

# Étude statistique et dynamique de la propagation d'épidémies dans un réseau de *petit monde*

Nouredine Zekri<sup>a,b</sup>, Jean Pierre Clerc<sup>a\*</sup>

<sup>a</sup> École polytechnique universitaire de Marseille, Technopôle Château Gombert, 5, rue Enrico Fermi, 13453 Marseille, France

<sup>b</sup> USTO, département de physique, LEPM, BP 1505 El Mnaouer, Oran, Algérie

Reçu le 12 mars 2001 ; accepté après révision le 6 mai 2002

Note présentée par Pierre-Gilles de Gennes.

## Résumé

Nous étudions numériquement dans ce travail les propriétés statistiques et dynamiques des amas dans un modèle unidimensionnel de petit monde. Les paramètres choisis dans cette modélisation correspondent à un réseau réaliste d'enfants d'âge scolaire où une épidémie comme la rougeole peut se propager. Plusieurs résultats sur le comportement statistique des amas autour du seuil de la percolation ainsi que leur évolution avec le temps sont discutés. *Pour citer cet article* : N. Zekri, J.P. Clerc, C. R. Physique 3 (2002) 741–747.

© 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

épidémie / petit monde / percolation / exposants critiques

## Statistical and dynamical study of the epidemics propagation in a *small world* network

## Abstract

We study numerically in this work the statistical and dynamical properties of the clusters in a one dimensional small world model. The parameters chosen correspond to a realistic network of children of school age where a disease like measles can propagate. Extensive results on the statistical behavior of the clusters around the percolation threshold, as well as the evolution with time, are discussed. *To cite this article* : N. Zekri, J.P. Clerc, C. R. Physique 3 (2002) 741–747.

© 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

epidemics / small world / percolation / critical exponents

## Abridged English version

Social networks require both clustering (excluding the random graphs [1]) and the feature of connecting any two sites within a finite number of steps (called Small World Network (SWN) [2,3]). Recently Watts and Strogatz [4,5] proposed a SWN model consisting of a low dimensional network (with a periodic boundary conditions for convenience) of occupied sites with a probability  $p$  (the remaining sites being empty) and where, in addition to the nearest neighbor connexions (NN) each site is connected randomly to  $\phi$  other sites of the network (see Fig. 1 for an example of one dimensional SWN).

\* Correspondance et tirés à part.

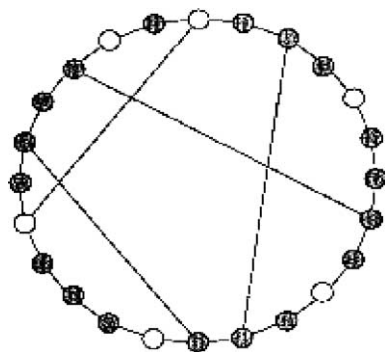
Adresses e-mail : zekri@mail.univ-usto.dz; zekri@polytech.univ-mrs.fr (N. Zekri); Jean-Pierre.Clerc@polytech.univ-mrs.fr (J.P. Clerc).

This model was shown to be useful for studying diseases transmission [2–5] where the epidemic outbreak corresponds to the percolation threshold. This threshold was studied both analytically and numerically for fixed numbers of NN and  $\phi$  (see Eq. (1)). However, in a real social network these numbers are randomly distributed and can affect sensitively the percolation threshold. This is the first aim of this paper where we apply this network to model a school age disease and examine the effect of a normal distribution of NN and CS on the percolation threshold. We found that the width of these distributions decreases the threshold (see Fig. 2). On the other hand, we found that this threshold is a phase transition and that the exponent  $x$  of the power-law behavior of the largest cluster with the  $p - p_c$ , seems to be universal within the statistical errors (as shown in Fig. 3).

The second aim was to study the dynamical behavior of the number of infected cases within the largest cluster. This number was found to increase up to a maximum before decreasing. The increase is exponential in the epidemic phase (with a rate corresponding to the characteristic time of the epidemics) while it is constant or power-law increasing in the endemic phase (the power-law increase corresponds to the percolation threshold). This result allows us to make a criterium for deciding whether it is or is not an epidemic phase.

## 1. Introduction

Pour modéliser la propagation d’une épidémie, il est nécessaire de définir un *réseau social* reliant deux individus. Les propriétés attendues de ce réseau doivent d’une part permettre la création d’amas (ce qui exclu les graphes aléatoires [1]), et d’autre part permettre la connexion de deux individus aléatoirement choisis après un nombre fini d’étapes (ce qui exclu les réseaux réguliers avec seulement les plus proches voisins). En effet, l’expérience de Milgram [2] montre que la connexion entre deux individus pour une transmission par courrier nécessite en moyenne six étapes (*six degrés de séparation*) et conduit à la définition du *Réseau du Petit Monde (RPM)* [3]. Watts et Strogatz ont proposé récemment une alternative pour le RPM [4,5]. Ils considèrent un réseau de basse dimension (unidimensionnel par exemple) de sites occupés de concentration  $p$  répartis aléatoirement, les reste des sites étant vide. En utilisant des conditions aux limites périodiques (par convenance), ce réseau devient un cercle avec un nombre de proches voisins  $k = 2$  (nous pouvons choisir un plus grand nombre ;  $k = 4$ , etc.). En outre, ils ont rajouté un désordre à ce modèle en incluant aléatoirement pour chaque site  $\phi$  liens vers une nouvelle position choisie aléatoirement de l’ensemble du réseau (voir Fig. 1 pour un exemple unidimensionnel). Pour de faibles valeurs de  $\phi$  ceci correspond encore à un réseau régulier mais avec quelques connexions à longues distances (appelées *Court-Circuits (CS)*) à travers le réseau. Notons que la dimension du système augmente lorsqu’on introduit des CS [6]. Une revue de ces modèles peut être trouvée dans les références [3,7]. Tous les travaux réalisés sur ce modèle utilisent un nombre moyen de proches voisins et de CS fixe. Or, en réalité ces nombres fluctuent



**Figure 1.** Réseau unidimensionnel de *petit monde* avec 24 sites,  $k = 2$  et  $\phi = 4/24$  (4 CS). Les sites gris représentent les individus exposés.

**Figure 1.** One dimensional small world network with 24 sites,  $k = 2$  and  $\phi = 4/24$  (4 CS). The gray sites represent susceptible individuals.

autour de leur moyennes ce qui pourrait ainsi affecter le seuil de percolation et la distribution des tailles d'amas.

D'autre part, des études de la dynamique de percolation ont été récemment faites par Pastor-Satorras et al. [8] et Moore et al. [9] respectivement pour des modèles de type *Exposés–Contaminés–Exposés (ECE)* et *Exposé–Contaminés–Guéris (ECG)* [10] sur des réseaux de type *RPM*. Mais quoique ces études touchent l'aspect le plus appliqué de ces modèles (en pratique seule l'évolution du nombre de nouveaux cas peut être mesurée), elles n'ont examiné que qualitativement la forme de l'évolution dans le temps de la densité de sites infectés dans les deux cas : épidémique et endémique.

Dans cet article, nous abordons aussi bien les statistiques des amas pour examiner les transitions de phase, que l'effet des fluctuations du nombre des proches voisins et des *CS* sur le seuil de percolation dans un *RPM* unidimensionnel [5] (comme indiqué ci-haut, ce réseaux a une dimension supérieure à l'unité et, par conséquent, la percolation existe pour une concentration finie). Nous appliquerons ensuite ces réseaux à un échantillon réaliste d'enfants d'âge scolaire pour étudier l'évolution dans le temps du nombre de nouveaux sites infectés du plus grand amas pour différentes concentrations de sites exposés (occupés), autour du seuil de percolation. Nous examinons en particulier dans ce dernier cas la forme d'évolution du nombre de cas pour caractériser le début d'une épidémie, afin de proposer un critère de détection d'une épidémie.

## 2. Description du modèle

Le modèle de percolation ont été appliqués à plusieurs domaines [11]. Nous nous intéressons ici à son application à la propagation d'épidémies en utilisant le *RPM* de Newman et Watts [5] avec pour chaque site un nombre moyen de proches voisins  $k$  (fixé à 2) et un nombre moyen de court circuits  $\phi$  engendrés aléatoirement (suivant une distribution uniforme) parmi les autres sites du réseau. Dans le cas où ces nombres ne sont pas fixés, ils sont déterminés aléatoirement suivant une Gaussienne centrée sur  $\bar{k}$  et  $\bar{\phi}$  respectivement avec un écart type de  $\delta k$  et  $\delta\phi$  respectivement. Cette distribution est choisie uniquement pour introduire l'effet des fluctuations et repose sur l'hypothèse d'un nombre total de connexions indépendant pour chaque site. D'autre part, puisque  $k$  et  $\phi$  sont des entiers positifs ou nuls, et à cause de la limitation de mémoire de calcul, nous n'utilisons que les nombres générés dans les intervalles

$$0 \leq k \leq k_{\max} = \frac{M\bar{k}}{k + \bar{\phi}}, \quad (1)$$

$$0 \leq \phi \leq \phi_{\max} = \frac{M\bar{\phi}}{k + \bar{\phi}} \quad (2)$$

avec  $M = 60$ . Les sites occupés sont engendrés aléatoirement avec une concentration  $p$ . Pour  $k$  et  $\phi$  fixés, Newman et al. [12] ont trouvé analytiquement que le seuil de percolation  $p_c$  est donné par

$$\phi = \frac{(1 - p_c)^{k/2}}{kp_c}. \quad (3)$$

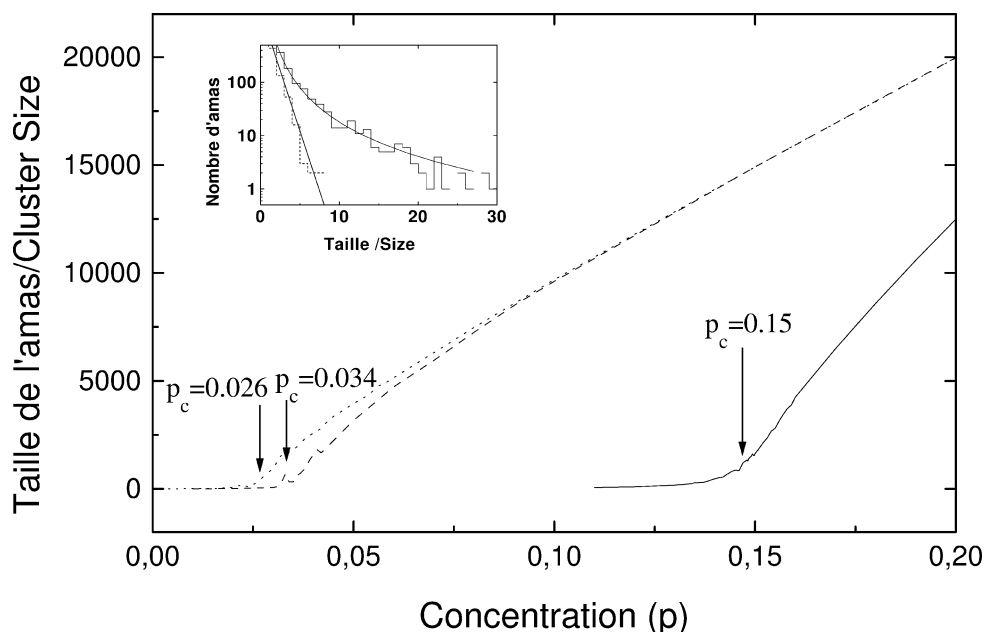
Du point de vue épidémiologique, puisque le contact entre personnes est le facteur de transmission de maladies, et que nous nous intéressons à des maladies d'enfants (comme la rougeole ou toute autre maladie à déclaration obligatoire permettant d'obtenir une statistique précise de cas connus), les proches voisins représentent les frères, soeurs et voisins alors que les *CS* représentent toutes les rencontres avec les autres enfants à l'école. Le problème étudié dans ce cas est une percolation de site [13] puisque tout site  $j$  exposé à la maladie (occupé) la contracte automatiquement s'il est directement connecté à un site malade  $i$  (occupé aussi). Le seuil de percolation  $p_c$  est la concentration minimale de sites exposés au-delà de laquelle la taille moyenne  $\xi$  du plus grand amas de sites occupés se comporte en puissance ( $\xi = (p - p_c)^x$ ), alors qu'elle diverge dans un réseau régulier [13] (en fait la percolation dans les *RPM* n'a pas le même sens que dans les réseaux réguliers). Ce seuil correspond à une transition d'une situation endémique en dessous de  $p_c$  vers une autre épidémique au-delà de cette concentration (le nombre moyen d'amas présente aussi un

maximum à ce seuil). L'exposant  $x$  est universel pour des réseaux réguliers [13]. Ce type de transition ainsi que l'universalité seront examinées sur le présent réseau.

### 3. Seuil de percolation et distribution d'amas

Dans cette section, nous réalisons 100 configurations du réseau décrit précédemment avec une taille fixée à 100 000 sites. Le nombre moyen de proches voisins est fixé à  $k = 2$  (dans le cas d'une distribution de proches voisins nous choisissons un écart type de 2). Nous varions le nombre moyen de court circuits  $\phi$  (et éventuellement l'écart type) et déterminons le seuil de percolation  $p_c$  ainsi que l'exposant  $x$ . Cette statistique est suffisante pour déterminer  $p_c$  jusqu'au troisième chiffre significatif. Nous étudions aussi la distribution des tailles d'amas autour du seuil de la percolation afin de caractériser le début d'une épidémie en fonction des propriétés statistiques des amas (ceux-ci sont définis par l'ensemble de sites occupé connectés entre eux).

Dans la Fig. 2 nous montrons la variation de la taille moyenne du plus grand amas en fonction de la concentration des sites occupés pour trois cas différents,  $\phi = 6, 30$  et pour des fluctuations  $\delta\phi = 15$  autour d'un nombre moyen  $\bar{\phi} = 30$ . Le dernier cas représente un échantillon réaliste d'enfants dont le nombre de frères, soeurs et voisins varie entre 0 et 4 avec une moyenne de 2 alors que le nombre d'enfants rencontrés à l'école est en moyenne de 30 (environ un classe d'école) avec une fluctuation de 15 suivant l'activité des enfants et leurs contacts à l'école. Nous voyons clairement de cette figure que dans tous les cas de  $k$  et  $\phi$ , les tailles des amas deviennent non nulles au-delà du seuil  $p_c$  et varient en puissance avec  $p$ . La valeur de  $p_c$  est en bon accord avec la prédiction analytique de Newman et al. [12] (Eq. (3)) pour le cas où  $k$  et  $\phi$  sont fixés. Dans le cas de fluctuations de  $k$  et  $\phi$  le seuil de percolation décroît sensiblement en augmentant



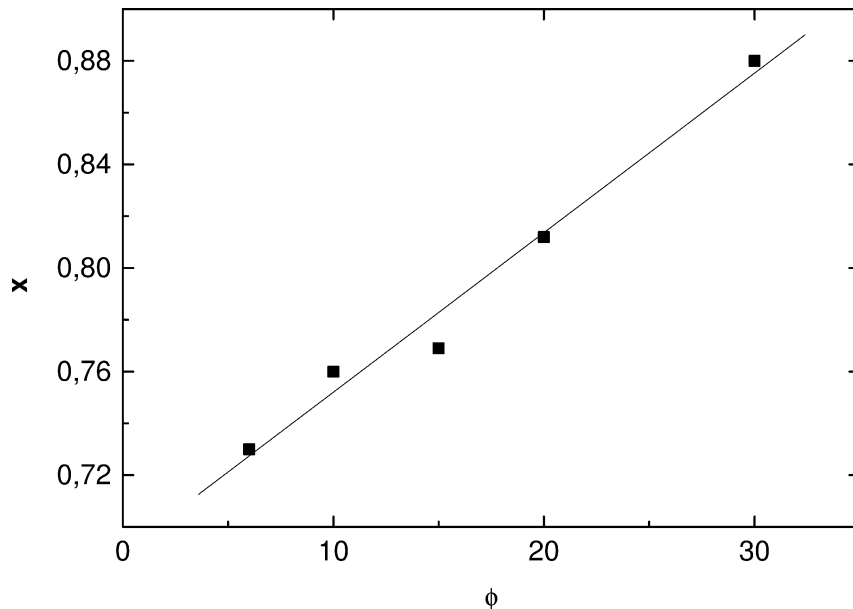
**Figure 2.** Variation de la taille moyenne du plus grand amas avec  $p$  pour 3 valeurs de  $\phi$  : 6 (pleine) ; 30 (tirés) et 30 avec  $\delta\phi = 15$  (pointillés). Le graphe inséré représente la distribution (semi-logarithmique) des tailles d'amas dans le premier cas pour  $p = 0,01$  (pleine) et 0,15 (tirés).

**Figure 2.** Variation of the average size of the largest cluster with  $p$  for 3 values of  $\phi$  : 6 (solid curve); 30 (dashed) and 30 with  $\delta\phi = 15$  (dots). The insert represents the cluster size distribution (semi-log plot) for the first case with  $p = 0,01$  (solid) and 0,15 (dashed).

l'écart type, comme si le nombre de coordination moyen  $\bar{k} + \bar{\phi}$  avait augmenté. Le nombre moyen de coordination n'est donc pas suffisant pour caractériser le début d'une épidémie. Il est aussi important de noter que les fluctuations des tailles d'amas présentent un maximum au seuil de la percolation (non montrés ici). Ce maximum est fini pour un système fini mais pourrait diverger pour un échantillon infini. En effet, la distribution des tailles d'amas (voir le graphe inséré de la Fig. 2) décroît exponentiellement en dessous de  $p_c$  alors qu'elle suit une loi en puissance au seuil de percolation (en accord avec les résultats de Castellano et al. [14] sur d'autres systèmes) caractérisant les distributions généralisées de [15]. L'exposant de cette distribution est légèrement supérieur à 2 impliquant la divergence des fluctuations de tailles d'amas. Ainsi ce seuil correspond bien à une transition de phase de second ordre mais la grandeur qui diverge est la fluctuation des tailles d'amas alors qu'elle est la taille du plus grand amas pour un réseau régulier. La taille du plus grand amas est le paramètre d'ordre dans ce cas.

Nous avons aussi examiné le comportement de l'exposant  $x$  en variant le nombre moyen de CS (fixé ici). Pour les réseaux réguliers ( $\phi = 0$ ), l'exposant critique  $x$  est universel [13] (indépendant de  $k$  et  $\phi$ ). Mais cette universalité n'est pas vérifiée dans certains cas particuliers de réseaux réguliers [16]. Le réseau actuel avec des court circuits pourrait donc constituer un bon test d'universalité de cet exposant. On remarque dans la Fig. 3 que  $x$  augmente linéairement mais très légèrement avec  $\phi$  (la pente est 0,0062). Cette faible variation est compatible avec les erreurs statistiques dans la détermination de  $x$ . Il est donc raisonnable de penser que  $x$  ne dépend pas de  $\phi$  à la limite des valeurs considérées dans le présent calcul.

Cependant, même si le modèle utilisé dans ce travail s'approche plus d'un réseau réaliste, il paraît impossible en épidémiologie de vérifier ces résultats dans la réalité. En effet, il n'y a d'une part aucun moyen de déterminer les amas dans un échantillon de population et d'autre part, il est impossible de connaître la concentration des enfants séronégatifs (exposés à une maladie) sauf au prix d'un examen sérologique appliqué à tous les enfants (d'une ville par exemple). Donc les résultats sur le comportement de  $p_c$  ne



**Figure 3.** Dépendance en  $\phi$  de l'exposant  $x$  au seuil de percolation. La ligne pleine représente la fonction  $6,2 \times 10^{-3}\phi$ .

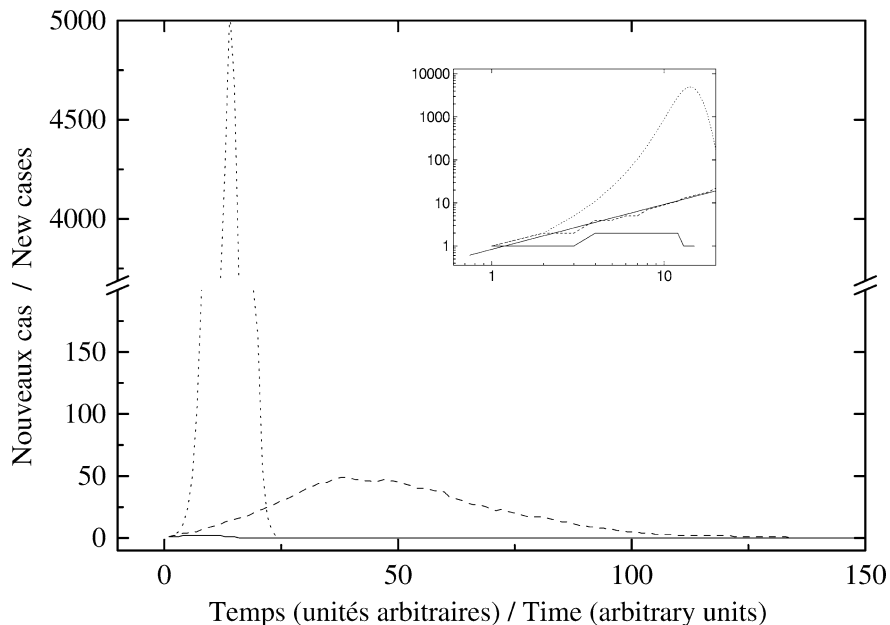
**Figure 3.**  $\phi$  dependence of the exponent  $x$  at the percolation threshold. The solid line represents the function  $6.2 \times 10^{-3}\phi$ .

semblent pas directement applicables en épidémiologie. Par contre le comportement temporel du nombre de cas ou en d’autres termes, la vitesse de propagation de l’épidémie est plus directement accessible aux techniques épidémiologiques. Une investigation de la dynamique de la propagation est donc indispensable.

#### 4. Étude dynamique de la propagation d’une maladie

Supposons que l’épidémie commence par se propager à partir d’un seul site infecté. Alors, dans un temps  $\Delta t$  il transmettra la maladie à tous les sites occupés (exposés) qui lui sont connectés, après  $2\Delta t$  ces sites la transmettrons aux seconds connectés et ainsi de suite. En supposant que le temps de transmission de l’épidémie  $\Delta t$  est constant, nous pouvons ainsi déterminer l’évolution du nombre de cas contaminés en fonction du temps pour différentes phases ; endémique ( $p < p_c$ ) et épidémique ( $p \geq p_c$ ). Nous choisirons comme unité de temps  $\Delta t$  et nous nous restreignons à  $\phi = 6$  sans tenir compte des fluctuations ( $p_c = 0,15$ ).

Dans la Fig. 4 nous montrons l’évolution du nombre de nouveaux cas en fonction du temps pour trois différentes situations :  $p = 0,1$  (situation endémique avant  $p_c$ ),  $p = p_c = 0,15$  et  $p = 0,2$  (situation épidémique au-delà de  $p_c$ ). Nous remarquons que dans le cas endémique le nombre de cas est quasi stationnaire. Au seuil de percolation ce nombre commence par augmenter en puissance avec le temps (voir graphe inséré dans cette figure). Dans ce cas, l’exposant est autour de l’unité et semble ne pas dépendre de  $\phi$  aux erreurs statistiques près (nous l’avons vérifié pour deux valeurs différentes de  $\phi$  : 6 et 30). Ce comportement signifie une diffusion normale du virus au seuil de percolation. Au-delà de  $p_c$  le nombre de cas augmente exponentiellement. Le nombre de contaminés diminue par la suite (ce qui n’est pas le cas dans les résultats de Pastor-Satorras [8] où il sature, car dans leur cas un site ayant été contaminé peut l’être plusieurs fois). L’exposant de cette variation dépend du nombre de proches voisins et des court circuits. Ce résultat pourrait permettre aux services de prévention de détecter une épidémie en évaluant systématiquement le nombre de cas en fonction du temps et de les ajuster à une exponentielle.



**Figure 4.** Évolution temporelle du nombre de nouveaux infectés dans le cas  $\phi = 6$  pour 3 situations :  $p = 0,1$  (pleine),  $p = 0,15$  (tirés) et  $p = 0,2$  (pointillés). Le même graphe (en log–log) est inséré.

**Figure 4.** Temporal evolution of the number of new infected in case  $\phi = 6$  for 3 situations:  $p = 0.1$  (solid),  $p = 0.15$  (dashed) and  $p = 0.2$  (dotted.) The insert is a log–log frame of the same plot.

En effet, les données (qui proviennent du plus grand amas dans le cas d'une épidémie) sont en général relevées quotidiennement et enregistrées chaque semaine. Donc si le temps caractéristique (l'inverse du taux d'augmentation exponentielle) est supérieur à la période d'enregistrement des données, il est possible de détecter l'épidémie et d'intervenir rapidement (campagnes de vaccination etc.).

## 5. Conclusion

Dans cet article nous avons étudié les statistiques des tailles d'amas dans un *RPM* unidimensionnel en tenant compte des fluctuations du nombre de proches voisins et de connexions à longue distances (court circuits). Nous avons trouvé que ces fluctuations affectent sensiblement le seuil de percolation et que les court circuits ne semblent pas détruire l'universalité des exposants critiques. D'autre part, et afin de rendre ce modèle plus directement applicable aux mesures épidémiologiques, nous avons étudié l'évolution en fonction du temps du nombre de cas contaminés (sites occupés). Nous avons trouvé que le début d'une épidémie est caractérisé par une augmentation au moins en puissance (sinon exponentielle) du nombre de cas avec le temps. Nous remarquons finalement que le réseau étudié ici est figé, c'est-à-dire que les contacts sont engendrés à un temps précis. Ceci est une bonne approximation puisque la période d'une épidémie est négligeable devant le temps d'évolution du système (naissances, immigrations etc.). Une étude plus étendue décrivant un système évolutif où le nombre de contact change avec le temps tiendrait compte mieux de ces paramètres. Ceci fera l'objet d'une investigation future.

**Remerciements.** Un des auteurs (NZ) remercie le Fond Arabe pour le Développement Economique et Social pour son support financier et l'encadrement de l'École Polytechnique Universitaire de Marseille pour leur hospitalité pendant l'évolution de ce travail. Les auteurs remercient aussi Dr. Marc Barthélémy pour les discussions fructueuses.

## Références bibliographiques

- [1] B. Bollobas, Random Graphs, Academic Press, New York, 1985.
- [2] S. Milgram, Psychology Today 2 (1967) 60.
- [3] D.J. Watts, Small Worlds, Princeton University Press, Princeton, 1999.
- [4] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440.
- [5] M.E.J. Newman, D.J. Watts, Phys. Lett. A 263 (1999) 341.
- [6] C.-P. Zhu, S.-J. Xiong, Phys. Rev. B 63 (2001) 193405.
- [7] M.E. Newman, Model of the small world, cond-mat/0001118.
- [8] R. Pastor-Satorras, A. Vespignani, cond-mat/0102028.
- [9] Moore, Newman, Phys. Rev. E 61 (2000) 5678.
- [10] N.T.J. Bailey, The Mathematical Theory of Infectious Diseases, Griffin, Londres, 1975.
- [11] P.G. de Gennes, La percolation : un concept unificateur, La Recherche 331 (2000).
- [12] M.E. Newman, D.J. Watts, Phys. Rev. E 60 (1999) 7332.
- [13] D. Stauffer, A. Aharony, Introduction to Percolation Theory, Taylor & Francis, Londres, 1992.
- [14] C. Castellano, M. Marsili, A. Vespignani, Phys. Rev. Lett. 85 (2000) 3536.
- [15] M. Shlesinger, G.M. Zaslavsky, U. Frish, Lévy Flights and Related Topics in Physics, Springer, Berlin, 1995.
- [16] J.P. Straley, J. Phys. C 15 (1982) 2343.