New Journal of Physics

The small world yields the most effective information spreading

Linyuan Lü[1,2], **Duan-Bing Chen**[1] **and Tao Zhou**[1,3]

[1] Web Sciences Center, University of Electronic Science and Technology of China, 610054 Chengdu, People's Republic of China
[2] Department of Physics, University of Fribourg, Chemin du Musée 3, CH-1700 Fribourg, Switzerland
E-mail: zhutou@ustc.edu

Abstract. The spreading dynamics of information and diseases are usually analyzed by using a unified framework and analogous models. In this paper, we propose a model to emphasize the essential difference between information spreading and epidemic spreading, where the memory effects, the social reinforcement and the non-redundancy of contacts are taken into account. Under certain conditions, the information spreads faster and broader in regular networks than in random networks, which to some extent supports the recent experimental observation of spreading in online society (Centola D 2010 *Science* **329** 1194). At the same time, the simulation result indicates that the random networks tend to be favorable for effective spreading when the network size increases. This challenges the validity of the above-mentioned experiment for large-scale systems. More importantly, we show that the spreading effectiveness can be sharply enhanced by introducing a little randomness into the regular structure, namely the small-world networks yield the most effective information spreading. This work provides insights into the role of local clustering in information spreading.

[3] Author to whom any correspondence should be addressed.

IOP Institute of Physics ⦶ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Contents

1. Introduction

Understanding the dynamics of epidemic spreading is a long-term challenge and has attracted increasing attention in recent years. Firstly, the fast development of database technology and computational power has made more data available and analyzable to the scientific community. Secondly, many new objects of study have come into the horizon of epidemiologists: for example computer viruses, opinions, rumors, behavior, innovations, fads and so on. Lastly, in addition to the compartment model and population dynamics [1], novel models and tools appeared recently, inspired by empirical discoveries about network topology [2, 3], temporal regularities of human activities [4–6] and scaling laws in human mobility [7, 8].

In the simplest way, we can roughly divide the human-activated spreading dynamics into two classes according to the disseminules: one is the spreading of infectious diseases requiring physical contacts and the other is the spreading of information, including opinions, rumors and so on (here we mainly consider information whose value and authenticity need judgement and verification by individuals, in contrast to information about jobs, discounts, etc). In the early stage, scientists tried to describe these two classes by using a unified framework and analogous models (see, e.g., [9, 10]), emphasizing their homology and yet overlooking their essential differences. Very recently, scientists started to take into serious consideration the specific features of information spreading [11, 12], as well as the different mechanisms across different kinds of information [13]. Dodds and Watts [14] studied the effects of limited memory on contagion, yet did not consider the social reinforcement. Some recent works indicate that social reinforcement plays an important role in the propagation of opinions, news, innovations and fads [15–19].

In this paper, we propose a variant of the susceptible–infected–recovered (SIR) model for information spreading, which takes into account three different spreading rules from the standard SIR model: (i) memory effects, (ii) social reinforcement and (iii) non-redundancy of contacts. The main contributions are twofold. Firstly, we show that when the spreading rate λ is smaller than a certain value λ^*, the information spreads more effectively in regular networks than in random networks, which to some extent supports the experiment reported by Centola [20]: behavior spreads faster and can infect more people in a regular online social network than in a random one (with not more than 200 people in the experiment). We further show that with the increasing of the network size, the value of λ^* will decrease, which challenges the validity of Centola's experiment [20] for very-large-scale networks. Secondly, the effectiveness of information spreading can be remarkably enhanced by introducing a little randomness into the regular structure, namely the small-world networks [21] yield the most effective information

spreading. This result is complementary to the traditional understanding of epidemic spreading on networks where infectious diseases spread faster in random networks than in small-world networks.

2. Model

Consider a network with N nodes and E links representing the individuals and their interactions, respectively. Hereinafter, for convenience, we use the language of news spreading, but our model can be applied to the spreading of many kinds of information such as rumors and opinions, not limited to news. At each time step, each individual adopts one of four states: (i) *Unknown*—the individual has not yet heard the news, analogous to the susceptible state of the SIR model. (ii) *Known*—the individual is aware of the news but not willing to transmit it, because she is suspicious of the authenticity of the news. (iii) *Approved*—the individual approves the news and then transmits it to all her neighbors. (iv) *Exhausted*—after transmitting the news, the individual will lose interest and never transmit this news again, analogous to the recovered state in the SIR model.

At the beginning, one node is randomly chosen as the 'seed' and all others are in the unknown state. This seed node will transmit the news to all her neighbors and then become exhausted. Once an individual (in either an unknown or a known state) receives news, she will judge whether it is true depending on the number of times she has heard it—news or a rumor is more likely to be approved if heard many times (a very recent model allows the infectivity and/or susceptibility of hosts to be dependent on the number of infected neighbors [22]). The present rules imply two features of information spreading, namely memory effects and social reinforcement, which are usually neglected in the standard SIR model and its variants for rumor propagation.

In our model, we assume that for a given individual if she receives the news at least once at the tth time step and she has received this news $m(t)$ times until time t ($m(t)$ is a cumulative number), the probability that she will approve it at time t is $P(m) = (\lambda - T)e^{-b(m-1)} + T$, where $\lambda = P(1)$ is the approving probability for the first receipt. $T \in (0, 1]$ is the upper bound of the probability indicating maximal approving probability. Here we do not consider the interest decay, and we assume that the time scale of news spreading is much faster than our memory decay. After approval, she will transmit the news to all her neighbors in the next time step and then become exhausted. If an individual, in either an unknown or a known state, does not receive any news in the tth time step, nothing will happen no matter how many times this individual has received the news. The memory effects are embodied by $m(t)$, which is a cumulative number instead of the independent spreading rates for different contacts in the standard SIR model. With increasing m, $P(m)$ will infinitely approach T and the speed is determined by the parameter $b > 0$, which reflects the social reinforcement effect. Figure 1 shows the approving probability as a function of m, given different b. Larger b indicates stronger social reinforcement. For example, $P(2)$ is equal to 0.227 when $b = 0.2$ and is equal to 0.486 when $b = 0.8$. Since an individual who has transmitted the news will immediately become exhausted, our model ensures that each link is used at most once without any redundancy of contacts. The spreading process comes to an end when no new individual approves the news and spreads it.

We perform our model on three kinds of networks with an identical node degree k. (i) *Regular networks*. A regular network is a one-dimensional ordered network with periodic boundary conditions, where each node is connected to its k nearest neighbors, namely to the
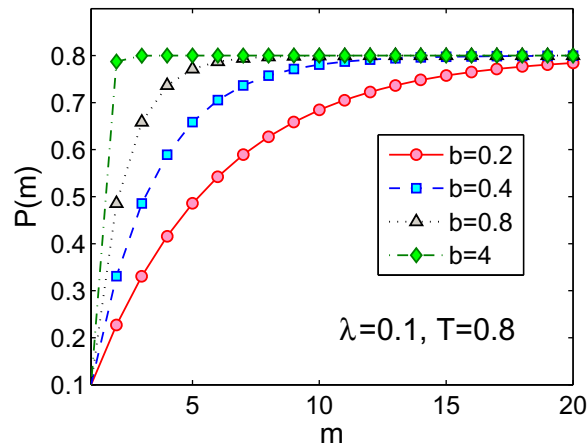
Figure 1. The approving probability as a function of m.

$k/2$ nearest neighbors clockwise and counterclockwise [21]. Note that, in the literature on graph theory, the term 'regular networks' usually stands for networks whose nodes are of the same degree and thus the following homogeneous small-world networks are also regular. In this paper, we follow the literature on complex networks and use the term 'regular networks' to represent networks with ordered structure. (ii) *Homogeneous small-world networks*. The homogeneous small-world network is constructed by randomly reshuffling links of a regular network, while keeping the degree of each node unchanged [23]. According to the link exchanging method [24], at each time step, we randomly select a pair of edges A–B and C–D. These two edges are then rewired to be A–D and B–C. To prevent multiple edges connecting the same pair of nodes, if A–D or B–C already exists in the network, this step is aborted and a new pair of edges is randomly selected. We implement pE steps, where p indicates the randomness of the network. (iii) *Random networks*. Repeating the above rewiring operations many times leads to a homogeneous random network. Theoretically speaking, a homogeneous random network is obtained only for $p \rightarrow \infty$; we here consider $p \in [0, 10]$ and when $p > 1$, the topological statistics are very close to those of random networks. In all simulations, the node degree is set to $k = 6$, and we have carefully checked that the results are not sensitive to the node degree unless k is very large or very small.

3. Results

We denote by R the number of approved nodes of the news. Larger R at the final state indicates broader spreading. We first compare the spreading processes on regular and random networks. Figure 2 reports four typical examples with different λ values and fixed $b = 0.8$. Surprisingly, for small λ (e.g. figure 2(a)), the spreading on regular networks is faster and broader than that on random networks. These results are in accordance with the online social experiment of Centola [20] and yet against the traditional understanding of network spreading [28]. With the increasing of λ, the random networks will be favorable for faster and broader spreading. Figure 3 shows the dependence of the number of approved nodes at the final state on the parameter λ. There is a crossing point at about $\lambda_c \approx 0.145$, after which R of random networks exceeds that of regular networks. The inset shows the difference between the number of final approved nodes
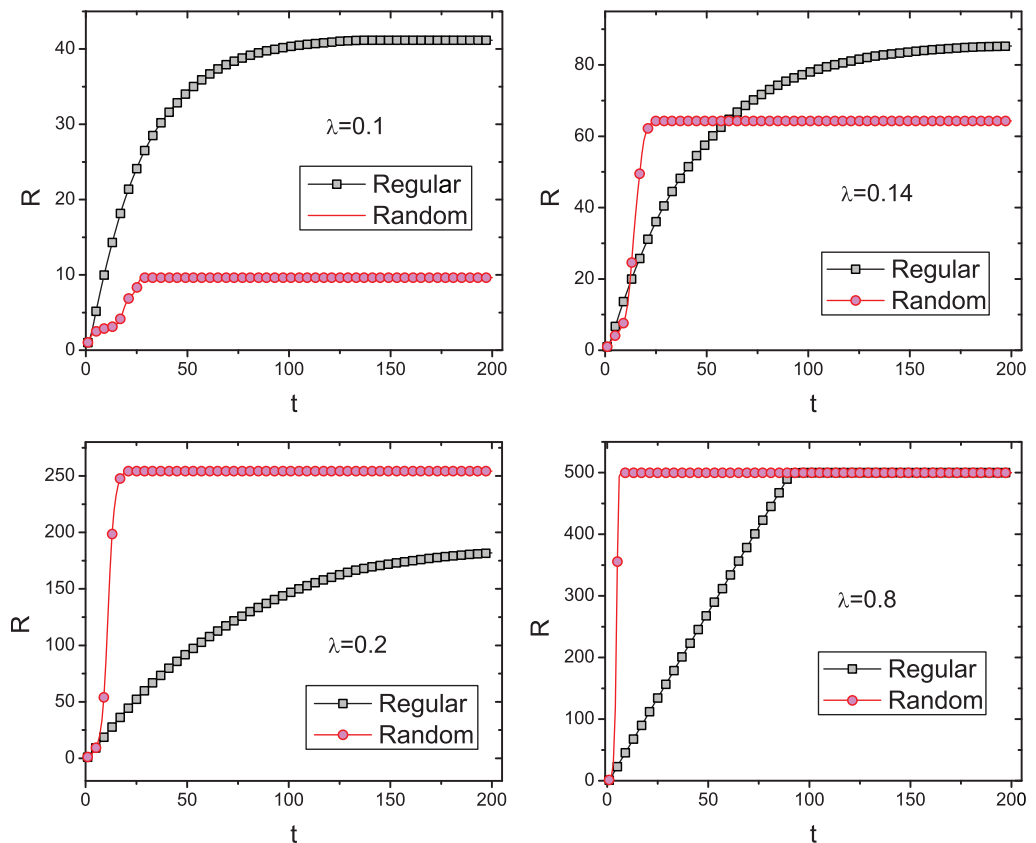
Figure 2. The number of approved nodes as a function of time on a regular network (black squares) and on a random network (red circles). The parameters are $N = 500$, $k = 6$, $b = 0.8$ and $T = 1$. The results are obtained by averaging over 500 independent realizations.

on regular and random networks, namely $R^{\mathrm{re}} - R^{\mathrm{ra}}$ against λ. With very large λ, almost every node will run into the approved state and thus R is not sensitive to the network structure, but the spread on random networks is still faster than that on regular networks (see, e.g., figure 2(d)).

Figure 4 displays the crossing point λ_c as a function of the network size N. When N is small, λ_c decreases sharply with increasing N, whereas as N gets larger, λ_c becomes insensitive to N. As a whole, λ_c shows non-increasing behavior versus N. Note that the phenomenon that spreading on regular networks is faster and broader than that on random networks will be more remarkable and easier to observe if λ_c is large. Therefore, our result about $\lambda_c(N)$ indicates that for large-scale systems, Centola's experimental results may not hold or will be weakened to some extent.

In a previous study on the SIR model, it was pointed out that the number of recovered nodes at the end of evolution increases with increasing randomness p in small-world networks [29]. In contrast, our simulations show that the number of approved nodes in the final state does not monotonically increase with increasing p; instead, an optimal randomness p^* exists subject to the highest R. Figure 5 shows the dependence of the number of final approved nodes on the randomness p given $b = 0$ (triangles), $b = 0.4$ (squares) and $b = 0.8$ (circles). With strong social reinforcement, even very small randomness can bring a remarkable improvement in the number of final approved nodes, R. Take the case $b = 0.8$ for example; on the regular networks (i.e.
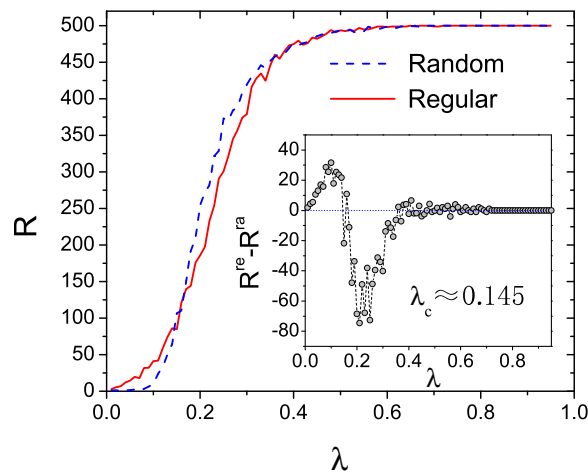
Figure 3. The dependence of the number of approved nodes at the final state on the parameter λ for regular (red solid line) and random (blue dash line) networks. The parameters are $N = 500$, $k = 6$, $b = 0.8$ and $T = 1$, the same as those for figure 2. The inset shows the number of final approved nodes on regular networks R^{re} minus that on random networks R^{ra}, against λ. The results are obtained by averaging over 500 independent realizations.
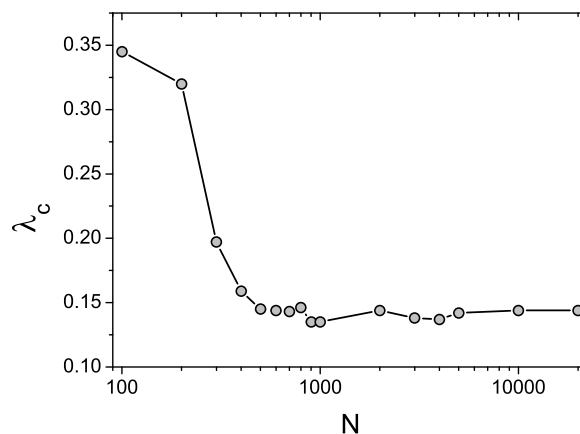

Figure 4. The dependence of λ_c on the network size N. The parameters are $k = 6$, $b = 0.8$ and $T = 1$. The results are obtained by averaging over 500 independent realizations.

$p = 0$), R is 205, whereas by introducing a tiny randomness $p = 0.02$, this number will suddenly increase to 6593, which is also higher than for the random networks (i.e. $p = 1$, $R = 4049$). We also plot the clustering coefficient C as a function of p in figure 5. As expected, C decreases with increasing p. The results indicate that local clustering can to some extent enhance the approving rate of information, which refines the completely negative valuation of the clustering coefficient in epidemic spreading [25–27].

The dependence of optimal randomness p^* on the strength of social reinforcement b given different N values is shown in figure 6, where one can observe that stronger social reinforcement (i.e. larger b) results in smaller p^*. In the presence of weak social reinforcement (i.e. small b),
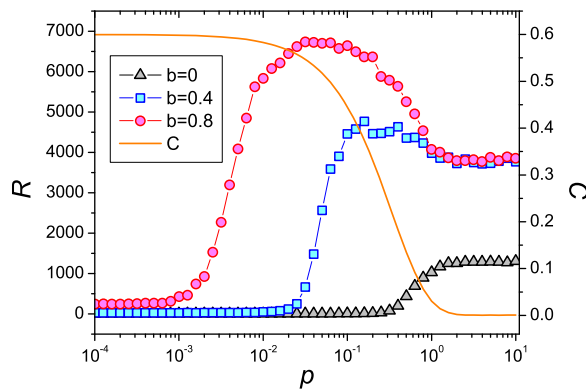
Figure 5. The number of final approved nodes against the randomness p given $b = 0$ (triangles), $b = 0.4$ (squares) and $b = 0.8$ (circles). Other parameters are $N = 10^4$, $k = 6$, $\lambda = 0.2$ and $T = 1$. The results are obtained by averaging over 10 000 independent realizations. The clustering coefficient C, as a monotonic function of p, is also displayed.
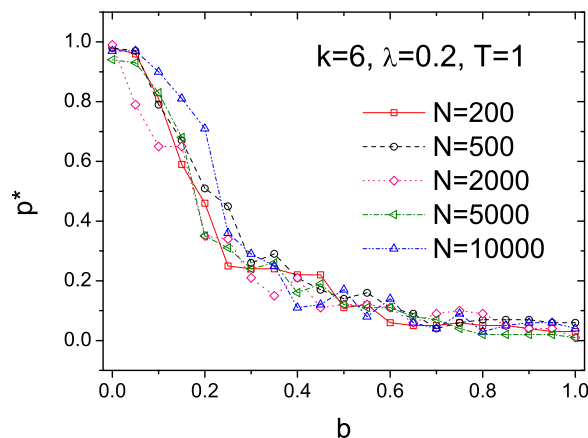

Figure 6. The dependence of optimal randomness p^* on the strength of social reinforcement b given different N values. The results are obtained by averaging over 500 independent realizations.

our result (p^* is close to 1) is analogous to the well-known one [28, 29] that the speed and range of spreading obey the relation 'Random > Small-World > Regular'. In contrast, the small-world networks yield the most effective spreading when social reinforcement plays an important role (i.e. large b).

To further investigate the advantages of small-world networks for information spreading, we calculate the complementary cumulative distribution $p(R > R_c)$, namely the probability that in a realization the information has reached more than R_c individuals. As shown in figure 7, the advantages of small-world networks are twofold. Compared with random networks, there is higher probability to spread out (see the region when R_c is small). For example, in small-world networks, $p(R > 10) = 0.703$, whereas for random networks, this number is only 0.460. If the information can spread out, like an epidemic disease, then in both kinds of networks it can reach a majority of the population. Comparing with regular networks, information in small-world
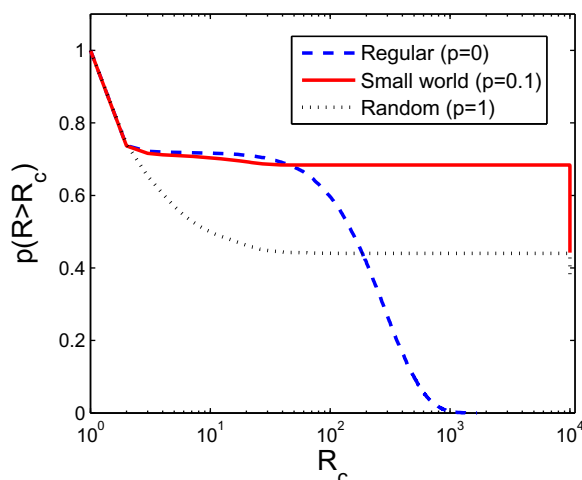
IOP Institute of Physics Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Figure 7. The cumulative probability that in a realization, the information reaches more than R_c individuals for regular, random and small-world networks. The parameters are $\lambda = 0.2$, $k = 6$, $T = 1$ and $N = 10\,000$. These distributions are obtained from 10 000 realizations.

networks can spread wider. According to figure 7, the maximum R in regular networks is only 1680, while in small-world networks it can reach 9900 individuals with a probability of 0.684.

4. Conclusion and discussion

Thanks to the fast development of database technology and computational power, detailed analysis of information spreading in large-scale online systems has become feasible nowadays. In our opinion, the similarity between information spreading and epidemic spreading is overemphasized in previous studies (see, e.g., the models summarized in the review article [30]), and currently we should turn to the other side of the matter: revealing the essential difference between them. The significant difference may include: (i) *Time decaying effects*. An infectious disease can exist for more than thousands of years in human society and still remain active, but no one is willing to spread news one year old. Actually, our attention to information decays very fast [31], and thus when we model information spreading, especially if it involves multiple information competing for attention, we have to consider the time decaying effects. (ii) *Tie strength*. It is well known that in social networks, ties with different strengths play different roles in maintaining the network connectivity [32], information filtering [33], information spreading [34] and so on. We guess that the weak ties provide faster paths for information spreading, while the strong ties provide trust paths (i.e. with high infectivity). However, this point is still unclear to date. (iii) *Information content*. Information with different contents may have very different spreading paths, and even with the same content, different expressions may lead to very different performances. Some of them are born with fashionable features, while others are doomed to be kept from being known. Whether these two kinds of information are different only quantitatively or they follow qualitatively different dynamic patterns is still under investigation [35]. (iv) *Role of spreaders*. A recent analysis on Twitter shows that different kinds of spreaders, such as media, celebrities, bloggers and formal organizations, play remarkably different roles in network construction and information

IOP Institute of Physics **Φ** DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

spreading [36], which may result in different spreading paths and outbreaking mechanisms from epidemic spreading. (v) *Memory effects*. Previous contacts could impact the information spreading in current time [14]. Such memory effects can be direct, since an agent may tend to be interested in or disgusted with objects heard many times, and/or indirect since previous contacts could change the tie strength that further impacts the current interactions. (vi) *Social reinforcement*. If more than one neighbor approved the information and transferred it to you, you have high probability of approving it. Generally speaking, if an agent twice receives an information item recommended by her neighbors, the approval probability should be much larger than twice the approval probability with a single recommending. (vii) *Non-redundancy of contacts*. People usually do not transfer an information item more than once to the same guy, which is very different from sexually transmitted diseases, to name one example.

In this paper, we propose a simple model for information spreading in social networks that considers the memory effects, social reinforcement and non-redundancy of contacts. Under certain conditions, the information spreads faster and broader in regular networks than in random networks, which to some extent supports Centola's experiment [20]. At the same time, we show that random networks tend to be favorable for effective spreading when the network size increases, which challenges the validity of Centola's experiment for large-scale systems. Furthermore, simulation results suggest that by introducing a little randomness into the regular structure, the small-world networks yield the most effective information spreading. Although this simple model cannot take into account all the above-mentioned features of information spreading, it largely refines our understanding of spreading dynamics. For example, traditional spreading models on complex networks show that diseases spread faster and wider in random networks than in small-world networks [28, 29], yet our results suggest that the small world may be the best structure for effective spreading under the consideration of social reinforcement. Indeed, information in small-world networks has much higher probability to spread out than in random networks, and can spread much broader than in regular networks. In addition, local clustering is well known to play a negative role in spreading [25–27], while our model indicates that local clustering is very helpful in facilitating the acceptance/approval of information for individuals and thus can, to some extent, speed up the spreading.

Acknowledgments

References

[1] May R L and Anderson R M 1991 *Infectious Diseases of Humans: Dynamics and Control* (Oxford: Oxford University Press)
[2] Pastor-Satorras R and Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200
[3] Barrat A, Barthelemy M and Vespignani A 2008 *Dynamical Processes on Complex Networks* (Cambridge: Cambridge University Press)
[4] Vazquez A, Rácz B, Lukács A and Barabási A-L 2007 *Phys. Rev. Lett.* **98** 158702
[5] Iribarren J L and Moro E 2009 *Phys. Rev. Lett.* **103** 038702

IOP Institute of Physics ❶ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

[6] Yang Z, Cui A-X and Zhou T 2011 *Physica* A **390** 4543

[7] Wang P, González M C, Hidalgo C A and Barabási A-L 2009 *Science* **324** 1071

[8] Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco J J and Vespignani A 2009 *Proc. Natl Acad. Sci. USA* **106** 21484

[9] Liu Z, Lai Y-C and Ye N 2003 *Phys. Rev.* E **67** 031911

[10] Moreno Y, Nekovee M and Pacheco A F 2004 *Phys. Rev.* E **69** 066130

[11] Hill A L, Rand D G, Nowak M A and Christakis N A 2010 *PLoS Comput. Biol.* **6** e1000968

[12] House T 2011 *J. R. Soc. Interface* **8** 909

[13] Romero D M, Meeder B and Kleinberg J 2011 *Proc. 20th Int. Conf. WWW* (New York: ACM)

[14] Dodds P S and Watts D J 2004 *Phys. Rev. Lett.* **92** 218701

[15] Centola D, Eguíluz V M and Macy M W 2007 *Physica* A **374** 449

[16] Medo M, Zhang Y-C and Zhou T 2009 *Eur. Phys. Lett.* **88** 38005

[17] Cimini G, Medo M, Zhou T, Wei D and Zhang Y-C 2011 *Eur. Phys. J.* B **80** 201

[18] Wei D, Zhou T, Cimini G, Wu P, Liu W and Zhang Y-C 2011 *Physica* A **390** 2117

[19] Prapivsky P L, Redner S and Volovik D 2011 arXiv:1104.4107

[20] Centola D 2010 *Science* **329** 1194

[21] Watts D J and Strogatz S H 1998 *Nature* **393** 440

[22] Pérez-Reche F J, Ludlam J J, Taraskin S N and Gilligan C A 2011 *Phys. Rev. Lett.* **106** 218701

[23] Santos F C, Rodrigues J F and Pacheco J M 2005 *Phys. Rev.* E **72** 056128

[24] Maslov S and Sneppen K 2002 *Science* **296** 910

[25] Eguíluz V M and Klemm K 2002 *Phys. Rev. Lett.* **89** 108701

[26] Petermann T and De Los Rios P 2004 *Phys. Rev.* E **69** 066116

[27] Zhou T, Yan G and Wang B-H 2005 *Phys. Rev.* E **71** 046141

[28] Zhou T, Fu Z-Q and Wang B-H 2006 *Prog. Nat. Sci.* **16** 452

[29] Zanette D H 2001 *Phys. Rev.* E **64** 050901

[30] Castellan C, Fortunato S and Loreto V 2009 *Rev. Mod. Phys.* **81** 591

[31] Wu F and Huberman B A 2007 *Proc. Natl Acad. Sci. USA* **104** 17599

[32] Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kerész J and Barabási A-L 2007 *Proc. Natl Acad. Sci. USA* **104** 7332

[33] Lü L and Zhou T 2010 *Eur. Phys. Lett.* **89** 18001

[34] Miritello G, Moro E and Lara R 2011 *Phys. Rev.* E **83** 045102

[35] Crane R and Sornette D 2008 *Proc. Natl Acad. Sci. USA* **105** 15649

[36] Wu S, Hofman J M, Mason W A and Watts D J 2011 *Proc. 20th Int. Conf. WWW* (New York: ACM)