PHYSICAL REVIEW E **73**, 035103(R) (2006)

Information propagation on modular networks

Liang Huang, Kwangho Park, and Ying-Cheng Lai
Department of Electrical Engineering, Arizona State University, Tempe, Arizona 85287, USA
(Received 21 October 2005; revised manuscript received 17 January 2006; published 16 March 2006)

Networks with a community (or modular) structure underlie many social and biological phenomena. In such a network individuals tend to form sparsely linked local communities, each having dense internal connections. We investigate the dynamics of information propagation on modular networks by using a three-state epidemic model with a unit spreading rate (i.e., the probability for a susceptible individual to be "infected" with the information is one). We find a surprising, resonancelike phenomenon: the information lifetime on the network can be maximized by the number of modules. The result can be useful for optimizing or controlling information spread on social or biological networks.

DOI: 10.1103/PhysRevE.73.035103 PACS number(s): 89.75.Hc, 87.23.Ge

Networks with a community structure, or modular networks, are relevant to many social and biological phenomena [1–6]. A modular network consists of a number of groups, where nodes within each group are densely connected, but the linkage among the groups is sparse. Such is indeed the case in many social networks, where individuals in a society tend to form groups according to their social characteristics. Within a group, each member is directly connected to most other members, but connections among different groups are relatively rare. Among the many outstanding problems concerning modular networks, the propagation of information, such as rumor, news, or facts, is of great interest.

Research on epidemic in networks started with the work of Sudbury [7] on completely random networks and has received increasing attention [8–22] after the discoveries of complex networks such as the small-world [23] and the scale-free [24] networks. The problem of epidemic deals with whether an initially localized seed infection can spread to a substantial part of the network [25,26]. The pioneering work by Pastor-Satorras and Vespignani [8] considered a two-state model, where nodes can be either susceptible (S) or infected (I). A susceptible node can become infected and an infected node can recover and return to the susceptible state—hence the *SIS* model. They found that for scale-free networks, there is no intrinsic epidemic threshold in the thermodynamic limit. The result was extended by Lloyd and May [9] to the *SIR* model, a three-state model where a node in the network can be in one of the three states: susceptible, infected, and refractory (R), and an infected node can become refractory and is no longer susceptible to the infection. For a comprehensive review on the epidemic in complex networks, see Refs. [18,22].

In this Rapid Communication we investigate the *SIR* dynamics on modular networks. Our interest is in information propagation, which may be particularly important for social networks. The information can be, for instance, rumor, news, or facts. In general, once an "ignorant" is contacted with a piece of information, there is a high probability that the individual will spread the information. In the *SIR* framework, a convenient way to model this situation is to set the spreading rate to be one, which is the probability that a susceptible node is infected when contacted. This situation also applies to a very virulent epidemic where a contacted individual is almost certainly infected. In the *SIR* framework, the work of Zanette [10] and Newman *et al.* [15] suggested the existence of a propagation threshold for small-world networks but modular networks typically possess the small-world feature [1–3]. Our focus here is on *for how long a piece of information on a modular network may last*. A key parameter characterizing a modular network is the *number of modules*, M. Intuitively, one would expect the information lifetime to increase with M. However, we find, surprisingly, that the lifetime can be maximized for a specific value of M. In particular, as M is increased initially, the time increases but reaches a maximum for some value of M, and then decreases as M is increased further, which is basically a *resonant phenomenon*. The implication can be quite striking: the information lifetime is relatively short for modular networks having either a small or a large number of modules. In the case of the spread of an extremely virulent disease in a human society, assuming the size of a city is proportional to the number of modules in the underlying social network, the epidemic may last long not for cities of small or large size, but for those of medium size. In the remainder of this Rapid Communication, we shall present analysis and numerical evidence to substantiate our finding.

We consider a modular network with $N \gg 1$ nodes and M modules, where $M \ll N$. Each module is thus a subnetwork of $n = N/M \gg 1$ nodes, which can be either scale-free, small-world, or random. For convenience, each module is assigned an integer, say, from 1 to M, and all modules are placed on a topological ring with the periodic boundary condition. For each pair of adjacent modules, one node is chosen randomly from each module and a link is added between the two nodes. At this stage all modules are connected through a next-neighbor type of links. Links of shortcut type are generated by randomly selecting pairs of modules of distance l apart along the ring according to the probability $P(l) \sim e^{-\alpha l}$ and linking them, where α is a control parameter. For $\alpha \sim 0$, random long-range links are highly probable, making the whole modular network small-world-like. Because the linkage among the modules is sparse comparing with the linkage within each module, and because for α not close to zero the links among modules are mostly local or diametri-

cal, large-scale propagation is more unlikely in these cases. It thus makes sense to focus attention on networks with $\alpha \sim 0$.

The *SIR* dynamics for information propagation is implemented on the network, as follows. Initially all nodes are susceptible. At $t=0$, a piece of information is generated at a randomly chosen node (seed). At the next time step, one of its neighbors is randomly picked up, and becomes "infected" with the information if it is susceptible; otherwise, the original infected node itself becomes refractory. This process continues until there is no longer any infected node in the network, and the time the whole process takes is the information lifetime T. The number of refractory nodes for $t \geqslant T$ thus represents the number of nodes in the network that have been infected.

For a piece of information to spread on a modular network, the number of links among the modules needs to be large. The minimally required average number of modular links can be estimated, as follows. First, recall that each module is effectively a subnetwork that can be random, small-world, or scale-free. For a random network, the mathematical theory of *SIR* dynamics with unit spread rate [7] indicates that the fraction of nodes that can be infected approaches a universal constant of about 0.8 as the number of nodes goes to infinity. For scale-free networks and more general networks that contain both random and scale-free components, there is numerical evidence that the fraction is slightly below 0.8 (Ref. [19]). In any case, given a network of reasonably large size, the fraction of nodes that can be infected under the *SIR* dynamics is approximately a constant $r_0 \lesssim 0.8$. Next, let k_M be the average number of shortcut type of links. Taking into account the next-neighbor type of connections between the modules and the fact that $n \gg k_M$, the average number of nodes in a module with links going outside is $k_e = 2k_M + 2$. Thus, within an infected module, the average number of such nodes that carry the information is $r_0 k_e$. Finally, let $\langle k \rangle$ be the average number of internal links per node in the subnetwork. If a node with an outgoing link is infected, the probability that the link is chosen to spread the information is $1/(\langle k \rangle + 1)$. Thus, on average, the number of nodes that carry the information *and* spread it to a different module is $r_0 k_e/(\langle k \rangle + 1)$. For information to spread over the entire network, we must have $r_0 k_e/(\langle k \rangle + 1) \geqslant 1$, yielding $k_M^{\min} = (\langle k \rangle + 1)/(2r_0) - 1$. For instance, if $\langle k \rangle = 10$, using $r_0 = 0.8$ we obtain $k_M^{\min} = 6$. This agrees quite well with numerics, as shown in Fig. 1, where the fraction of infected nodes on the whole modular network is calculated as a function of k_M, for six different values of N. Apparently, the fraction becomes substantial for $k_M \geqslant 6$, indicating a large-scale information spread on the modular network.

We now examine the dependence of the lifetime T on the number of modules. Figures 2(a) and 2(b) show, for a network of $N = 4 \times 10^4$ nodes, the fraction r of infected nodes and T vs M, respectively. The subnetwork in each module is random. We see that as M varies over 2 orders of magnitude (from 10 to 1000), r remains approximately constant (about 0.54). Since r is substantially above zero, a large-scale information spread on the network occurs. The surprising phenomenon is that the lifetime T, as shown in Fig. 2(b), is apparently nonmonotonic and in fact exhibits a bell-shape

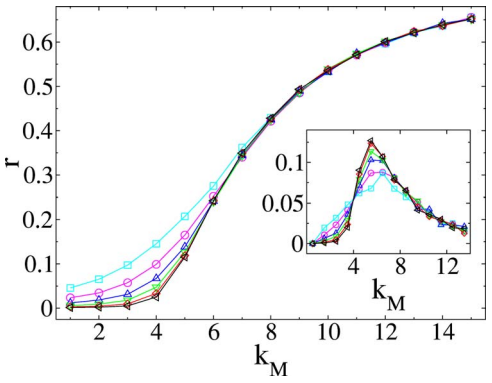

FIG. 1. (Color online) For a modular network with a ring topology, the fraction of infected nodes vs k_M. Parameters are $\alpha = 0$, $\langle k \rangle = 10$, and $n = 200$. The six curves correspond to a network size of $N = 5 \times 10^3$ (squares), 10^4 (circles), 2×10^4 (up triangles), 4×10^4 (down triangles), 8×10^4 (diamonds), and 16×10^4 (left triangles), respectively. Each data point is the result of averaging over 10^4 random realizations of the network. The inset shows dr/dk_M vs k_M, which is indicative of a continuous phase transition.

behavior. There exists a value of M for which the time reaches maximum, indicating a resonance-type of phenomenon. The phenomenon persists when each subnetwork is scale-free as shown in Figs. 2(c) and 2(d).

In Fig. 2, squares are the data for fixed $\langle k \rangle$ and k_M. As the number of module M is increased, the average degree of the network $\bar{k} = \langle k \rangle + 2(k_M + 1)M/N$ also increases. One may wonder whether the resonance is caused by this increase. To address this issue, we fix the value of $\bar{k}$ by reducing $(k_M + 1)M$ inner edges, i.e., edges that connect nodes in the same module, while keeping the network fully connected. The results for fixed $\bar{k}$ and k_M are also shown in Fig. 2 (circles). We see that the results for fixed $\langle k \rangle$ and for fixed $\bar{k}$ are essentially the same, indicating that the observed resonant phenomenon is not a numerical artifact, but more likely an intrinsic property of modular networks.

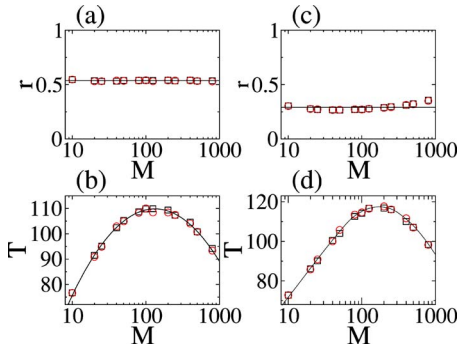

FIG. 2. (Color online) For modular network with a ring topology, (a) the fraction of infected nodes and (b) the information lifetime T vs the number of modules, for random subnetworks. Squares: $\langle k \rangle = 10$, the average degree $\bar{k}$ varies as M increases; circles: $\bar{k} = 10$. Other parameters are $\alpha = 0$, $N = 4 \times 10^4$, and $k_M = 10$. Each data point is the average over 10^4 random network realizations. (c) and (d) are the corresponding plots when the subnetwork within each module is scale-free.

To establish the credence and the generality of the observed resonant phenomenon, we seek a theoretical explanation. To gain insight, we consider the spread of information from a seed node on a two-dimensional square lattice. By the assumptions of the *SIR* dynamics, once a node is infected, it will become refractory or stay infected, and cannot be infected again. Assume that a node at the point $\mathbf{r}=(x,y)$ is infected at time $t-\tau$ from the node at $(x,y-a)$, where τ and a are the time step and lattice constant, respectively. At time t this newly infected node infects one of its nearest-neighbor nodes, if it is susceptible. If all nearest-neighbor nodes of the newly infected nodes are susceptible except for the node at $(x,y-a)$, the probability for any of these susceptible nodes to be infected at time $t+\tau$ is 1/3. Let $P(\mathbf{r},t)$ be the probability that a node at the point $\mathbf{r}$ is infected at time t. We have

$$P(\mathbf{r},t) = (1/3)[P(x+a,y,t+\tau) + P(x-a,y,t+\tau) + P(x,y+a,t+\tau)].$$

Subtracting $P(\mathbf{r},t+\tau)$ from both sides and dividing by τ, we get, in the continuum limit $a\to 0$ and $\tau\to 0$, $\partial P(\mathbf{r},t)/\partial t = D\partial^2 P(\mathbf{r},t)/\partial x^2 + \mu\partial P(\mathbf{r},t)/\partial y$, where $D=-a^2/3\tau$ and $\mu=-a/3\tau$. Since $|D|\ll|\mu|$, the diffusion term can be neglected, yielding $\partial P(\mathbf{r},t)/\partial t \approx \mu\partial P(\mathbf{r},t)/\partial y$. In this equation, the term on the right-hand side is derived by taking into account only the unidirectional spreading of the information along the y axis. Since unidirectional spreading can also occur in the x direction, the equation governing the propagation of infection in the two-dimensional lattice is

$$\frac{\partial P(\mathbf{r},t)}{\partial t} = \nu\left[\frac{\partial P(\mathbf{r},t)}{\partial x} + \frac{\partial P(\mathbf{r},t)}{\partial y}\right], \tag{1}$$

where ν is a constant. Equation (1) is invariant under the scaling transformation $\mathbf{r}=(x,y)\to l\mathbf{r}=(lx,ly)$, $t\to l^z t$, and $P(\mathbf{r},t)\to l^\alpha P(\mathbf{r},t)$, where l is a dilatation factor. Comparing all terms in Eq. (1) under the transformation, we have $z=1$. This means that, if the seed node is at $\mathbf{r}=0$ and $t=0$, there is a nonzero probability that a node at distance L will be infected at time T, where $T\sim L$. For a complex network, although we were not able to derive a similar equation, the basic dynamical process for infection spreading is the same. Since the relevant distance is the network diameter d, we expect the information lifetime to be proportional to d: $T\sim d$, which has been confirmed numerically for both random and scale-free networks, as shown in Fig. 3(a).

Now consider a modular network of M modules, where the subnetwork of n nodes within each module is either random or scale-free. The average network diameter of each subnetwork is of the order of $\ln n$ [27,28]. Assume that the links among the modules are randomly distributed. If each module is regarded as a node in a network, the network diameter is of the order of $\ln M$. For two randomly selected nodes in the modular network, on average their distance is of the order of $D_M(a_1+a_2\ln n)$, where D_M is the average number of modules that the shortest path between the two nodes passes, which is of the order of $\ln M$, and a_1 and a_2 are constants. The diameter of the modular network can thus be written as $d=(a_3+a_4\ln M)(a_1+a_2\ln n)$, where a_3 and a_4 are constants. Since $n=N/M$, we have

$$d = a + b\ln M + c(\ln M)^2, \tag{2}$$

where a, b, and c are constants. A numerical verification of Eq. (2) is shown in Fig. 3(b). The quadratic dependence of

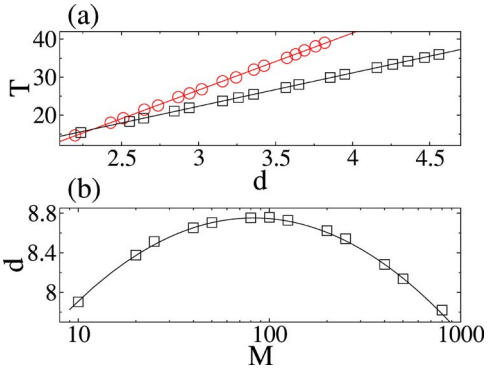

FIG. 3. (Color online) (a) Relation between information lifetime T and network diameter d for both random (lower trace) and scale-free (upper trace) networks. For both networks, the average degree is 10 and the network size varies from 100 to 2×10^4. Each value of d is obtained from 10 network realizations and each value of T is the average over 10^5 realizations. (b) For a ring modular network of $N=4\times 10^4$ nodes, the relation between the network diameter and the number of modules. Parameters are $\alpha=0$, $\langle k\rangle=10$, and $k_M=10$. Each data point is the average over 10 random network realizations. The solid curve is the theoretical fit.

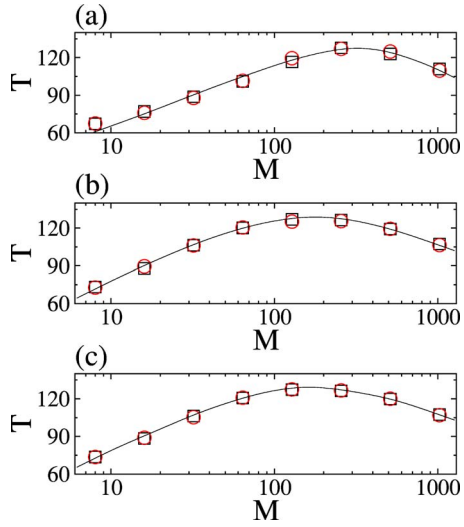

FIG. 4. (Color online) For Zachary networks of $N=52\,000$ nodes, the resonant behavior between the information lifetime and the number of modules for three cases: (a) scale-free subnetworks and preferential modular links, (b) random subnetworks and preferential modular links, and (c) random subnetworks and random modular links. In each subgraph, squares: $\langle k\rangle=10$, the average degree $\bar{k}$ varies as M increases; circles: $\bar{k}=10$. Other parameters are $\beta=0$ and $k_M=10$. Each data point is obtained by averaging the lifetime over 10^4 random network realizations.

the network diameter d on M, together with the linear relation between the information lifetime and the diameter, suggests a quadratic relation (resonant behavior) between the lifetime and the number of modules, as observed numerically.

Can the resonant behavior occur in more realistic modular networks? To address this question we have also studied another class of modular networks, the Zachary networks [29], which were originally proposed as a model of social networks. To construct a Zachary network of N nodes, we first divide all nodes into M modules, each having $n \gg 1$ nodes. Next, the modules are organized into levels, where each group in level 1 consists of two modules, and each level-2 group consists of two level-1 groups, and so on. Finally, random links among modules are added according to the probability $P(l) \sim e^{-\beta l}$, where $l \geqslant 0$ is the level distance between two random nodes in the network and $\beta \geqslant 0$ is a control parameter. In particular, a node (say, node i) is chosen randomly and a link is added between this node and another node from a *different* module (target module) according to $P(l)$. Once the target module is determined, the node (say, node j) in the module to which node i will connect is determined either randomly or by a preferential attachment rule within the target module. For the latter, the probability that node j is picked up is proportional to k_j, the number of links this node already has within the module. The process is repeated until the number of links among modules reaches the prescribed number $k_M M$. Implementing the *SIR* dynamics on the Zachary network, we have again observed the resonant phenomenon, as shown in Fig. 4, where squares are for fixed $\langle k \rangle$ and circles are for fixed $\bar{k}$. Note that there is essentially no difference between the two cases, suggesting that the resonant phenomenon is generic for modular networks.

In summary, our investigation of the *SIR* dynamics on complex, modular networks leads to the finding of an interesting resonance like phenomenon: the information lifetime typically exhibits a quadratic dependence on the number of modules. Thus, a piece of information will last shorter for networks having either a small number or a large number of modules. The same result holds for extremely virulent epidemics. In particular, our result may be useful for a social network where such an epidemic has just emerged. Knowing for how long the epidemic can potentially last can help in key decision making such as resources distribution in order to suppress the epidemic.

This work is supported by the NSF under Grant No. ITR-0312131 and by the AFOSR under Grant No. F49620-01-1-0317.

[1] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2002).
[2] E. Ravasz, A. L. Somera, D. A. Mongru, Z. Oltvai, and A.-L. Barabási, Science **297**, 1551 (2002).
[3] V. Spirin and L. A. Mirny, Proc. Natl. Acad. Sci. U.S.A. **100**, 12123 (2003).
[4] A. E. Motter, T. Nishikawa, and Y.-C. Lai, Phys. Rev. E **68**, 036105 (2003).
[5] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature (London) **435**, 814 (2005).
[6] E. Oh, K. Rho, H. Hong, and B. Kahng, Phys. Rev. E **72**, 047101 (2005).
[7] A. Sudbury, J. Appl. Probab. **22**, 443 (1985).
[8] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001); Phys. Rev. E **63**, 066117 (2001); Phys. Rev. E **65**, 035108(R) (2002).
[9] A. L. Lloyd and R. M. May, Science **292**, 1316 (2001); R. M. May and A. L. Lloyd, Phys. Rev. E **64**, 066112 (2001).
[10] D. H. Zanette, Phys. Rev. E **64**, 050901(R) (2001); **65**, 041908 (2002).
[11] M. Kuperman and G. Abramson, Phys. Rev. Lett. **86**, 2909 (2001).
[12] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 036104 (2002).
[13] Z. Dezsö and A.-L. Barabási, Phys. Rev. E **65**, 055103(R) (2002).
[14] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **26**, 521 (2002).
[15] M. E. J. Newman, I. Jensen, and R. M. Ziff, Phys. Rev. E **65**, 021904 (2002).
[16] V. M. Eguíluz and K. Klemm, Phys. Rev. Lett. **89**, 108701 (2002).
[17] R. Cohen, S. Havlin, and D. ben-Avraham, Phys. Rev. Lett. **91**, 247901 (2003).
[18] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[19] Z.-H. Liu, Y.-C. Lai, and N. Ye, Phys. Rev. E **67**, 031911 (2003).
[20] J. Balthrop, S. Forrest, M. E. J. Newman, and M. M. Williamson, Science **304**, 527 (2004).
[21] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **92**, 178701 (2004).
[22] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, UK, 2004).
[23] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[24] A.-L. Barabási and R. Albert, Science **286**, 509 (1999); A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[25] R. M. May and R. M. Anderson, Philos. Trans. R. Soc. London, Ser. B **321**, 565 (1988).
[26] R. M. Anderson and R. M. May, *Infectious Diseases of Humans* (Oxford University Press, Oxford, 1992).
[27] B. Bollobas, *Random Graphs* (Cambridge University Press, Cambridge, UK, 2001).
[28] B. Bollobas and O. Riordan, Combinatorica **24**, 5 (2004).
[29] W. W. Zachary, J. Anthropol. Res. **33**, 452 (1977).