Distribution of infected mass in disease spreading in scale-free networks

Lazaros K. Gallos, Panos Argyrakis*

Department of Physics, University of Thessaloniki, Thessaloniki 54124, Greece

Abstract

We use scale-free networks to study properties of the infected mass M of the network during a spreading process as a function of the infection probability q and the structural scaling exponent γ. We use the standard SIR model and investigate in detail the distribution of M. We find that for dense networks this function is bimodal, while for sparse networks it is a smoothly decreasing function, with the distinction between the two being a function of q. We thus recover the full crossover transition from one case to the other. This has a result that on the same network, a disease may die out immediately or persist for a considerable time, depending on the initial point where it was originated. Thus, we show that the disease evolution is significantly influenced by the structure of the underlying population.
© 2003 Elsevier B.V. All rights reserved.

PACS: 89.75.−Hc; 87.23.Ge

Keywords: Scale-free networks; Spreading; SIR model; Infected mass

There has been a growing interest recently in the network structure [1–8] and dynamics [9,10] of real-life organized systems. Many such systems, covering an extremely wide range of applications, have been recently shown [1–4,8] to exhibit scale-free character in their connectivity distribution, meaning that they obey a simple power law. Thus, the distribution of the connectivity of nodes, follows a law of the form

$$P(k) \sim k^{-\gamma}, \tag{1}$$

where k is the number of connections that a node has and γ is a network parameter which determines the degree of its connectivity. These networks have some unusual

---

* Corresponding author.
  E-mail address: panos@physics.auth.gr (P. Argyrakis).

properties, thus justifying the heavy interest in recent years. For example, spreading processes in scale-free networks show a dynamic behavior which is different than in other classes of networks. These processes are plausible models for the spreading of diseases, epidemics, etc. Several models for spreading exist in the literature built on different algorithms, such as the SIR model [11], the SIS model [9], the SIRS model [12], etc.

In a recent paper, Newman [11] studied analytically the SIR model in scale-free networks. The connectivity distribution had an exponential cutoff of the form

$$P(k) \sim k^{-\gamma}e^{-k/\kappa}, \tag{2}$$

where κ is an arbitrary cutoff value for k. This work gave a closed-form solution for the epidemic size and the average outbreak size as a function of the infection probability. It showed that there is a critical infection threshold (q_c) only for small finite values of κ, but as κ increases q_c decreases, apparently resulting in $q_c \to 0$ (no critical threshold at all) in the limit of $\kappa \to \infty$. The same result was also shown for the SIS model by Pastor-Satorras and Vespignani [9].

The absence of a critical threshold is not a universal network property. Actually, in the well-studied cases of lattice networks and small-world networks, the opposite is true [13]. Such a threshold [14,15] is always present, which separates the infected from the uninfected regions. This has as a result that these models do not offer a very realistic picture. Recently, Warren et al. [13] have used a heterogeneous distribution for the infection probabilities both for lattices and small-world networks. They model the variability in a population which results in a broadening of the transition regime; however, a threshold still exists and the behavior of the transition is qualitatively similar to the case of the simpler SIR model on a lattice. Because of this, scale-free networks are distinctly different regarding the predictions on the rate and efficiency of spreading. This is clearly much closer to what is intuitively expected, and can provide useful estimates for the properties of epidemics of any kind. The same type of model could also describe a diverse set of networks, such as social networks, virus spreading on the Web, rumor spreading, signal transmission etc.

In the present study, we calculate the distribution of the epidemic size, i.e., the distribution of the infected mass, for several different γ values. This property helps us to better understand the importance of the starting point (origin) of the disease. It turns out that this distribution is not a smooth function for all networks, but depends strongly on the network density, i.e., the value of γ.

We use a simulation algorithm to construct a scale-free network comprised of $N=10^5$ nodes. We follow a network generation method which enables us to freely vary the connectivity distribution of the network. We assign a number of edges k for each node by using a power-law distribution $P(k) \sim k^{-\gamma}$. We find the highest connected node and we start by connecting this node to k other randomly chosen distinct nodes. We then continue by randomly choosing open links from uncompleted nodes and connect them to another random node. This process is similar to, e.g. the one used in Ref. [16], with the difference that for $\gamma = 2$, we first fully connect the highest-connected node, since otherwise it is difficult to create such a network. Care is taken that no duplicate links are established between the same two nodes and once a node has reached the number
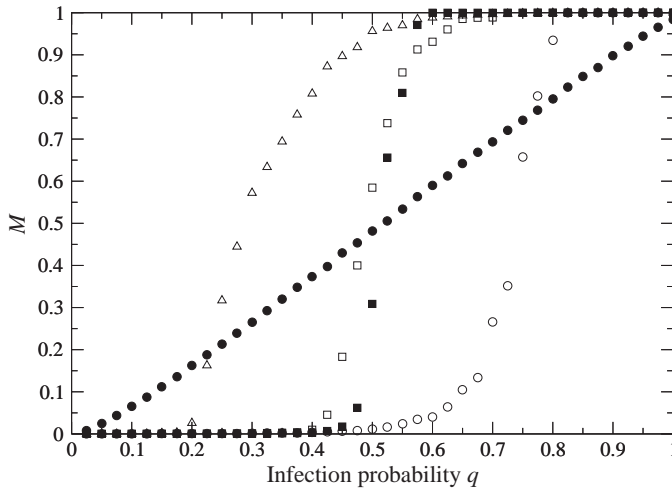
Fig. 1. Percentage M of infected sites as a function of q, for the different models studied (●: scale-free network, $\gamma = 2.0$, filled square: two-dimensional lattice, ○: small-world $p = 0.0$, □: small-world $p = 0.01$, △: small-world $p=1.0$, where p is the probability for rewiring a link in a small-world network). The absence of a threshold is evident for scale-free networks.

of edges initially assigned to it, it no longer accepts any new connections. The cutoff value for the maximum possible connectivity of a node was fixed to $N/2$.

The spreading of a disease follows the standard SIR (susceptible, infective, recovered) model. Initially, all nodes are in the susceptible (S) status, and a random node is infected (I). During the first time step it tries to infect with probability q, the nodes linked to it, and when the attempt is successful, the status of the linked node switches from S to I. The process is repeated with all infected nodes trying to influence their susceptible (S) neighbors during each time step. After trying to infect its neighbors, the status of an infected site changes to recovered (R) and can no longer be infected. The simulation stops when there are no infected nodes in the system or when all nodes have been infected. Small-world networks were constructed as described in Ref. [17].

We monitor the percentage M of nodes infected and the duration of a disease (i.e., the time needed for the disease to either disappear or cover the entire network). In Fig. 1, we show for comparison purposes the percentage M of infected sites as a function of q, for three different network types. The form of the curves for the lattice and the small world networks is similar, i.e., in all cases there is a sharp transition and a critical point. However, this behavior is unrealistic, as it does not follow the majority of actual situations, such as disease spreading in real-life networks [13]. If this were to happen, e.g. the Web would be in a state of either no virus present or the entire Web (all computers in the world) would be infected, with a very small probability of having an intermediate situation with only a certain fraction of computers infected, which is the realistic picture. Scale-free networks with a high degree of connectivity ($\gamma = 2.0$) follow a much smoother spreading evolution, as the mass of the infected
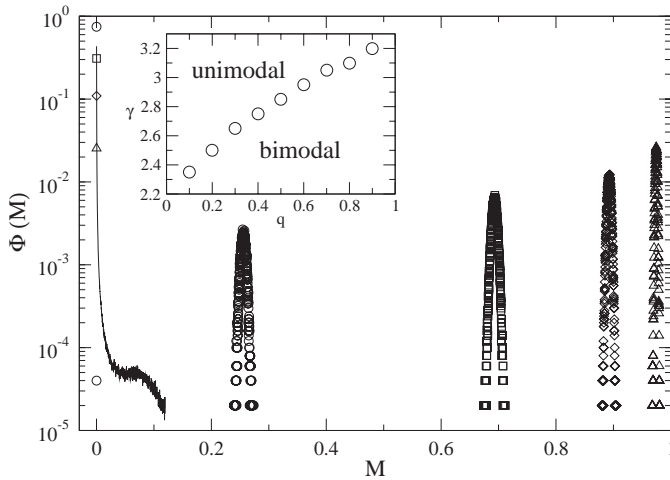
Fig. 2. Infected mass distribution $\Phi(M)$ of scale-free networks for $\gamma = 2.0$, $N = 10^5$ and different infection probabilities: $q = 0.1$ (○), 0.5 (□), 0.8 (◇), 0.95 (△). To each gaussian part corresponds a single point at a value higher than $M = 1/N$, except for the case of $q = 0.1$ which has two points (one at $1/N$ and one at $2/N$). Solid line: $\gamma = 2.9$ and $q = 0.5$. Inset: Approximate values of γ where the distribution turns from bimodal to unimodal, as a function of q.

population increases almost linearly with the infection probability q and there is no transition regime. This is in agreement with the recent formalism of Newman [11]. The linear behavior can be understood as follows: On a scale-free network, there exist nodes with a wide range of connectivity. For fixed infection probability q, the average probability for an infected node with k links to spread the disease is kq. If this number is greater than 1, it is statistically certain that a neighbor node will be infected. If it is significantly less than 1, the disease will die out. On the other hand, for a small-world network and for a lattice network, there is a characteristic mean number of links $\langle k \rangle$ assigned to each node, which fixes the $\langle k \rangle q$ factor for all nodes.

Here, for a scale-free network, we look at the distribution of the quantity M. This is shown in Fig. 2, where we see that for different γ values we have two opposite situations. First, for $\gamma = 2.0$ we see that for fixed q, we can have both cases coexisting, depending on the connectivity of the initially infected node. For this case, the distribution $\Phi(M)$ of the infected population for fixed q is bimodal and comprises of two distinct parts, a strong peak at $M = 1/N$ (only one node is infected) attributed to initially infected nodes with $k=1$, or more accurately $kq \ll 1$, and a Gaussian-shaped part at some higher value of M. The almost linear increase of M with q in Fig. 1 is due to the fact that for higher q, the peak of $1/N$ decreases (kq increases), and the average value of the Gaussian distribution increases accordingly. We also notice in Fig. 2 that the gaussian part of the distribution has an average value which is larger than q. For example, for $q = 0.1$ the peak of the distribution is close to 0.25, and when $q = 0.8$ it is closer to 0.9. This means that if the peak in $1/N$ did not exist, the behavior of the curve in Fig. 1 would be superlinear, i.e., M would always be larger

than q as a result of the complex connectivity of the network, but it would not reach the value of $M = 1$ for infection probabilities less than 1. The interplay between the peak at $1/N$ and the gaussian part yields the final curve which follows roughly a linear increase.

This result is a superposition of the two possible states present in the case of a simple lattice, where below the threshold, the distribution is a simple peak at $M = 1/N$, while above the threshold, the distribution peaks around $M = 1$. On a scale-free network, there exists a finite probability for the disease to be either eliminated immediately or cover a considerable portion of the network. Thus, for small γ values, the prediction on the future of a disease largely depends on the place where it originates, because of the existence of the gap.

This bimodality implies that every disease which survives the initial step(s) spreads over a non-zero portion of the network population. This is not the case for large γ values, where recently [18], results for the SIR model on a scale-free network (with $\gamma = 3.0$) were presented. The distribution of M did not show any bimodality, similarly to our solid curve on the left of Fig. 2, which corresponds to $\gamma = 2.9$ and $q = 0.5$. We see that the bimodality is now lost, and it is replaced by a smoothly decreasing function with increasing M. This result is in agreement with the work of Moreno et al. [18]. Thus, the distribution changes from bimodal to unimodal as we go from a dense network to a sparse network. The crossover from one to the other takes place as it is given in the inset in Fig. 2, where we plot γ vs. q, i.e., each (γ, q) pair is exactly at the corresponding transition point. The distinction between the two regimes is not very clear, since the transition is not sharp, as can also be seen by the existence of a secondary 'hump' in the continuous line of Fig. 2. This is not due to fluctuations, and for this reason we choose to locate the crossover at the point where the peak at low M values is separated by the second peak through a region where M is exactly equal to 0. Note also that the unimodal–bimodal transition does not directly correspond to the threshold values of each network.

These observations show that in scale-free networks, nodes of high connectivity act as "boosters" to the disease spreading; even if very few nodes remain infected, by the time a high connectivity node is infected, it spreads the disease over a significant number of its neighbors, even for small q. This fact stresses the importance of the 'hubs', as it has also been observed in the past in studies of the static properties for such networks [6,19,16]. Similarly, the low-connectivity nodes may serve to isolate large clusters of the network. These clusters are effectively screened by the disease via the presence of the low-connectivity node, especially in the case of high γ, where the network is loosely connected.

Similar conclusions can be drawn for the duration of a disease. For small-world networks and regular lattices, the duration of a disease when epidemics takes place is practically constant for infection probabilities greater than the threshold (Fig. 3). For scale-free networks, on the contrary, there is a slow increase of the duration as a function of q. However, the duration is much smaller now, which implies that even for considerable infection probabilities, a disease cannot last for a long time, and a considerable portion of the network can be infected in a practically small and constant time. The duration of the disease at $q = 1$ is also a measure of the network 'diameter',
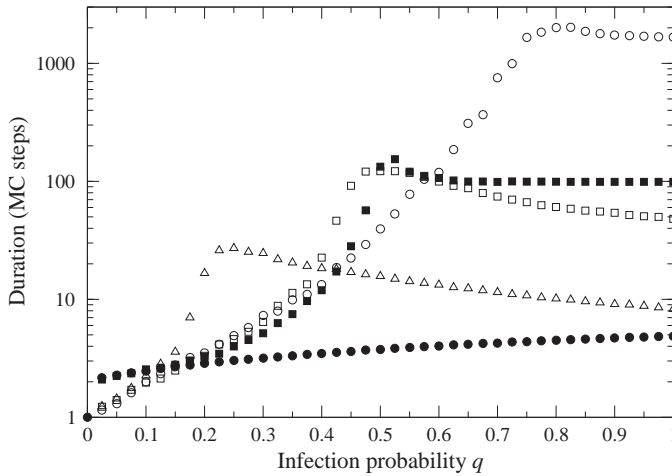
Fig. 3. Disease duration on different topologies. Symbols are the same as in Fig. 1.

since it represents the average number of links needed to cross, before reaching all the system nodes.

Upon monitoring the distribution of uninfected sites as a function of time, we observed that it followed a power law at all times. The exponent of this power law was always the same as the one used for the initial connectivity distribution, with the curve scaled down by a constant factor. Thus, sites of different connectivity are infected with the same relative rate.

Summarizing, we have investigated spreading properties on scale-free networks. For the SIR model we studied, we find that the starting point of the disease is very important, because it can either stop the disease or facilitate its spreading. This phenomenon for dense networks yields a bimodal distribution for the infected mass, with a peak close to zero and a Gaussian part around a finite value of M. Despite the smooth increase of the infected mass with q, the disease spreads rapidly on the network in a practically constant time, almost independently of q. This rapid spreading manifests the compactness of the network (as compared to lattice and small-world networks) and its small diameter which is related to the short path length from any site of the network to another. For sparse networks this is not the case, as it also has been previously observed.

References

[1] S.H. Strogatz, Nature 410 (2001) 268.
[2] R. Albert, A.-L. Barabasi, Rev. Mod. Phys. 74 (2002) 47.
[3] S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. 51 (2002) 1079.
[4] A.-L. Barabasi, R. Albert, Science 286 (1999) 509.
[5] R. Albert, A.-L. Barabasi, Phys. Rev. Lett. 85 (1999) 5234.
[6] R. Albert, H. Jeong, A.-L. Barabasi, Nature 406 (2000) 378.

[7] P.L. Krapivsky, S. Redner, F. Leyvraz, Phys. Rev. Lett. 85 (2000) 4629.
[8] F. Liljeros, et al., Nature 411 (2001) 907.
[9] R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. 86 (2001) 3200.
[10] R. Pastor-Satorras, A. Vazquez, A. Vespignani, Phys. Rev. Lett. 87 (2001) 258 701.
[11] M.E.J. Newman, Phys. Rev. E 66 (2002) 016 128.
[12] M. Kuperman, G. Abramson, Phys. Rev. Lett. 86 (2001) 2909.
[13] C.P. Warren, et al., Math. Biosci. 180 (2002) 293.
[14] C. Moore, M.E.J. Newman, Phys. Rev. E 61 (2000) 5678;
      C. Moore, M.E.J. Newman, Phys. Rev. E 62 (2001) 7059.
[15] D.H. Zanette, Phys. Rev. E 64 (2001) 050901(R).
[16] R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, Phys. Rev. Lett. 85 (2000) 4626.
[17] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440.
[18] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Eur. Phys. J. B 26 (2002) 521.
[19] D.S. Callaway, M.E.J. Newman, S.H. Strogatz, D.J. Watts, Phys. Rev. Lett. 85 (2000) 5468.