

Available online at www.sciencedirect.com



www.elsevier.com/locate/physa

PHYSICA

Physica A 387 (2008) 623-630

How community structure influences epidemic spread in social networks

Xiaoyan Wu, Zonghua Liu*

Institute of Theoretical Physics and Department of Physics, East China Normal University, Shanghai, 200062, China

Received 15 August 2007 Available online 2 October 2007

Abstract

Two key features of social networks are the community structure and the high clustering coefficient. For understanding their influences on dynamical processes, we present a model with both an adjustable clustering coefficient and an adjustable degree of community. This model has an invariant degree distribution when its clustering coefficient is being adjusted. We find that the efficiency of epidemic spreading in this model depends mainly on the degree of community and decreases with increase of the degree of community. For a fixed degree of community, the efficiency will decrease with increase of the clustering coefficient. Numerical simulations have confirmed the theoretic analysis.

© 2007 Elsevier B.V. All rights reserved.

PACS: 89.75.Hc; 87.23.Ge; 87.19.Xx

Keywords: Social networks; Adjustable clustering coefficient; Epidemic spreading; Degree of community

1. Introduction

Both the network structure and dynamics on complex networks have been well studied recently [1–3]. Two most important quantities for characterizing the structure of a complex network are the degree distribution and the clustering coefficient. It is found that most of the realistic networks are scale-free (SF) networks which has a power-law degree distribution and low clustering coefficient. To understand its mechanism, Barabasi and Albert presented a growing network model according to the preferential attachment principle which says that a node with more links has larger possibility of getting a new link [4]. After that, a lot of amended SF network models have been given, which stress different aspects of growing, such as nonlinear preferential attachment, initial attractiveness, internal edges and rewiring, fitness model, etc. [5–8].

However, there is another kind of network, i.e., the social network, which has a high clustering coefficient [9–11]. It has been revealed that social networks have a community structure where the links are dense in a community but sparse between communities. Moreover, it is possible for an individual in a social network to belong to one or several communities [12,13]. To explain the mechanism of social networks, several models have been presented,

^{*} Corresponding author. Tel.: +86 21 62233216. E-mail address: zhliu@phy.ecnu.edu.cn (Z. Liu).

which include the mutual friendship approach [9,11], shared nodes among groups [12], dense links in community and sparse links between communities [10,14], and growing groups [15] etc. These models usually have exponential degree distribution. Although these achievements have allowed people to understand social networks much better than before, the mechanism is still not completely clear because of the diversity of communities. For example, it is possible for a community to have a SF degree distribution [16]. We here present a novel model reflecting the features of both community structure and high clustering coefficient in social networks. This model has a power-law degree distribution, i.e., its degree distribution is non-exponential. A main advantage of this model is that both its clustering coefficient and its degree of community can be continuously changed, which makes it a paradigmatic model for studying the influence of community structure on the dynamics of social networks.

One of the important dynamics on complex networks is epidemic spreading. Here an important issue is, for a given structure of complex network, how the infection propagates and whether there exist effective control strategies for preventing or suppressing the spread of infection [1–3,17]. In efforts to solve this problem, a number of approaches/models have been presented [18–25]. Among them, an interesting result pointed out by Paster-Satorras and Vespignani is that a virus could spread on the Internet and WWW even when the infection probabilities are vanishingly small [18]. Recently, some preliminary work has been done on epidemic spreading in social networks [9,14]. We know that one important feature of community is its high clustering coefficient. How the clustering coefficient influences the epidemic is an open question. Here we will use our new model to study this problem. We find that epidemic spreading on a community network is mainly determined by its degree of community and a lower degree of community favors spreading, while for a fixed degree of community, increase of the clustering coefficient will reduce the efficiency of epidemic spreading.

An epidemic process can be generally characterized by two typical models, i.e., the SIS (susceptible–infected–susceptible) model [18,26–28] and the SIR (susceptible–infected–refractory) model [21,29–31]. In the SIS model, an agent has two possible statuses: susceptible and infected. A susceptible agent may become infected once it contacts an infected one. After a time step, the infected agents recover and return to the susceptible state. However in the SIR model, an agent has three possible statuses: susceptible, infected, and refractory. The infected agents cannot go back to the susceptible status but become refractory, which describes the phenomenon of long-time immunity. In addition to these two typical models, there are other models, such as the SI (susceptible–infected) model [32] and SIRS (susceptible–infected–refractory–susceptible) model [33] etc. In this paper we will choose the SIR model for studying the influence of community structure on the spreading of epidemic.

The paper is organized as follows. In Section 2, we introduce a community network model with both an adjustable clustering coefficient and an adjustable degree of community, and show the confirmation of numerical experiments. Then in Section 3, we study the influence of community structure on epidemic spreading using both theoretical analysis and numerical simulations. Finally, the conclusions are given in Section 4.

2. Community network model with both an adjustable clustering coefficient and an adjustable degree of community

Although the possible mechanism of forming a community network has been partially revealed by several models [9–11,14,15], how the high clustering coefficient shows up is still not completely understood. Moreover, how the degree of community influences the epidemic spreading is an open question. Observing the fact that a new friendship is formed between two people through an introducer of a common old friend in a community activity, the clustering of friends in a community can be modelled by a triad formation approach [16] in which a link is put between two neighbors of one node. This approach also fits for the links between different groups. Considering the fact that a friendship may also come from introducing oneself, the preferential attachment principle should also be taken. In addition, different social networks may have different degrees of community, which represent the relative densities of links in a group and between groups. After considering all these aspects, we here present a social network model with a multiple-community structure. Our algorithm is as follows:

- (1) Initially, there are m_0 groups. In each group, there are m_1 nodes which are completely connected to each other. In this paper, we let $m_0 = m_1 = 3$.
- (2) At every time step, each group has added a new node with probability p, i.e., the total added nodes are m_0p . Each added new node will emit m links to the existing nodes of the same group. Here we choose m=2. The first link will be preferentially attached to a node-i with probability $k_i / \sum k_j$ where k_i represents the links of node-i, and

j in the sum is for all the nodes in the group. The second link will be randomly connected to one of the neighbors of node-i with probability q and be preferentially connected to anyone in the group with probability 1-q. Therefore, the evolution of k_i can be given by

$$\frac{\partial k_i}{\partial t} = p \left[\frac{k_i}{\sum_j k_j} + q k_i \left(\frac{k_n}{\sum_j k_j} \frac{1}{k_n} \right) + (1 - q) \frac{k_i}{\sum_j k_j} \right]$$

$$= 2p \frac{k_i}{\sum_j k_j} \tag{1}$$

where the term $k_i(\frac{k_n}{\sum_i k_i} \frac{1}{k_n})$ comes from the k_i neighbors of node-*i*.

(3) At each time step, we preferentially choose a node-i in each group with probability 1-p and let it emit m links to m nodes in other groups, i.e., the total added links are $mm_0(1-p)$. The m nodes can be chosen as follows: We firstly choose one group randomly from the other two groups. Then we preferentially choose a node from the chosen group for the first link, and the second link will randomly go to one neighbor of the chosen node with probability q and preferentially go to anyone in the chosen group with probability 1-q. Their contribution to the evolution of k_i is

$$\frac{\partial k_i}{\partial t} = (1 - p) \left[2 \frac{k_i}{\sum_j k_j} + \left(\frac{k_i}{\sum_j k_j} + q k_i \left(\frac{k_n}{\sum_j k_j} \frac{1}{k_n} \right) + (1 - q) \frac{k_i}{\sum_j k_j} \right) \right]$$

$$= 4(1 - p) \frac{k_i}{\sum_j k_j} \tag{2}$$

where the term $(\frac{k_i}{\sum_j k_j} + qk_i(\frac{k_n}{\sum_j k_j} \frac{1}{k_n}) + (1-q)\frac{k_i}{\sum_j k_j})$ comes from the other two groups.

(4) Repeat the steps (2) and (3) until the nodes in each group amount to N_0 . That is, the total of nodes in the network is $N = m_0 N_0$ and the evolution will be stopped when $t = (N_0 - m_0)/p$.

By adding Eqs. (1) and (2) we obtain

$$\frac{\partial k_i}{\partial t} = 2(2 - p) \frac{k_i}{\sum_j k_j}.$$
(3)

When t is large, we have $\sum_j k_j = m_1(m_1 - 1) + 2mt \approx 2mt$. For a node added at time t_i , its initial condition is $k_i(t_i) = m = 2$. Substituting these into Eq. (3) we obtain the solution

$$k_i(t) = 2\left(\frac{t}{t_i}\right)^{\frac{2-p}{m}}.\tag{4}$$

The probability of a node with degree $k_i(t) < k$ can be written as

$$P[k_i(t) < k] = P\left[t_i > \left(\frac{2}{k}\right)^{\frac{m}{2-p}} t\right] = 1 - \left(\frac{2}{k}\right)^{\frac{m}{2-p}} \frac{t}{t + m_0}.$$
 (5)

Thus, we have

$$P(k) = \frac{\partial P[k_i(t) < k]}{\partial k} = 2^{\frac{m}{2-p}} \frac{m}{2-p} k^{-\frac{m}{2-p}-1}.$$
 (6)

Obviously, it is a power-law degree distribution.

Another important quantity for characterizing the network structure is the clustering coefficient, C, which characterizes the possibility for one's friends to become friends each other. For a node-i with degree k_i , its clustering coefficient $C_i = \frac{2E_i}{k_i(k_i-1)}$ where E_i represents the total links among the k_i neighbors of node-i. The clustering

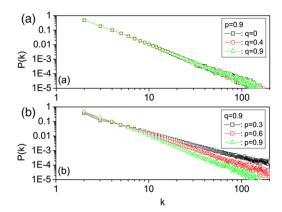


Fig. 1. (Color online.) Degree distributions for $m_0 = m_1 = 3$, m = 2 and N = 3000. (a) denotes the situation of fixed p = 0.9 and varying q where the "squares", "circles" and "triangles" represent the cases of q = 0, 0.4, and 0.9, respectively; (b) denotes the situation of fixed q = 0.9 and varying p where the "squares", "circles" and "triangles" represent the cases of p = 0.3, 0.6, and 0.9, respectively.

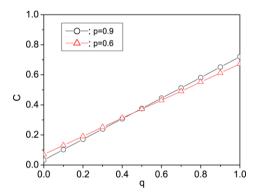


Fig. 2. (Color online.) Clustering coefficient C versus q where the "circles" and "triangles" represent the cases of p = 0.9 and 0.6, respectively.

coefficient of the network equals the average of all the C_i , i.e., $C = \frac{1}{N} \sum C_i$. From the algorithm (1) to (4) we know that E_i is proportional to q. Thus, we have

$$C \sim q$$
. (7)

In conclusion, p controls the degree distribution and q controls the clustering coefficient. Therefore, both P(k) and C can be continually changed through p and q. A key feature of this model is that the degree distribution P(k) in Eq. (6) is independent of the parameter q! Therefore, we can keep the P(k) unchanged and study the influence of C on the epidemic spreading through adjusting q.

For confirming the theoretical predictions (6) and (7), we firstly construct a community network with m=2, $m_0=m_1=3$ and total nodes N=3000 according to the above rules (1)–(4). Then we calculate its degree distributions for different q and p. Fig. 1 shows the results where (a) represents the case of changing q and (b) the case of changing p. From Fig. 1(a) it is easy to see that the P(k) for different q are overlapped, confirming that P(k) is independent of q. From Fig. 1(b) we see that the slopes for different p are different and are exactly what was predicted from Eq. (6).

We have also calculated the clustering coefficient. We find that C is sensitive to q but robust against p, i.e., C is mainly determined by q. Fig. 2 shows the results where the "circles" and "triangles" represent the cases of p=0.9 and 0.6, respectively. Obviously, C is linearly increasing with q, confirming Eq. (7).

3. Influence of community structure on epidemic spreading

For a given network, the process of epidemic spreading is as follows [17,21,29]: Suppose initially only one node ("seed") is infected and the remaining nodes are susceptible. The infection will spread along the links to the

neighboring nodes of the infected one with probability λ . With time going on, more nodes will be infected and then it is possible for a susceptible node to have more than one infected neighbor. When a susceptible node has $k_{\rm inf}$ infected neighbors, it will become infected with probability $[1-(1-\lambda)^{k_{\rm inf}}]$. At the same time, the infected node will decay into the refractory one with probability μ . Without loss of generality, we set $\mu=1$ since it only affects the definition of the time scale of the epidemic spreading. In the early stage of evolution, the number of infection nodes will increase. At sufficiently large time, this number will begin to decrease until there are no longer any infected nodes in the network at time τ ; then the process is over. During the evolution process with $t<\tau$, a susceptible node may become infected and then refractory; thus the number of refractory nodes will keep increasing. The number of the final refractory nodes represents the epidemic prevalence.

Because the constructed community network is not homogeneous but heterogeneous, different nodes will have different abilities to spread the epidemic. In order to take into account the heterogeneity induced by the nodes with different degree k, we divide the nodes into different groups where the nodes in each group have the same degree k. We let $s_k(t)$, $i_k(t)$, and $r_k(t)$ be the densities of susceptible, infected, and refractory nodes of degree k at time t, respectively. Thus, we have

$$s_k(t) + i_k(t) + r_k(t) = 1.$$
 (8)

Because of the clustering structure, the speed of epidemic spreading will be reduced by the inner links in a cluster. Thus, we cannot directly write the evolution equations of $s_k(t)$, $i_k(t)$, and $r_k(t)$, as in Ref. [29]. This can be illustrated through an example of $\lambda=1$ as follows: For a tree-like network with the same degree k at every node, an infected node will make k infected nodes in the first step and k(k-1) infected nodes in the second step. But for a clustered network with the same degree k and a clustering coefficient C, an infected node will make k infected nodes in the first step and $k(k-1-\frac{2E}{k})=k(k-1)(1-C)$ infected nodes in the second step where E is the links among the neighbors of a node. When C=1, the epidemic spreading will stop at t=1. Therefore, the epidemic spreading is linearly reduced by a factor 1-C. Returning to our model, we use f(C) to characterize the influence of high clustering on the speed of epidemic spreading, where f(C) satisfies 0 < f(C) < 1 and df(C)/dC < 0. On the basis of this analysis, we have

$$\frac{\mathrm{d}s_{k}(t)}{\mathrm{d}t} = -\lambda k s_{k}(t) f(C) \sum_{k'} \frac{k' P(k') i_{k'}(t)}{\langle k \rangle}$$

$$\frac{\mathrm{d}i_{k}(t)}{\mathrm{d}t} = -i_{k}(t) + \lambda k s_{k}(t) f(C) \sum_{k'} \frac{k' P(k') i_{k'}(t)}{\langle k \rangle}$$

$$\frac{\mathrm{d}r_{k}(t)}{\mathrm{d}t} = i_{k}(t)$$
(9)

where the mean-field approach is used, and the sum $\sum_{k'} k' P(k') i_{k'}(t)/\langle k \rangle$ is the probability that a randomly chosen link points to an infected node. Suppose, at the beginning, every node has the same possibility of being infected. Hence the initial condition is $r_k(0) = 0$, $i_k(0) = 1/NP(k)$, and $s_k(0) = 1 - i_k(0)$. In the limit $i_k(0) \to 0$, $s_k(0) \approx 1$. Integrating the first equation of Eq. (9) we obtain

$$s_k(t) = e^{-\lambda k f(C)\phi(t)}$$
(10)

where the auxiliary variable $\phi(t)$ is defined as

$$\phi(t) = \int_0^t \sum_{k'} \frac{k' P(k') i_{k'}(t')}{\langle k \rangle} dt' = \frac{\sum_{k'} P(k') r_{k'}(t)}{\langle k \rangle}.$$
(11)

The derivative of $\phi(t)$ can be expressed as

$$\frac{\mathrm{d}\phi(t)}{\mathrm{d}t} = \frac{\sum k' P(k')(1 - s_{k'}(t) - r_{k'}(t))}{\langle k \rangle}$$

$$= 1 - \phi(t) - \frac{\sum k' P(k') \mathrm{e}^{-\lambda k'} f(C)\phi(t)}{\langle k \rangle}.$$
(12)

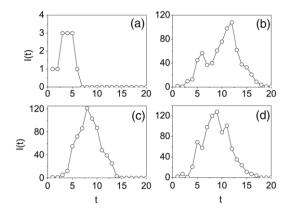


Fig. 3. Four specific evolution processes for parameters N = 3000, $\lambda = 0.2$, p = 0.9, q = 0.9 and different initial conditions.

At the end of the infection process, we have $t = \tau$ and $d\phi(t)/dt = 0$, yielding

$$\phi(\tau) = 1 - \frac{\sum k' P(k') e^{-\lambda k' f(C)\phi(\tau)}}{\langle k \rangle}.$$
(13)

The total epidemic prevalence is

$$R \equiv \sum P(k)r_k(\tau) = 1 - \sum P(k)e^{-\lambda kf(C)\phi(\tau)}.$$
(14)

From Fig. 1(b) one can see that dP(k)/dp > 0 for k < 5 and dP(k)/dp < 0 for k > 5. This can also be obtained from Eq. (6). Combining this result with the exponential decay of $e^{-\lambda k f(C)\phi(\tau)}$ with k, we have the result that the second term in Eq. (14) increases with p, which gives

$$\frac{\mathrm{d}R}{\mathrm{d}p} < 0. \tag{15}$$

Recalling that $C \sim q$ in Eq. (7), then from Eq. (14) one can easily get that

$$\frac{\mathrm{d}R}{\mathrm{d}q} < 0. \tag{16}$$

Hence, the prevalence R is a monotonically decreasing function of both p and q.

For confirming the predictions (15) and (16), let us make numerical simulations. Remember that Eq. (14) is based on a mean-field approximation; thus meaningful results should be averages over different configurations and different initial conditions. For a specific realization, the number of final infected nodes depends on the initial condition. This property has been confirmed by our numerical simulations; see Fig. 3 for four specific evolution processes where I(t) denotes the total of infected nodes at time t. Obviously, Fig. 3(b) and (d) reflect the community structure where the multiple peaks come from the epidemic spreading in different communities.

The integration of I(t) gives the final infected nodes N_r in a realization. From Fig. 3 it is easy to see that N_r depends sensitively on the concrete processes or initial conditions. We also notice that the epidemic cannot be spread out for a quite finite part of initial conditions when λ is not large, such as $\lambda = 0.2$. Thus, a meaningful result should be an average over a large number of realizations. Let us see how the average $\langle I(t) \rangle$ changes with time t for different structures. Fig. 4(a) shows the results for p = 0.9 where the "circles" and "triangles" represent the cases of q = 0.9 and 0.6, respectively. Obviously, there is a finite difference between the cases of q = 0.9 and 0.6. For confirming the influence of structure parameters p and q on the epidemic spread, we let $\langle N_r \rangle = \sum_{t=1}^{\tau} \langle I(t) \rangle$. As $\langle N_r \rangle = NR$, we here use $\langle N_r \rangle$ to replace R to check Eqs. (15) and (16). Fig. 4(b) and (c) show how $\langle N_r \rangle$ changes with q and p, respectively. Obviously, $\langle N_r \rangle$ decreases with both p and q, confirming the theoretical predictions (15) and (16). Moreover, comparing Fig. 4(b) with (c) we can see that p has a larger influence on $\langle N_r \rangle$ than q, indicating that the efficiency of epidemic spreading depends mainly on the degree of community.

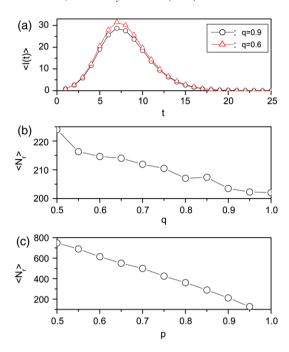


Fig. 4. (Color online.) How the structure parameters p and q influence the epidemic prevalence for N=3000 and $\lambda=0.2$. (a) $\langle I(t)\rangle$ versus t for p=0.9 where the "circles" and "triangles" represent the cases of q=0.9 and 0.6, respectively; (b) $\langle N_r \rangle$ versus q for p=0.9; and (c) $\langle N_r \rangle$ versus p for q=0.9. The results are averaged over 10^4 realizations.

4. Discussion and conclusions

In the community network, the links between groups are sparse but important. These links make the local clusters have more chance of contacting other clusters, i.e., they behave as liaison agents. Thus, they make the network become less heterogeneous and this results in larger epidemic prevalence. This result is consistent with the previous result that a homogeneous network is more efficient than a heterogeneous network in spreading an epidemic [21,22].

In conclusion, we have constructed a community network model to study the influence of community structure on epidemic spreading. This model has an invariant degree distribution for different clustering coefficients and its degree of community and clustering coefficient can be continuously adjusted. We find that the epidemic spreading will decrease with both the parameters p and q, indicating that a high degree of community and high clustering coefficient are disadvantages for epidemic prevalence.

Acknowledgements

This work was supported by the NNSF of China under Grants No. 10475027 and No. 10635040, by NCET-05-0424, and by National Basic Research Program of China (973 Program) under Grant No. 2007CB814800.

References

- [1] R. Albert, A.-L. Barabasi, Rev. Modern Phys. 74 (2002) 47.
- [2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Phys. Rep. 424 (2006) 175.
- [3] M.E.J. Newman, SIAM Rev. 45 (2003) 167.
- [4] A.-L. Barabasi, R. Albert, Science 286 (1999) 509; Phys. Rev. Lett. 85 (2000) 5234.
- [5] P.L. Krapivsky, S. Redner, F. Leyvaz, Phys. Rev. Lett. 85 (2000) 4629.
- [6] S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Phys. Rev. Lett. 85 (2000) 4633.
- [7] G. Bianconi, A.-L. Barabasi, Europhys. Lett. 54 (2001) 436.
- [8] Z. Liu, Y.-C. Lai, N. Ye, P. Dasgupta, Phys. Lett. A 303 (2002) 337; Phys. Rev. E 66 (2002) 036112.
- [9] M.E.J. Newman, Phys. Rev. Lett. 89 (2002) 208701; Phys. Rev. E 67 (2003) 026126;
 M.E.J. Newman, J. Park, Phys. Rev. E 68 (2003) 036122.
- [10] M.E.J. Newman, M. Girvan, Phys. Rev. E 69 (2004) 026113.

- [11] E.M. Jin, M. Girvan, M.E.J. Newman, Phys. Rev. E 64 (2001) 046132.
- [12] G. Palla, I. Derenyi, I. Farkas, T. Vicsek, Nature 435 (2005) 814.
- [13] G. Palla, A.-L. Barabasi, T. Vicsek, Nature 446 (2007) 664.
- Z. Liu, B. Hu, Europhys. Lett. 72 (2005) 315;
 Y. Zhou, Z. Liu, J. Zhou, Chin. Phys. Lett. 24 (2007) 581.
- [15] J.D. Noh, H. Jeong, Y. Ahn, H. Jeong, Phys. Rev. E 71 (2005) 036131.
- [16] P. Holme, B.J. Kim, Phys. Rev. E 65 (2002) 026107.
- [17] R.M. Anderson, R.M. May, Infectious Diseases in Humans, Oxford University Press, Oxford, 1992.
- [18] R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. 86 (2000) 3200; Phys. Rev. E 65 (2002) 035108.
- [19] J. Joo, J.L. Lebowitz, Phys. Rev. E 69 (2004) 066105.
- [20] M. Barthelemy, A. Barrat, J. Theoret. Biol. 235 (2005) 275.
- [21] Z. Liu, Y.C. Lai, N. Ye, Phys. Rev. E 67 (2003) 031911.
- [22] D.H. Zanette, Phys. Rev. E 65 (2002) 041908; 64 (2001) 050901(R).
- [23] E.B. Naim, Phys. Rev. E 69 (2004) 050901(R).
- [24] Z.G. Shao, J.P. Sang, Physica A 351 (2005) 662.
- [25] L.A. Meyers, B. Pourbohloul, M.E.J. Newman, J. Theoret. Biol. 232 (2005) 71.
- [26] V.M. Eguiluz, K. Klemm, Phys. Rev. Lett. 89 (2002) 108701.
- [27] M. Boguna, R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. 90 (2003) 028701.
- [28] R. Olinky, L. Stone, Phys. Rev. E 70 (2004) 030902(R).
- [29] Y. Moreno, J.B. Gomez, A.F. Pacheco, Phys. Rev. E 68 (2003) 035103(R);
 - Y. Moreno, M. Nekovee, A. Vespignani, Phys. Rev. E 69 (2004) 055101(R);
 - Y. Moreno, R. Pastor-Satorras, A. Vespignani, Eur. Phys. J. B 26 (2002) 521;
 - H. Zhang, Z. Liu, W. Ma, Chin. Phys. Lett. 23 (2006) 1050.
- [30] D. Zheng, P. Hui, S. Trimper, B. Zheng, Physica A 352 (2005) 659.
- [31] R. Yang, B. Wang, J. Ren, W. Bai, Z. Shi, W. Wang, T. Zhou, Phys. Lett. A 364 (2007) 189.
- [32] P. Crepey, F.P. Alvarez, M. Barthelemy, Phys. Rev. E 73 (2006) 046131.
- [33] M. Kuperman, G. Abramson, Phys. Rev. Lett. 86 (2001) 2909.