Statistical and dynamical study of disease propagation in a small world network

Nouredine Zekri[1,2,*] and Jean Pierre Clerc[1,†]

[1]*IUSTI, Technopôle Château Gombert, Université de Provence, 13453 Marseille, France*
[2]*USTO, Département de Physique, LEPM, Boîte Postale 1505 El M'Naouar, Oran, Algeria*

Statistical properties and dynamical disease propagation have been studied numerically using a percolation model in a one dimensional small world network. The parameters chosen correspond to a realistic network of school age children. It has been found that percolation threshold decreases as a power law as the shortcut fluctuations increase. It has also been found that the number of infected sites grows exponentially with time and its rate depends logarithmically on the density of susceptibles. This behavior provides an interesting way to estimate the serology for a given population from the measurement of the disease growing rate during an epidemic phase. The case in which the infection probability of nearest neighbors is different from that of short cuts has also been examined. A double diffusion behavior with a slower diffusion between the characteristic times has been found.

I. INTRODUCTION

To model disease propagation, it is necessary to define the corresponding social network connecting any two individuals in the world. The expected properties of such a network should be both the clustering (which excludes models of disorder such as the random graphs [1]), and to allow a connection between any two individuals within a finite number of steps (which excludes the regular networks with only nearest neighbor connections). Indeed, for the latter feature Milgram showed in 1967 that the average number of steps connecting any two individuals is six (also called six degrees of separation) [2]. This behavior recently led Watts and Strogatz to propose the model of small world network (SWN) [3,4]. They considered a low dimensional network with periodic boundary conditions for convenience (a ring, for example) where they rewired some bonds with a probability ϕ to a new site randomly chosen from the network. For small values of ϕ, this still corresponds to a regular network but with few long range connections called shortcuts (SC). A more recent work on the SWN was proposed by Newman and Watts [5], where the number k of nearest neighbors (NN) is conserved but instead of rewiring, they added an average density ϕ of new bonds from each site i to other randomly chosen nodes (except its nearest neighbors). A review of these models and their application to various fields and particularly epidemics can be found in Refs. [3,6]. In these networks the percolation threshold was extensively investigated and its dependence to the NN and SC was found to satisfy the following equation [5,7]:

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

This threshold corresponds in epidemics to the smallest concentration of susceptibles leading to the outbreak [5]. However, the statistical behavior of percolative SWN networks is different from that of regular systems [8]. In particular, at the percolation threshold, there is no diverging cluster for the SWN because a SC between the two ends of the system has a finite probability to occur for any nonvanishing number of these bonds. On the other hand, the characteristic length scale in such networks (which corresponds to the correlation length in regular lattices) behaves as $\phi^{-1/d}$ at the percolation threshold, d being the euclidean dimension of the system [7]. It is then obvious that this characteristic length does not diverge for such networks. Therefore, a further investigation of the cluster statistics and the phase transition around this percolation threshold for such networks seems to be necessary.

Let us now consider the application of this model to epidemics, which seems to be one of its main aims. From the large amount of works using SWN, there is no direct comparison with the existing data, may be due to the complexity of the diseases features (infection and latent periods, birth and death rates, etc.). Furthermore, the parameters used (mainly ϕ) are very small and do not simulate the real connections between individuals. They use also commonly average values of the NN and the SC while these quantities strongly fluctuate in the real life (the number of contacts, friends, family members, etc., varies from 0 to few tens), which can influence sensitively the results on the density of susceptibles at the percolation threshold (epidemic outbreak). On the other hand, it is impossible in practice to measure the density of susceptibles systematically (it needs an extensive serological investigation in the epidemic phase). Generally, for large population samples epidemiologists measure the evolution with time of the number of cases for a given disease. It is then necessary to study the dynamical behavior of the propagation of the disease and relate it to the density of susceptibles. There are only few works that examined (only qualitatively) the dynamical behavior of the disease on social networks [9–11]. The aim of these works was to show how the density of infected behaves in the endemic and epidemic phases.

---

*Email address: zekri@mail.univ-usto.dz

†Email address: clair@iusti.univ-mrs.fr

In this paper, we use a site percolation on a SWN with parameters (k and ϕ) representing a sample of school age children to study the effect of the fluctuations of NN and SC on the percolation threshold. Furthermore, in order to propose a formula for determining the serology of the sample from the rate of increase of the number of cases, we investigate also extensively the dynamical behavior of an infectious disease as well as its effect on the density of susceptibles below and above the percolation threshold. A ''superdiffusion'' is found above the percolation threshold when the cluster is initially infected by one or a small number of infectious sites and its characteristic time dependence on the density of susceptibles is determined. We examined also the case where the infection probability of the NN is different from that of SC, showing a double diffusion with two characteristic times. In the next section we describe the model and then present the results on the cluster statistics and the percolation threshold. The results on the dynamical behavior of the disease are presented in Sec. IV.

II. MODEL DESCRIPTION

We consider the one dimensional SWN described by Newman and Watts [5] but ϕ represents the total number of SC generated for each site uniformly from all the other sites of the network. In the case where k and ϕ are not fixed they are generated randomly within a normal distribution centered at their average values with fluctuations δk and $\delta \phi$, respectively. The coordination number is the total number of bonds to a given site ($z = k + \phi$). We study in this network a site percolation problem [8] by assuming each susceptible site j (occupied) contracts the disease if it is connected with an ill site i (occupied also). The occupied sites (susceptibles) are randomly generated with a concentration p while the empty sites correspond to refracted individuals. For k and ϕ fixed, the percolation threshold p_c is related to k and ϕ by Eq. (1) [7]. This threshold corresponds to a transition from the endemic phase below p_c to the epidemic one above this point [10]. In SWN networks, p_c is the minimum concentration of occupied sites above which the average largest cluster size ξ of the occupied sites becomes power-law increasing with the concentration [$\xi = (p - p_c)^x$], while it diverges in a regular network [8] (note here that the exponent x is positive). By analogy with the regular lattices [8], we will check the universality of the exponent x.

We are interested in the application of such a model to a childhood disease like measles. In such diseases, epidemiological investigations on school age children can be easily controlled and provide data with a minimum bias. We choose parameter values corresponding to such a disease by taking $k = 2$ to be the average number of brothers, sisters, and neighbors, while $\phi = 30$ represents the average number of children one can meet at the school. These parameters should correspond to a topology closer to that encountered in a real social network. Regarding the dynamical study, we assume the major contribution to the epidemics provided by the largest cluster. We restrict ourselves then to this cluster and start the infection with one or few infectious sites at time 0. These sites will infect all the connected sites in the next
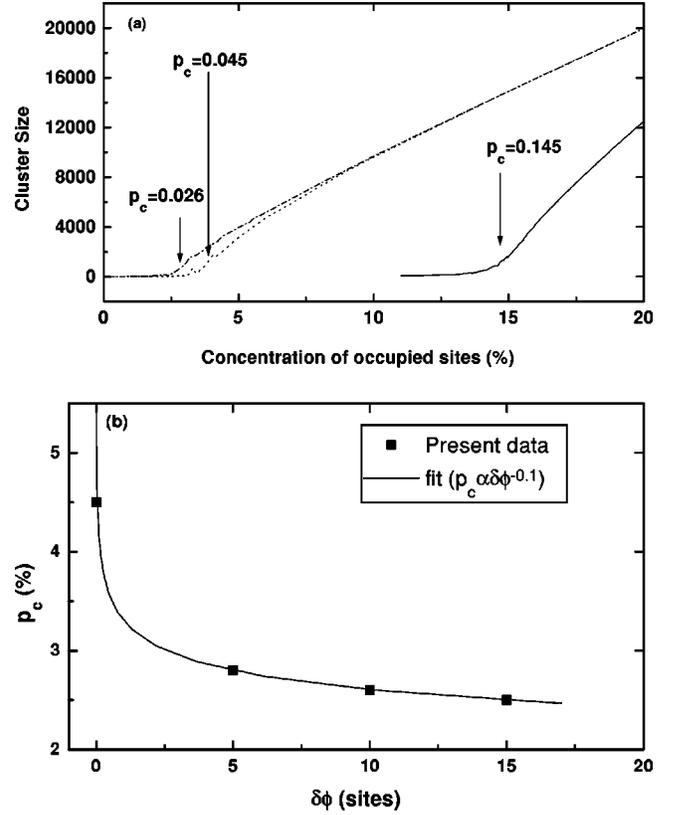

FIG. 1. (a) Cluster size (number of sites in the cluster) versus concentration of the occupied sites for three cases: $\phi = 6$ (solid curve), $\phi = 30$ (dotted curve), and $\phi = 30$ with fluctuations $\delta k = 2$, $\delta \phi = 15$ (dash-dotted curve). (b) p_c versus $\delta \phi$ (sites) for $\phi = 30$ sites. The solid line is a fit of the data.

step (after a time Δt), which themselves infect their connected sites after $2\Delta t$ and so on. We assume the latent and infection periods to be smaller than Δt, which is taken in the rest of this paper as a unit time. The number of infected sites in each step is averaged by varying the initial infectious site position through the whole cluster.

III. PERCOLATION THRESHOLD AND CLUSTER DISTRIBUTION

In this section, we realize 100 configurations of the network described in the previous section with a size fixed at 100 000 sites. We examine the effects of ϕ and its fluctuations on the average cluster sizes, p_c and x. Finally, we investigate the cluster size distribution around p_c in order to determine the main contribution to the propagation of the disease.

In Fig. 1(a) we show the variation of the cluster size with the concentration of occupied sites for three different cases: $\phi = 6,30$ (fixed values) and for k and ϕ randomly generated with a normal distribution centered at 2 and 30, respectively, with a fluctuation of 2 and 15, respectively. We see clearly from this figure that in all cases the cluster sizes vary as a power law of $(p - p_c)$ above p_c. For fixed k and ϕ the value of p_c is in good agreement with the analytical predictions of Newman and Watts [7] [Eq. (1)]. However, in the case of

fluctuations of k and ϕ, this threshold decreases sensitively [about 50% in the case shown in Fig. 1(a)] as the fluctuations increase. Therefore, the average values of k and ϕ are not sufficient to characterize an epidemic outbreak. The SC fluctuations $\delta\phi$ decrease p_c as a power law with an exponent 0.1 [as shown in Fig. 1(b)], indicating a sensitive participation of the larger values of ϕ to built the largest cluster. Therefore, the percolation threshold behaves as

$$p_c \simeq \phi^{-1} \delta\phi^{-0.1}. \tag{2}$$

From this behavior, we can estimate the percolation threshold in a real sample of school age children to be in the range 2.3% to 2.8%.

Now let us restrict ourselves to the case of fixed $k=2$ and $\phi=6$ in order to examine the statistical behavior of the clusters around p_c (without loss of generality, these values are chosen only because p_c is large enough to enable sufficient cluster statistics for such a sample size). We found that the cluster size fluctuations are maximum at this threshold [see Fig. 2(a)] implying a divergence of this quantity at p_c, which seems to be the signature of a phase transition. The cluster size distribution [see Fig. 2(b)] confirms this divergence since it decreases exponentially below p_c while it is power-law decreasing at this threshold (this power-law behavior is in agreement with the results of Castellano *et al.* [12] on other systems). Indeed, at p_c this corresponds to a Lévy distribution [13] with an exponent of 2.13 indicating the divergence of all its moments. We notice here that only the higher sizes (rare events) contribute to the outbreak at p_c (as expected in such distributions). Above p_c the small size clusters are *absorbed* by the largest one and we have again an exponentially decreasing distribution for small clusters while there is only one very large cluster [not shown in Fig. 2(b)].

Since the cluster size does not diverge at p_c, it is obvious that x is not universal (because it is not a critical exponent), but it is interesting to know how it depends on ϕ in such lattices. In Fig. 2(c), the exponent x seems to vary only linearly for larger values of ϕ but with a very small slope (about 5.6×10^{-3}). It is difficult to predict its behavior for very small values of ϕ because in this case the network tends to a regular one and the cluster size becomes very large so that the sample sizes considered here do not allow us to measure this exponent accurately.

However, even if the parameters chosen in this model are close to those of a real social network, it seems impossible for epidemiologists to check these results. Indeed, as explained below, they cannot measure the density of susceptibles, except if they investigate systematically the serology of a sufficiently large sample of school age children (e.g., for a city sample). Therefore, the behavior of p_c should be checked for measurable quantities. In the case of disease propagation, the time dependence of the number of cases can be directly measured by epidemiological techniques. We will investigate this dynamical behavior in the following section.
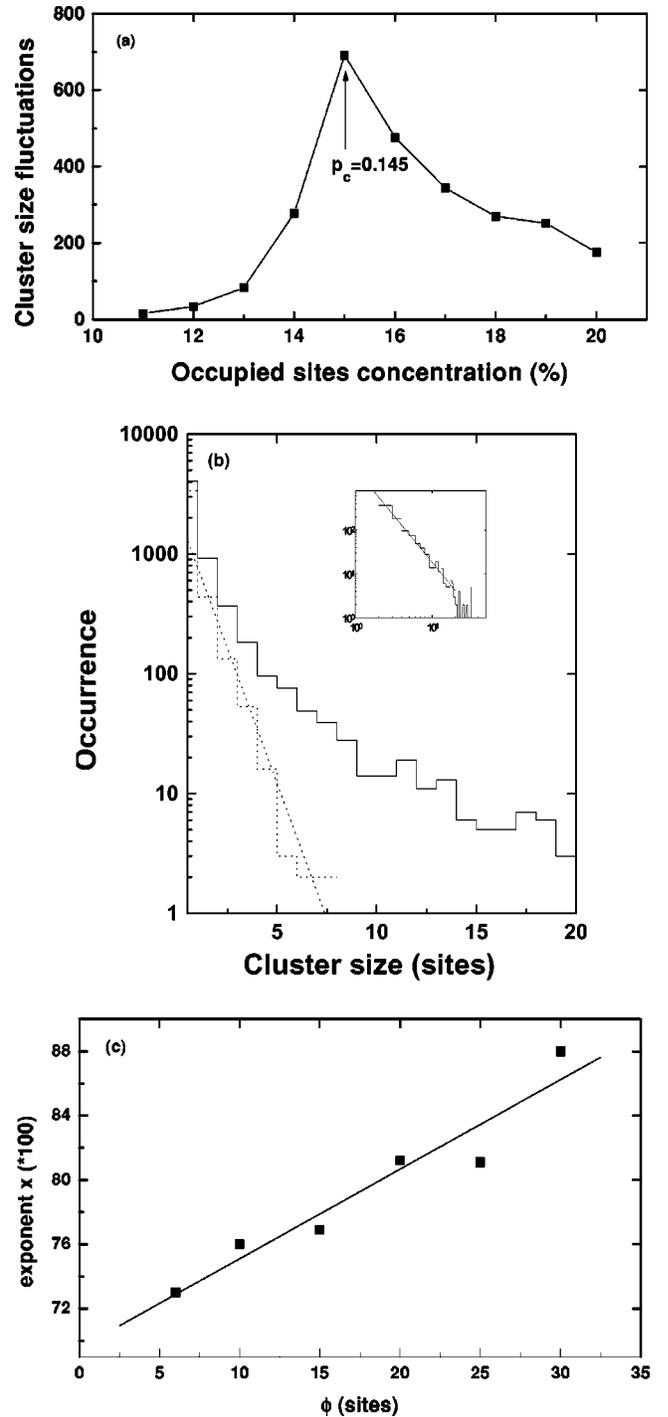

FIG. 2. (a) Cluster size fluctuations (sites) versus the occupied sites concentration ($\phi=6$ sites). The solid curve is a guide for the eyes. (b) Distribution of the cluster size (sites) for fixed ($k=2$, $\phi=6$) in a semilog plot at p_c (solid curve) and $p=1\%$ (dotted curve). The dotted line is a linear fit of the data below p_c. The inset is a log-log plot of the distribution at p_c with a linear fit. (c) Variation of the exponent x with the number of short cuts ϕ (sites). The solid line is a linear fit of the data to 5.6×10^{-3}.

IV. DYNAMICAL STUDY OF THE PROPAGATION
OF A DISEASE

In this section we restrict ourselves to the fixed values of k and ϕ (2 and 30, respectively) to simulate a sample of school age children. From the results of Fig. 2(b), we assume that the main growing effect of the infection comes from the largest cluster and estimate the propagation time of the epidemics only from this cluster. We determine the evolution with time of the number of cases for both phases endemic ($p<p_c$) and epidemic ($p \geqslant p_c$). As shown in Fig. 2(a) the cluster size at p_c strongly fluctuates and, therefore, the time behavior of the number of cases also fluctuates. The variation of the number of cases with time is shown in Fig. 3(a) for three cases (just below p_c, at p_c, and above p_c) with only one initial infectious site. In both cases, the number of cases increases up to a maximum and then decrease because the number of susceptibles decreases. In the endemic phase, the number of connections between occupied sites in the cluster is mostly one and does not allow a significant increase of the number of cases (the behavior in this case is underestimated since all the clusters should contribute to this increase). For susceptible densities around p_c this situation persists for a long time and the number of cases increases linearly with time showing a normal diffusion of the disease. In the epidemic phase the increase becomes exponential indicating a ''superdiffusion'' [13,14] of the disease, due mainly to the increasing number of connections in the cluster [as shown in Fig. 3(b)]. This exponential growth is also observed for SIR models [15], where the rate is proportional to the basic reproduction rate R_0, which correponds in our case to the average number of connections in the cluster. We have also performed a Monte-Carlo simulation to the measles propagation in a more realistic sample (births, deaths, infection, and latent periods, etc.), where the average infections is two for each infectious individual and found also an exponential growth of the infected cases [16]. Therefore, this exponential growth does not seem to depend on the topology of the sample but the rate is sensitive to the geometry of the network. Note in the present work that in the case of more than one initial infectious site [see Fig. 3(c)] the exponential growth behavior does not change but the growing rate fluctuates due to the fluctuating number of connections. The average rate of the exponential growth γ (corresponding to the characteristic time of the epidemics) is shown in Fig. 4 to increase as $Ln(p)$ above p_c, while the period of this epidemic behavior decreases. From this figure we can conclude that when the characteristic time decreases below five (or γ increases above 0.2), the epidemic behavior takes over. This behavior seems to have a direct application in epidemiology since it provides a method for the estimation of the serological situation (density of susceptibles) from the characteristic time that is easily measurable. Therefore, this result stimulates a proposal for a serological examination for a given childhood disease in a sample of school age children, but during an epidemic period to compare a realistic behavior with that obtained in this paper.

Now let us examine the case of adding different infection probabilities to this system. We consider that a site i infects
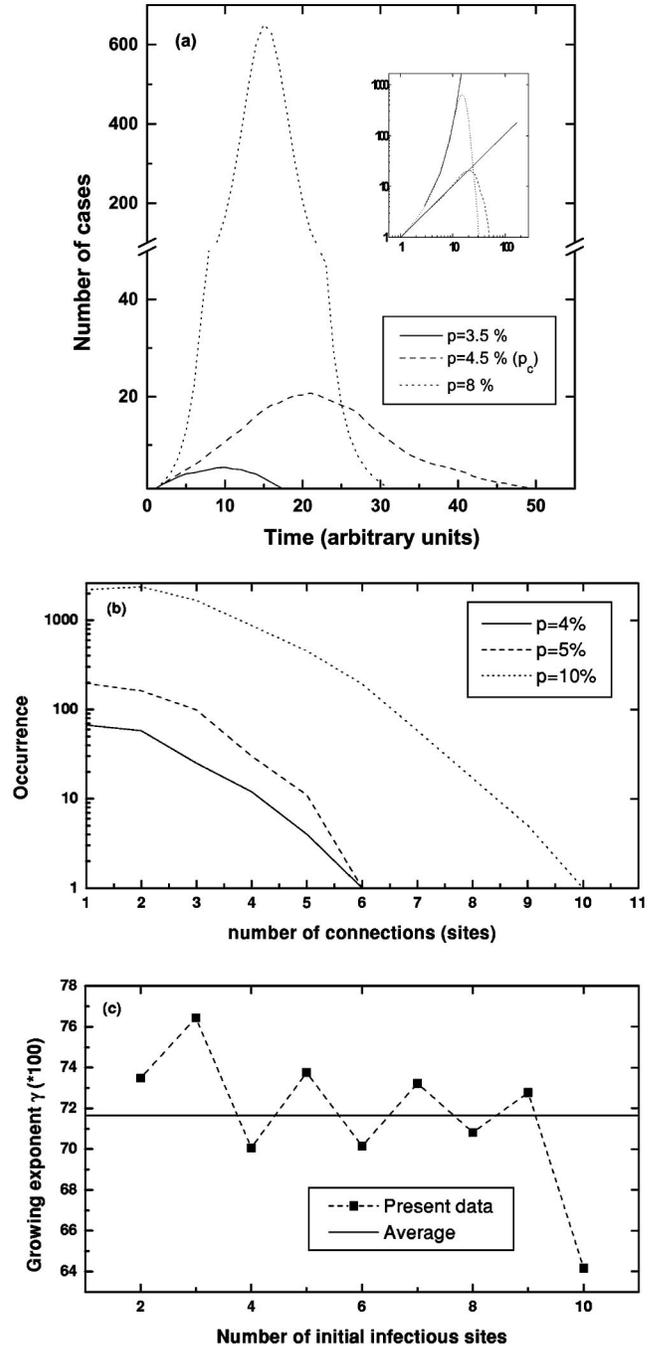

FIG. 3. (a) Number of cases versus time for three different cases: $p=3.5\%$ (solid curve), $p=4.5\%$ (dashed curve), and $p=8\%$ (dotted curve). Inset: log-log plot with a power-law fit of $p=4.5\%$ and an exponential fit of $p=8\%$. (b) Distribution of the number of connections (acquaintances) in the largest cluster for $p=3.5\%$ (solid curve), $p=5\%$ (dashed curve), and $p=10\%$ (dotted curve). (c) The rate of the exponential growth (in arbitrary units) versus number of initial infectious sites. The horizontal line is the average rate.

another site j with a probability p_n if j is a neighbor of i and p_{sc} if it is a shortcut. The motivation of this investigation is that a susceptible child has a different probability to be infected by his brothers (or sisters) than by the other children
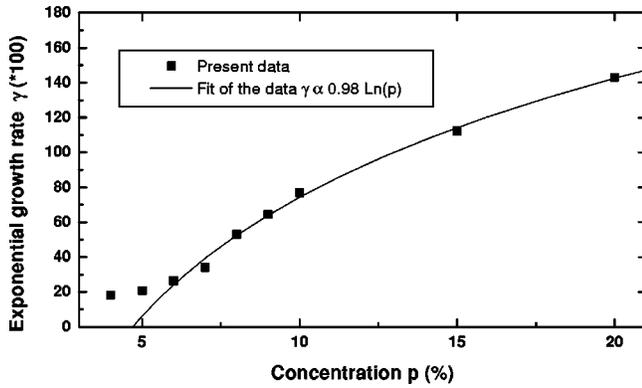
FIG. 4. The rate of the exponential growth versus p. The solid line is a fit of the curve linearly with $Ln(p)$.
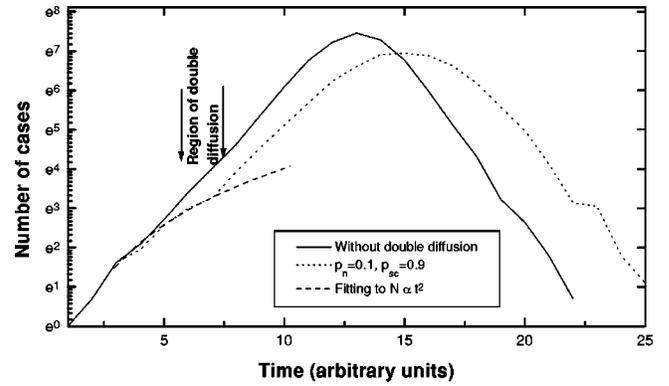

FIG. 5. The number of infected cases versus time (in arbitrary units) for one initial infectious site and an infection probability one (solid curve), and the probabilities of infection: $p_n = 0.1$ and $p_{sc} = 0.9$ (dotted curve). The dashed curve is a power-law fit of the second data in the region of the double diffusion.

meeting him at the school. We see clearly a double diffusion behavior in Fig. 5 (for $p_n = 0.1$ and $p_{sc} = 0.9$), where the number of infected starts growing exponentially up to the characteristic time $(1/\gamma)$, then it increases as a power law up to a new characteristic time from which it grows again exponentially with the same rate. The slow diffusion is due to the small contact probability for the neighbors ($p_n = 0.1$) and has been observed in other fields [17]. This slow diffusion appears very short because the number of NN is very small ($k = 2$). It should be interesting to investigate this double diffusion for larger k (which is the case of animal diseases).

V. CONCLUSION

We have investigated in this paper, the statistics of the cluster sizes in a one dimensional SWN by taking into account the NN and SC fluctuations. We found that these fluctuations decrease p_c as a power law with a small exponent leading to an expression for the percolation threshold. We found also that cluster size fluctuations is the quantity governing the phase transition in such a network. On the other hand, in order to apply our results to the measured quantities in epidemiology, we have studied the dynamics of the disease propagation in such clusters. We found in epidemic

phases a ''superdiffusion'' with an exponentially growing number of infected sites, while at p_c this number increases as a power law. The growing characteristic time is larger than five in the endemic situation and decreases linearly with $Ln(p)$ in the epidemic phase. This result provides a way to estimate the density of susceptibles in the epidemic phase. We propose then a serological investigation in epidemic situations to check this behavior. Finally, we examined the case where the infection probability is very small in the *NN* compared to the SC. The dynamical behavior of infected cases shows a double diffusion with two characteristic times, and a power-law increase (deceleration) between them. We think that this effect is useful for samples with large NN and shows a way to stop the propagation of the epidemic for other diseases.

ACKNOWLEDGMENTS

[1] B. Bollobas, *Random Graphs* (Academic Press, New York, 1985).

[2] S. Milgram, Psychology Today **2**, 60 (1967).

[3] D.J. Watts, *Small Worlds* (Princeton University Press, Princeton, 1999).

[4] D.J. Watts and S.H. Strogatz, Nature (London) **393**, 440 (1998).

[5] M.E.J. Newman and D.J. Watts, Phys. Lett. A **263**, 341 (1999).

[6] M.E. Newman, J. Stat. Phys. **101**, 819 (2000).

[7] M.E. Newman and D.J. Watts, Phys. Rev. E **60**, 7332 (1999).

[8] D. Staufer and A. Aharony, *Percolation Theory*, 2nd ed. (Taylor and Francis, London, 1994); D.J. Bergman and D. Stroud, Solid State Phys. **46**, 147 (1992).

[9] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).

[10] C. Moore and M.E.J. Newman, Phys. Rev. E **61**, 5678 (2000); **62**, 7059 (2000).

[11] M. Kuperman and G. Abramson, Phys. Rev. Lett. **86**, 2909 (2001).

[12] C. Castellano, M. Marsili, and A. Vespignani, Phys. Rev. Lett. **85**, 3536 (2000).

[13] P. Lévy, *Théory de l'Addition des Variables Aléatoire*

(Gautier-Villars, Paris, 1937); M. Shlesinger, G.M. Zaslavsky, and U. Frish, *Lévy Flights and Related Topics in Physics* (Springer, Berlin, 1995); J.P. Bouchaud and A. Georges, Phys. Rep. **195**, 12 (1990).

[14] S.N. Evangelou and D.E. Katsanos, Phys. Lett. A **164**, 456 (1992).

[15] R.M. Anderson and R.M. May, *Infectious Diseases of Humans. Dynamics and Control* (Oxford University Press, Oxford, 1991).

[16] N. Zekri and J.P. Clerc (unpublished).

[17] A.M. Dykhne (private communication).