Epidemics in networks: a master equation approach

M Cotacallapa and M O Hase

Escola de Artes, Cincias e Humanidades, Universidade de Sao Paulo, Avenida Arlindo Bettio 1000, 03828-000, Sao Paulo, SP, Brazil

E-mail: moshe@usp.br and mhase@usp.br

Received 10 July 2015, revised 4 November 2015
Accepted for publication 2 December 2015
Published 5 January 2016

Abstract
A problem closely related to epidemiology, where a subgraph of ‘infected’ links is defined inside a larger network, is investigated. This subgraph is generated from the underlying network by a random variable, which decides whether a link is able to propagate a disease/information. The relaxation timescale of this random variable is examined in both annealed and quenched limits, and the effectiveness of propagation of disease/information is analyzed. The dynamics of the model is governed by a master equation and two types of underlying network are considered: one is scale-free and the other has exponential degree distribution. We have shown that the relaxation timescale of the contagion variable has a major influence on the topology of the subgraph of infected links, which determines the efficiency of spreading of disease/information over the network.

Keywords: complex networks, epidemics, master equations

1. Introduction
The analysis of the structure and dynamics of many ‘real’ networks, like the world wide web [1], the internet [2], and many biological [3] and social [4] networks, has inspired a number of works in recent years. Nowadays, from the massive amount of data available, one can recognize a heterogeneous structure in many of these (complex) networks, usually following a power-law degree distribution. From the standpoint of their representation as graphs, the links between vertices show the interconnections between different nodes, and the degree distribution is a relevant aspect of transmission of any kind of information over the network. The spreading of information between vertices is closely related to the classical problem of percolation [5, 6], which was first examined on Bravais lattices or other regular structures like Cayley trees [7].
An important question that arises is associated with the problem of epidemiology on complex networks. First of all, in the study of epidemiology, the disease should propagate between people who are interconnected in a fashion which resembles a complex network, and not a regular lattice. Moreover, since critical phenomena on complex networks display a richer behavior than classical statistical models on regular lattices (for instance, the first results of the ferromagnetic Ising model on a Barabási–Albert model can be found in [8, 9]; see also [10] for a review of critical phenomena on complex networks), one expects richer results when dealing with problems defined on complex networks.

In the literature, it is known that a susceptible–infective–susceptible (SIS) epidemic model on a scale-free network displays persistence, and no positive critical value $\lambda_c$, below which the epidemic is confined on a finite fraction of the infinite network, exists [11–14]. This problem is closely related to the problem of the robustness of a network where the connectivity of a graph is examined as vertices are removed, and this problem can be mapped on a percolation problem [15–18]. It has been known since then that scale-free networks are robust against failures (removal of random vertices), but weak against attacks (removal of targeted nodes).

In this work, the problem of the propagation of a disease or information will be revisited in the framework of a master equation combined with a disordered variable that mimics the propagation rate. This random variable will be analyzed in two extreme regimes: annealed and quenched. These terms are commonly present in disordered systems (in particular, spin glass theory [19]), and the problem of epidemiology (or propagation of information over the graph) on a complex network will be viewed as a disordered problem in this work.

Given a network, an edge that links two vertices represents an interaction between them; nevertheless, this connection is a potential channel of communication, but does not imply necessarily that these two nodes share or transmit a piece of information or a disease. In this work, we characterize the subgraph of this network that effectively propagates information/disease, and see how this subgraph relates to the underlying network. In this sense, our work differs from the traditional studies of epidemiology on complex networks commented on in the previous paragraph, since we are not investigating the emergence of a giant component of infected nodes and establishing a threshold in the infection rate as a percolation problem. We did not consider the traditional dynamical process of disease/information spread, but characterized the subgraph of the network composed by nodes that can effectively communicate with each other to transmit disease/information when it appears. Therefore, we are interested in the statistics of degrees of the nodes that can participate in the process of propagation of information/disease (if it appears) in the network. This is a measure of how information/disease can propagate over the network, and we have introduced the ‘reduced degree distribution’ (see section 2) to quantify it.

Moreover, we have examined the graph which effectively transmits information/disease in two cases, as mentioned before. When an edge is recognized as a true transmitter of information/disease between two vertices, we considered a case where this connection remains so and does not close the flow of information/disease. This is the case where the random variable that dictates if a link is ‘functional’ or not is a quenched variable. We have also considered the other extreme situation, where the change of the state of links (transmit/not transmit) is rapid (the annealed case).

The layout of this work is as follows: in section 2 we present the basic setup for the models that are analyzed in two different regimes of contagion process in sections 3 and 4. The results are discussed in the last section.
2. Epidemic models

In this section, we give a summary of the models that will be considered in this work. To develop the epidemic models, two questions are asked in the construction of the network: ‘which node will we link to?’ and ‘will we be infected by this node?’. In the present case, the first question is answered by the rules of the network dynamics, and the second question is where this paper introduces the contagion process.

The main setup is a growing network where a single vertex is added at each time step to the graph, and this new incoming vertex links with just one single old pre-existing one with probability \( \Pi(k, t) \) at time \( t \), where \( k \) is the degree of the old node that receives the link from the new one. Once this connection is made, we draw a random variable \( r_t \) to decide if this link can transmit disease (or information) or not. This two-step dynamical process can be described by the discrete time master equation

\[
p_r(k, s, t + 1) = r\Pi(k - 1, t)p_r(k - 1, s, t) + \left[1 - \Pi(k, t)\right]p_r(k, s, t),
\]

where \( p_r(k, s, t) \) stands for a measure of the site \( s \) having degree \( k \) at time \( t \) in the subgraph of vertices that can propagate information/disease. Note that the label \( s \) of a vertex stands also for the instant when this vertex joined the network. The boundary condition associated with this quantity for a new incoming vertex is, then, \( p_r(k, s, t = s) = \delta_{k,1} \). For the random variable \( r_t \), which mimics the contagion process, we choose a Bernoulli distribution,

\[
\mathcal{P}(r_t) = (1 - c)\delta_{r_t,0} + c\delta_{r_t,1},
\]

where \( \delta \) is the Kronecker symbol and \( c \in [0, 1] \subset \mathbb{R} \) controls the contagion rate. For notational convenience, we adopt two vertices that have two connections between them at time \( t = 2 \) as the initial condition, but this particular choice is not expected to have a serious impact on the properties we are interested in after a long time. In this setup, the total degree of the graph at time \( t \) is simply \( 2t \).

We will now discuss some dynamics chosen for the growing network above in this work. In the first one, at each unit of time a new node is added and randomly links to another old node chosen with equal probability. In other words, at time \( t \), one has \( \Pi(k, t) = \frac{1}{k} \). After that, it decides whether or not to infect the old node. The second dynamics is based on a preferential attachment scheme, and \( \Pi(k, t) \propto k^\gamma \) [20]: vertices that have higher degrees (infectious or not) are more likely to receive new connections.

Furthermore, we will investigate the role of the contagion process, represented by the set of random variable \( \{r_t\} \). Let us examine the equation (1), which provides information on the subgraph of vertices that can propagate information/disease. A vertex \( s \) of this subgraph can have degree \( k \) at time \( t + 1 \) if:

- The vertex \( s \) of this subgraph has \( k - 1 \) links at time \( t \), and receives a link from the new incoming vertex with probability \( \Pi(k - 1, t) \). This is the term \( \Pi(k - 1, t)p_r(k - 1, s, t) \).

Then, two situations should be considered:

a. This new link allows the propagation of information/disease, \( r_t = 1 \).

b. This new link does not allow the propagation of information/disease, \( r_t = 0 \).

- The vertex \( s \) of this subgraph has already \( k \) links at time \( t \); therefore, it must not receive the connection from the new incoming vertex to keep degree \( k \), and this happens with probability \( 1 - \Pi(k, t) \).
We will consider two limit cases of its relaxation time: the annealed and quenched ones. In the former, the contagion variable \( r \) varies much faster than a typical ‘observational time’ of the whole process, and this fast fluctuation leads us to treat it as effectively having its mean value, \( \langle r \rangle = c \), in respect to the distribution (2). In this annealed case, all the links transmit information/disease with probability \( c \). In the other limit, the random variable \( r \) is considered to be quenched. This means that during a typical ‘observational time’, it remains constant (with \( r = 1 \) or \( r = 0 \)) without any fluctuation. The analysis should then consider many realizations of the process, and one should evaluate \( \langle p_i(k, s, t) \rangle \) in the end. It is important to stress that in the present work, vertices have the state ‘transmits disease/information’ (when it appears) and ‘does not transmit disease/information’ (even if appearing). Nodes may be connected, but they do not always share or transmit information (or diseases), and by ‘contagion’ between two vertices we mean that not only are they connected, but there is also a flow of information (disease) allowed between them. In this sense, this work is different from the previous ones treated in the literature [11–18]. It is worth mentioning that, in the context of opinion dynamics, a dynamics with another timescale (time-dependent flipping rate of the voter model) was considered [21].

Our goal is to examine the ‘reduced degree distribution’ \( \tilde{P}_r(k) \), which is defined as

\[
\tilde{P}_r(k) = \lim_{t \to \infty} \frac{1}{t} \int_0^t ds \ p_r(k, s, t).
\]

Let us first consider the case without the contagion variable to understand (3): if \( p(k, s, t) \) is the probability of a vertex \( s \) having \( k \) links at time \( t \), the time-dependent degree distribution \( P(k, t) \) is the average value of \( p(k, s, t) \) over the sites, \( P(k, t) = \sum_{s=1}^t p(k, s, t)/t \), and the degree distribution is its stationary value, \( P(k) = \lim_{t \to \infty} P(k, t) \). The ‘reduced degree distribution’ \( \tilde{P}_r(k) \) above is inspired in this definition. This quantity, \( \tilde{P}_r(k) \), is related to the degree distribution of the subgraph of vertices that allow propagation of information/disease, and we will analyze different systems to examine if it differs (and how it differs) from the original degree distribution of the underlying network. Nevertheless, it should be noted that the function \( \tilde{P}_r(k) \) is not normalized for \( c \neq 1 \) (therefore, it is not a probability distribution), because it counts the fraction of nodes that have transmitting links with respect to the whole network (and not to the vertices of the infected subgraph only). However, it gives a measure of the distance to the process where the contagion always happens (which is the case \( r_s = 1 \) for any \( 1 \leq s \leq t \), or \( c = 1 \)).

3. Fixed contagion process

As a preliminary analysis, we will consider the random variable \( r_s (1 \leq s \leq t) \), which decides the flow of infection on an edge, as being the same for every link and no changes will be made over this variable while the network is growing. In this section, therefore, we will denote this random variable as \( r = r_s \). This crude approximation will be compared later to a more realistic situation where each link has an independent configuration of propagation. From this setup, two limit cases of the random variable \( r \) will be considered as follows.

In the ‘annealed’ approximation, the fluctuation of the random variable \( r \) is regarded as being fast such that its value at equation (1) can be replaced by its mean value, \( \langle r \rangle \), where the mean is taken with respect to the Bernoulli distribution

\[
\mathcal{P}(r) = (1 - c)\delta_{r,0} + c\delta_{r,1},
\]

which resembles (2) except for the fact that here \( r_s = r \) for any link. Consequently, in the annealed case, one should examine the master equation
where one replaces the variable $r$ with the number $r_c$, according to (4). The equation can be solved by standard techniques [22] to obtain $p_{(s)}(k, s, t)$ and then the reduced degree distribution $\tilde{P}_r(k)$.

On the other hand, in the ‘quenched’ limit, the variable $r$ does not change during the analysis, which means that one requires the solution of the master equation (5) with $r$ fixed, and the average $\langle p_{(s)}(k, s, t) \rangle$ is taken at the end in order to determine the stationary quantity $\langle \tilde{P}_r(k) \rangle$.

We will now analyze two dynamics for the growing network. In the first case, all of the pre-existing nodes to which the new incoming vertex links have the same probability of being connected. In the second dynamics, the new incoming vertex attaches to another according to the linear preferential linking.

3.1. Random network with equiprobable attachment

In the present case, the probability of linking at time $t$ is $\Pi(k, t) = 1/t$, and the solution of the master equation can be obtained by the usual techniques [22] to yield

$$\tilde{P}_r(k) = \left( \frac{r}{2} \right)^k.$$  \hfill (6)

Therefore, in the annealed case, one has

$$\tilde{P}_{(s)}(k) = c^k 2^{-k} \quad \text{(annealed)},$$  \hfill (7)

which is showed in figure 1, while in the quenched case, the reduced distribution is

$$\langle \tilde{P}_r(k) \rangle = c2^{-k} \quad \text{(quenched)},$$  \hfill (8)

and the reduced distribution can be seen in figure 2.
One can see that while in the quenched case the reduced distribution differs only by a factor from the degree distribution, in the annealed case the difference is more significant in the sense that the basis of the exponent changed. For a given degree \( k > 1 \), one can also see that \( \tilde{P}_{\gamma}(k) < \langle \tilde{P}_r(k) \rangle \), showing that the quenched variable for contagion is more efficient than the annealed one in this model.

### 3.2. Preferential attachment network

In this section, the linking probability is \( \Pi(k, r) = \frac{k}{2r} \) according to the model introduced before, where the denominator \( 2r \) is the normalization factor of the number of degrees. Solving the master equation in this case leads to

\[
\tilde{P}_{\gamma}(k) = \frac{4c^k}{k(k + 1)(k + 2)},
\]

in the annealed situation (illustrated in figure 3), while one has

\[
\langle \tilde{P}_r(k) \rangle = \frac{4c}{k(k + 1)(k + 2)},
\]

in the quenched case, shown in figure 4.

Note that the effect of the annealed variable has major consequences in this case: the reduced degree distribution is not a power-law function any more. This means that the subgraph of contagion has a different structure from the underlying network. On the other hand, the quenched variable still preserves the power-law structure. As in the previous case (equiprobable attachment model), \( \tilde{P}_{\gamma}(k) < \langle \tilde{P}_r(k) \rangle \) for \( c = 1 \) and \( k > 1 \).
4. Link-dependent contagion process

In this section, the master equation

$$p_{r}(k, s, t + 1) = r_{s} \Pi(k - 1, t)p_{r}(k - 1, s, t) + (1 - \Pi(k, t))p_{l}(k, s, t)$$

(11)

will be analyzed when the underlying network follows a preferential linking prescription and the contagion linking, governed by the random variable $r_{s} \in [0, 1]$ ($1 \leq s \leq t$), depends on

---

**Figure 3.** Reduced distribution of epidemic network using annealed approach over a preferential attachment network, where $c = \{1, 0.75, 0.5, 0.25\}$.  

**Figure 4.** Reduced distribution of epidemic network using quenched approach over a preferential attachment network, where $c = \{1, 0.75, 0.5, 0.25\}$.  

4. Link-dependent contagion process

In this section, the master equation

$$p_{r}(k, s, t + 1) = r_{s} \Pi(k - 1, t)p_{r}(k - 1, s, t) + (1 - \Pi(k, t))p_{l}(k, s, t)$$

(11)
the particular edge considered. From the structure of the equation, it is clear that the annealed approximation leads to the same results as before, and one needs to analyze the quenched case only. In order to determine the reduced distribution \( P_{kr}(\tilde{r}) \), one should first evaluate

\[
\langle p_r(k, s, t) \rangle := \int dr_1 \cdots dr_t \mathcal{P}(r_1, \ldots, r_t) p_r(k, s, t).
\]  

(12)

Since the random variables \( \{r_s\} \) are independent and identically distributed, the global distribution can be factorized as \( \mathcal{P}(r_1, \ldots, r_t) = \prod_{s=1}^t \mathcal{P}(r_s) \), where

\[
\mathcal{P}(r_s) = (1 - c) \delta_{r_s, 0} + c \delta_{r_s, 1}
\]

as in (2), and, fixing the site \( t \), one has

\[
\langle p_r(k, s, t) \rangle := \int dr \mathcal{P}(r) p_r(k, s, t),
\]

where the \( p_r \) stands for the marginal distribution

\[
p_r(k, s, t) = \int dr_1 \cdots dr_{t-1} p_r(k, s, t).
\]

(15)

Furthermore, from the obvious equality

\[
p_r(k, s, t) = (1 - r_t) p_r(k, s, t) + q_r(k, s, t), \quad \text{where } q_r(k, s, t) := r_t p_r(k, s, t),
\]

the desired quantity is evaluated as

\[
\langle p_r(k, s, t) \rangle = \int_0^1 dr \mathcal{P}(r)(1 - r_t) p_r(k, s, t) + \int_0^1 dr \mathcal{P}(r) q_r(k, s, t).
\]

(17)

Nevertheless, instead of equation (17), let us consider

\[
\int_0^R dr \mathcal{P}(r) p_r(k, s, t) = \int_0^R dr \mathcal{P}(r)(R - r_t) p_r(k, s, t) + \int_0^R dr \mathcal{P}(r) q_r(k, s, t).
\]

(18)

It is evident that (18) recovers (17) when \( R = 1 \). By invoking the Laplace transform on (18), one has

\[
\psi^L(\zeta) = \left( \frac{\zeta}{\zeta - 1} \right) \chi^L(\zeta),
\]

(19)

where \( \psi^L \) and \( \chi^L \) are Laplace transforms of the functions

\[
\psi(R) := \int_0^R dr \mathcal{P}(r) p_r \quad \text{and} \quad \chi(R) := \int_0^R dr \mathcal{P}(r) q_r,
\]

(20)

respectively. The function (12) is determined by first obtaining \( \psi(R) \) and letting \( R \to 1 \). One may solve the equation (19) and find that

\[
\psi(R) = \int_0^\infty dr \mathcal{P}(r) q_r e^{R - r} \theta(R - r_t),
\]

(21)

where \( \theta \) stands for the Heaviside function with \( \theta(0) = 1 \).

In order to continue the calculations, one needs to evaluate the function \( q_r \), above, and this can be done from the master equation (11). Taking advantage of the fact that the random variable \( r_t \) is binary (note that \( r_t = r_t^0 \)), multiplying the master equation by \( r_t \) and integrating in order to analyze the marginal distribution \( p_r \) yields
If the underlying graph obeys the linear preferential linking rule, 
\[ \Pi(k, t) = \frac{k}{2t} \], and from the

the recurrence equation

the recurrence equation (22) can be cast as

with the boundary condition \( Q_z(z, s = t = s) = r_z \) for large times. This equation can be solved by the methods of characteristics and inverting the Z-transform, one has

Finally, inserting (13) and (25) into (21), and letting \( R \rightarrow 1 \), one has

for \( t \gg 1 \). We have then the reduced degree distribution, which is a stationary quantity, by the procedure described in (3). It turns out that the reduced distribution is just a constant multiplying the original one,

\[ \langle \bar{P}_t(k) \rangle = \frac{4c}{k(k + 1)(k + 2)}. \]
As a way to support this analytical result, we performed a simulation of this model, where we generated $10^3$ networks with $10^4$ nodes each (figure 5). The uncertainty bars in the figure are smaller than the size of the points. Works like [23] study in detail the uncertainty in the output by relating it with the model input, which can be done in many different contexts [24, 25].

Following the same procedure, one can also show that

$$\left\langle P_c(k) \right\rangle = c2^{-k}$$  \hspace{1cm} (28)

if the underlying graph is made by uniform random linking, $\Pi(k, t) = 1/t$.

Note that the results (27) and (28) are the same as (8) and (10), respectively. Although the reduced distributions are the same, the possible configurations of the contagion network are clearly different.

5. Discussion

In this work, we have shown an analysis based on analytical results supported by some simulations. Firstly, the quenched approach has better performance in spreading an epidemic in all scenarios. This means, in simple words, that ‘it is better to share, spread or infect sometimes a complete object, information or disease, than propagating all the time, but with just a part of the information, disease or any element to spread’. We should also note that the present scheme of defining propagation could capture distinct behaviors in contagion and its underlying graph: even with the connections between vertices following a power-law distribution, the subgraph of the contagion process may not follow the same distribution, as shown in the annealed case (see figure 3). Nevertheless, one should remember that in the present setup, the problem of propagation of information over the network is associated with the presence of contagion links only. A system where the vertices are spins (therefore, they have states that can influence other vertices), and the links can be favorable/unfavorable to align them, constitutes a different way to deal with the contagion problem which can be mapped to a spin glass problem. In this case, one can see that spreading due to an annealed random contagion variable is more efficient than the quenched approach [26].

Another important result is the reduced degree distribution between link-variable (section 4) and fixed contagion (section 3) being the same. This coincidence can be connected to the distribution of the contagion variable being an independent and identically distributed random variable and the absence of interaction between vertices. As a consequence, a single variable that decides the contagion of the entire graph (fixed contagion case) and the situation where each one of the links obeys the same distribution independently have similarities, despite the fact that the configuration of infected links is clearly different in both cases.

Acknowledgments

MC acknowledges the OEA scholarship program for financial support.

References

[1] Barabási A L, Albert R and Jeong H 2000 *Physica A* **281** 69–77
[2] Faloutsos M, Faloutsos P and Faloutsos C 1999 *SIGCOMM Comput. Commun. Rev.* **29** 251–62
[3] Jeong H, Tombor B, Albert R, Oltvai Z N and Barabasi A L 2000 *Nature* **407** 651–4
[4] Barabási A, Jeong H, Néda Z, Ravasz E, Schubert A and Vicsek T 2002 *Physica A* **311** 590–614
[5] Flory P J 1941 *J. Am. Chem. Soc.* **63** 3083–90
[6] Broadbent S R and Hammersley J M 1957 Math. Proc. Camb. Phil. Soc. 53 629–41
[7] Stauffer D and Aharony A 1994 Introduction To Percolation Theory: Revised 2nd edn (London: Taylor and Francis)
[8] Aleksiejuk A, Holyst J A and Stauffer D 2002 Physica A 310 260–6
[9] Bianconi G 2002 Phys. Lett. A 303 166–8
[10] Dorogovtsev S N, Goltsev A V and Mendes J F F 2008 Mod. Phys. Rev. 80 1275–335
[11] Pastor-Satorras R and Vespignani A 2001 Phys. Rev. Lett. 86 3200–3
[12] Pastora Satorras R and Vespignani A 2001 Phys. Rev. E 63 66117
[13] May R M and Lloyd A L 2001 Phys. Rev. E 64 66112
[14] Boguñá M, Pastor-Satorras R and Vespignani A 2003 Phys. Rev. Lett. 90 28701
[15] Albert R, Jeong H and Barabasi A L 2000 Nature 406 378–82
[16] Cohen R, Erez K, Ben-Avraham D and Havlin S 2000 Phys. Rev. Lett. 85 4626–8
[17] Cohen R, Erez K, Ben-Avraham D and Havlin S 2001 Phys. Rev. Lett. 86 3682–5
[18] Callaway D S, Newman M E J, Strogatz S H and Watts D J 2000 Phys. Rev. Lett. 85 5468–71
[19] Mézard M, Parisi G and Virasoro M A 1987 Spin Glass Theory and Beyond (World Scientific Lecture Notes in Physics) (Singapore: World Scientific)
[20] Barabási A L and Albert R 1999 Science 286 509–12
[21] Baxter G J 2011 J. Stat. Mech. 2011 P09005
[22] Dorogovtsev S N and Mendes J F F 2003 Evolution of Networks: From Biological Nets to the Internet and WWW (Oxford: Oxford University Press)
[23] Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M and Tarantola S 2007 Introduction to sensitivity analysis Global Sensitivity Analysis: The Primer (Chichester: Wiley) ch 1
[24] Convertino M, Muñoz Carpena R, Chu-Agor M, Kiker G and Linkov I 2014 Environ. Modelling Softw. 51 296–309
[25] Lüdtke N, Panzeri S, Brown M, Broomhead D S, Knowles J, Montemurro M A and Kell D B 2008 J. R. Soc. Interface 5 223–35
[26] Hase M O and Mendes J F F 2008 J. Phys. A: Math. Theor. 41 145002