Strong ties promote the epidemic prevalence in susceptible-infected-susceptible spreading dynamics

Ai-Xiang Cui, Zimo Yang, and Tao Zhou

Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, People’s Republic of China

(Dated: November 26, 2013)

Understanding spreading dynamics will benefit society as a whole in better preventing and controlling diseases, as well as facilitating the socially responsible information while depressing destructive rumors. In network-based spreading dynamics, edges with different weights may play far different roles: a friend from afar usually brings novel stories, and an intimate relationship is highly risky for a flu epidemic. In this article, we propose a weighted susceptible-infected-susceptible model on complex networks, where the weight of an edge is defined by the topological proximity of the two associated nodes. Each infected individual is allowed to select limited number of neighbors to contact, and a tunable parameter is introduced to control the preference to contact through high-weight or low-weight edges. Experimental results on six real networks show that the epidemic prevalence can be largely promoted when strong ties are favored in the spreading process. By comparing with two statistical null models respectively with randomized topology and randomly redistributed weights, we show that the distribution pattern of weights, rather than the topology, mainly contributes to the experimental observations. Further analysis suggests that the weight-weight correlation strongly affects the results: high-weight edges are more significant in keeping high epidemic prevalence when the weight-weight correlation is present.

PACS numbers: 87.95.-k, 87.23.Ge, 89.65.-s, 05.10.-a

I. INTRODUCTION

Early before the classification of social ties proposed, in 1954, the Russian mathematical psychologist Anatol Rapoport[1] has been aware of “well-known fact that the likely contacts of two individuals who are closely acquainted tend to be more overlapping than those of two arbitrarily selected individuals”. This argument became one cornerstone of social network theory. In 1973, ties in social networks, generally, come in two varieties: strong and weak, which has been first proposed by the American sociologist Mark Granovetter[2]. Different relationships can be measured in the currency of tie strength. According to the closeness, connections to close friends have been said to be “strong” ties, while those to acquaintances have been called “weak” ties[2,3].

Strong ties connect with the people you really trust, people whose social circles tightly overlap with your own. Often, they are also the people most like you. Weak ties, conversely, connect with merely acquaintances and often provide access to novel information. Tie strength usually plays an vital role in many real networks and is crucial to understand dynamical processes on the networks[6,7]. Weak ties display an important bridging function[8,9], while strong ties are more likely to activate the flow of referral information and more influential than weak ties[10,11]. In addition, weak ties could play a more significant role than strong ties to keep the stability[12], maintain the connectivity[13] and uncover the missing information[14,15], while strong ties can be better utilized to enhance the human resource flexibility[16], to provide accurate recommendation[17,18], and so on.

In despite of the qualitative distinction between strong and weak ties, tie strength could be quantitatively described by edge weight—the edges with high weights are considered to be strong. In a number of social networks, edges are often associated with weights that differentiate them in terms of their strength, intensity, capacity or the frequency of recent contacts[2,20]. For non-social networks, weights often refer to the functions performed by edges, e.g., the amount of traffic flowing along connections in world-wide airport networks[21], the number of joint papers of two coauthors in scientific collaboration networks[22], the number of synapses and gap junctions in neural networks[23], the carbon flow between species in food webs[24].

Weight plays a significant role in disparate network-based dynamics, such as transportation[23,24], synchronization[25,27], percolation[28,29], and so on. In this article, we concentrate on the effects of weights on epidemic spreading. Yan et al.[30] investigated the epidemic spreading in weighted scale-free networks and the simulation results indicated that the more homogeneous weight distribution of the network is, the more quickly epidemic spreads on it. This finding was further demonstrated by an edge-based mean-field solution[31]. Chu et al.[32] showed that weight distribution has strong impacts on both epidemic threshold and prevalence. Karsai et al.[33] considered the contact process in weighted scale-free networks, in which the weight of an edge connecting two high-degree nodes is relatively small. Yang et al.[34] further proved that in the contact process, the epidemic prevalence can be maximized by setting the edge weight inversely proportional to the degree of the receiving node. Baranchelli and Pastor-Satorras...
considered the diffusive dynamics on weighted networks and shed light on the validity of mean-field theory on weighted networks. Rattana et al. proposed a pairwise-type approximation for epidemic dynamics on weighted networks, showing a more accurate solution than traditional methods.

In this article, we propose a weighted susceptible-infected-susceptible (SIS) epidemic spreading model, with a tunable parameter controlling the preference of spreading: whether or not an infected individual prefers to contact others through edges with high weights. Experimental results on six real networks show that the preferential contacts through strong ties could largely improve the epidemic prevalence. We compare such results with two statistical null models, where the topological structure and weight distribution are randomized respectively. The aforementioned strong ties effects is qualitatively the same under randomized topology while is vanished if the weights are randomly redistributed, indicating that the distribution pattern of weights mainly contributes to the experimental observations. Further analysis suggests that the weight-weight correlation strongly affects the results, with the optimal value of the controlling parameter monotonously depending on the correlation strength.

This article is organized as follows. In Section II, we present the details of the SIS model. The data description, together with simple statistics, are shown in Section III. Section IV reports the experimental results and Section V provides theoretical insights by comparing the experimental results with statistical null models. The main conclusions are summarized in Section VI.

**II. MODEL**

The SIS model is suitable to describe the cases when individuals cannot acquire immunity after recovering from the disease, such as influenza, pulmonary tuberculosis and gonorrhea. With disease of this kind individuals that are cured usually catch again. In the networked SIS model, nodes are in two discrete states, “susceptible” or “infected”. Each infected node will contact all its neighbors once at each time step, and therefore the infectivity of each node is proportional to its degree. In the real world, individuals may be only able to contact limited population within one time step. For example, salesmen in network marketing processes will not make referrals to all his acquaintances due to the limited money and time. In sexual contact networks, although a few individuals have hundreds of sexual partners, their sexual activities are not far beyond a normal level due to the physiological limitations. Therefore, in the present model we assume every individual has the same infectivity. Without the lose of generality, at each time step, each infected node will select one of its neighbors to contact. If the selected neighbor has been infected already, nothing happens, otherwise it will be infected with probability \(\alpha\). Meanwhile, each infected node will become susceptible in the next time step with probability \(\beta\). Since each infected individual only selects one neighbor at each time step, the disease can spread out only when \(\alpha > \beta\). In the following study, we fix \(\alpha = 0.4\) and \(\beta = 0.1\), and we have already checked that the specific choices of \(\alpha\) and \(\beta\) will not change the qualitative results reported in this article.

The probability that an infected node \(i\) will select its neighbor \(j\) is

\[
P_{ij} = \frac{s_{ij}^b}{\sum_{l} s_{il}^b},
\]

where \(\Gamma_i\) is the set of neighbors of node \(i\), \(s_{ij}\) denotes the tie strength between \(i\) and \(j\), and \(b\) is a tunable parameter. If \(b = 0\), the infected node randomly selects a neighbor to contact, equivalent to an unweighted SIS

**TABLE I. Basic structural features of the largest connected components of the studied six real networks.**

| Networks   | \(N\)  | \(E\)  | \(k_{max}\) | \(\langle k \rangle\) | \(d\)    | \(C\)    | \(r_D\)  |
|-----------|--------|--------|-------------|-----------------|---------|---------|---------|
| FSN       | 1893   | 13835  | 255         | 14.6            | 3.06    | 0.11    | -0.188  |
| EEN       | 33696  | 180811 | 1383        | 10.7            | 4.03    | 0.509   | -0.116  |
| SSN       | 74444  | 382456 | 2517        | 10.3            | 4.21    | 0.06    | -0.067  |
| ESN       | 75868  | 405729 | 3044        | 10.7            | 4.31    | 0.138   | -0.041  |
| GPN       | 8842   | 31837  | 88          | 7.2             | 4.60    | 0.007   | 0.015   |
| OAS       | 10670  | 22002  | 2312        | 4.1             | 3.64    | 0.297   | -0.186  |

**FIG. 1.** (Color online) Spreading processes for different values of the parameter \(b\). These six plots show how the number of infected individuals, \(I(t)\), changes with time on the six real networks: (a) Facebook-like social network, (b) Enron email network, (c) Slashdot social network, (d) Epinions social network, (e) Gnutells peer-to-peer network, (f) Oregon autonomous systems. Results are obtained by averaging over 1000 independent realizations.
model \cite{43}. If \( b > 0 \), strong ties are favored to constitute the paths of spreading, while if \( b < 0 \), weak ties are favored.

In different contexts, the strength of a tie may have different definitions and measures \cite{13,44}, which may depend on external information to network topology. For general networks, one may not be aware of external information and thus it is meaningful to give a natural definition solely based on network topology. According to Rapoport’s theory \cite{1} and other supportive observations \cite{[15–17] and models \cite{48–50}, we define the tie strength \( s_{ij} \) in spite of \( i \) and \( j \)'s common neighbors, as follows:

\[
s_{ij} = \frac{n_{ij} + \delta}{\sqrt{k_i k_j}}. \tag{2}
\]

where \( n_{ij} \) is the number of common neighbors of \( i \) and \( j \), \( k_i \) is the degree of \( i \), and \( \delta \) is a constant that gives chance to the tie connecting two nodes without common neighbors. For simplicity, we set \( \delta = 1 \). The dynamical process starts with randomly selecting a certain number of infected nodes, we set it as 20. This initial number has no effect on the stable state.

### III. DATA

To see the impacts of tie strength on the spreading dynamics, experiments are carried out on six real networks: (1) Facebook-like Social Network (FSN) \cite{51}; it originates from an online community for students at University of California. The data set includes the users that have sent or received at least one message, and an undirected edge is set between \( i \) and \( j \) if \( i \) has sent (received) an online message to (from) \( j \). (2) Enron Email Network (EEN) \cite{52,53}; it covers about half million emails. Nodes of this network are email addresses and if an address \( i \) sent at least one email to address \( j \), an undirected edge is established between \( i \) and \( j \). (3) Slashdot Social Network (SSN) \cite{54}; nodes in this network are the users in Slashdot, which is a technology-related news website, and edges represent friendships between users. (4) Epinions Social Network (ESN) \cite{55}; this is a who-trust-whom online social network. Nodes are the members of the general consumer review site Epinions.com, and edges represent the trust relationships between two members. (5) Gnutella Peer-to-peer Network (GPN) \cite{56}: Nodes represent hosts and edges stand for connections between the hosts. (6) Oregon Autonomous Systems (OAS) \cite{57}; this is an AS-level Internet topology graph obtained by the Route Views Project. In order to guarantee the connectivity, we use the largest connected components of these networks. Table \ref{table:1} presents the basic statistics of the largest connected components of the six networks.

### IV. EXPERIMENTAL RESULTS

Figure 1 gives us an intuition of the spreading processes following the proposed model, showing how the number of infected nodes, \( I(t) \), changing with time, \( t \). Clearly, the epidemic dynamics is considerably affected by the parameter \( b \), in a complicated and non-monotonous way. Define the epidemic prevalence as the fraction of infected nodes in the stable state

\[
\rho = \frac{I(\infty)}{N}. \tag{3}
\]

Figure 2 displays how the epidemic prevalence \( \rho \) changes with \( b \) for the six real networks. For each case, there is a
model. Given a real network with degree sequence promoted when strong ties are favored in the spreading is promoted remarkably. Sometimes, it almost gets doubled case \( \rho \) underlying contributors for the nontrivial peaks in optimal value of \( \rho \).

FIG. 4. The epidemic prevalence \( \rho \) as a function of the parameter \( b \) based on the null model with randomly redistributed weights that correspond to (a) Facebook-like social network, (b) Enron email network, (c) Slashdot social network, (d) Epinions social network, (e) Gnutells peer-to-peer network, and (f) Oregon autonomous systems. Results are obtained by averaging over 1000 independent realizations.

well-defined peak for \( \rho(b) \). For convenience, we call the value of \( b \) corresponding to the largest prevalence \( \rho^* \) the optimal value of \( b \), denoted by \( b^* \).

As shown in figure 2, for all the six real networks, the optimal value \( b^* > 0 \). Comparing with the unweighted case \( b = 0 \), the epidemic prevalence for each real network is promoted remarkably. Sometimes, it almost gets doubled. In a word, the epidemic prevalence can be largely promoted when strong ties are favored in the spreading process.

V. ANALYSIS

To validate the significance of our findings and dig out underlying contributors for the nontrivial peaks in \( \rho(b) \) curves, we further compare our results with two statistical null models. The first one is called configuration model \( \text{CM} \). Given a real network with degree sequence \( \{k_1, k_2, \cdots, k_N\} \), where \( k_i \) is the degree of the \( i \)th node. In the configuration model, the \( i \)th node is assigned \( k_i \) stubs and the network is constructed by randomly choosing stubs and connecting them to form edges, avoiding multiple connections and self-connections. Detailed rules can be found in Ref. \( \text{CM} \). Under the configuration model, the topology, except for the degree heterogeneity, is randomized, while the correlation between local structure and weight is held, since an edge’s weight is still determined by the common neighborhood and degree of its two endpoints, by Eq. (2). As shown in figure 3, in spite of the slightly different values of \( b^* \) in original and randomized networks, the \( \rho(b) \) curves in configuration models are qualitatively the same to original ones.

In the second null model, for a given real network, we keep its topology unchanged and calculate edge weights according to Eq. (2), then we exchange the weights of two randomly selected edges \( e \). After sufficient number of exchanges (we set it as \( 10^E \) in this article), the weights are randomly redistributed. Under this null model, the topology is held while the correlation between local structure and edge weight is vanished. As shown in figure 4, after the redistribution of weights, the optimal value is exactly \( b^* = 0 \) for all six cases. At this point, the weighted SIS model degenerates to the unweighted one, or it is equivalent to the most homogeneous case with all weights are the same. Therefore, the results are in accordance with the known conclusion \( \text{CM} \) that in the absence of correlation between structure and weight, the most homogeneous weight distribution leads to the widest epidemic spreading.

In a word, by comparing with these two statistical null models, we conclude that the distribution pattern of weights, or say the correlation between local structure and weight, rather than the topology itself, mainly contributes to the experimental observations.

The correlation of weights of two adjacent edges continuously changes in the randomizing process of the second null model. Here we use the well-known Pearson correlation coefficient \( r \), similar to the degree-degree correlation coefficient \( \text{CM} \), to quantify the weight-weight correlation. Denote by \( \Gamma = \{(e_A, e_B), (e_{A_2}, e_{B_2}), \cdots, (e_{A_M}, e_{B_M})\} \) the set of all edge pairs sharing a common endpoint (i.e., two adjacent edges), where \( M = \sum_{i=1}^{N} k_i (k_i - 1) \) is the number of these edge pairs. Then the weight-weight correlation
just adjustable, this is indeed a good news when considering the

lence by subtly adjusting the system. If the system is ad-

ditionally, in the real networks, that the nontrivial peak comes from the local correlation

monotonously change with $r$

is possible that, if $(e_{A}, e_{B})$ (corresponding to $\rho$)

Clearly, the value of $r$ lies in the range $[-1, 1]$: $r > 0$

where the sum runs over all $M$ pairs and $s_{A}$ and $s_{B}$

denote the weights of edges $e_{A}$ and $e_{B}$, respectively. Note

that, if $(e_{A}, e_{B}) \in \Gamma$, then $(e_{B}, e_{A}) \in \Gamma$ too. Therefore

$\bar{s} = \frac{1}{M} \sum_{i=1}^{M} s_{A_i} = \frac{1}{M} \sum_{i=1}^{M} s_{B_i}$. For some $i$ and $j$, it

is possible that $e_{A_i} \equiv e_{A_j}$, and for an edge $e$, if its two

associated nodes are respectively of degree $k_{1}^{2}$ and $k_{2}^{2}$,

the weight $w_{i}$ has been counted $(k_{1}^{2} - 1) + (k_{2}^{2} - 1)$ times.

Clearly, the value of $r$ lies in the range $[-1, 1]$: $r > 0$

indicates a positive correlation (i.e., high-weight edges tend to be adjacent to other high-weight edges), $r < 0$

indicates a negative correlation (i.e., high-weight edges tend to be adjacent to low-weight edges), and $r = 0$ is

the case for no correlation.

For all the six real networks, the weight-weight correlation

coefficients are all considerably larger than zero, and during the randomizing process of the second null

model, $r$ almost monotonously decays to zero. As shown

in figure 5 and figure 6, both the maximal epidemic prevalence $\rho^*$ (corresponding to $b^*$) and the optimal value $b^*$

monotonously change with $r$. In particular, when $r = 0$, the

optimal value of $b$ is also equal to zero, indicating

that the nontrivial peak comes from the local correlation

of weight distribution. Additionally, in the real networks, compared with the null model, we can enlarge the prevalence

by subtly adjusting the system. If the system is ad-

justable, this is indeed a good news when considering the spreading of valuable information and innovation, while

we can assign $b$ a very large absolute value to against

the infectious disease. Real systems never work in such

a perfectly mathematical way, but our analysis provides

some theoretical insights that may benefit the design of interventions of real spreading dynamics.

VI. CONCLUSION AND DISCUSSION

Experimental results on six real networks demonstrated

that the preferential contacts through high-

weight edges can promote the epidemic prevalence in

the stable state. Further analysis showed that this phe-

nomenon results from the correlation between structure and weight, rather than the topological structure itself.

Specifically speaking, it is contributed by the local correlation

of weight distribution pattern. Although the non-

zero weight-weight correlation requires a heterogeneous distribution of weights, the heterogeneity of weight distri-

bution alone will not lead to the aforementioned non-

trivial phenomenon.

Although the spreading processes in real systems are far different from the present ideal model, this study gives

some inspirations on how to prevent infectious diseases and facilitate socially responsible information. However,

even if in an ideal system, the design of interventions is very challenging since the different goals may conflict to

each other. In the present model, the preference towards

strong ties could enlarge the prevalence while it may slow
down the spreading process (details will be reported else-

where). Analogous examples are numerous: by bypassing

main intersections and arterial roads could enhance the

network throughput while increase the delivering time

especially in uncongested states [24], and removing di-

rected loops could enhance the synchronizability while

slow down the synchronizing process [64].

Different nodes (e.g., hub nodes and peripheral nodes) and different edges (e.g., strong ties and weak ties) are

supposed to play different roles in the evolution and func-

tions of the network. By introducing the structure-based

weights of edges, we can to some extent distinguish the

roles of different edges in the spreading dynamics. We

hope this work could contribute to the long journey to-

wards fully understanding the relations between struc-

tural features and functional performances.

ACKNOWLEDGMENTS

We acknowledge Pak-Ming Hui for valuable discuss-

ions. This work is partially supported by the National Natural Science Foundation of China under Grant No.

11222543 and the Special Project of Sichuan Youth Scien-

cle and Technology Innovation Research Team under

Grant No. 2013TD0006. A.-X.C. acknowledges the Fun-

damental Research Funds for the Central Universities un-

der Grant No. ZYGX2012YB027 and T.Z. acknowledges

the Program for New Century Excellent Talents in Uni-

versity under Grant No. NCET-11-0070.
[60] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, Phys. Rev. E 71, 027103 (2005).
[61] M. Li, Y. Fan, J. Chen, L. Gao, Z. Di, and J. Wu, Physica A 350, 643 (2005).
[62] M. Li, Y. Fan, D. Wang, D. Li, J. Wu, and Z. Di, Phys. Lett. A 364, 488 (2007).

[63] J. L. Rodgers and W. A. Nicewander, The American Statistician 42, 59 (1988).
[64] T. Zhou, M. Zhao, and C. Zhou, New J. Phys. 12, 043030 (2010).