The Spreading of Epidemics in Complex Networks

Xiangyu Song

PHY 563 Term Paper, Department of Physics, UIUC

May 8, 2017

Abstract

The spreading of epidemics in complex networks has been extensively studied in the last few decades. Depending on the nature of the disease and the network it spreads on, there are different critical behaviors. In this essay I provide a brief introduction to the study of complex networks and the susceptible-infectious-susceptible(SIS) epidemic model. I discuss the different critical behaviors when SIS model is applied to exponential network and scale-free network. The existence of non-zero epidemic threshold in exponential networks and the lack of such threshold in scale-freee networks can help understanding computer virus epidemics.

Contents

1 Introduction

The study of networks has a long history and can be dated back to Leonard Euler's study of Königsberg Bridge Problem in 1736. Later on people developed a language called graph theory to describe the properties of networks. Starting in the 1950s, a growing interest in quantitative methods in sociology encourages people to borrow, or adapt graph theory for social network analysis. Until now, many social, biological, and communication systems can be properly described by complex networks whose nodes represent individuals or organizations and links mimic the interactions among them. Interesting behaviors of epidemics spreading has been studied in metabolic networks, food networks, and most importantly, the Internet and world-wide-webs, because they are of immense technical and economical value.

Network analysis is powerful because of its breadth. By abstracting away the details of a problem and mapping it onto a network, we can describe the important topological features with a clarity that would be impossible were all the details retained[1]. Since all the results for one network model is general, we can apply them to a special problem for useful information - which is why network analysis has spread beyond its original domain of pure mathematics, and become highly active in engineering, biology, sociology, and computer science.

However, a real world network usually contains too much details and could be hard to describe. Strogatz listed six complications in his work[2]: structural complexity, network evolution, connection diversity, node diversity, dynamical complexity, and, what's more, these complications can influence one another[1]. Naturally, an approach to such a complex network problem is to assume only one of them is important and the others can be ignored. This assumption must be reasonable, though. For example, in this essay I consider an epidemic model and assume all the networks are static, because the spreading of this disease is very quick comparing with the network's evolution over time.

In its simplest form, a network is a set of discrete elements (called vertices, points,

[1] 1. Structural complexity: the wiring diagram could be an intricate tangle. 2.Network evolution: the wiring diagram could change over time. 3. Connection diversity: the links between nodes could have different weights, directions and signs. 4. Dynamical complexity: the the nodes could be nonlinear dynamical systems. 5. Node diversity: there could be many different kinds of nodes. 6. Meta-complication: The various complications can influence each another. Cited from[2]

or nodes), and connections (called lines, edges, or links). The simplest characteristic of a vertex in a graph is its degree, that is the number of its nearest neighbors. In physics this is often called connectivity (denoted k). Knowing connectivity of a vertex is one way to tell how important it is in a network.

Vertices linked by edges form paths. The concept of a path is a sequence of edges which connect a sequence of vertices. A connected network is such a network that every vertex can be reached by any other vertex through a path. In a disconnected network this is impossible, because there are distinct components that forbids some vertices to reach an arbitrary vertex in the network. A loop (simple cycle) is a closed path visiting each its vertex only once. By definition, trees are graphs without loops.

A network can be described by adjacency matrix A. For a simple graph with vertex set V, the adjacency matrix is a square $V \times V$ matrix A such that its element A_{ij} is 1 when there is an edge from vertex i to vertex j, and 0 when there is no edge. The diagonal elements of the matrix are all zero, since edges from a vertex to itself (loops) are not allowed in simple graphs.

By adding features to simple graphs, we can evaluate more complicated systems. If we add weight to links, the elements of the adjacency matrix can take a range of values, instead of being binary. Also we may add directions to each link, and the adjacency matrix will be no more symmetric. Sometimes we need to model a system using bilayer or multilayer networks. In this case, the interactions are not only within a layer but also between different networks.

In the remainder of this essay, I will briefly introduce study of complex networks in section2, and talk about standard models of epidemics in section3. In section4, I show the different results on applying SIS epidemic model to homogeneous and inhomogeneous networks.

2 Two main groups of complex networks

Complex networks can be classified in two main groups, depending on their connectivity properties: the exponential network, which is homogeneous, and scale-free network, which is inhomogeneous.

2.1 Exponential network

The exponential connectivity network, or exponential network, is probably the most studied network. As its name suggests, the nodes' connectivity distribution (the probability P(k) of finding a node connected to other k nodes) is exponentially bounded. Random graph model is a typical example.

According to Erdös and Rényi[3], to form this network, we start with n vertices, so that there are $\binom{n}{2}$ possible pairs of vertices. To build a graph with N edges, we choose N edges from all these possible edges. Define p as the probability of finding an edge between distinct pairs of vertices. The probability is described by binomial distribution:

$$Pr(N) = \left(\begin{pmatrix} n \\ 2 \end{pmatrix} \right) p^N (1-p)^{\begin{pmatrix} n \\ 2 \end{pmatrix} - N} \tag{1}$$

We can define the mean value of N, and the mean connectivity of a vertex with p

$$\langle N \rangle = \begin{pmatrix} n \\ 2 \end{pmatrix} p,\ c \equiv \langle k \rangle = \frac{\langle N \rangle}{n/2} = (n-1)p \tag{2}$$

the probability P(k) of finding a node connected to other k nodes

$$P(k) = \begin{pmatrix} n-1 \\ k \end{pmatrix} p^k (1-p)^{n-1-k} \tag{3}$$

In large n limit, we obtain a Poisson distribution

$$P(k) = e^{-c} \frac{c^k}{k!} \tag{4}$$

2.2 Scale-free network

Scale-free networks exhibit a power-law connectivity distribution:

$$P(k) \sim k^{-\gamma} \tag{5}$$

here, the parameter γ must be larger than zero to ensure a finite average connectivity $\langle k \rangle$. For $2 < \gamma \leq 3$ this fact implies that each node has a statistically significant probability of having a very large number of connections compared to the

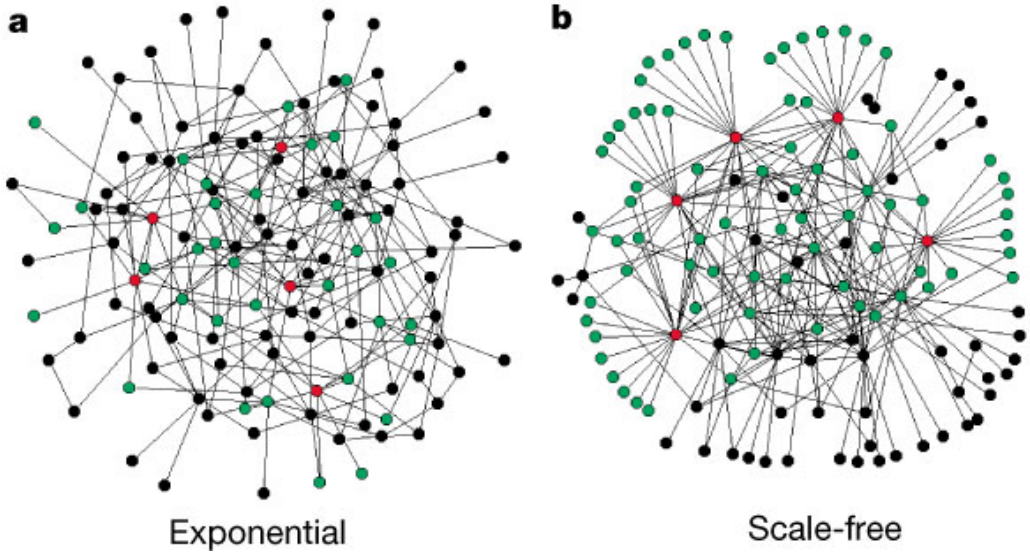

Figure 1: Exponential and Scale-free networks, cited from[5]. Both contain 130 nodes and 215 links ($\langle k \rangle = 3.3$). Red nodes are the ones with highest number of links; green nodes are their nearest neighbors. a, The exponential network is homogeneous: most nodes have approximately the same number of links. b, The scale-free network is inhomogeneous: a few nodes have large number of links, while the majority of the nodes have only one or two links.

average connectivity of the network[4].

Figure2.2 is an illustration of the two typical networks. In the exponential network only 27% of the nodes are reached by the five most connected nodes, in the scale-free network more than 60% are reached[5]. It is clear that the connected nodes are very important in scale-free network.

One special case of scale-free network is the Barabási-Albert (BA) model. In this model, $P(k) \sim k^{-3}$. Several natural and human-made systems, including the Internet and the world wide web, are approximately scale-free and contain few nodes with unusually high connectivity compared with other nodes in the network[6]. The γ factor for the Internet is about 2.5[7].

3 The susceptible-infectious-susceptible (SIS) model

Epidemic models are many, but in this essay we consider only one of them: The susceptible-infectious-susceptible (SIS) model. It is predominantly used for sexually transmitted diseases(STDs): people can be cured and then repeat infections. It is

different from the susceptible-infectious-recovered (SIR) model, which is appropriate for infectious diseases that confer lifelong immunity.

To describe this progress we neglected many details, but this model has been proved successful for a long time.

$$\begin{aligned} \frac{dS}{dt} &= gI - \phi S, \\ \frac{dI}{dt} &= \phi S - gI, \end{aligned} \tag{6}$$

In these equations, S, I refer to the number of susceptible and infectious individuals, respectively, in a population of size N. Here we assume only two discrete states exist. These states completely neglect the details of the infection mechanism within each individual. g is the rate of recovery, ϕ is the force of infection, which tells us the rate at which susceptible individuals become infected. This parameter is a function of the number of infectious individuals. It also contains information about the interactions between individuals that lead to the spreading of epidemics[8].

In reality population usually don't mix at random with infectious individuals, but if we assume it to be true, then each individual has an equal chance of coming into contact with one another, the force of infection is then given by:

$$\begin{gathered} \phi = \text{ effective number of contacts per unit time} \\ \times \text{transmission rate} \times \text{ proportion of contacts infectious} \end{gathered}$$

We will use this model and discuss its spreading in exponential and scale-free networks.

4 The spreading of epidemics

Romualdo Pastor-Satorras and colleague[7] defined an effective spreading rate to describe the epidemic transmission in complex network: At each time step, each susceptible node is infected with probability ν if it is connected to one or more infected nodes. At the same time, infected nodes are cured and become again susceptible with probability δ, defining an effective spreading rate.

$$\lambda = \frac{\nu}{\delta} \tag{7}$$

We can always rescale the time, so without loss of generality, we can set $\delta = 1$.

4.1 Epidemic spreading in exponential network

As we already seen in Figure2.2, exponential networks can be considered homogeneous. We further assume all individuals are equivalent, irrespective of their connectivity, each individual has the same number of linkes($k \simeq \langle k \rangle$), the connectivity fluctuations are very small($\langle k^2 \rangle \sim \langle k \rangle$), also the mixing is homogeneous. Thus, we can use mean field theory to calculate its behavior. Define total prevalence $\rho(t)$, which is the density of infected nodes present at time t.

$$\partial_t \rho(t) = -\rho(t) + \lambda \langle k \rangle \rho(t)[1 - \rho(t)] \tag{8}$$

Here we use this model for an endemic state with a dynamically stationary value for the density of infected individuals. That is to say $\rho(t) \ll 1$.

Assume infected individuals become healthy by a unit rate, the first term in the equation above gives the value. The second term represents the density of newly infected individuals, following the idea of infection force in section3.It is proportional to the infection spreading rate (λ), number of links from each individual ($\langle k \rangle$), and the probability that a given link is connected to a healthy individual ([1-$\rho(t)$]). The $\rho(t)$ in second term is because of the homogeneous mixing hypothesis, which is indeed equivalent to a mean-field treatment of the model[4].

There is no birth rate or natural death rate in this equation, because we assume the time scale of the disease is much smaller than the lifespan of individuals.

After imposing the stationarity condition $\partial_t \rho(t) = 0$, we obtain the equation for the behavior of the system at large times,

$$\rho[-1 + \lambda \langle k \rangle (1 - \rho)] = 0 \tag{9}$$

for the steady state density ρ of infected nodes. This equation defines an epidemic threshold $\lambda_c = \langle k \rangle^{-1}$, and

$$\rho = 0, \, if \, \lambda < \lambda_c \tag{10}$$

$$\rho = (\lambda - \lambda_c)/\lambda, \, if \, \lambda \geq \lambda_c \tag{11}$$

Thus, SIS model in homogeneous network predicts the existence of a nonzero epi-

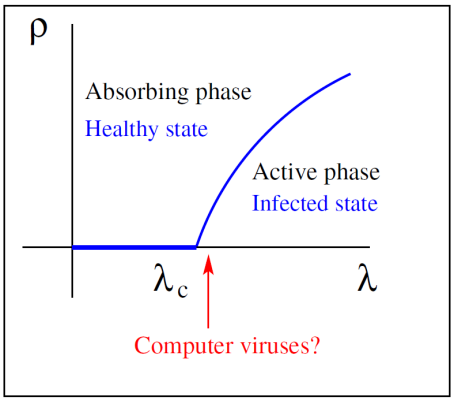

Figure 2: Schematic phase diagram for the SIS model in homogeneous networks, cited from[4]. It is clear the epidemic threshold is of a positive value. When the value of λ is larger than the threshold, the infection spreads. Otherwise the infection dies fast. The SIS model in this network is going through an absorbing-state phase transition at the threshold. It separates two phases: active, or infected phase, from absorbing, or healthy phase.

demic threshold λ_c, proportional to the inverse of the average number of neighbors of every node, $\langle k \rangle$, below which the epidemics always dies, and endemic states are impossible. A sketch of such behavior is in Figure4.1. The epidemic threshold separates infected phase with finite prevalence from healthy phase with null prevalence.

Is this a good prediction for computer virus? Probably not, because the computer virus data that has been observed has long lifetimes and very small prevalence. In Figure4.1, it corresponds to a value of λ that is infinitesimally close to the threshold. In fact, if we consider the propagation of computer virus, the network they spread on is a scale-free one. Virus prefer computers that are actively communicating to others because they exchange more data.

4.2 Epidemic spreading in scale-free network

When moving from exponential network to scale-free network, we can no long assume the connectivity fluctuation is negligible. Thus, for an infected node $\rho_k(t)$ at time t with connectivity k, we can write

$$\frac{d\rho_k(t)}{dt} = -\rho_k(t) + \lambda k[1 - \rho_k(t)]\Theta[\{\rho_k(t)\}] \tag{12}$$

here in the first term we also consider a unitary recovery rate without losing generality. The second term is almost the same as in homogeneous case, and considers

the probability that a node with k links is healthy ([1-$\rho_k(t)$]) and get infected by a neighbor. This probability is proportional to the infection rate (λ), the number of links for this node (k), and the probability $\Theta[\{\rho_k(t)\}]$ that any given link points to an infected node.

Assume Θ is a function of the partial densities of infected nodes $\{\rho_k(t)\}$. In the steady state, the ρ_k are functions of λ. Thus, the probability Θ is also an implicit function of λ, and by imposing the stationarity condition, we obtain

$$\rho_k = \frac{k\lambda\Theta(\lambda)}{1 + k\lambda\Theta(\lambda)} \tag{13}$$

which is reasonable, because this equation means as connectivity of a node gets higher, the probability to be in an infected state rise as well. As the network is no longer homogeneous, the computation of $\Theta(\lambda)$ for a general complex network can be difficult. For a random scale-free network, the probability that a link points to a node with s connections is equal to $sP(s)/\langle k\rangle$[7].

$$\Theta(\lambda) = \frac{1}{\langle k\rangle}\sum_k kP(k)\rho_k \tag{14}$$

Since ρ_k is on its turn a function of $\Theta(\lambda)$, we obtain a self-consistency equation that allows to find $\Theta(\lambda)$ and an explicit form. Finally, we can evaluate the order parameter (persistence) ρ using the relation. In the limit of small $\Theta(\lambda)$, for any scale-free connectivity distribution.

$$\rho = \sum_k P(k) \tag{15}$$

To find the epidemic threshold in this case, just noticing λ_c is the value of λ above which it is possible to obtain a nonzero solution for $\Theta(\lambda)$. In fact, we obtain

$$\Theta(\lambda) = \frac{1}{\langle k\rangle}\sum_k kP(k)\frac{k\lambda\Theta(\lambda)}{1 + k\lambda\Theta(\lambda)} \tag{16}$$

where $\Theta(\lambda)$ is a function of λ only. The trivial solution is always satisfying the equation. As for non-trivial result, note a non-zero stationary prevalence ($\rho_k \neq 0$) is obtained when the right-hand-side and the left-hand-side of the above equation, expressed as function of $\Theta(\lambda)$, cross in the interval $0 < \Theta(\lambda) \leq 1$. A non-trivial solution is thus allowed, which corresponds to

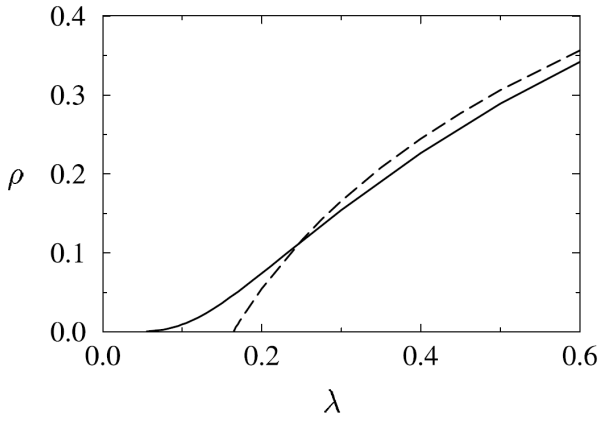

Figure 3: Total prevalence ρ for the SIS model in a Barabási-Albert network (full line) as a function of the spreading rate λ, compared with the theoretical prediction for a homogeneous network (dashed line), cited from[7]

$$\frac{d}{d\Theta}(\frac{1}{\langle k\rangle}\sum_k kP(k)\frac{k\lambda\Theta(\lambda)}{1+k\lambda\Theta(\lambda)})|_{\Theta=0} \geq 1 \tag{17}$$

The epidemic threshold λ_c when left hand side and right hand side are equal, and we have

$$\frac{\sum_k kP(k)\lambda_c k}{\langle k\rangle} = \frac{\langle k^2\rangle}{\langle k\rangle} = 1, \rightarrow \lambda_c = \frac{\langle k\rangle}{\langle k^2\rangle} \tag{18}$$

This results implies that in scale-free networks with connectivity exponent $2 < \gamma \leq 3$, for which $\langle k^2\rangle \rightarrow \infty$ in the limit of a network of infinite size, we have $\lambda_c = 0$.

An analytical result for Barabási-Albert network is showed in Figure4.2. This network is a toy model of scale free network, and has a connectivity distribution of $P(k) \sim k^{-3}$. The result shows the absence of epidemic threshold or critical point in the model. Compare with the non-zero epidemic threshold in exponential network, it is clear that scale-free networks are prone to the spreading of epidemics, regardless of spreading rate. Although Barabási-Albert network is of $\gamma = 3$, it has been proved that this result can be generalized for networks with $2 < \gamma \leq 3$, since the Internet has $\gamma = 2.5$, it shares the same critical behavior as BA network[7].

5 Conclusions

As for epidemic spreading on complex networks, the connectivity fluctuations of the network play a major role by strongly enhancing the infection's incidence. Exponential networks have homogeneous connectivity, and nodes in a network are almost equivalent. In scale-free networks, the fluctuation of connectivity could be very large, and the network has inhomogeneous connectivity. Technological networks, such as the Internet and the world-wide-web are scale-free networks.

Of course, the epidemiological picture in this essay is an simplified one. The evolving of networks is completely ignored, but in reality pages and links are created and destroyed in the Internet every second. Still, we can see that scale-free networks are in fact very weak in the face of infections. The lack of epidemic threshold in scale-free networks indicates that infections in such networks can increase proliferate, no matter what the spreading rate is. It is worth mentioning that although I only reviewed the SIS model on epidemics spreading, the SIR model, when applied to exponential and scale-free networks, gives the same result[9].

References

[1] Newman, Mark, Albert-Laszlo Barabasi, and Duncan J. Watts. The structure and dynamics of networks. Princeton University Press, 2011.

[2] Strogatz, Steven H. "Exploring complex networks." Nature 410.6825 (2001): 268-276.

[3] ERDdS, P., and A. R&WI. "On random graphs I." Publ. Math. Debrecen 6 (1959): 290-297.

[4] Bornholdt, Stefan, and Heinz Georg Schuster, eds. Handbook of graphs and networks: from the genome to the internet. John Wiley & Sons, 2006.

[5] Albert, Rka, Hawoong Jeong, and Albert-Lszl Barabsi. "Error and attack tolerance of complex networks." nature 406.6794 (2000): 378-382.

[6] Albert, Rka, and Albert-Lszl Barabsi. "Statistical mechanics of complex networks." Reviews of modern physics 74.1 (2002): 47.

[7] Pastor-Satorras, Romualdo, and Alessandro Vespignani. "Epidemic dynamics and endemic states in complex networks." Physical Review E 63.6 (2001): 066117.

[8] Keeling, Matt J., and Ken TD Eames. "Networks and epidemic models." Journal of the Royal Society Interface 2.4 (2005): 295-307.

[9] Moreno, Yamir, Romualdo Pastor-Satorras, and Alessandro Vespignani. "Epidemic outbreaks in complex heterogeneous networks." The European Physical Journal B-Condensed Matter and Complex Systems 26.4 (2002): 521-529.