résultat indique donc que la force dynamique des liens peut efficacement modéliser la force des interactions humaines dans les réseaux sociaux.

2.1.2 Processus de Bellman-Harris

Nous venons de déterminer une nouvelle matrice caractérisant un réseau, sur lequel nous pouvons simuler une propagation via percolation. Cependant, le phénomène de percolation s'applique dans le cadre d'une propagation statique, et il serait intéressant de pouvoir développer une généralisation de ce processus dans le cadre dynamique.

Un moyen de généraliser cette idée de percolation sur graphe statique (avec matrice de poids modifiée comme vu précédemment) est le modèle de branchement de Bellman-Harris [9] qui permet de prendre en considération les variabilités à la fois de l'intensité des contacts et du temps d'activation des noeuds. Ce développement a principalement été développé dans le cadre de l'étude de la propagation d'information en raison du comportement non-poissonien des êtres humains.

La dynamique d'information est souvent expliquée par des quasi-arbres. En d'autres termes, la propagation de l'information s'effectue généralement en cascades et les éventuelles boucles du graphe sont négligeables car leur longueur est telle que le temps que l'information en fasse un tour complet (ou au moins strictement plus d'un demi-tour dans chaque direction) est beaucoup plus grand que la durée totale

de "l'épidémie", donc l'information n'a pas le temps de "remonter". Nous proposons donc d'étudier la propagation sur une série d'arbres dont les sommets sont les sources de la contamination et nous appellerons seconds propagateurs les autres individus infectés. Nous décomposons le processus d'infection en deux étapes : l'individu décide d'abord d'envoyer l'information à un nombre aléatoire $r > 0$ de personnes suivant une certaine loi de probabilité p_r de moyenne \bar{r} puis les personnes contactées décident selon une probabilité λ d'accepter l'information et de la propager à leur tour, participant ainsi à une nouvelle génération d'informés, la propagation prenant fin lorsque toutes les personnes contactées décident de ne pas transférer l'information. Ce mécanisme est illustré par la figure 2.1

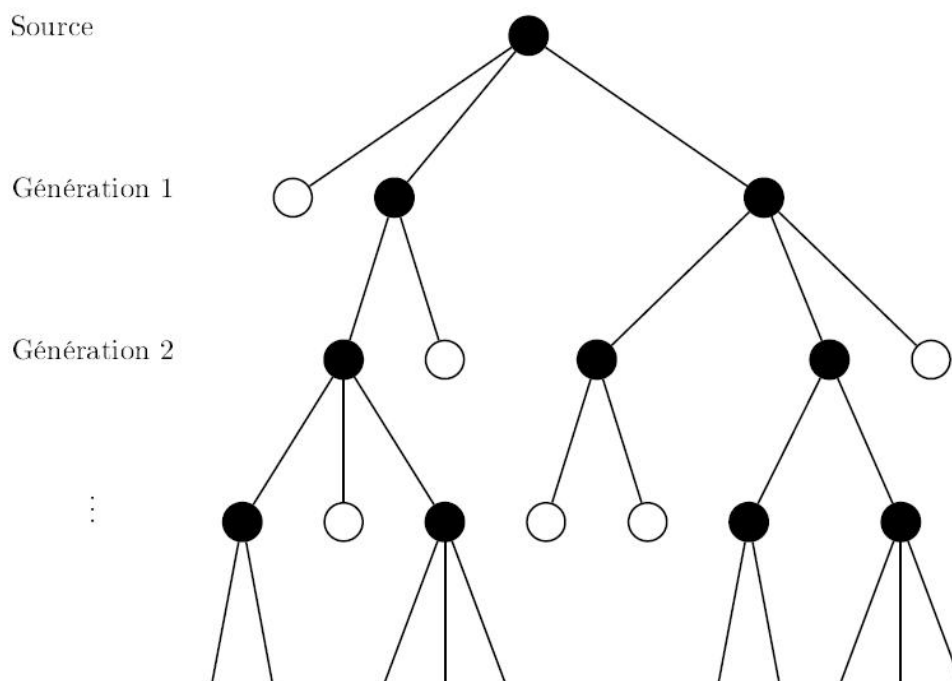


FIGURE 2.1 – Propagation selon Bellman-Harris. Chaque noeud est contaminé (colorié) avec une probabilité λ . Le nombre r de tentatives de contamination de chaque infecté (*i.e.* le nombre de voisins de la génération suivante) est déterminé selon la distribution p_r et le temps d'attente τ avant que ces tentatives soient simultanément effectuées par l'individu infecté est déterminé par la distribution $G(\tau)$.

Notons que dans l'article [9], tous les termes ci-dessous sont indicés par 0 (pour les sources) ou 1 (pour les autres). Cela provient d'une double cause. D'une part,

on peut intuitivement considérer -et les données collectées par l'étude le confirment- que les individus ont tendance à propager plus facilement une information provenant d'une source sûre qu'une information obtenue par hasard ; d'autre part cela se justifie théoriquement par une propriété connue sous le nom de paradoxe de l'amitié. Celui-ci déclare que, en moyenne, la moyenne du nombre d'amis que possèdent mes amis est supérieure à mon nombre d'amis lui-même, en d'autres termes, que mes amis sont en moyenne plus populaires que moi ! Bien que contre-intuitif, ce paradoxe est facilement compréhensible puisqu'il est plus probable d'être ami avec des gens très populaires plutôt qu'avec des gens qui le sont très peu. Du point de vue mathématique, le nombre moyen d'amis est simplement donné par

$$\langle k \rangle = \sum kP(k),$$

alors que le nombre moyen d'amis des amis est donné par

$$\begin{aligned} \sum k \frac{kP(k)}{\langle k \rangle} &= \frac{\langle k^2 \rangle}{\langle k \rangle} \\ &\leq \langle k \rangle, \end{aligned}$$

où l'inégalité est stricte pour toute distribution sauf pour une fonction delta, ou distribution de Dirac, auquel cas l'égalité se produit puisque le nombre de voisins est constant.

Cependant, cette distinction n'est nécessaire que pour la toute première infection, et a une importance très relative quant au comportement global de la propagation qui est surtout régie par le comportements des seconds propagateurs. Ainsi, afin d'alléger les notations, nous négligerons cette distinction.

Nous pouvons construire à partir de la distribution p_r une distribution \tilde{p}_r du nombre de personnes contaminées

$$\tilde{p}_r = \begin{cases} 1 - \lambda & \text{si } r = 0 \\ \lambda p_r & \text{sinon.} \end{cases}$$

Nous ne considérons alors plus le graphe entier mais seulement le réseau de propagation en supposant l'existence d'une arête uniquement si l'information envoyée a été acceptée et réenvoyée. Cette modification de graphe étudié est illustrée par la figure 2.2

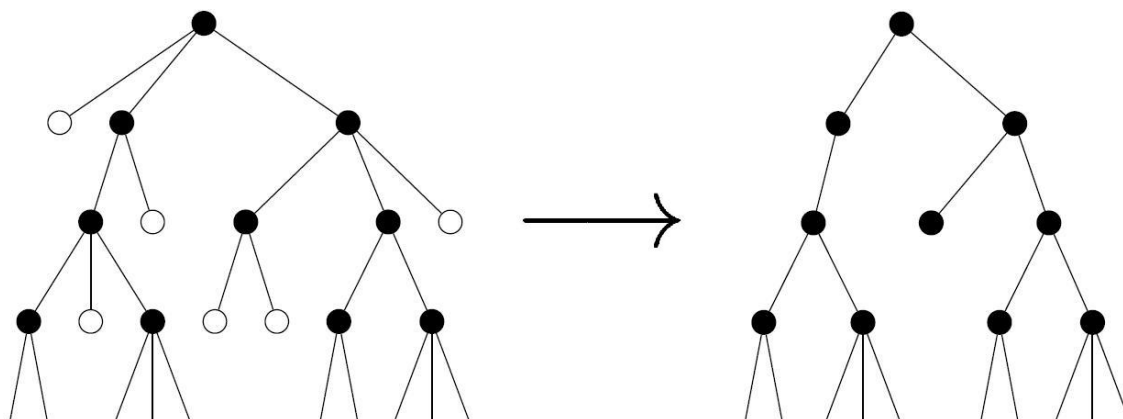


FIGURE 2.2 – À gauche : le graphe initial, les noeuds pleins représentent les individus acceptant l'information, le degré de chaque noeud est déterminé par la distribution p_r ; à droite : le graphe dont on n'a gardé que les noeuds qui acceptent l'information, le degré des noeuds est cette fois déterminé par la distribution \tilde{p}_r .

Notons $f(x)$ la fonction génératrice de \tilde{p}_r définie par

$$f(x) = \sum_{r=0}^{\infty} \tilde{p}_r x^r,$$

et $G(t)$ la distribution des temps d'attente avant transmission du message d'un individu vers ses cibles.

Ce que nous voulons connaître, c'est la distribution $I(t)$ du nombre de personnes informées au temps t et celle $S(t)$ de la taille d'une cascade enclenchée en $t = 0$.

Un raisonnement similaire s'applique pour obtenir ces deux valeurs, nous allons l'exposer pour ce qui concerne $I(t)$.

Le principe est le suivant : à chaque infection, de nouvelles branches se créent, dans lesquelles une propagation est possible. Ces différentes nouvelles branches subissent chacune des propagations indépendantes, copies du même processus de contamination ; elles sont donc statistiquement équivalentes. Partant d'un noeud unique informé ($I(0) = 1$) et en supposant qu'il faut un temps τ avant l'infection de la génération suivante, nous pouvons obtenir une relation auto-référentielle pour $I(t)$ disant que le nombre de personnes infectées en un temps t est donné par le nombre de personnes infectées par un nombre de processus équivalent au nombre de personnes infectées par ce noeud et qui se dérouleraient en un temps $t - \tau$, pour autant que l'infection ait pu avoir lieu, *i.e.* $t > \tau$; dans le cas contraire le noeud infecté serait le seul à

détenir cette information.

Mis en équation, cela donne

$$I(t) = \begin{cases} 1 & \text{si } t < \tau \\ \sum_{i=1}^r I(t - \tau) & \text{sinon.} \end{cases}$$

Définissons $F(s, t) = \sum_{k \geq 0} P[I(t) = k] s^k$ la fonction génératrice de $I(t)$.

Si $t < \tau$, alors on obtient facilement la valeur de $F(s, t)$:

$$\begin{aligned} F(s, t) &= \sum_{k \geq 0} P[1 = k] s^k \\ &= s \end{aligned}$$

Si on n'est plus à un moment trivial, alors le calcul de la valeur de $F(s, t)$ est un peu plus compliqué, et l'on retrouve forcément une valeur via auto-référence :

$$\begin{aligned} F(s, t) &= \sum_{k \geq 0} P[I(t) = k] s^k \\ &= \sum_{k \geq 0} P \left[\sum_{i=1}^r I(t - \tau) = k \right] s^k \\ &= \sum_{k \geq 0} \tilde{p}_k \left(P \left[\sum_{r \geq 0} P[I(t - \tau) = r] s^r \right]^k \right) \\ &= \sum_{k \geq 0} \tilde{p}_k [F(s, t - \tau)]^k \\ &= f[F(s, t - \tau)] \end{aligned}$$

Donc, pour un τ tiré aléatoirement,

$$F(s, t) = \begin{cases} s & \text{si } t < \tau \\ f[F(s, t - \tau)] & \text{sinon.} \end{cases}$$

Il ne reste plus qu'à moyenniser pour toutes les valeurs de τ via la fonction de densité

$G(\tau)$ définie plus haut :

$$\begin{aligned}
F(s, t) &= \int_0^{+\infty} F(s, t) dG(\tau) \\
&= \int_0^t F(s, t) dG(\tau) + \int_t^{+\infty} F(s, t) dG(\tau) \\
&= \int_0^t f[F(s, t - \tau)] dG(\tau) + \int_t^{+\infty} s dG(\tau) \\
&= \int_0^t f[F(s, t - \tau)] dG(\tau) + s[G(\tau)]_t^{+\infty} \\
&= \int_0^t f[F(s, t - \tau)] dG(\tau) + s[1 - G(t)]
\end{aligned}$$

Nous pouvons refaire exactement le même raisonnement pour obtenir $S(t)$.

En posant $\phi(s, t) = \sum_{k=1}^{+\infty} P[S(t) = k] s^k$, nous obtenons presque la même équation :

$$\phi(s, t) = s \int_0^t f[\phi(s, t - \tau)] dG(\tau) + s[1 - G(t)]$$

Et nous avons à nouveau décrit la propagation via des équations intégro-différentielles, prenant ainsi en compte le passé de la propagation.

Enfin, nous pouvons calculer la probabilité a que la propagation s'arrête (dans le cas théorique d'un réseau infini). Pour cela, il y a deux possibilités : soit que la source n'ait pas propagé l'information (\tilde{p}_0), soit qu'elle s'arrête après. Cette seconde probabilité q représente la probabilité que le nombre de nouveaux noeuds infectés soit nul pour n'importe quel temps fini, au-delà d'un certain temps fini. En terme de fonction génératrice, cela revient à

$$\begin{aligned}
q &= \lim_{t \rightarrow +\infty} P[I(t) = 0] \\
&= \lim_{t \rightarrow +\infty} F(0, t) \\
&= 0 + \int_0^{+\infty} f \left[\lim_{t \rightarrow +\infty} F(0, t) \right] dG(\tau) \\
&= f(q)
\end{aligned}$$

Donc la probabilité a que la propagation ne touche qu'un nombre fini de personnes en un temps infini est donnée par

$$a = 1 - \lambda + q,$$

avec q solution de $q = f(q)$.

2.2 Marche aléatoire

La seconde approche que nous développons maintenant se base sur les marches aléatoires. Considérons un marcheur présent initialement à un noeud d'origine et souhaitant se promener sur le réseau. A chaque instant il emprunte la première arête qui lui est accessible, et cela peu importe où elle le mènera,. Nous pouvons alors par exemple nous intéresser au nombre de noeuds qu'il a parcourus, à celui où il se situe en un temps donné ou au temps qu'il lui faudrait pour parcourir tout le réseau. Cette approche nous permettra notamment de prendre en considération le temps qui s'est écoulé depuis le début de la propagation, et d'aller ainsi au-delà des hypothèses de Markov. Pour cela, nous considérons que l'activation d'une arête reliant i à j s'effectue selon une distribution de temps d'attente notée ψ_{ij} . Il y a diverses manières de considérer ces distributions ψ_{ij} ; nous considérons ici qu'elles régissent la probabilité qu'apparaisse une arête reliant j à i en un temps compris entre t et $t + dt$ après que le noeud j ait été atteint. En guise d'illustration, nous pouvons imaginer le marcheur comme un motard impatient d'utiliser son véhicule pour le simple plaisir de rouler. Il est présent en un noeud et chaque route qu'il peut prendre, les arêtes, sont bloquées par un feu rouge. Dès que l'un de feux devient vert, il s'engage sur l'arête correspondante et attendra sur le noeud suivant qu'un autre feu passe au vert. La distribution ψ_{ij} correspond alors au temps que devra attendre le motard pour que, une fois arrivé au noeud j , le feu bloquant la route reliant i à j passe au vert.

En général, pour modéliser le phénomène de propagation de maladie - typiquement du modèle comportemental le plus élémentaire SI (Suceptible-infecté)- nous utilisons ce processus de marche aléatoire. Nous observons simplement le trajet emprunté par un grand nombre de marcheurs dont les points de départs correspondent aux noeuds sources de la maladie, et en agrégeant ces marches nous pouvons déterminer les caractéristiques qui nous intéressent concernant la propagation de l'infection.

2.2.1 Détermination de la densité de marcheurs

Considérons un marcheur partant du noeud source de l'épidémie qui se promène de noeud en noeud dès qu'il en a l'occasion, c'est-à-dire dès qu'une arête partant de son noeud de résidence apparaît. Le temps d'attente est alors donné par la distribution ψ_{ij} .

De la définition de ψ_{ij} on a

$$\int_0^{+\infty} \psi_{ij}(t) dt = 1.$$

La probabilité que cette arête apparaisse en un temps t est donnée par

$$\int_0^t \psi_{ij}(t') dt',$$

et donc, celle que ce lien n'apparaisse pas en un délai t est

$$\chi_{ij}(t) \equiv 1 - \int_0^t \psi_{ij}(t') dt'.$$

Notons $T_{ij}(t)$ la probabilité d'effectivement faire le trajet depuis j vers i en un temps t . Pour qu'un tel événement se produise, deux conditions sont nécessaires. D'une part, il faut qu'une arête apparaisse entre ces deux noeuds en un temps t . D'autre part, il faut que le marcheur soit toujours présent en j une fois que cette arête apparaît, ce qui signifie qu'il faut qu'aucune arête partant de j à un de ses voisins autre que i n'apparaisse. La distribution de probabilité $T_{ij}(t)$ vaut alors

$$\begin{aligned} T_{ij}(t) &= \psi_{ij}(t) \times \prod_{k \neq i} \chi_{kj}(t) \\ &= \psi_{ij}(t) \times \prod_{k \neq i} \left(1 - \int_0^t \psi_{kj}(t') dt' \right). \end{aligned}$$

Cette dernière équation met en avant l'importance de l'ordre d'apparition des arêtes et donne une importance accrue aux arêtes qui ont tendance à s'activer plus vite et qui sont de ce fait plus susceptibles d'être parcourues par le marcheur.

Concentrons-nous sur la trajectoire empruntée par le marcheur et essayons de déterminer la probabilité $n_i(t)$ de le trouver au noeud i en un temps t . Si nous notons $q_i(\tau)$ la probabilité d'arriver au noeud i en un temps τ inférieur à t , et $\phi_i(t')$

la probabilité de ne pas quitter le noeud i en un délai t' après l'arrivée sur ce noeud, alors nous avons

$$n_i(t) = \int_0^t \phi_i(t-t')q_i(t') dt'.$$

Il est plus simple de manipuler cette égalité dans l'espace de Laplace associé. Nous noterons $\hat{f}(s)$ la transformée de Laplace d'une fonction quelconque $f(t)$, définie par

$$\hat{f}(s) = \int_0^{+\infty} f(t)e^{-st} dt.$$

Dans l'espace de Laplace, ce produit de convolution se résume alors à un simple produit, bien plus aisément manipulable :

$$\hat{n}_i(s) = \hat{\phi}_i(s)\hat{q}_i(s).$$

Commençons par obtenir le premier facteur de ce produit. La probabilité de rester immobile sur le noeud i pour une durée t est donné par

$$\phi_i(t) = 1 - \int_0^t \underbrace{\sum_{j=1}^N T_{ji}(t)}_{T_i(t)} dt',$$

dont la transformée de Laplace est

$$\hat{\phi}_i(s) = \frac{1}{s} \left(1 - \hat{T}_i(s) \right).$$

L'expression du second facteur du produit qui nous intéresse est un peu plus délicate à trouver.

La probabilité $q_i(t)$ d'arriver en i en un temps exactement t est l'agrégation de toutes les probabilités $q_i^{(k)}$ d'y arriver en un nombre k d'étapes. Une relation de récurrence relie ces probabilités, car la probabilité d'arriver dans un timing parfait en i en $k+1$ étapes est celle d'arriver en k étapes à l'un quelconque de ses voisins j , pondérée par la probabilité de faire le dernier pas dans le délai restant imparti. Ce qui peut s'écrire

$$q_i^{(k+1)}(t) = \int_0^t \sum_j T_{ij}(t-t')q_j^{(k)}(t') dt'.$$

Ce qui donne dans l'espace de Laplace

$$\hat{q}_i^{(k+1)}(s) = \sum_j \hat{T}_{ij}(s) \hat{q}_j^{(k)}(s).$$

Nous obtenons dès lors

$$\begin{aligned} \hat{q}_i(s) &= \hat{q}_i^{(0)} + \sum_{k=0}^{+\infty} \hat{q}_i^{(k+1)}(s) \\ &= \hat{q}_i^0 + \sum_j \hat{T}_{ij}(s) \sum_{k=0}^{+\infty} \hat{q}_j^{(k)}(s) \end{aligned}$$

La notation matricielle allège cette expression :

$$\hat{q}(s) = \hat{T}(s) \hat{q}(s) + \hat{q}^{(0)}(s)$$

Il reste alors à remarquer que $q^{(0)}(t) = n(0)\delta(t)$ pour obtenir une expression pour $\hat{q}(s)$:

$$\hat{q}(s) = \left(I - \hat{T}(s) \right)^{-1} n(0)$$

Ce qui nous permet de trouver l'expression finale de la transformée de Laplace de $n(t)$:

$$\hat{n}(s) = \frac{1}{s} \left(1 - \hat{D}_T(s) \right) \left(I - \hat{T}(s) \right)^{-1} n(0), \quad (2.1)$$

où $\left(\hat{D}_T \right)_{ij}(s) = \hat{T}_i(s) \delta_{ij}$

Nous avons ainsi déterminé une solution formelle dans l'espace de Laplace de la densité de marcheurs aléatoires présents à chaque noeud d'un graphe dont la dynamique est gouvernée par les distributions de temps d'attente $\psi_{ij}(t)$.

2.2.2 Equation intégro-différentielle

En prenant la transformée de Laplace de la dérivée de $n(t)$ ¹ et en utilisant l'équation (2.1), nous allons obtenir -après quelques manipulations algébriques- une forme

1. Cette transformée de Laplace est bien définie car la transformée de Laplace de $n(t)$ et sa dérivée existent.

de l'équation intégro-différentielle dans l'espace initial. La formule théorique est donnée par

$$\mathcal{L}\left(\frac{dn}{dt}\right) = s\hat{n}(s) - n(0). \quad (2.2)$$

Théoriquement, on peut alors intégrer numériquement cette équation pour obtenir la densité de marcheurs en fonction du temps.

Posons

$$\begin{aligned} A &\equiv (I - \hat{D}_T(s)) \\ B &\equiv (I - \hat{T}(s)). \end{aligned}$$

Montrons que A est inversible. Ce serait le cas si toutes les entrées de $\hat{D}_T(s)$ étaient strictement inférieures à 1 pour tout s fini, car nous aurions alors une matrice diagonale dont tous les termes sont strictement positifs. Par définition, cela revient à montrer que les éléments $T_j(s)$ le sont. Or c'est bien le cas :

$$\begin{aligned} \hat{T}_j(s) &= \int_0^{+\infty} e^{-st} T_j(t) dt \\ &< \underbrace{\int_0^{+\infty} T_j(t) dt}_1 \end{aligned}$$

En utilisant les définitions de A et B dans l'équation (2.2), on obtient :

$$\begin{aligned} \mathcal{L}\left(\frac{dn}{dt}\right) &= s \frac{1}{s} AB^{-1} n(0) - n(0) \\ &= (AB^{-1} - I) n(0) \\ &= (AB^{-1} - I) \underbrace{(sBA^{-1}) \left(\frac{1}{s} AB^{-1} n(0)\right)}_I \\ &= (A - B) sA^{-1} \hat{n}(s) \\ &= \left(\hat{T}(s) - \hat{D}_T(s)\right) \frac{s}{I - \hat{D}_T(s)} \hat{n}(s) \\ &= \left(\hat{T}(s) \hat{D}_T^{-1}(s)\right) \hat{K}(s) \hat{n}(s), \end{aligned}$$

où $K \equiv \frac{s\hat{D}_T(s)}{I-\hat{D}_T(s)}$ est appelé le "noyau de mémoire" (*Kernel memory*) et caractérise l'importance de la mémoire dans la dynamique.

En prenant la transformation de Laplace inverse, nous obtenons :

$$\frac{dn}{dt} = \left(T(t) * \mathcal{L}^{-1}(\hat{D}_T^{-1}(s)) - \delta(t) \right) * K(t)$$

2.3 Différences entre marche aléatoire et propagation SI

La littérature scientifique relative à la marche aléatoire relève très souvent la similarité entre celle-ci et la propagation de maladie. On se contente alors souvent d'effectuer un grand nombre de marches aléatoires, et en agrégeant ces différentes marches, on détermine une probabilité d'occupation, c'est-à-dire d'infection, pour chaque noeud. La figure 2.3 illustre la comparaison entre une propagation SI et l'agrégation de plusieurs marches aléatoires.

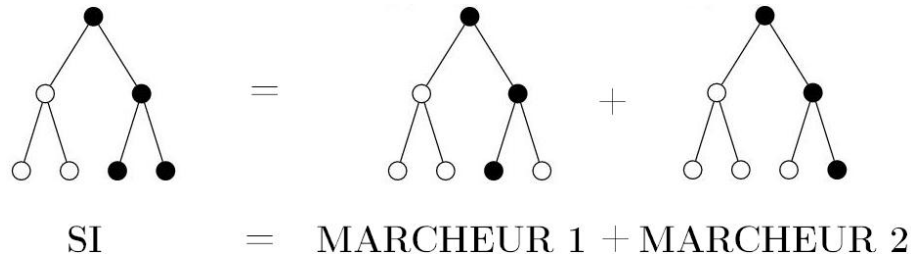


FIGURE 2.3 – Une propagation SI est considérée comme équivalente à plusieurs marches aléatoires indépendantes.

Cependant on mentionne très rarement les différences fondamentales qui peuvent exister entre ces deux processus. Prenons pour illustration l'équation (2.1) qui nous donne une probabilité d'occupation de chaque noeud, donc une probabilité d'infection pour chaque noeud.

Les principales différences qui résident, à mes yeux, sont les corrélations entre les différents propagations. Et *a fortiori*, les corrélations entre les probabilités d'infection qu'ont deux noeuds qui possèdent le même "parent", *i.e* le même contaminateur commun. Cela est dû au fait que la probabilité d'infection déterminée par une marche aléatoire résulte de l'agrégation du trajet de nombreux marcheurs qui partagent le même point de départ, alors que celle déterminée par le modèle SI est le fruit d'une

unique propagation qui se répand dans tout le réseau.

Considérons à titre illustratif un arbre dans lequel chaque noeud possède "deux enfants". La question qui nous intéresse est de voir quelle est la corrélation d'infection entre deux enfants issus d'un même parent, c'est-à-dire, sachant que l'un est contaminé en un temps t , quelle est la probabilité que l'autre le soit également. Nommons arbitrairement le noeud parent Adam, et ses deux enfants Caïn et Abel. Dans le cas de la marche aléatoire, cette corrélation est nulle car la contamination de Caïn et Abel est le produit de deux processus indépendants, le premier étant la marche du premier marcheur qui, partant de la source, est arrivé chez Caïn; le second celle du second marcheur aboutissant chez Abel. Pour cela, à chaque marche et à chaque niveau, un "test de décision" a été effectué pour décider quelle arête serait empruntée, et cela depuis le départ. Après chaque test, le marcheur quitte le noeud où il était pour atteindre le noeud choisi.

Dans le cas de la propagation de l'épidémie, par contre, chaque noeud reste contaminé après avoir transmis la maladie. Donc dans le cas de l'infection des deux enfants, il aura fallu que le processus atteigne le parent. Une fois contaminé, Adam contaminera Caïn en un temps déterminé par $\psi_{\text{Adam-Caïn}}$ et Abel par $\psi_{\text{Adam-Abel}}$. Cependant, le fait qu'Adam contamine d'abord Caïn ou Abel n'a pas d'influence quant au délai d'infection du second, car bien qu'ayant contaminé un enfant, Adam reste contaminé et donc porteur de la maladie pour le second enfant.

Une deuxième conséquence très importante de ces corrélations est liée à l'impact des communautés. En effet, si le réseau possède de fortes communautés, comme c'est le cas dans de nombreux réseaux sociaux, nous pouvons assister à une contamination très rapide intra-communautaire car les individus d'une communauté sont fortement connectés les uns aux autres, ce qui augmente les risques d'infections. Dans le cas d'une simulation via marche aléatoire dans un réseau comportant des communautés, la contamination globale sera sous-estimée. Cela est dû au fait qu'une communauté peut capturer un marcheur et ne lui laisser qu'une très faible probabilité d'en sortir. Cette occasion n'apparaît en effet que lorsque le marcheur arrive sur un noeud frontière de la communauté, auquel cas il n'a qu'une chance sur le nombre de voisins de ce noeud de quitter la communauté, si on suppose qu'il n'existe qu'un seul voisin n'appartenant pas à la communauté.

La figure 2.3 montre un exemple de graphe comportant deux communautés pour lequel ce phénomène peut se produire : alors qu'une propagation compartimentale aura pour conséquence d'infecter les deux communautés, un marcheur aléatoire se

retrouvera capturé dans une communauté avec très peu de chances d'en sortir. L'aggrégation de différentes marches aura pour effet de réduire la taille globale de l'épidémie.

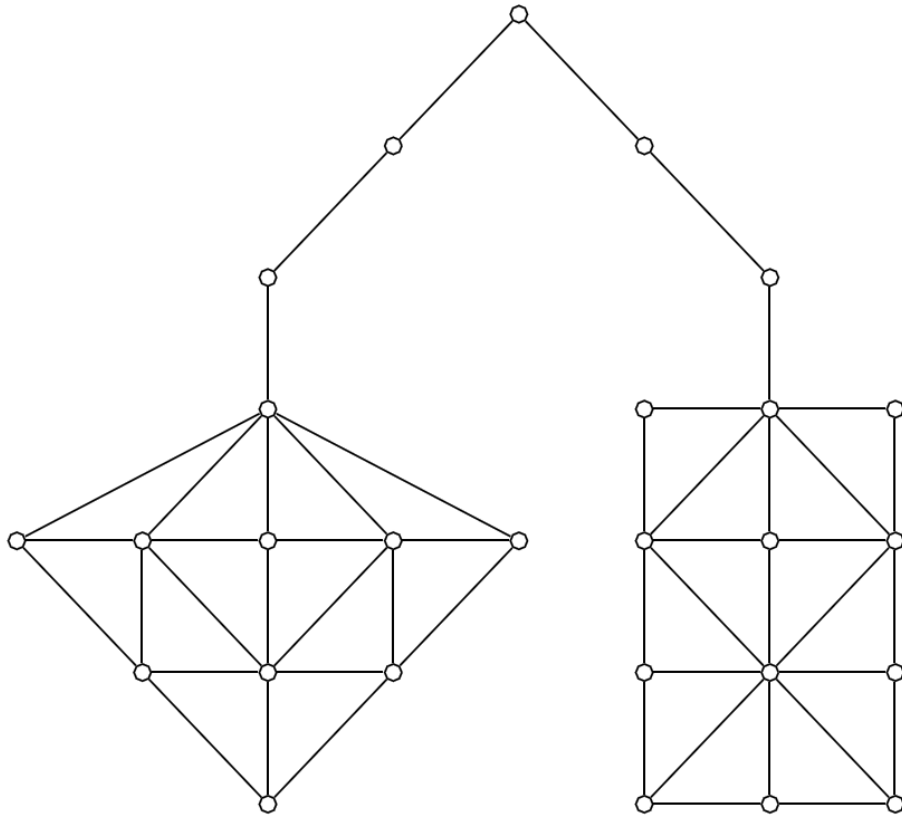


FIGURE 2.4 – La présence de communautés dans le graphe peut provoquer des différences entre une propagation de maladie et son approximation via une agrégation de marches aléatoires. Les communautés capturent un marcheur et ne lui laissent que très peu de chances de quitter la communauté.

2.3.1 Variation sur les résultats obtenus avec la marche aléatoire

Reprenons le développement effectué dans le cadre de la marche aléatoire et qui a abouti à l'obtention de l'équation (2.1). La distinction fondamentale que nous avons effectuée dans la sous-section précédente peut être appliquée à l'obtention de cette équation. Dans le cas d'une propagation SI , il n'est pas nécessaire de s'assurer que l'arête reliant i à j soit la première partant de j pour qu'il y ait contamination. Ainsi, pour le modèle SI , la probabilité de transmission effective de j à i en un temps t , notée $T_{ij}(t)$, est simplement égale à la probabilité qu'une arête reliant ces deux noeuds s'active dans le délai imparti, soit $\psi_{ij}(t)$. De même, la probabilité $n_i(t)$ d'avoir le noeud i infecté au temps t ne dépend plus que de la probabilité $q_i(\tau)$ qu'il ait été contaminé en un temps τ inférieur ou égal à t . On a donc la relation

$$n_i(t) = \int_0^t q_i(t') dt',$$

dont la transposition dans l'espace de Laplace donne

$$\hat{n}_i(s) = \frac{1}{s} \hat{q}_i(s).$$

Le raisonnement que nous avons effectué pour trouver la valeur de $\hat{q}_i(s)$ reste valable et nous obtenons donc le même résultat :

$$\hat{n}(s) = \left(I - \hat{T}(s) \right)^{-1} n(0)$$

Ce qui nous donne pour $\hat{n}(s)$:

$$\begin{aligned} \hat{n}(s) &= \frac{1}{s} \left(I - \hat{T}(s) \right)^{-1} n(0) \\ &= \frac{1}{s} \left(I - \hat{\psi}(s) \right)^{-1} n(0) \end{aligned}$$

On constate ainsi que l'expression obtenue diffère de celle obtenue dans le cas de la marche aléatoire.

Chapitre 3

Application numérique

La dernière partie de ce mémoire est consacrée à une application numérique relative aux propagations dans des graphes. Tous les codes ont été écrits dans le programme Matlab[®] et sont fournis en annexe.

3.1 Motivations

Dans le cadre de propagations sur réseaux, nous avons vu que la distribution du temps d'attente d'activation des arêtes a une importance cruciale sur la dynamique du processus. Considérons plusieurs propagations régies par un même loi de distribution mais qui diffèrent par le ou les paramètres associés à cette loi. Ce ou ces paramètres ont une influence sur la moyenne et la variance de la distribution. S'il est évident qu'augmenter la moyenne du temps d'attente des arêtes avant leur activation a pour effet de ralentir la propagation, l'impact de l'augmentation de la variance n'est quant à lui pas trivial à déterminer. Pire : alors que dans l'article [10] l'augmentation de la variance a pour effet de ralentir la propagation sur les réseaux étudiés (appels téléphoniques en Europe et échange de courriers électroniques), l'article [11] met en évidence l'effet opposé pour un autre réseau étudié (les rapports sexuels d'*escort girls* au Brésil). Ces deux effets opposés s'expliquent par des différences dans la topologie des réseaux complexes étudiés. L'objectif de la simulation numérique est d'étudier l'impact de l'augmentation de la moyenne et de la variance sur un graphe dont les propriétés topologiques sont assez simples et connues.

3.2 Description du processus

Nous avons étudié dans un arbre de taille infinie (voir figure 3.1) une propagation inspirée de [9], mais simplifiée. La structure du graphe est régie comme suit : chaque noeud interne, excepté l'origine o , possède trois voisins : un prédécesseur et deux successeurs.

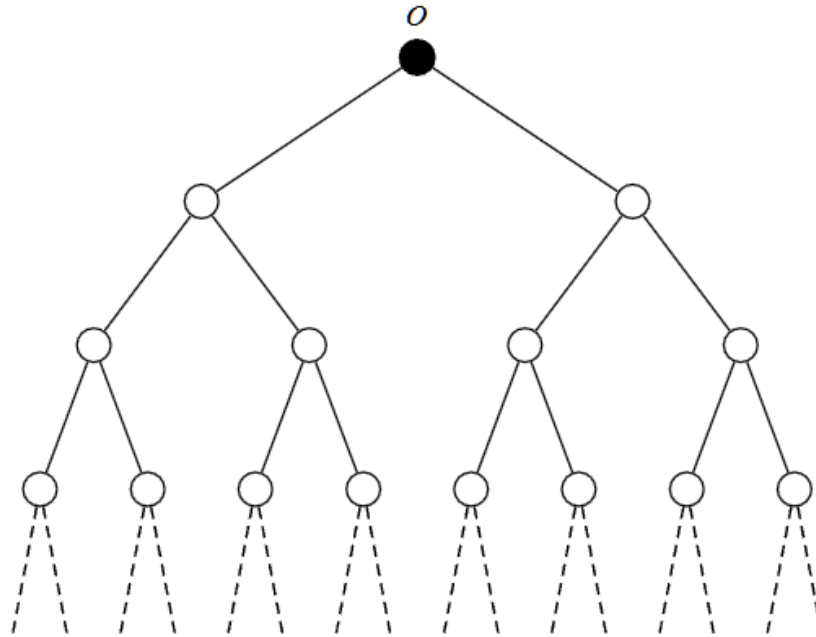


FIGURE 3.1 – Réseau considéré pour l'application numérique. Chaque noeud a un prédécesseur et deux successeurs.

En considérant la distance $d(x)$ comme la longueur du plus court chemin reliant le noeud x au noeud d'origine o , on définit le prédécesseur *papa* d'un noeud x comme étant celui parmi ses trois noeuds voisins y, z, t tel que $papa \in \{y, z, t\}$ et $d(x) = d(papa) + 1$. Grâce à la structure particulière du graphe, il est possible de numéroter chaque noeud de manière unique, en prenant comme première coordonnée sa génération, c'est-à-dire sa distance avec l'origine, et en seconde coordonnée sa place dans la génération. Nous pouvons dès lors numéroter un noeud sans au préalable devoir numéroter tous les noeuds des autres générations : seul importe de connaître la numération de son prédécesseur. En effet, si un noeud est numéroté (i, j) , alors ses enfants seront numérotés $(i + 1, 2j - 1)$ et $(i + 1, 2j)$. Forts de cette numé-

tation unique, nous pouvons également indiquer les différents temps d'activation des arêtes tirés depuis la distribution déterminée : le temps d'attente associé à une arête reliant un noeud (i, j) à son prédécesseur est noté $t_{i,j}$. La figure 3.2 présente cette numérotation qui a été très utile pour l'implémentation numérique.

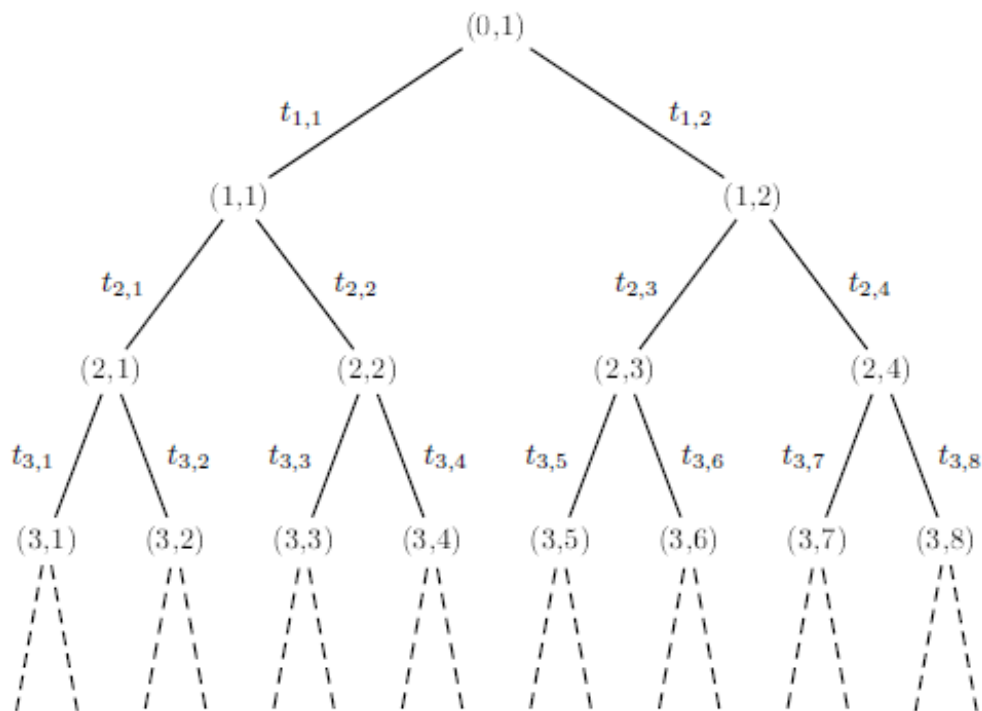


FIGURE 3.2 – Chaque noeud du réseau est dénoté par un couple dont le premier élément indique la génération à laquelle il appartient, et le second, la place qui lui est attribuée sur cette génération. Le temps d'activation de l'arête liant un noeud (i, j) à son prédécesseur est noté $t_{i,j}$. Les temps $t_{i,j}$ sont déterminés aléatoirement à partir d'une même distribution aléatoire.

Le processus stochastique étudié consistait en une propagation SI telle que dès qu'une arête s'active, il y a contamination. Initialement, seul le noeud d'origine est contaminé, et les temps d'attente des arêtes adjacentes au noeud d'origine sont tirés arbitrairement d'une loi aléatoire. Ce temps d'attente représente donc le temps qu'il faut pour infecter un successeur à partir du moment où la maladie a été contractée. Les distributions considérées sont la loi exponentielle, qui correspond à un processus markovien en temps continu, et la loi uniforme, qui induit un processus non-markovien.

Tant pour la moyenne que pour la variance, nous nous sommes intéressés à deux choses : l'évolution du nombre d'infectés au cours du temps et le nombre de personnes infectées en un temps donné en fonction de différentes valeurs de ces paramètres.

Puisque nous considérons des processus stochastiques, il est important de ne pas se focaliser sur une seule réalisation mais d'effectuer une moyennisation sur un ensemble de réalisations. Nous avons à chaque fois moyennisé les résultats sur 100 réalisations. Chaque barre d'erreur a une hauteur égale à l'écart-type de l'ensemble des valeurs moyennisées.

3.3 Justification des choix

Avant d'étudier les résultats obtenus, quelques choix relatifs au processus simulé peuvent poser question. Nous allons nous empresser de les justifier immédiatement.

Modèle SI

Le modèle SI est le plus basique des modèles compartimentaux d'infection. Cependant il n'est pas forcément mal adapté aux propagations par réseaux. En effet, si le temps de guérison d'un individu est beaucoup plus élevé que les temps d'attente avant activation de ses liens, le modèle SIR se résume à un modèle SI. De plus, nous sommes intéressés par l'évolution dynamique de l'épidémie, son comportement aux premiers instants. Or un modèle SIR peut être efficacement approximé par un modèle SI dans justement son comportement initial. Certes, dans un réseau statique fini, tous les noeuds seront finalement infectés dans le cadre modèle SI. Cependant, ce n'est pas forcément le cas dans un réseau dynamique comme celui que nous étudions, et cela rend le modèle SI davantage réaliste dans le cas temporel que dans le cas statique ([11]).

Arbre

Une objection au fait de considérer une propagation dans un arbre est de dire que les réseaux sociaux sont organisés en communautés et qu'ils ne sont donc pas des quasi-arbres. Nous pouvons répondre à cela que les infections virales sont généralement presque des arbres purs, avec très peu de boucles courtes et de triangles fermés ([9]). De plus, les chemins de contamination empruntés par les virus ne sont pas forcément les plus courts ([10]). Enfin, malgré les petites approximations que cela entraîne sur de grands réseaux, cela nous permet de considérer un graphe dont nous connaissons parfaitement la topologie, et ainsi de nous focaliser sur la dynamique

déterminée par le temps d'activation des arêtes.

Loi uniforme

Puisque nous voulons étudier un processus non-markovien, nous devons rejeter l'étude approfondie d'un processus déterminé par une loi exponentielle, puisque cela correspond justement à un processus de Markov¹. Il faut donc en choisir une autre. L'avantage de considérer pour les valeurs possibles du temps d'activation des arêtes une loi uniforme définie sur l'intervalle $[a, b]$, notée $U(a, b)$, est double : d'une part garantir que le temps d'attente tiré sera positif (il suffit que a le soit) et d'autre part déterminer facilement la valeur de la moyenne μ et de la variance σ^2 de la loi de probabilité, contrairement à une loi normale ou à une loi de puissance, par exemple. La densité de probabilité d'une telle loi est donnée par

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$$

La moyenne vaut donc

$$\begin{aligned} \mu &= \int_{-\infty}^{+\infty} x f(x) dx \\ &= \left[\frac{x^2}{2(b-a)} \right]_{x=a}^{x=b} \\ &= \frac{a+b}{2}. \end{aligned}$$

Le second moment m_2 est donné par

$$\begin{aligned} m_2 &= \int_{-\infty}^{+\infty} x^2 f(x) dx \\ &= \left[\frac{x^3}{3(b-a)} \right]_{x=a}^{x=b} \\ &= \frac{a^2 + ab + b^2}{3}. \end{aligned}$$

1. Nous avons cependant considéré ce cas à titre comparatif dans le cas de l'étude du nombre d'infectés en fonction de la moyenne, car nous pouvons facilement déterminer les équations régissant le processus dans ce cas.

Donc la variance vaut

$$\begin{aligned}\sigma^2 &\equiv m_2 - \mu^2 \\ &= \frac{a^2 + ab + b^2}{3} - \frac{(a + b)^2}{4} \\ &= \frac{(b - a)^2}{12}.\end{aligned}$$

3.4 Résultats : description

3.4.1 Processus Markovien

Avant de nous intéresser au cas de la distribution uniforme, attardons-nous un instant sur le cas où la distribution est exponentielle, ce qui implique un processus de Poisson. Comme cette distribution n'est qu'à un paramètre, il est impossible d'étudier de manière séparée l'impact de la variation de la moyenne et de la variance.

La densité de probabilité d'une distribution exponentielle de paramètre λ est donnée par

$$f(t, \lambda) = \begin{cases} 0 & \text{si } t < 0 \\ \lambda e^{-\lambda t} & \text{sinon.} \end{cases}$$

L'espérance μ d'une variable aléatoire suivant une telle loi vaut $\frac{1}{\lambda}$ et sa variance $\frac{1}{\lambda^2}$, soit μ^2 .

Nombre d'infectés en un temps fixé

Nous avons représenté le nombre d'infectés après 20 secondes en fonction de la moyenne μ (figure 3.3) et observons que le nombre de personnes infectées en un temps donné diminue lorsque la valeur de la moyenne augmente.

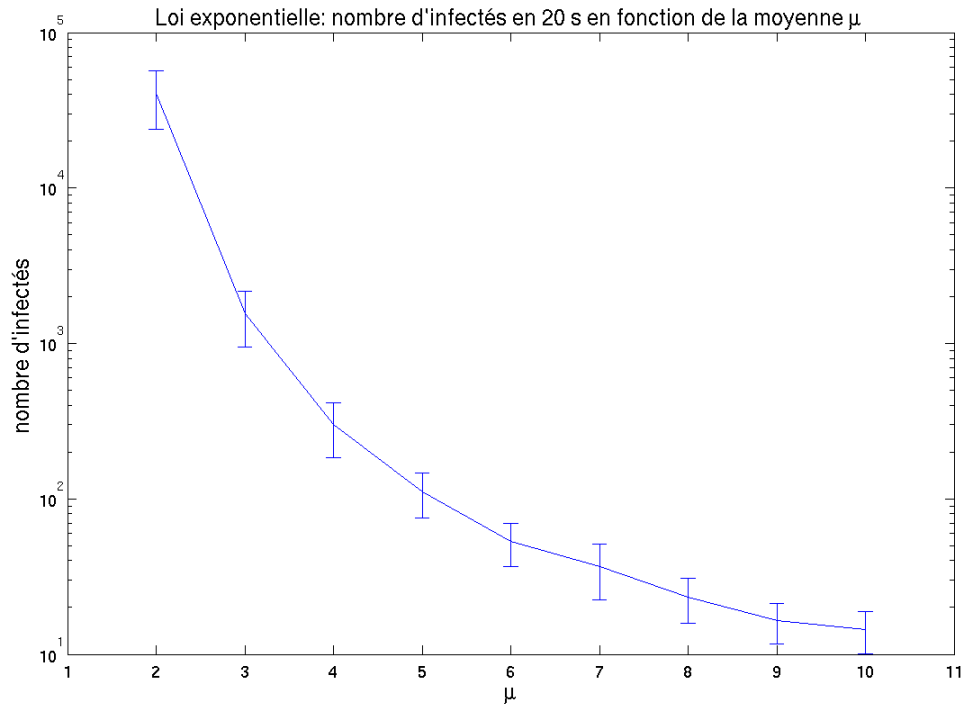


FIGURE 3.3 – Nombre d'infectés en 20 secondes en fonction de la moyenne μ : lorsque la moyenne -et donc la variance- augmente, le nombre d'infectés diminue.

Evolution du nombre d'infectés en fonction du temps

Nous avons également représenté l'évolution du nombre d'infectés au cours du temps pour différentes valeurs de μ (figure 3.4). La taille de l'épidémie augmente exponentiellement et nous pouvons à nouveau constater que d'une augmentation de la moyenne résulte une diminution de la taille de l'épidémie.

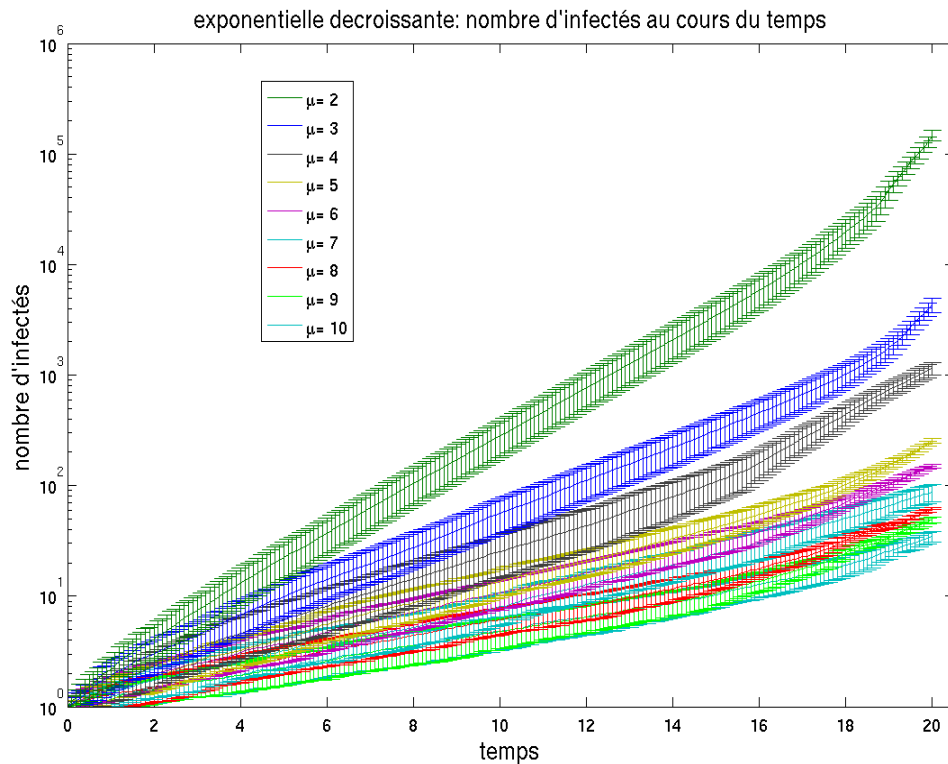


FIGURE 3.4 – Evolution du nombre d’infectés au cours du temps lorsque le processus est markovien. La croissance est exponentielle, et la taille de l’épidémie augmente lorsque la moyenne -et donc la variance- diminue.

Nous nous intéressons maintenant au cas où la distribution est uniforme, et donc le processus non-markovien.

3.4.2 Etude en fonction de la moyenne

Etudions d’abord l’impact qu’a l’augmentation de la moyenne du temps d’attente sur la taille de l’épidémie, en précisant que nous n’avons pas imposé une variance fixe par souci graphique. Les intervalles de temps admis pour une simulation où la moyenne désirée est μ sont de la forme $[\epsilon, 2\mu - \epsilon]$, où ϵ a été fixé à 10^{-15} . Dès lors la variance vaut $\frac{(\mu - \epsilon)^2}{3} \approx \frac{\mu^2}{3}$. Nous pouvons nous autoriser cela car, comme nous le verrons dans la partie consacrée à l’étude des variations conjointes de la moyenne

et de la variance, la moyenne a beaucoup plus d'impact que la variance sur la taille globale de l'épidémie.

Les observations faites sur l'impact de la modification de la moyenne dans le cas d'une distribution exponentielle sont valides ici (figures 3.5 et 3.6). Remarquons que la taille de l'épidémie est plus élevée, à moyenne égale, dans le cas markovien que dans ce cas non-markovien.

Nombre d'infectés en un temps fixé

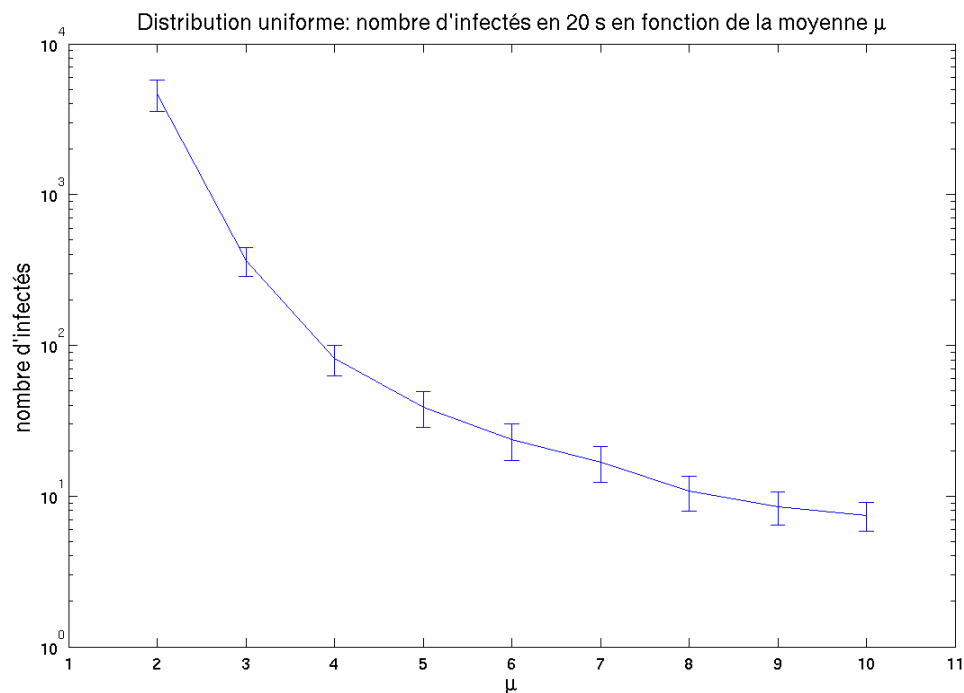


FIGURE 3.5 – Distribution uniforme. Nombre d'infectés en 20 secondes en fonction de la moyenne μ : lorsque la moyenne augmente, le nombre d'infectés diminue.

Evolution du nombre d'infectés en fonction du temps

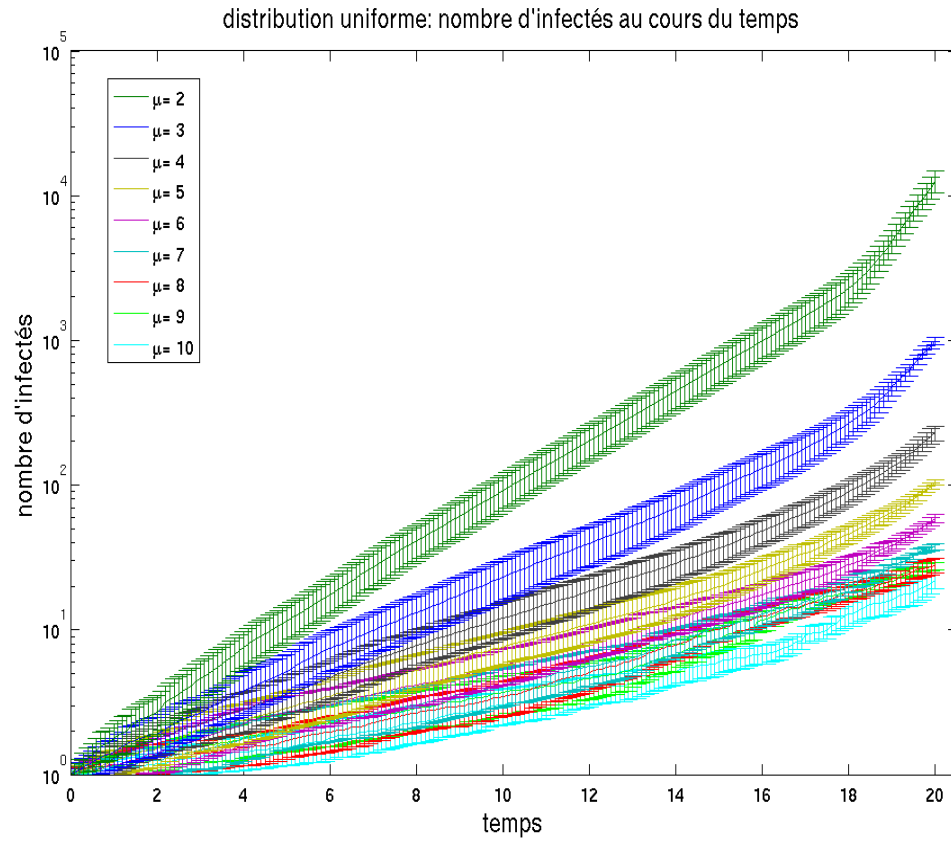


FIGURE 3.6 – Distribution uniforme. Evolution du nombre d'infectés au cours du temps pour différentes valeurs de μ . La croissance est exponentielle, et la taille de l'épidémie augmente lorsque la moyenne diminue.

3.4.3 Etude en fonction de la variance

Observons maintenant l'effet d'une variation de la variance tout en gardant une moyenne constante, fixée ici à 10. Considérons une distribution de probabilité uniforme sur $[a, b]$ de moyenne μ . Si la variance est nulle, nous avons un processus déterministe, et tous les noeuds d'une génération sont simultanément infectés en un temps μ après ceux de la génération précédente. Ainsi, après un temps compris entre $k\mu$ et $(k + 1)\mu$, k générations auront été infectées.

La taille s de l'infection sera alors égale à :

$$\begin{aligned}
 s &= \underbrace{1}_{\text{source}} + \underbrace{2}_{\text{première génération infectée}} + \underbrace{4}_{\text{deuxième génération infectée}} \\
 &+ \dots + \underbrace{2^k}_{k^{\text{e}} \text{ génération infectée}} \\
 &= \sum_{i=0}^k 2^i \\
 &= \frac{1 - 2^{k+1}}{1 - 2} \\
 &= 2^{k+1} - 1.
 \end{aligned}$$

La fonction décrivant le nombre de personnes infectées au cours du temps est ainsi une fonction en escaliers.

La variance augmente lorsque l'intervalle $[a, b]$ grandit, toutefois nous devons nous restreindre à choisir une valeur strictement positive pour a . Nous fixons à nouveau la valeur minimale de a à $\epsilon = 10^{-15}$, ce qui donne comme valeur maximale pour la variance :

$$\begin{aligned}
 \sigma_{max}^2 &= \frac{(\mu - \epsilon)^2}{3} \\
 &\approx \frac{100}{3} \text{ dans notre cas.}
 \end{aligned}$$

Evolution du nombre d'infectés en fonction du temps

Les figures 3.7 et 3.8 représentent l'évolution du nombre d'infectés au cours du temps pour différentes valeurs de la variance respectivement sans et avec barres d'erreurs. Nous pouvons observer qu'entre deux valeurs non-nulles de la variance, la taille de l'épidémie est plus grande dans le cas où la variance est la plus élevée. Cependant, pour certains temps, le cas déterministe présente une taille de l'épidémie supérieure

aux cas stochastiques : plus la variance est faible, plus est grand le temps à partir duquel la taille de l'épidémie devient définitivement supérieure pour le cas stochastique (par rapport au cas déterministe). Pour des valeurs plus basses de la variance, il n'est pas évident que le nombre de personnes infectées est toujours supérieur au cas déterministe (*i.e* lorsque la variance est nulle).

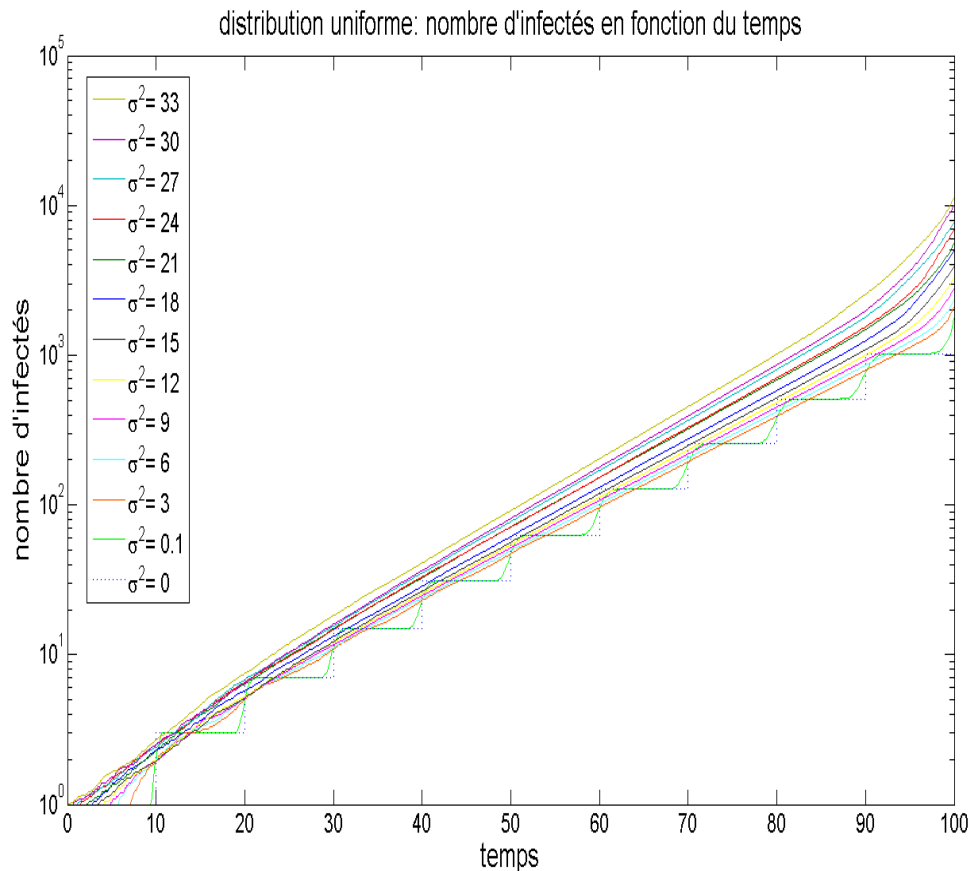


FIGURE 3.7 – Distribution uniforme. Evolution du nombre d'infectés au cours du temps pour différentes valeurs de la variance lorsque $\mu = 10$. La taille de l'épidémie dans le cas stochastique semble augmenter en fonction de la variance. Plus la variance du processus stochastique est faible, et plus le temps pour que la taille de l'épidémie soit supérieure au cas déterministe est élevé.

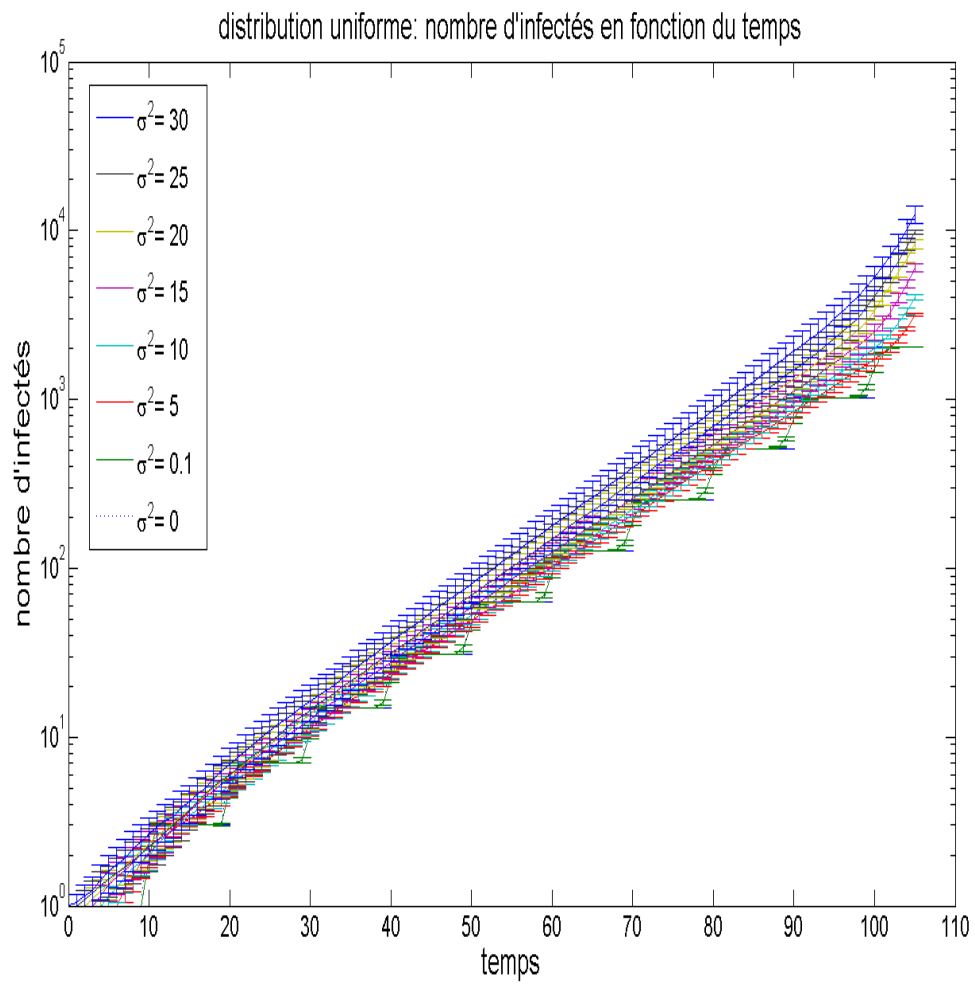


FIGURE 3.8 – Distribution uniforme. Evolution du nombre d'infectés au cours du temps pour différentes valeurs de la variance lorsque $\mu = 10$ avec barres d'erreurs.

Evolution du nombre d'infectés après 105s en fonction de la variance

La figure 3.9 représente le nombre d'infectés en fonction de la variance après 105 secondes. Même si de proche en proche les barres d'erreurs se chevauchent, nous pouvons voir une différence significative entre des valeurs assez éloignées les unes des autres (par exemple pour $\sigma^2 \in \{0, 5, 15, 25\}$).

Ainsi, pour un temps de 105 secondes, un processus stochastique crée une plus grande épidémie que le cas dynamique.

Notons ici que l'échelle de l'axe vertical est linéaire et non logarithmique comme c'était le cas pour l'étude de la moyenne (figure 3.5), ce qui démontre un plus faible impact de la modification de la variance par rapport à une modification de la moyenne.

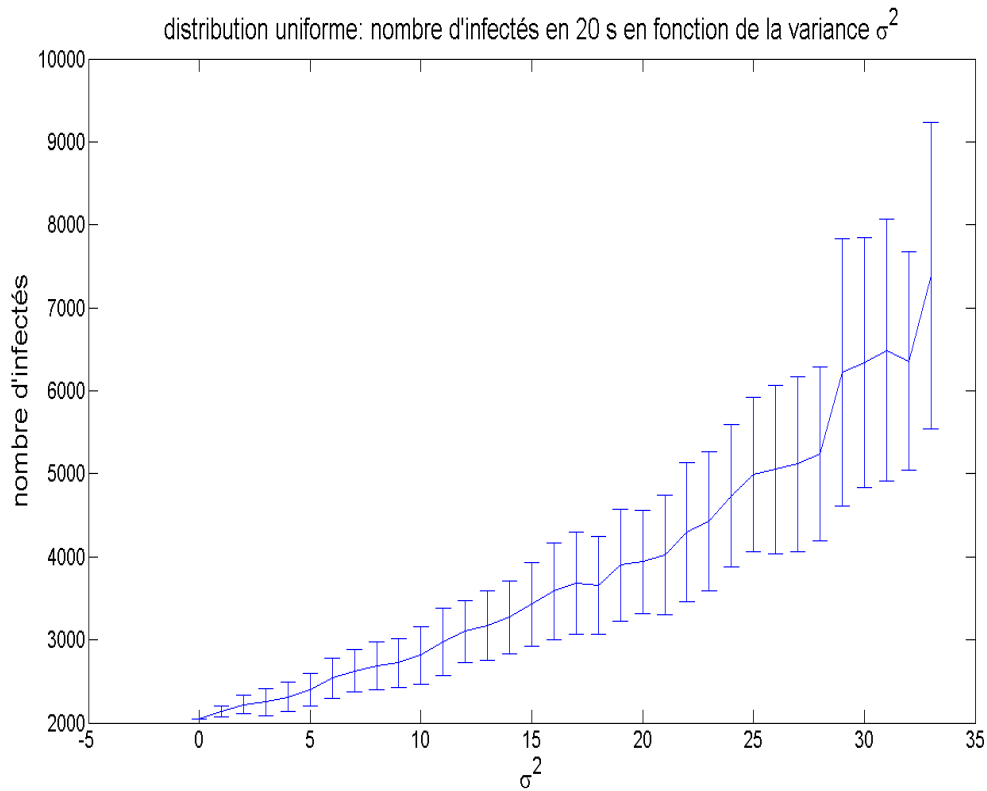


FIGURE 3.9 – Nombre de personnes infectées en fonction de la variance après 105s. Une augmentation de la variance entraîne une augmentation de la taille de l'épidémie.

Evolution du nombre d'infectés après 100s en fonction de la variance

La figure 3.10 représente le nombre d'infectés en fonction de la variance après cette fois 100 secondes. A nouveau, nous observons que la taille de l'épidémie est supérieure pour des valeurs très grandes de la variance.

Cependant, nous observons une discontinuité à l'origine. La taille de l'épidémie lorsque la variance est nulle est à nouveau atteinte aux environs de $\sigma^2 = 11$. Donc, le processus stochastique peut avoir pour impact de diminuer la taille de l'épidémie si la variance n'est pas suffisamment élevée.

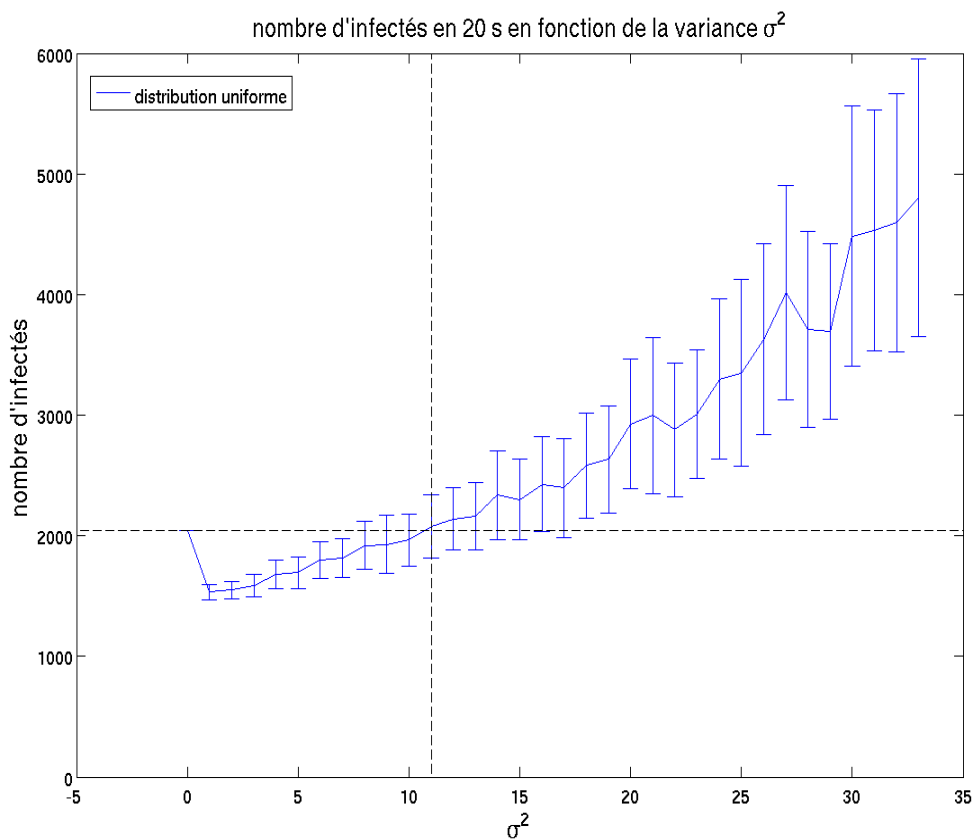


FIGURE 3.10 – Nombre de personnes infectées en fonction de la variance après 100 secondes. Une augmentation de la variance entraîne une augmentation de la taille de l'épidémie dans le cas stochastique. Pour de faibles valeurs de la variance (environ $\sigma^2 < 11$), la taille de l'épidémie est inférieure à celle du cas déterministe.

La discontinuité dans le graphe se justifie intuitivement. En effet, le temps auquel nous étudions la taille de l'épidémie est égal à exactement 10 fois la moyenne. Ainsi, la contamination de la dixième génération vient juste de se produire et le nombre d'infectés dans le cas déterministe vaut $2^{10+1} - 1 = 2047$.

Considérons maintenant le cas où la variance est très petite, par exemple prenons $\sigma^2 = 0.1$. Donc l'intervalle de temps dans lequel sont tirées les valeurs aléatoires est $[10 - \sqrt{\frac{3}{10}}, 10 + \sqrt{\frac{3}{10}}] \approx [9.5, 10.5]$. Globalement, lorsque nous nous plaçons à la moitié de l'intervalle de temps séparant deux contaminations de générations dans le cas déterministe, c'est-à-dire 5 secondes, la propagation de la maladie dans l'arbre s'effectue de manière similaire pour le cas stochastique : la première contamination survient entre 9.5 et 10.5 secondes, la deuxième entre 19 et 21 secondes, la neuvième entre 85.5 et 94.5 secondes et la dixième s'effectuera entre 95 et 105 secondes. Ainsi, après 100 secondes, il est certain que tous les noeuds de la neuvième génération seront contaminés. Par contre, la densité de probabilité de la distribution uniforme étant symétrique, il y a autant de chances que la contamination du dixième noeud ait lieu avant ou après 100 secondes exactement. L'espérance du nombre de personnes infectées après 100 secondes est alors de $2^{9+1} - 1 + \frac{2^{10}}{2} = 1523$, ce qui est inférieur à celle du cas déterministe.

Une question que l'on pourrait se poser est de déterminer, en fonction de la moyenne μ et de l'entier k , une valeur minimale de la variance à partir de laquelle la taille de l'épidémie prise en un temps égal à $k\mu$ secondes soit supérieure au cas déterministe. Calculons la condition minimale pour qu'au moins un individu de la $k+1^e$ génération soit infecté avant que toute la génération k soit infectée, donc :

$$(k+1)a < kb.$$

En notant

$$\begin{aligned} a &= \mu - x \\ b &= \mu + x, \end{aligned}$$

d'où on tire

$$\begin{aligned} \sigma_{\min}^2 &= \frac{(b-a)^2}{12} \\ &= \frac{x^2}{3}. \end{aligned}$$

On peut résoudre l'inéquation :

$$\begin{aligned} (k+1)a &< kb \\ \mu &< (2k+1)x \\ \frac{\mu}{2k+1} &< \sqrt{3\sigma_{\min}^2} \\ \frac{\mu^2}{3(2k+1)^2} &< \sigma_{\min}^2. \end{aligned}$$

Cependant, cette valeur nous donne la possibilité qu'au moins un individu de la génération $k+1$ soit infecté avant le dernier de la génération k . Or nous voulons pouvoir compenser la non-contamination de 2^{k-1} individus, donc nous devons permettre la contamination de 2^{k-1} individus avant celle du dernier de la génération k , ce qui est plus compliqué à déterminer. Le tableau suivant recense les différentes valeurs obtenues expérimentalement pour le cas $\mu = 10$.

k	2	3	4	5	6	7	8	9	10
valeur estimée expérimentalement	31	27	25	21	18	15	13	12	11

TABLE 3.1 – Valeurs minimales expérimentales de la variance pour lesquelles la taille de l'épidémie devient plus grande dans le cas stochastique que dans le cas déterministe à partir de la génération k .

Taille du plus long chemin

En lien direct avec ce que nous venons d'évoquer, nous pouvons nous intéresser à la profondeur de l'épidémie, c'est-à-dire taille de la plus longue cascade d'infection en fonction du temps et de la variance, en fixant la valeur de l'un de ces paramètres et en faisant varier l'autre. Dans le cas d'un arbre, la longueur d'une cascade est indiquée par la génération de son extrémité.

La figure 3.11 représente l'évolution de la longueur maximale d'une cascade en fonction du temps. Nous observons une augmentation de cette longueur en fonction de la variance. La figure 3.12 représente la même chose pour de faibles valeurs de la variance. La longueur maximale moyenne est toujours plus grande (au sens large) dans le cas stochastique que dans le cas déterministe. Cependant, le temps nécessaire pour que cette supériorité demeure stricte est élevé pour de faibles valeurs de σ^2 et diminue lorsque la variance augmente.

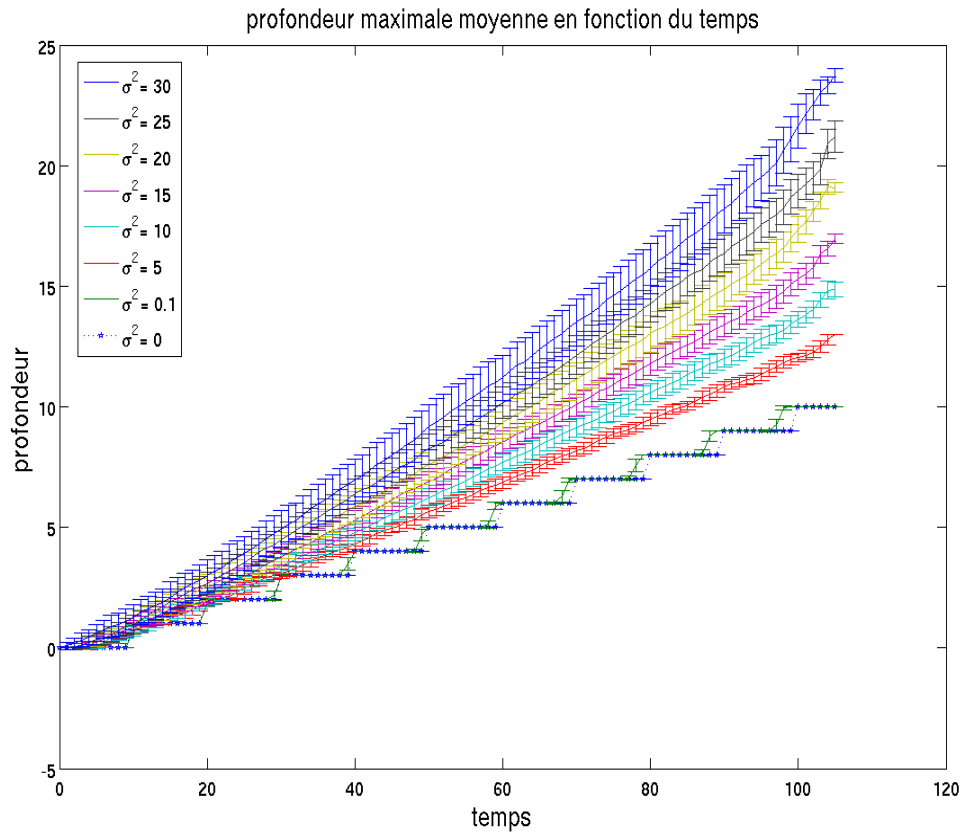


FIGURE 3.11 – Taille de la plus longue cascade en fonction du temps. Une augmentation de la variance entraîne une augmentation de la génération du noeud le plus éloigné de la source.

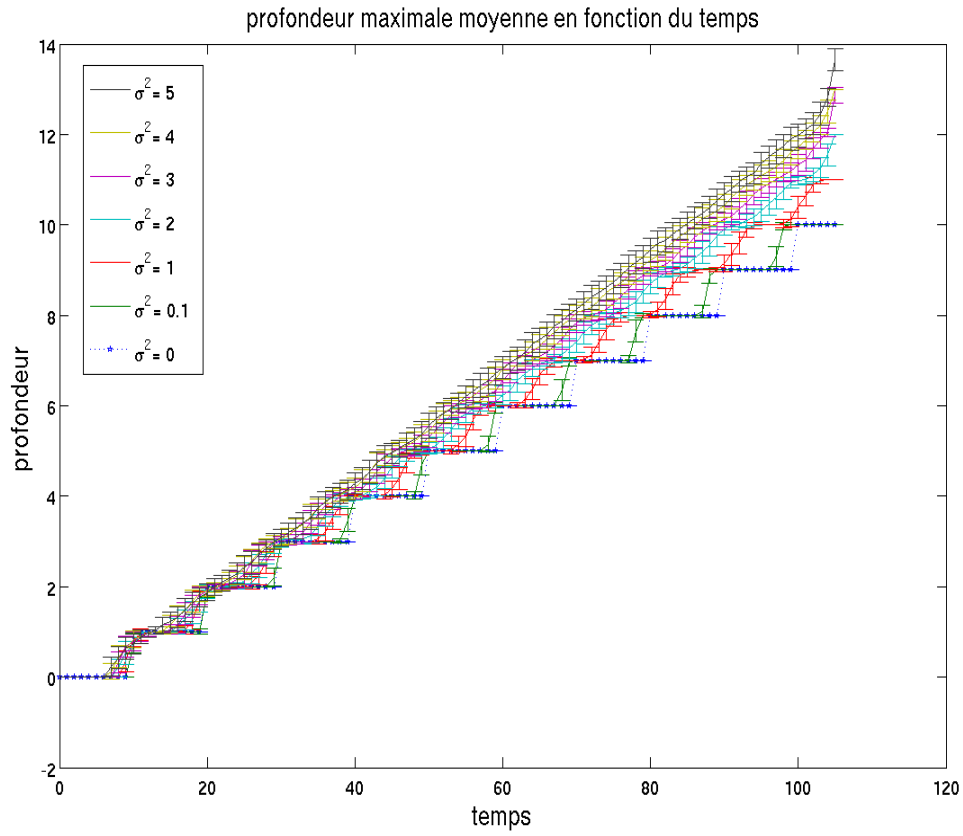


FIGURE 3.12 – Taille de la plus longue cascade en fonction du temps pour de petites valeurs de σ^2 . Le temps nécessaire pour que la distance maximale soit toujours supérieure au cas déterministe diminue avec le temps. Elle reste élevée pour de faibles valeurs de σ^2 .

Fixons maintenant le temps à respectivement 40 (figure 3.13), 100 (figure 3.14) et 105 secondes (figure 3.15). Nous observons bien la croissance en fonction de la variance. Lorsque le temps considéré n'est pas un multiple entier de la moyenne, nous observons un saut entre la valeur en 0 et la valeur suivante non-nulle, ce saut est moins prononcé dans le cas contraire. La distance maximale peut aller jusqu'à doubler pour les plus grandes valeurs de la variance.

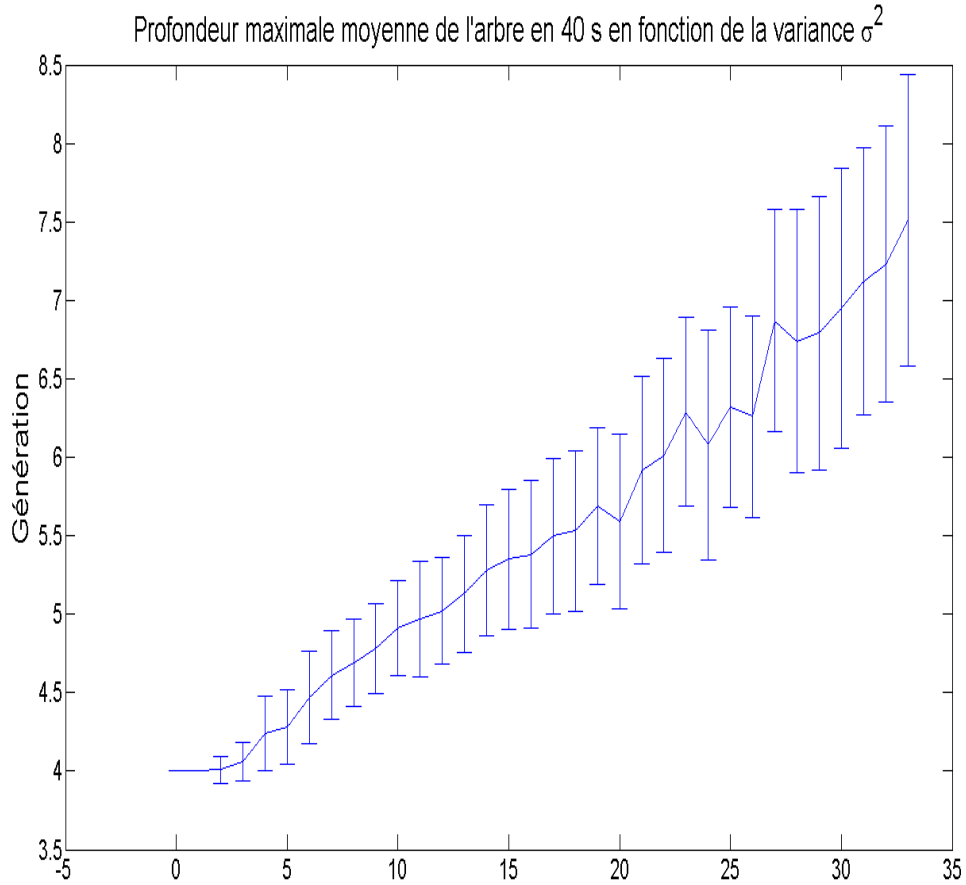


FIGURE 3.13 – Taille de la plus longue cascade après 40 s en fonction de la variance σ^2 .

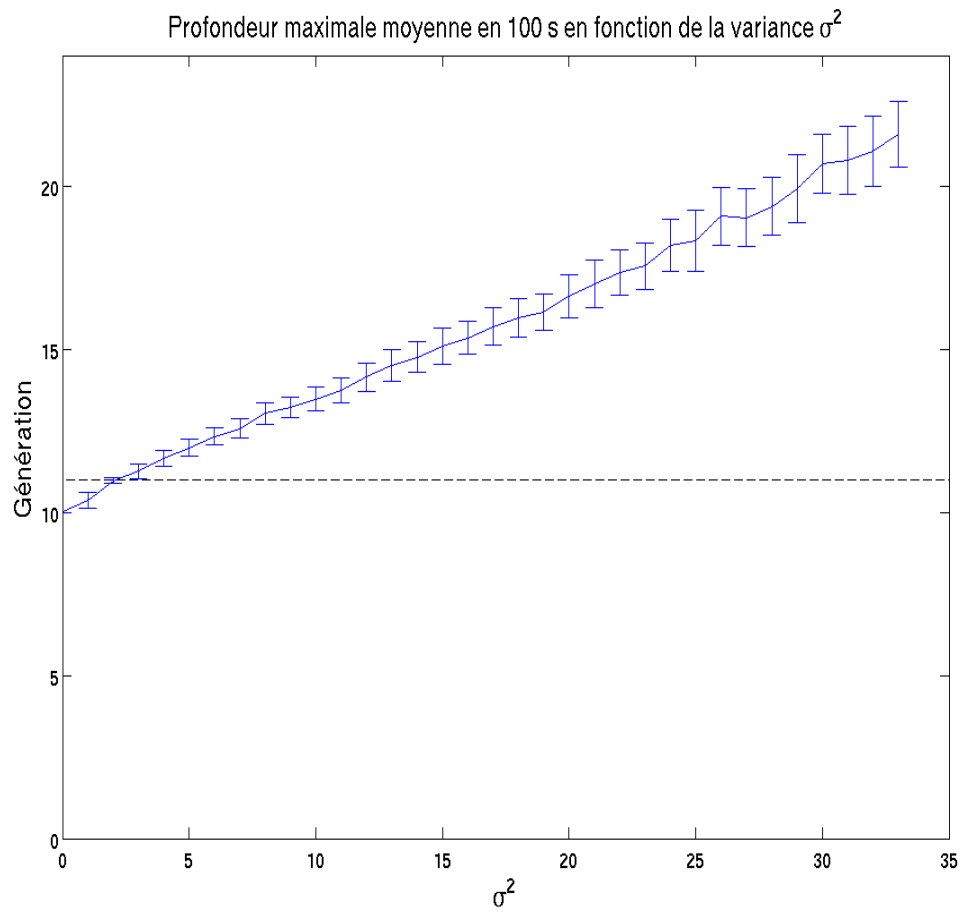


FIGURE 3.14 – Taille de la plus longue cascade après 100 s en fonction de la variance σ^2 .

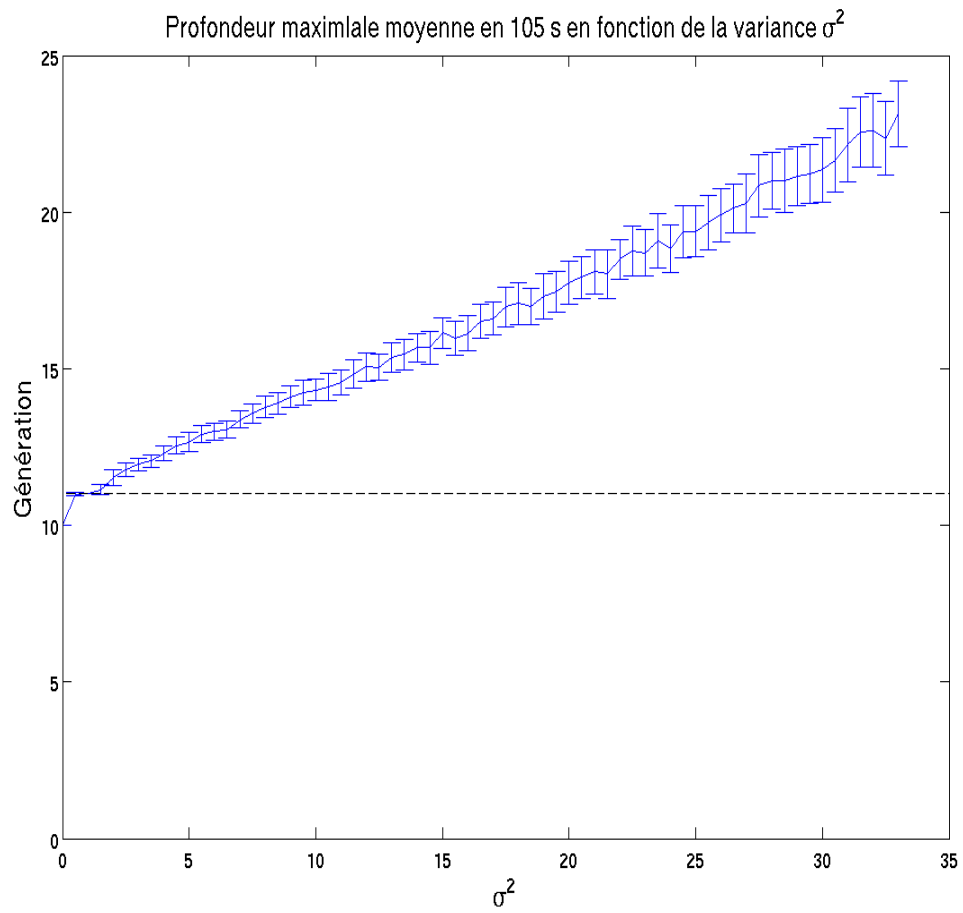


FIGURE 3.15 – Taille de la plus longue cascade après 105 s en fonction de la variance σ^2 .

3.4.4 Etude en fonction de la moyenne et de la variance

Nous pouvons représenter le nombre d'infectés en fonction de la moyenne et de la variance. Pour cela, nous représentons sur une grille, ayant pour abscisse la moyenne et pour ordonnée la variance, la taille de l'épidémie par une couleur allant du rouge au mauve en passant par le vert et le bleu, le rouge correspondant correspondant aux plus petites valeurs et le mauve aux plus grandes.

La figure 3.16 présente un coloriage de la grille pour des temps égaux à 5, 10, 20 et 40 secondes. Nous observons une augmentation du nombre d'infectés lorsque la moyenne diminue et lorsque la variance augmente. Une modification de la moyenne a une bien plus grande influence sur la modification de la taille de l'épidémie qu'une modification de la variance. Pour des moyennes dont le temps observé est un multiple entier, nous observons l'effet décrit précédemment, à savoir que pour de faibles valeurs de la variance, la taille de la propagation dans le cas déterministe est supérieure à celle du processus stochastique.

La figure 3.17 présente quant à elle un coloriage de la grille pour des temps égaux à 5.375, 10.75, 21.5 et 43 secondes. Nous observons à nouveau une croissance de la taille de l'épidémie lorsque la moyenne diminue ou lorsque la variance augmente, avec une plus grande influence de la moyenne. Les temps d'observation n'étant pas des multiples entiers des moyennes, nous observons une valeur plus élevée, même pour une variance faible, de la taille de l'épidémie dans le cas stochastique que dans le cas dynamique.

Une autre visualisation des résultats est donnée par les figures 3.18 et 3.19 qui représentent en trois dimensions le nombre d'infectés. Nous voyons bien une plus forte descente suivant l'axe des moyennes que suivant celui des variances.

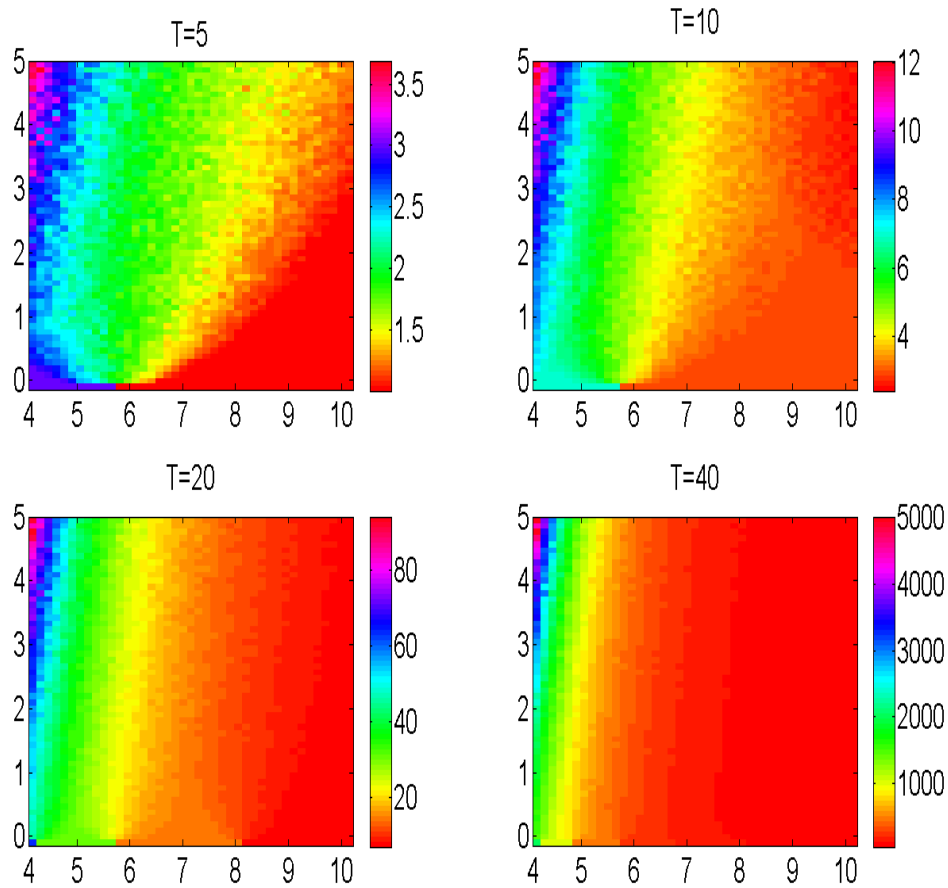


FIGURE 3.16 – Nombre de personnes infectées en fonction de la moyenne et de la variance. Une diminution de la moyenne ou une augmentation de la variance entraîne une augmentation de la taille de l'épidémie. La modification de la moyenne a un impact bien plus important que celle de la variance. La moyenne varie de 4 à 10, la variance de 0 à 5. Les temps d'observation valent 5, 10, 20 et 40 secondes. Lorsqu'il s'agit d'un nombre entier de fois la moyenne, nous observons que le nombre d'infectés pour une variance faible non nulle est inférieur au cas où la variance est nulle.

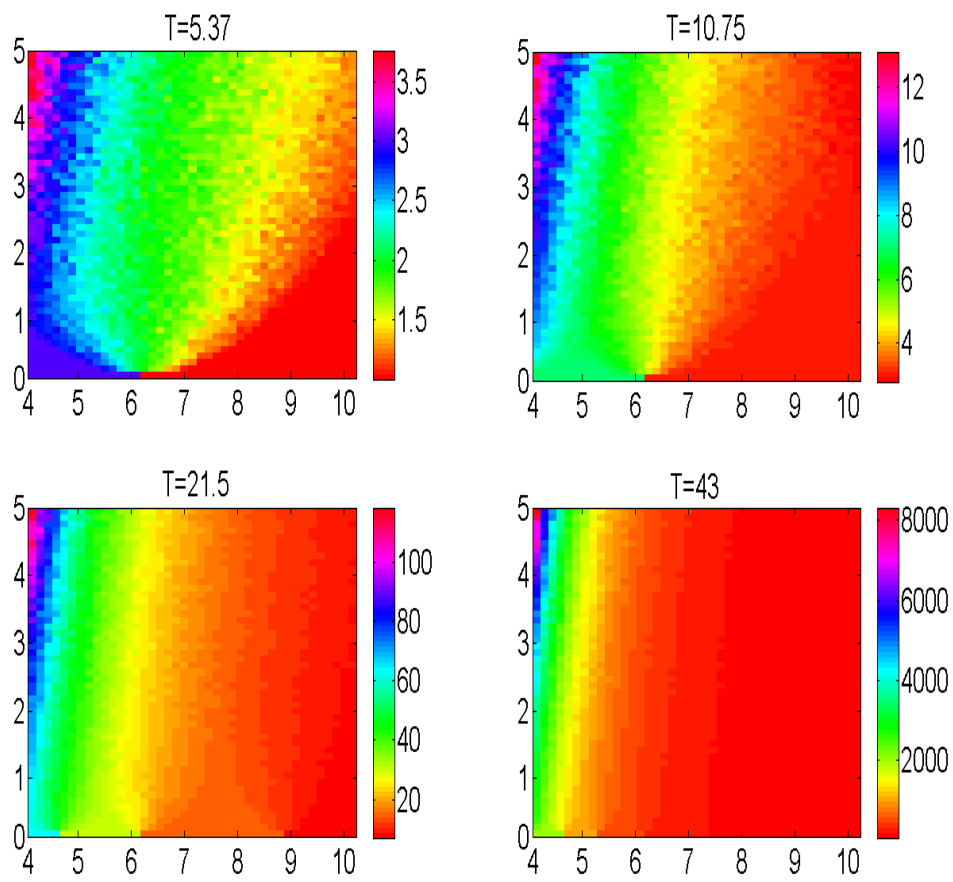


FIGURE 3.17 – Nombre de personnes infectées en fonction de la variance après 105s. La taille de l'épidémie augmente lorsque diminue la valeur de la moyenne ou augmente celle de la variance. La moyenne varie de 4 à 10 et la variance de 0 à 5. Les temps d'observation sont 5.375, 10.75, 21.5 et 43 secondes.

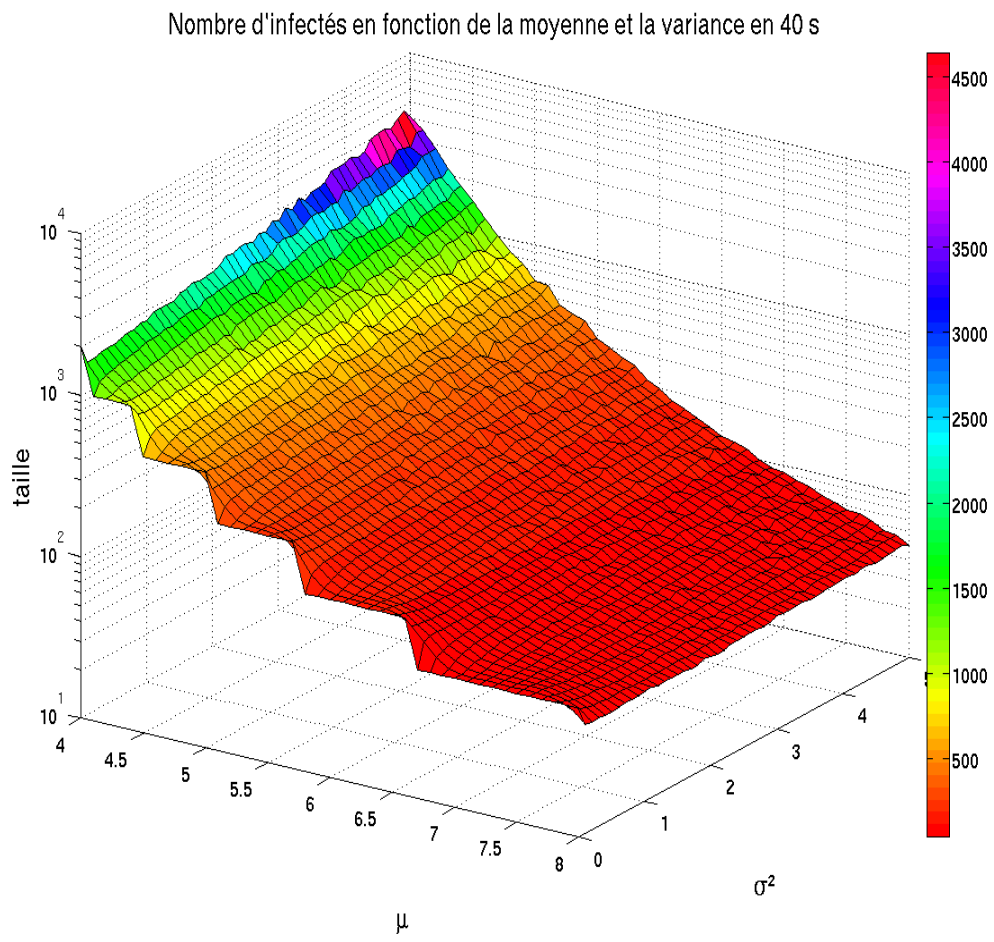


FIGURE 3.18 – Représentation en 3 dimensions après 40 secondes.

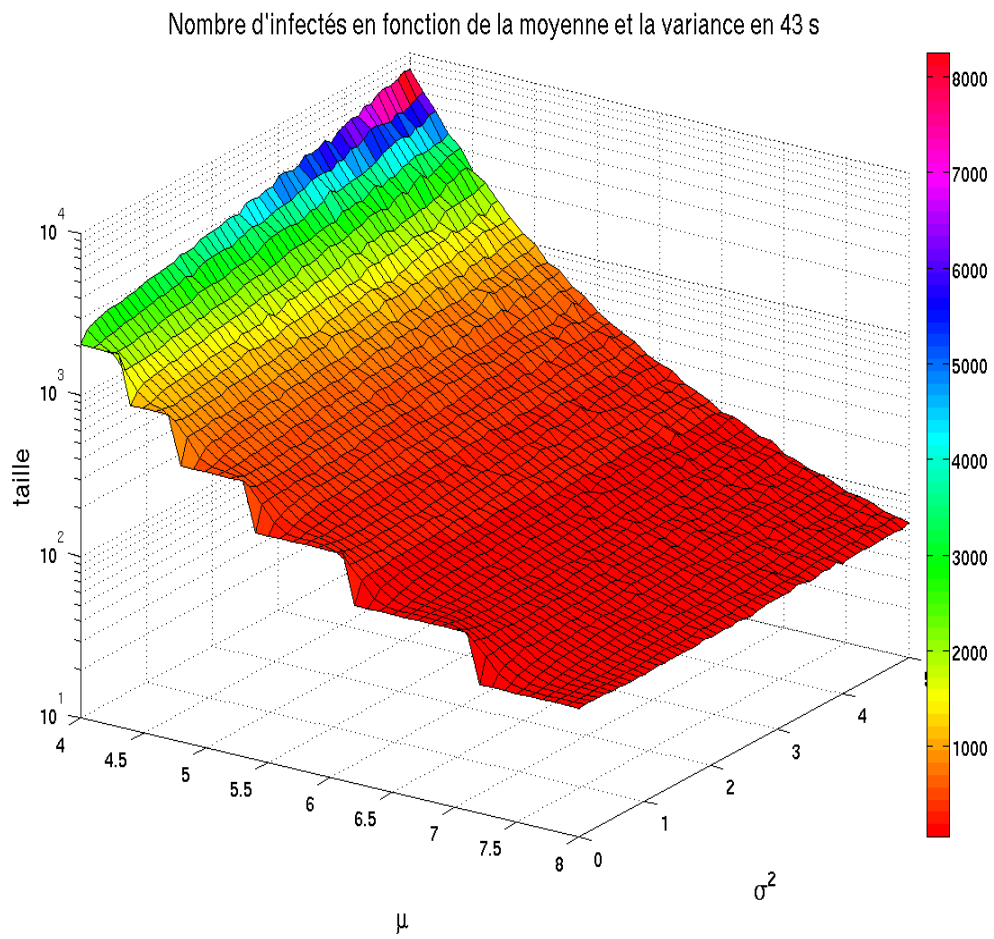


FIGURE 3.19 – Représentation en 3 dimensions après 43 secondes.

3.5 Résultats : conclusions

Nous avons tiré deux conclusions majeures de ces simulations :

1. Une augmentation de la moyenne entraîne une diminution de la taille de l'épidémie.

Cette première conclusion est intuitive : plus le temps moyen pour qu'une contamination ait lieu est élevé, plus la propagation sera lente.

2. Une augmentation de la variance entraîne une augmentation de la taille de l'épidémie.

Une explication à cette deuxième observation pourrait provenir du fait que le processus repose sur les valeurs minimales des tirages. En effet, pour déterminer quel noeud sera le suivant infecté, nous regardons parmi l'ensemble E des noeuds dont le prédécesseur est infecté celui dont le temps restant avant contamination est le plus petit. Ensuite, nous tirons deux nouveaux temps d'infection pour ses successeurs et ajoutons à E les deux successeurs du noeud nouvellement infecté, et réitérons la sélection. Parmi les temps tirés, ceux dont la valeur est petite ont plus d'influence que ceux dont la valeur est grande car ils seront plus vite sélectionnés pour transmettre la maladie et contamineront deux nouveaux noeuds, augmentant ainsi le réservoir d'individus directement contaminables.

Conclusion

Dans ce mémoire, nous avons effectué une revue de littérature portant sur un sujet de recherche moderne, mis en perspective les différents travaux et apporté deux contributions personnelles.

Nous avons d'abord montré que la théorie des processus sur réseaux statiques doit impérativement, si l'on veut considérer des modèles plus réalistes, être généralisée, voire reconsidérée, pour intégrer les aspects dynamique et non-markoviens observés empiriquement.

Nous avons étudié deux modèles allant dans ce sens. Le premier consiste en un modèle de propagation compartimental pour lequel on peut montrer que la force dynamique des liens est déterminée par une matrice qui diffère de celle d'adjacence. Nous avons ensuite étudié la dynamique de ce modèle. Le seconde étudie des marches aléatoires via des manipulations sur l'espace de Laplace associé, et permet d'obtenir une équation intégro-différentielle décrivant son évolution.

Comparant ces deux développements, nous avons apporté une première contribution personnelle en mettant en lumière des différences fondamentales entre une propagation type SI dans un graphe et l'agrégation de marches aléatoires, là où celles-ci ne sont pas évoquées dans la littérature. Ces différences résident principalement dans les corrélations de l'état de noeuds possédant un voisin commun, mais aussi dans l'impact des communautés.

Nous nous sommes également intéressés à une question dont la réponse varie en fonction du réseau étudié, celle de l'influence de la variance du temps d'attente avant infection sur la taille d'une épidémie. Contrairement aux résultats existant dans la littérature, principalement basés sur des simulations sur données empiriques, nous avons effectué des simulations numériques sur un graphe dont la topologie est simple et connue. Nous avons observé que l'augmentation de cette variance a pour effet, pour un temps donné, d'augmenter la taille de l'épidémie ainsi que celle de la plus

longue cascade d'infection.

Les recherches supplémentaires qui pourraient être menées dans la continuité de ce mémoire sont nombreuses. Deux axes majeurs semblent néanmoins se dégager : d'une part, développer d'autres processus stochastiques dans lesquels les hypothèses de Markov seraient rejetées ; d'autre part, considérer l'impact de la variance du temps d'attente avant infection sur d'autres données, telles que la taille moyenne de la plus courte cascade d'infection ou la distribution moyenne des longueurs des cascades au cours du temps. Il serait également intéressant d'élargir cette étude à des réseaux dont la topologie est plus complexe, ou encore de les comparer à des données réelles.

Bibliographie

- [1] M.E.J Newman. Epidemics on network. In *Networks : an introduction*, chapter 17. Oxford University Press, 2010.
- [2] Nicola Perra, Duygu Balcan, Bruno Gonçalves, and Alessandro Vespignani. Towards a characterization of behavior-disease models. *PLoS ONE*, July 2011.
- [3] Cristopher Moore and Mark E. J. Newman. Epidemics and percolation in small-world networks. *Phys. Rev. E*, pages 5678–5682, 2000.
- [4] Stanley Milgram. The small-world problem. *Psychology Today*, pages 61–67, 1967.
- [5] Brian Karrer and M. E. J. Newman. A message passing approach for general epidemic models. *Physical Review E*, July 2010.
- [6] E. Moro. Spreading on (dynamical) social networks. Slides from the Aalto Complex Networks Factory, june 2012.
- [7] Giovanna Miritello, Esteban Moro, and Rubén Lara. Dynamical strength of social ties in information spreading. *Phys. Rev. E*, Apr 2011.
- [8] Giovanna Miritello, Ruben Lara, and Esteban Moro. Time allocation in social networks : correlation between social structure and human communication dynamics. unpublished, 2012.
- [9] José Luis Iribarren and Esteban Moro. Branching dynamics of viral information spreading. *Phys. Rev. E*, 84 :046116, Oct 2011.
- [10] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A. L. Barabási, and J. Saramäki. Small but slow world : How network topology and burstiness slow down spreading. *Physical Review E*, (2), February 2011.
- [11] Luis E. C. Rocha, Fredrik Liljeros, and Petter Holme. Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts. *PLoS Computational Biology*, (3), March 2011.
- [12] Eben Kenah and James M. Robins. Second look at the spread of epidemics on networks. *Phys. Rev. E*, 76 :036113, Sep 2007.

- [13] Marc Barthélemy, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.*, 92 :178701, Apr 2004.
- [14] Herbert W. Hethcote and David W. Tudor. Integral equation models for endemic infectious diseases. *Journal of Mathematical Biology*, 1980.
- [15] Till Hoffmann, Mason A. Porter, and Renaud Lambiotte. Generalized master equations for non-poisson dynamics on networks. *Phys. Rev. E*, 86 :046102, Oct 2012.
- [16] S. Redner. *Random walk/Diffusion*, chapter 2.
- [17] S. Redner. *Growing networks*, chapter 9.
- [18] K. Chu. *Exemples of Saddle-Point Asymptotics*, chapter 8. 2003.
- [19] A. Clauset. *The Poisson process*, chapter 1. 2011.
- [20] Till Hoffmann, Mason A. Porter, and Renaud Lambiotte. Random walks on stochastic temporal networks. 2012.

Annexes

Les codes créés pour les simulations numériques sont fournis ici, agrémentés de quelques commentaires. Ils ont été implémentés en Matlab®.

Estimation de l'impact de la moyenne

```
% Propagation SI dans un arbre binaire: Détermination de l'influence de la
% moyenne
% Auteur: Martin Gueuning
% Annee: 2012-2013

clear all
clc
close all
format long
Tfin=20; % Temps final de la simulation
valmu=10:-1:2; % Valeurs de la moyenne
eps=10^{-15};
for choix=[1 2]
    % 1 pour tracer le nombre moyen de personnes infectées en fonction du
    % temps
    % 2 pour déterminer le nombre moyen de personnes infectées en un temps
    % T fixe (Tfin) en fonction de la moyenne mu
    tpmoy=0:0.1:Tfin; %tpmoy représente le temps d'infection moyen
    % d'un certain nombre de noeuds, utilise pour la representation du
    % nombre d'infectés en fonction du temps.
    e=0; % initialisation du vecteur d'erreurs

    for loi=[1 2] % loi 1: cas markovien: distribution exponentielle
                    % loi 2: distribution uniforme

        indmu=0; %indice de boucle

        for mu=valmu
            inf_reg=[]; % stocke le nombre total d'infectés de toutes les
                        % simulations de meme loi et meme moyenne
                        % à un pas regulier

            tps_j=[]; % Stocke les temps d'infection d'une simulation
            nb_j=[]; % Stocke le nombre d'infectés en tout temps pour
                    % une simulation
```

```

indmu=indmu+1;
nbtotinf(indmu)=0; % Initialisation

for j=1:100
j/100 % Indique le pourcentage de simulations accomplies
P=[];% P(i) représente la génération (ou étage) du noeud i
I=[]; % I(i) représente le numéro du noeud i sur son étage
rest=[]; %rest(i) indique le temps restant avant
% contamination du noeud i (éventuellement négatif)

    % Tirage des temps d'infection de deux successeurs
    switch loi
    case {1}
        R=random('exp',mu,[1,2]);
        % Distribution exponentielle negative
    case{2}
        R=random('unif',eps,2*mu-eps,[1,2]);
        % Loi uniforme
    end

% Contamination via le noeud source
P(1)=1;
P(2)=1;
I(1)=1;
I(2)=2;
i=2; % i représente le nombre de noeuds deja numérotés
Ttot=0; % durée totale de la propagation en cours
rest=R;
nbtotinf_j(j)=1; % nombre de personnes contaminées
                % lors de l'itération j
comptw=1; % compteur de boucle
tpsinf(1)=0;
tps_j(1)=0;
nb_j(1)=1;

% Sélection progressive du noeud responsable de
% la contamination et génération de ses
% successeurs jusqu'au temps final (Tfin)
while (Ttot+min(rest(rest>0)) <= Tfin)
    % [mu,Ttot,j/100]
    comptw=comptw+1;
    temp=rest;
    temp(temp<=0)=[]; % Vecteur ne reprenant que
                    % les temps d'attente >0
    tmin=min(temp); % Détermine le temps avant
                    % la prochaine infection
    nselect=find(rest<=tmin & rest>=tmin);
    % indice correspondant au noeud sélectionné

    rest=rest-tmin*ones(size(rest));

    for k=1:length(nselect)
        P(i+2*k-1)=P(nselect(k))+1;
        P(i+2*k)=P(nselect(k))+1;
        I(i+2*k-1)=2*I(nselect(k))-1;
        I(i+2*k)=2*I(nselect(k));
    end
end

```

```

% Tirage du temps d'attente avant infection sur les deux nouvelles arêtes

        switch loi
            case {1}
                R=random('exp',mu,[1,2]);
            case{2}
                R=random('unif',eps,2*mu-eps,[1,2]);
            end

        rest(i+2*k-1)=R(1);
        rest(i+2*k)=R(2);
    end

    nbtotinf_j(j)=nbtotinf_j(j)+length(nselect);
    Ttot=Ttot+tmin;
    i=i+2*length(nselect); % 2*length(nselect) noeuds
                          % viennent d'être numérotés

    tps_j(comptw)=Ttot;
    nb_j(comptw)=nbtotinf_j(j);
end % Fin de la propagation

% Stockage du nombre de contaminés en pas réguliers pour l'iteration j
comptk=0;
for k=tpmoy
    comptk=comptk+1;
    inf_reg(comptk,j)=nb_j(length(tps_j(tps_j<=k)));
end

end % Fin des 100 tirages similaires

switch choix
    case{1}
        nbmoy=mean(inf_reg,2);
        e=std(inf_reg',1)/2;
        plo(indmu)=errorbar(tpmoy,nbmoy,e);
        xlabel('temps')
        ylabel('nombre d''infectés')

        switch loi
            case {1}
                title('exponentielle décroissante')
            case{2}
                title('distribution uniforme');
            end

        end
        muuu=num2str(mu);
        msg=['\mu= ',muuu];
        set(plo(indmu),'userdata',msg)
        hold all
    case{2}
        nbtotinf(indmu)=mean(nbtotinf_j);
        e(indmu)=std(nbtotinf_j)/2;
end
end

```

```

end

switch choix %Legendes

    case{1}
        hc=get(gca,'children');
        s={};
        for tempo=hc'
            s={s{:},get(tempo,'userdata')};
        end
        legend(hc,s)
        figure

    case{2}
        plo2(loi)=errorbar(valmu,nbtotinf,e);
        title('nombre d''infectés en 20 s en fonction de la moyenne \mu')
        xlabel('\mu')
        ylabel('nombre d''infectés')
        hold all
        switch loi
            case {1}
                msg2='exponentielle décroissante';
            case{2}
                msg2='distribution uniforme';
        end
        set(plo2(loi),'userdata',msg2)
        hold all
        hc2=get(gca,'children');
        s={};
        for tempo2=hc2'
            s={s{:},get(tempo2,'userdata')};
        end
        legend(hc2,s)
    end
end
end
end

```


Estimation de l'impact de la variance

```
% Propagation SI dans un arbre binaire
% Auteur: Martin Gueuning
% Annee: 2012-2013

clear all
clc
close all
format long

Tfin=100; % Temps final de la simulation
mu=10;

for choix=[1 2 3 4]
    % 1 pour tracer le nombre moyen de personnes infectées en fonction du
    % temps pour différentes valeurs de la variance
    % 2 pour déterminer le nombre moyen de personnes infectées en un temps
    % T fixe (Tfin) en fonction de la variance
    % 3 pour déterminer la profondeur maximale moyenne en un temps donné en fonction
    % de la variance
    % 4 pour tracer la profondeur maximale moyenne en fonction du temps pour
    % différentes valeurs de la variance
    switch choix % Détermination des valeurs de la variance étudiées
        case{1,4}
            var=0:5:10*floor(mu^2/30);
            var(1)=0.1;
            var=[0 var];
            valx=sqrt(3)*var.^0.5;
        case{2,3}
            var=0:mu^2/3;
            valx=sqrt(3)*var.^0.5;
    end

    tpmoy=0:Tfin; % tpmoy représente le temps d'infection moyen
    % d'un certain nombre de noeuds,
    % utilisés pour la représentation du nombre d'infectés en fonction du temps.
    e=0; % réinitialisation du vecteur d'erreurs

    indvar=0; % indice de boucle

    for x=valx
        inf_reg=[]; % stocke le nombre total d'infectés de toutes les
        % simulations de meme loi et meme moyenne
        % à un pas regulier

        tps_j=[]; % Stocke les temps d'infection d'une simulation
        nb_j=[]; % Stocke le nombre d'infectés en tout temps pour
        % une simulation
        indvar=indvar+1;
        nbtotinf(indvar)=0; % variable temporaire, stocke le nombre d'infectés d'une itération
        profondeur(indvar)=0;
        for j=1:100
            j/100 % Indique le pourcentage de simulations accomplies
            P=[]; % P(i) represente la génération (ou étage) du noeud i
            I=[]; % I(i) represente le numero du noeud i sur son étage
            rest=[]; %rest(i) indique le temps restant avant contamination du noeud i
```

```

% (eventuellement negatif)
T=[]; % T(i) indique le temps de contamination nécessaire pour contaminer le noeud i
R=random('unif',mu-x, mu+x,[1,2]);

%Contamination via le noeud source
P(1)=1;
P(2)=1;
I(1)=1;
I(2)=2;
i=2; % i représente le nombre de noeuds déjà numérotés
Ttot=0; % durée totale de la propagation en cours
rest=R;
nbtotinf_j(j)=1; % nombre de personnes contaminées lors de la simulation j
profondeur_j(j)=1; % profondeur moyenne maximale lors de la simulation j
comptw=1; % compteur de boucle
tpsinf(1)=0;
nbinf(1)=1;
prof(1)=0;
tps_j(1)=0;
nb_j(1)=1;

% Sélection progressive du noeud responsable de
% la contamination et génération de ses
% successeurs jusqu'au temps final (Tfin)
while (Ttot+min(rest(rest>0)) <= Tfin)
% [(x/sqrt(3))^2,Ttot,j/100]
comptw=comptw+1;
temp=rest;
temp(temp<=0)=[]; % Vecteur ne reprenant que les temps d'attente >0
tmin=min(temp); % Détermine le temps avant la prochaine infection
nselect=find(rest<=tmin & rest>=tmin); % indice correspondant au noeud sélectionné
rest=rest-tmin*ones(size(rest));

for k=1:length(nselect)
P(i+2*k-1)=P(nselect(k))+1;
P(i+2*k)=P(nselect(k))+1;
I(i+2*k-1)=2*I(nselect(k))-1;
I(i+2*k)=2*I(nselect(k));
R=random('unif',mu-x, mu+x,[1,2]);

rest(i+2*k-1)=R(1);
rest(i+2*k)=R(2);
end

nbtotinf_j(j)=nbtotinf_j(j)+length(nselect);
Ttot=Ttot+tmin;
i=i+2*length(nselect); %2*length(nselect) noeuds viennent d'etre numérotés

tps_j(comptw)=Ttot;
nb_j(comptw)=nbtotinf_j(j);
prof_j(comptw)=max(P)-1;
end %Fin de la propagation

profondeur_j(j)=max(P)-1;

```

```

    comptk=0;
    switch choix
    case{1}
    % Stockage du nombre de contaminés en pas réguliers pour l'itération j

        for k=tpmoy
            comptk=comptk+1;
            inf_reg(comptk,j)=nb_j(length(tps_j(tps_j<=k)));
        end
    case{4}
    % Stockage de la profondeur maximale en pas réguliers pour l'itération j

        for k=tpmoy
            comptk=comptk+1;
            prof_reg(comptk,j)=prof_j(length(tps_j(tps_j<=k)));
        end
    end

    end % Fin des 100 tirages similaires

switch choix
case{1}
    nbmoy=mean(inf_reg,2);
    e=std(inf_reg',1)/2;
    plo(indvar)=errorbar(tpmoy,nbmoy,e);
    xlabel('temps')
    ylabel('nombre d''infectés')
    title('distribution uniforme');
    varrr=num2str((x/sqrt(3))^2);
    msg=['\sigma^2= ',varrr];
    set(plo(indvar),'userdata',msg)
    hold all

case{2}
    nbtotinf(indvar)=mean(nbtotinf_j);
    e(indvar)=std(nbtotinf_j)/2;

case{3}
    profondeur(indvar)=mean(profondeur_j);
    e(indvar)=std(profondeur_j)/2;

case{4}
    profmoy=mean(prof_reg,2);
    e=std(prof_reg',1)/2;
    plo(indvar)=errorbar(tpmoy,profmoy,e);
    xlabel('temps')
    ylabel('profondeur')
    title('profondeur en fonction du temps');
    varrr=num2str((x/sqrt(3))^2);
    msg=['\sigma^2= ',varrr];
    set(plo(indvar),'userdata',msg)
    hold all
end
end

switch choix

```

```

case{1,4}
    hc=get(gca,'children');
    s={};
    for tempo=hc'
        s={s{:},get(tempo,'userdata')};
    end
    legend(hc,s)
    figure

case{2}
    plo2=errorbar(var,nbtotinf,e);
    title('nombre d''infectes en 105 s en fonction de la variance \sigma^2')
    xlabel('\sigma^2')
    ylabel('nombre d''infectes')
    hold all
    msg2='distribution uniforme';
    set(plo2,'userdata',msg2)
    hold all
    hc2=get(gca,'children');
    s={};
    for tempo2=hc2'
        s={s{:},get(tempo2,'userdata')};
    end
    legend(hc2,s)
case{3}
    plo3=errorbar(var,profondeur,e);
    title('Profondeur en 100 s en fonction de la variance \sigma^2')
    xlabel('\sigma^2')
    ylabel('Génération')
    hold all
    a=[-5 5*(floor(var(length(var))/5)+1)];
    b=[profondeur(1)+1 profondeur(1)+1];
    plot(a,b,'--k')
end
end
end

```

Estimation de l'impact de la moyenne et de la variance

```
% Propagation SI dans un arbre binaire: Détermination de l'influence de la
% moyenne et de la variance sur la taille totale de l'épidémie
% Auteur: Martin Gueuning
% Année: 2012-2013
```

```
clear all
clc
close all
format long
```

```
Tfin=43;% Temps final de la simulation
eps=10-15;
valmu=4:0.1:8; % valeurs de la moyennes
valvar=floor(min(valmu)2/3):-0.1:0; % valeurs de la variance
valx=sqrt(3)*valvar.0.5; % Détermination des valeurs x
% telles que la distribution uniforme sur [mu-x,mu+x] aie la variance souhaitée
indmu=0;
compteur=0;
for mu=valmu
    indmu=indmu+1;
    indvar=0;
    for x=valx
        indvar=indvar+1;
        compteur=compteur+1;
        %Initialisations
        nbcontamine_T8=0;
        nbcontamine_T4=0;
        nbcontamine_T2=0;
        nbcontamine_T=0;

        for j=1:100
            P=[];% P(i) indique la génération (ou étage) du noeud i
            I=[]; % I(i) indique le numéro du noeud i sur son étage
            R=random('unif',mu-x,mu+x,[1,2]);

            % Contamination via le noeud source
            P(1)=1;
            P(2)=1;
            I(1)=1;
            I(2)=2;
            i=2; % i représente le nombre de noeuds déjà numérotés
            Ttot=0; % durée totale de la propagation en cours
            rest=R; % rest(i) indique le temps restant avant contamination du noeud i
            nbcontamine_j=1; % nombre de personnes contaminées au temps courant
            nbcontamine_T8_j=1;
            nbcontamine_T4_j=1;
            nbcontamine_T2_j=1;
            nbcontamine_T_j=1;
            comptw=1; % compteur de boucle

            % Sélection progressive du noeud responsable de la contamination
```

```

% et génération de ses successeurs jusqu'au temps final (Tfin)

while (Ttot+min(rest(rest>0)) <= Tfin)
    comptw=comptw+1;
    temp=rest;
    temp(temp<=0)=[]; % Vecteur ne reprenant que les temps d'attente >0
    tmin=min(temp); % Détermine le temps avant la prochaine infection
    nselect=find(rest<=tmin & rest>=tmin); % indice correspondant au noeud sélectionné
    rest=rest-tmin*ones(size(rest));

    for k=1:length(nselect)
        I(i+2*k-1)=2*I(nselect(k))-1;
        I(i+2*k)=2*I(nselect(k));
        R=random('unif',mu-x,mu+x,[1,2]);
        rest(i+2*k-1)=R(1);
        rest(i+2*k)=R(2);
    end

    nbcontamine_j=nbcontamine_j+length(nselect);
    Ttot=Ttot+tmin;
    i=i+2*length(nselect);

    if (Ttot+min(rest(rest>0)) > Tfin/8) && Ttot<=Tfin/8
        nbcontamine_T8_j=nbcontamine_j;
    else if (Ttot+min(rest(rest>0)) > Tfin/4) && Ttot<=Tfin/4
        nbcontamine_T4_j=nbcontamine_j;
    else if (Ttot+min(rest(rest>0)) > Tfin/2) && Ttot<=Tfin/2
        nbcontamine_T2_j=nbcontamine_j;
    else if (Ttot+min(rest(rest>0)) > Tfin) && Ttot<=Tfin
        nbcontamine_T_j=nbcontamine_j;
    end
    end
end

end
%Fin de la propagation

nbcontamine_T8=nbcontamine_T8+nbcontamine_T8_j/100;
nbcontamine_T4=nbcontamine_T4+nbcontamine_T4_j/100;
nbcontamine_T2=nbcontamine_T2+nbcontamine_T2_j/100;
nbcontamine_T=nbcontamine_T+nbcontamine_T_j/100;
end

grilleT8(indvar,indmu)=nbcontamine_T8;
grilleT4(indvar,indmu)=nbcontamine_T4;
grilleT2(indvar,indmu)=nbcontamine_T2;
grilleT1(indvar,indmu)=nbcontamine_T;
end
end

surf(valmu,floor(min(valmu)^2/3):-0.1:0,grilleT1)

title('Nombre d''infectés en fonction de la moyenne et la variance en 43 s')
xlabel('\mu','fontsize',20);
ylabel('\sigma^2','fontsize',20);

```

```
zlabel('taille','fontsize',20);  
colorbar
```

```
figure  
subplot(2,2,1)  
imagesc(grilleT8)  
subplot(2,2,2)  
imagesc(grilleT4)  
subplot(2,2,3)  
imagesc(grilleT2)  
subplot(2,2,4)  
imagesc(grilleT1)
```