Knowl Inf Syst (2012) 32:589–608
DOI 10.1007/s10115-011-0445-x

REGULAR PAPER

Information propagation in online social networks: a tie-strength perspective

Jichang Zhao · Junjie Wu · Xu Feng ·
Hui Xiong · Ke Xu

Received: 18 June 2011 / Revised: 28 September 2011 / Accepted: 22 October 2011 /
Published online: 12 November 2011

Abstract In this paper, we investigate the relationship between the tie strength and information propagation in online social networks (OSNs). Specifically, we propose a novel information diffusion model to simulate the information propagation in OSNs. Empirical studies through this model on various real-world online social network data sets reveal three interesting findings. First, it is the adoption of the information pushing mechanism that greatly facilitates the information propagation in OSNs. Second, some global but cost-intensive strategies, such as selecting the ties of higher betweenness centralities for information propagation, no longer have significant advantages. Third, the random selection strategy is more efficient than selecting the strong ties for information propagation in OSNs. Along this line, we provide further explanations by categorizing weak ties into positive and negative ones and reveal the special bridge effect of positive weak ties. The inverse quantitative relationship between weak ties and network clustering coefficients is also carefully studied, which finally gives reasonable explanations to the above findings. Finally, we give some business suggestions for the cost-efficient and secured information propagation in online social networks.

J. Zhao · X. Feng · K. Xu (✉)
State Key Laboratory of Software Development Environment, Beihang University,
Beijing, China
e-mail: kexu@nlsde.buaa.edu.cn

J. Zhao
e-mail: zhaojichang@nlsde.buaa.edu.cn

X. Feng
e-mail: xu.f@nlsde.buaa.edu.cn

J. Wu
School of Economics and Management, Beihang University, Beijing, China
e-mail: wujj@buaa.edu.cn

H. Xiong
Rutgers Business School, Rutgers University, New Brunswick, NJ, USA
e-mail: hxiong@rutgers.edu

Keywords Online social networks · Information diffusion/propagation · Tie strength · Social synchrony

1 Introduction

As a popular social media platform [1], online social networking sites (OSNSs) have been developed massively for business and political purposes, such as viral marketing, targeted advertising, political campaigns, and even terrorist activities. Indeed, these sites, such as Facebook [13], MySpace [26], LinkedIn [24], and Twitter [33], provide a powerful means of organizing contacts, publishing contents, and sharing interests [25]. The users of these sites and the friendships among them constitute the so-called online social networks (OSNs).

Recent years have witnessed increasing interests in studying the characteristics of information propagation in online social networks. For instance, there are studies which have a focus on measuring the topology of these social networks, understanding the patterns of user interactions, or investigating the features of user behaviors. Meanwhile, many concepts from the traditional field of social networks have been leveraged for studying online social networks. However, limited work has been done to unveil the following three questions:

– What is the key feature that makes OSNs distinctive for information propagation?

– How fast does the information propagate naturally in OSNs?

– What is the role of weak ties for the information diffusion in OSNs?

The work in this paper aims to address these questions from a tie-strength perspective. The key contributions of this study are summarized as follows:

1. We develop a novel information diffusion model $IP(\alpha, \beta, w)$ to simulate the information propagation in online social networks. This model provides flexibilities in controlling the preferences and the channels for information propagation and reflects the information pushing mechanism of OSNSs.

2. As a surprise, we find that the global but cost-intensive higher-BCT-first strategy (namely selecting the ties of higher betweenness centralities preferentially) shows only small advantages in information diffusion, but the random selection strategy is more efficient than selecting the strong ties for information propagation. Furthermore, we also find that the clustering coefficients of social networks have some impact on the efficiency of the information propagation.

3. Through both the theoretical and empirical studies, we show that: (1) There is a negative correlation between the clustering coefficient and the number of weak ties; (2) The positive weak ties are very important in connecting the isolated local clusters for the further spread of information in OSNs. These two points together provide reasonable explanations to the above surprise findings.

4. We give some business suggestions for the cost-efficient and secured information propagation in online social networking sites.

Overview The remainder of this paper is organized as follows. Section 2 gives the related work. In Sect. 3, we present some preliminaries. Section 4 introduces the proposed diffusion model for the analysis of information propagation. In Sect. 5, we show experimental results on real-world OSNs. Section 6 reveals the essential role of weak ties. Finally, in Sects. 7 and 8, we give some business insights and conclude the work.

2 Related work

Information propagation in online social sites has attracted great research interests recently. In [7], the authors collected traces of photos propagating in Flickr and tried to unveil how far a photo would spread and the role of the friendship in the diffusion. Their results showed that in Flickr, it was hard for a photo to propagate widely, even for the popular ones. They also found that the information exchanged between friends was likely to account for over 50% with a significant delay at each hop. In [36], the authors developed a linear influence model to model the information diffusion in online social media by predicting which node would influence which other nodes in the network. The role of long ties in social networks was studied in [6]. It was found that the long ties in social networks prohibited the complex cascade [5]. In the inspiring work [28], the authors found in a mobile communication network a coupling between interaction strengths and the network's local structure. It is interesting that if the weak ties are removed gradually, there will be a phase transition in the network.

The propagation control in social networks is another open problem of interest for many years. For instance, Guo et al. [21] considered the problem of maximizing social influences in social networks using two probability models called Independent Cascade (IC) model [19] and Linear Threshold (LT) model [17]. They proposed a greedy algorithm to find the minimum set of influential nodes. Bonneau et al. [8] improved an algorithm using a degree discount method. While Centola et al. [9] proposed the first scalable influence maximization algorithm tailored for the linear threshold model. Their simulations showed that this algorithm was scalable to networks with millions of nodes and edges and orders of magnitude faster than the greedy approximation algorithm. Different from the above methods, [20,30] solved the problem by constructing a layered graph under the susceptible-infected-susceptible (SIS) model. Recently, Goldenberg and Muller [22] addressed the problem of minimizing the propagation of contamination by blocking a limited number of links. Network immunization strategy is another possible solution of propagation control, especially for the virus. In [14], experiments had been performed to examine how the topology and human dynamics affect the virus propagation in email networks. Their results revealed that the most efficient immunization strategy is the node-betweenness strategy. These works, however, were mainly focus on how to select a minimum set of source nodes, without considering the mechanism of spread. In contrast, in this paper, we try to characterize the coupled dynamics between the mechanism of information propagation and the tie strength, with the assumptions that (1) the information is started randomly from one source node in the network and (2) different choices of spread paths will result in different information coverage.

Other research topics about online social sites mainly include topologies probing and measurement [1,25,32], human behavior in information generating and diffusing [2,18,31], mapping the data of interactions between users with the strength of friendships [15,34], and the privacy protection [3,4,23,37].

In summary, despite of the vast amount of research efforts on the online social network problem, further study is still needed to unveil the roles of tie strengths for the information propagation in online social networks. Our work in this paper just aims to fill this crucial void.

3 Preliminaries

In this paper, an undirected graph $G(V, E)$ is used to denote an online social network, where V is the set of nodes and E is the set of ties. The *averaged degree* is defined as

$$\langle k\rangle = 2|E|/|V|. \tag{1}$$

The *clustering coefficient* of a node i is used to respond to the closeness of its neighbors, which is defined as

$$c_i = \frac{2|E_i|}{k_i(k_i-1)}, \tag{2}$$

where E_i is the set of ties existing between its neighbors, and k_i is its degree. When $k_i = 1$, we simply let $c_i = 0$. The *averaged clustering coefficient* of a network can be defined as

$$C = \frac{1}{|V|}\sum_{i\in V} c_i. \tag{3}$$

A higher C means the network is highly clustered. Note that we may omit the word "averaged" if there is no confusion in the context.

The concept of *tie strength* was first introduced in [16], in which it was defined as the relative overlap of the neighborhood of two nodes in the networks. Based on this definition, Newman and Park [28] gave a simple but quantified definition to the overlap of neighbors of nodes i and j as follows:

$$s_{ij} = \frac{c_{ij}}{k_i - 1 + k_j - 1 - c_{ij}}, \tag{4}$$

where k_i and k_j are the degrees of i and j, respectively, and c_{ij} is the number of common acquaintances. In this paper, we adopt the definition of the tie strength in Eq. 4 and use it to characterize the strength of the relationship between two users of an online social network. The smaller s_{ij} is, the weaker the tie is.

Discussions In the literature, many researches have found that the tie strength indeed indicates the strength of the relationship between two nodes. For example, it has been found that in social networks, a pair of nodes has the propensity to be connected if they share a mutual neighbor [27,35]. Similarly, adjacent users in an online social site tend to trust each other [25], especially when they share a lot of common acquaintances. Recently, it has also been found that in mobile communication networks, the more two users' friends overlap, the stronger their contact would be [28]. These works well support the use of Eq. 4 for the measurement of tie strengths.

4 Modeling the information propagation

The problem of the information propagation in OSNSs has attracted great interests in various application domains because of the unique information pushing and republishing mechanisms [12,38].

Let us take Facebook for example. The applications named *News Feed* and *Live Feed* keep pushing all your friends' activities to your profile pages. Specifically, *News Feed* aggregates the most interesting contents that your friends posted, while *Live Feed* shows all the actions your friends are taking. Information pushing is more obvious in Twitter, in which your words will be pushed immediately to all your followers' terminals. Then, the information will be further propagated by republishing, including commenting, citing, and reprinting, supported by nearly all the online social networking sites.

Inspired by these observations, we describe the information diffusion in OSNSs as follows [38]: (1) A user i publishes the information I, which may be a message, a photo, a blog, etc.; (2) Friends of i will know I when they access the profile page of i or get some direct notifications from the online social site; (3) Some friends of i, may be one, many, or none, will comment on, cite, or reprint I, because they think that it is interesting, funny, or important; (4) The above steps will be repeated with i being replaced by each of those who have republished I.

Accordingly, to characterize the information propagation in OSNSs, we expand the model proposed in our previous work [38] to a new model called $IP(\alpha, \beta, w)$. In this model, α is a navigating parameter that determines how to select republishing nodes. $\beta \in [0, 1]$ indicates the strength of the information, i.e., how interesting, novel, important, funny, or resounding the information is. Finally, w represents the weights of the ties, which can be used for the channel selection of information propagation. The model is defined as follows:

- Step 1: Suppose a node i in the site publishes the information I with strength β at time $T = 0$. Set i to the state δ_1, which means i is aware of I. Set other nodes to the state δ_0, which means I is unknown to them.
- Step 2: Increase the time by one unit, i.e., $T = T + 1$. Set the state of each i's neighbor to δ_1. Add i to P, the set of nodes that have published or republished I, i.e., $P = P \cup \{i\}$.
- Step 3: Get the number of nodes that may republish I in the next round:

$$\xi_i = k_i \beta, \tag{5}$$

  where k_i is the degree of i.
- Step 4: Select one node j from the neighborhood of i with the probability

$$p_{ij} = \frac{w_{ij}^{\alpha}}{\sum_{m=1}^{k_i} w_{im}^{\alpha}}, \tag{6}$$

  where w_{ij} is the weight of the tie between i and j, which is determined by w. If $j \notin P$, then add it to Q, the queue of nodes that will republish I in the next round, i.e., $Q = Q \cup \{j\}$. Repeat this step ξ_i times.
- Step 5: Execute from Step 2 to Step 4 for each node in Q recursively until Q is null or all nodes have known I.

Some notable information regarding the details of the $IP(\alpha, \beta, w)$ model is summarized as follows:

First, it is obvious that in Step 2, the model will push the information I from i to all its friends. This responds to the information pushing mechanism in OSNSs.

Second, Eq. 5 indicates that the number of republishing nodes selected from the neighborhood of i is decided by k_i and β. This is consistent with the real situation that the user with more friends tends to attract more users to visit and republish the information. And the more interesting or important the information is, the higher the chance that it will be republished. Since it is found that only 1–2% friends will republish the information in Flickr [7], we set $\beta = 0.01$ in the experiments below.

Third, parameter α is used in Eq. 6 to associate the diffusion with the weights of the ties. When $\alpha = 1$, the model tends to select the ties with higher weights to republish the information, while the ties with lower weights are preferred when $\alpha = -1$. When $\alpha = 0$, the selection will be random, regardless of the weights of the ties. The introduction of α indeed provides great flexibilities to the model.

Table 1 Experimental data sets

Data set	$\|V\|$	$\|E\|$	$\langle k \rangle$	C
Caltech	762	16,651	43.70	0.41
Georgetown	9,388	425,619	90.67	0.22
Oklahoma	17,420	892,524	102.47	0.23
Princeton	6,575	293,307	89.22	0.24
UNC	18,158	766,796	84.46	0.20

Finally, in this study, the weights of ties defined by w are limited to the two categories: (i) ST: the strength of a tie defined by Eq. 4; (ii) BCT: the betweenness centrality of a tie, which is defined as the number of shortest paths between all pairs of nodes passing through the tie. The introduction of w indeed extends the applicable scope of the model, which is of great use for the comparison of information propagation in the experiments below.

5 Characterizing the information propagation

In this section, we characterize the information propagation in OSNs using the $IP(\alpha, \beta, w)$ model on five real-world Facebook data sets. Note that for the following simulations, when $s_{ij} = 0$, we simply let $s_{ij} = 1/2N$ to avoid the computation problem, where N is the network size. This guarantees that p_{ij} will not be zero in any case.

5.1 Experimental data sets

Five real-world data sets from Facebook were used for the experiments. They are the Facebook networks whose ties are within five American universities [32]: California Institute of Technology (`Caltech`), Georgetown University (`Georgetown`), Princeton University (`Princeton`), University of Oklahoma (`Oklahoma`), and University of North Carolina at Chapel Hill (`UNC`). All the data sets are anonymized and publicly available online [29]. In these data sets, each node represents a user, and each tie means there exists a friendship between two users.[1]

We summarize some key features of these data sets in Table 1. As shown in the table, `Caltech` is a highly clustered network with $C = 0.41$, whereas `Oklahoma` is the densest network with $\langle k \rangle = 102.47$.

5.2 A comparison of different information propagation strategies

Here, based on two types of weights, ST and BCT, we compare the performances of four information propagation strategies: higher-BCT-first, strong-tie-first, weak-tie-first, and random selection. The corresponding models for the four strategies are $IP(1, 0.01, BCT)$, $IP(1, 0.01, ST)$, $IP(-1, 0.01, ST)$, and $IP(0, 0.01, w)$, respectively. Note that for the random selection model $IP(0, 0.01, w)$, the ties for information propagation are selected with equal chances no matter what w is.

[1] In Facebook, creating a friendship between two users always needs mutual permissions, which guarantees the validity of the tie.

It is important to note that: (1) Since the betweenness centrality is always thought to be strongly related to the traffic through ties or nodes, we can expect that the higher-BCT-first strategy should perform much better than the other three strategies; (2) The strong-tie-first strategy is the most natural way for information propagation in OSNSs, since it is reasonable that your closest friends who have a lot of common friends with you would have similar interests as you, and may probably republish the information from you.

The results are shown in Fig. 1. We define the fraction of users who are aware of the information I as the *information coverage*, denoted by f_c. As can be seen, three observations are noteworthy as follows:

Observation 1 It is interesting to note that for all the network data sets except `Caltech`, the information coverage using the non-weak-tie-first strategies increases rapidly during the diffusion process, and reaches about 1 after only 10–30 hops. In fact, if we denote the number of new nodes that become aware of the information as n_{new}, then for the non-weak-tie-first strategies, n_{new} reaches the maximum even within 10 hops! Fig. 2 shows this apparent social synchrony phenomenon [10] in two networks, which also indicates the rapid and wide propagation of information in OSNs.

Remark 1 Indeed, it is the information pushing mechanism that greatly facilitates the information propagation in OSNs. We will validate this hypothesis in Sect. 5.3 below.

Observation 2 Although higher-BCT-first is indeed the fastest way for information diffusion in online social networks, its performance is far from dominant; that is, the performances of the random selection and strong-tie-first strategies are pretty close to it.

Remark 2 This observation implies that the natural information diffusion with the strong-tie-first strategy in OSNs is quite satisfying, since it is just slightly slower than the higher-BCT-first strategy, which is often expensive or even impractical for demanding a global view of network structures.

Observation 3 While the strong-tie-first strategy shows better performances than the weak-tie-first strategy, its performance is worse than that of the random selection strategy. To further illustrate this, we draw the publishing paths in `Georgetown` under different selections, as shown in Fig. 3. Clearly the random selection in Fig. 3b can find the most republishing nodes, which facilitates the fast information spread in the networks. In contrast, the weak-tie-first selection in Fig. 3a leads to very sparse publishing paths, which hinders the information from further spreading.

Remark 3 This observation reveals two facts. On one hand, compared with weak ties, strong ties are more favorable to the information diffusion in OSNs, which indeed agrees with our intuition. On the other hand, however, strong ties alone are not adequate for widening the spread of information. Actually, it is right the weak tie that makes the random selection strategy superior to the strong-tie-first strategy. We will detail the special role of weak ties in OSNs in the next section.

Observation 4 For the unique `Caltech` network, the performance of the weak-tie-first strategy is much closer to that of the strong-tie-first strategy.

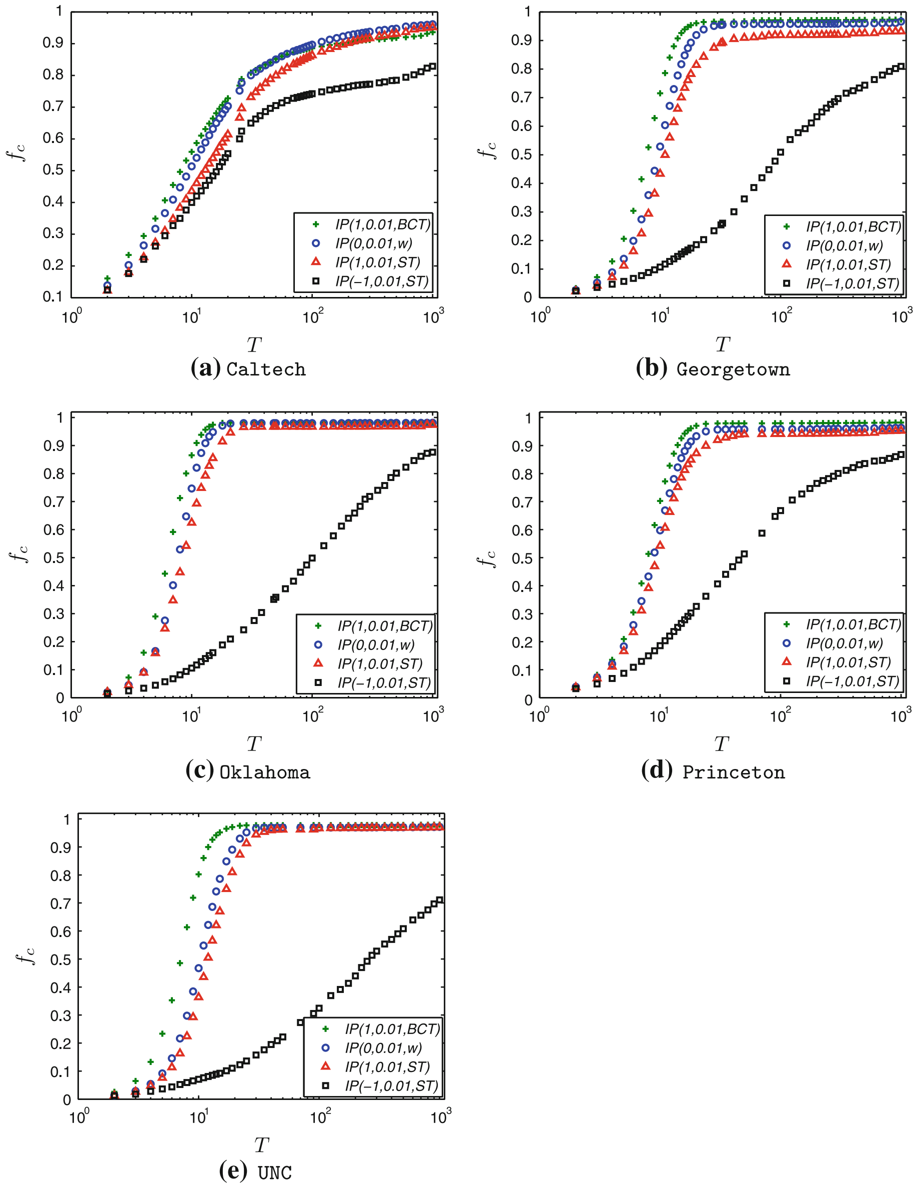
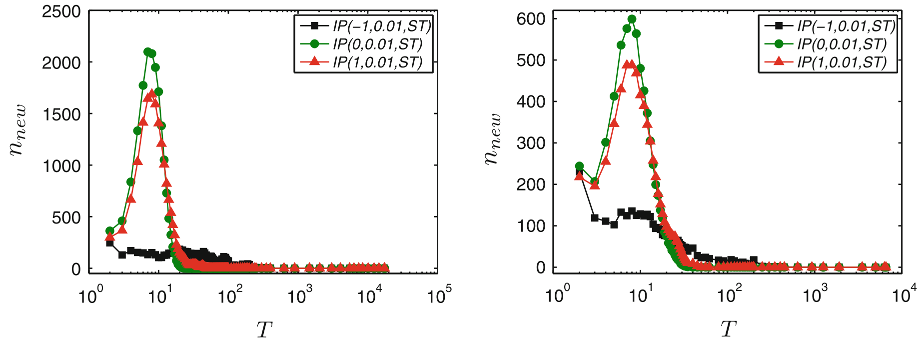


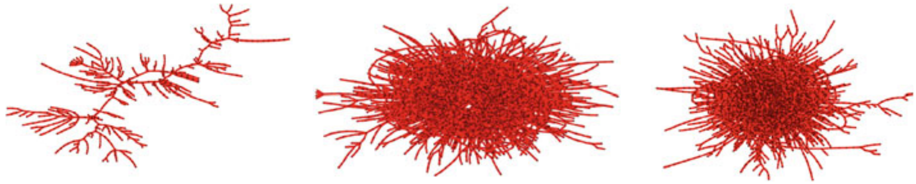
Fig. 1 A comparison of information coverage using different propagation strategies. **a** `Caltech`; **b** `Georgetown`; **c** `Oklahoma`; **d** `Princeton`; **e** `UNC`

Remark 4 If we recall the network characteristics in Table 1, we will find that one major factor that distinguishes `Caltech` from other networks is the clustering coefficient. That is, `Caltech` has a much higher C than the rest networks. This observation implies that the clustering coefficient has some relationships with the information diffusion in OSNs. We will revisit this in the next section.



(a) Oklahoma **(b)** Princeton

Fig. 2 Illustration of social synchrony. **a** Oklahoma; **b** Princeton



(a) $IP(-1, 0.01, ST)$ **(b)** $IP(0, 0.01, w)$ **(c)** $IP(1, 0.01, ST)$

Fig. 3 Publishing paths in Georgetown. (T=100) **a** $IP(-1, 0.01, ST)$; **b** $IP(0, 0.01, w)$; **c** $IP(1, 0.01, ST)$

5.3 The effect of information pushing mechanism

In this subsection, we highlight the impact of the information pushing mechanism. The $IP(1, 0.01, ST)$ model is selected for better simulating real-world information propagation scenarios. Specifically, we modify the $IP(1, 0.01, ST)$ model so that only f_p randomly selected friends can get the information pushed by the network systems, where $f_p = 5$, 10, 50 and 100%, respectively.

Figure 4 shows the information propagation results within the five networks. As can be seen, for all the networks, when $f_p = 5\%$, i.e., the information pushing is almost ineffective, the information diffusion becomes significantly slower and narrower. For instance, when $f_p = 5\%$, the information coverage in Caltech after $T = 10^3$ is not more than 0.5, far less than the 0.9 coverage when $f_p = 100\%$. Nonetheless, as f_p increases, the information diffusion efficiency improves greatly and rapidly. Indeed for all networks, when $f_p = 50\%$, the information diffusion statuses are already very close to the ones when the networks fully have the information pushing mechanism, namely $f_p = 100\%$.

In summary, this experiment well illustrates the importance of information pushing in online social networks. Indeed, information pushing speeds up the information exchange within an online social network, and thus spurs the formation of online communities.

5.4 The impact of the information strength

To illustrate the impact of information strength to information diffusion, we perform IP model on three sample data sets using different values of β. In the simulation, for each

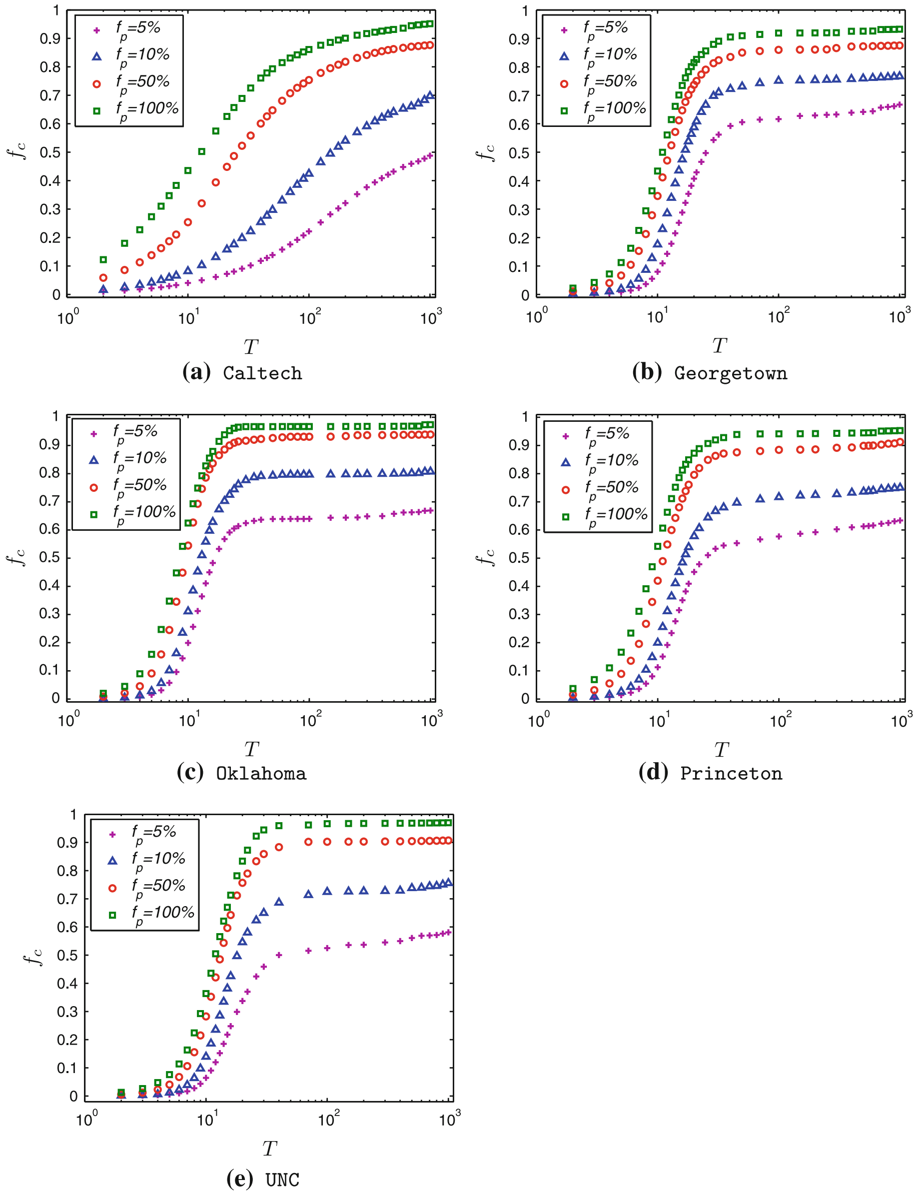
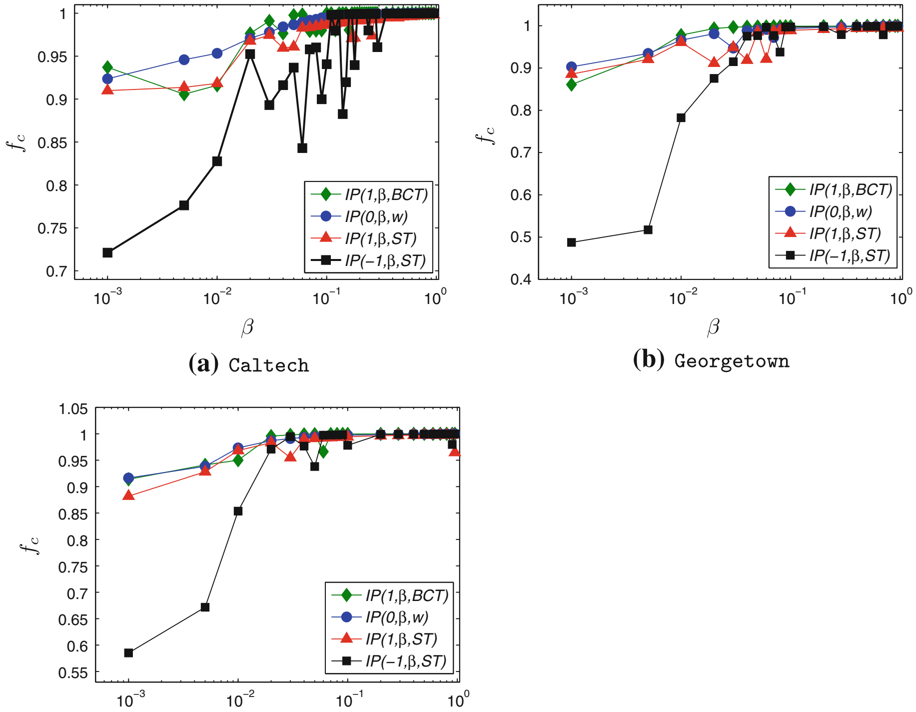


Fig. 4 The effect of information pushing in OSNs. **a** `Caltech`; **b** `Georgetown`; **c** `Oklahoma`; **d** `Princeton`; **e** `UNC`

configuration of parameters, we repeat the experiment 100 times and return the averaged f_c value as the final result, as shown in Fig. 5. As can be seen, it is obvious that for all the configurations, f_c grows as β increases, and f_c goes up to nearly 1 when $\beta > 0.1$. This indicates that the strength of information indeed has great and positive impact to the information diffusion in online social networks. Another observation is that no matter what the value of β is, the random selection strategy still performs the best as compared to the strategies of



(a) `Caltech` **(b)** `Georgetown` **(c)** `Princeton`

Fig. 5 Impact of β to information coverage. ($T = 1000$). **a** `Caltech`; **b** `Georgetown`; **c** `Princeton`

strong-tie-first and weak-tie-first, although the gap narrows when β goes up. Finally, again we can find that the performance of the higher-BCT-first strategy is still far from dominant as β grows.

6 Understanding the weak ties

In this section, we try to explain Observations 3 and 4 indicated in the previous section by unveiling the role of weak ties for the information diffusion in OSNs.

6.1 The special role of weak ties

According to Eq. 4, a weak tie can be formed due to several reasons, say for instance, a very small overlap of friends between two nodes, or a "star" node that has a very high degree. Here, we are more interested in the weak ties formed by one or two star nodes. Such nodes are usually the centers of different clusters or local communities in a network. As a result, the weak ties built on these nodes can serve as the "information bridges" that connect various isolated communities. We call such weak ties the "positive weak ties" for their positive effect for the information propagation in online social networks. To illustrate, four clusters are selected from the `Georgetown` network, as shown in Fig. 6. As can be seen, the three weak ties (red) indeed connect the four isolated clusters in the network.

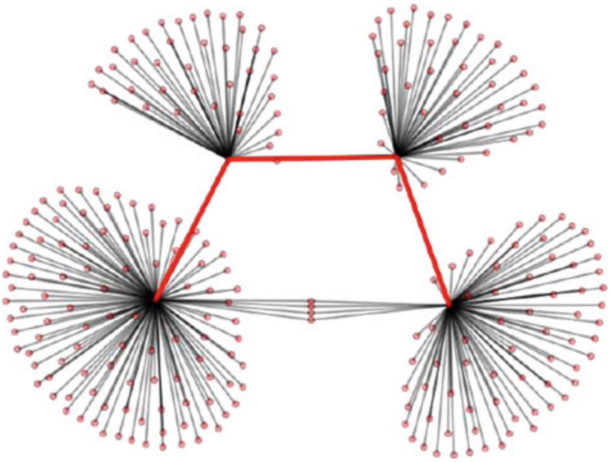


Fig. 6 Some symmetric weak ties (*red*) of Georgetown

In contrast to positive weak ties, there are also many "negative weak ties" in online social networks, which contain two low-degree nodes with very small overlaps on friends. Such nodes have few friends to republish the information, and therefore the information usually cannot spread further in the network. In other words, negative weak ties tend to show negative effect for the information propagation in online social networks.

An online social network typically contains both positive and negative weak ties. Roughly speaking, we can expect that there exist far more negative weak ties than positive ones in online social networks, for it is often very hard to maintain a large volume of friends as a star node. As a result, weak ties play a very special role for the information propagation in online social networks: First, due to the wide existence of negative weak ties, setting weak ties as the preferred channel (i.e., the weak-tie-first strategy) is often an ineffective way for the information propagation; Second, if we totally ignore the weak ties, the information may not reach some local communities without the helps of positive weak ties.

To illustrate the effect of positive weak ties, we employ the $IP(0, 0.01, w)$ model for the five networks in Table 1, remove the ties gradually in an increasing or decreasing order of tie strengths, and observe the changes of the information coverage when $T = |V|$. Figure 7 shows the results, where f_r denotes the fraction of removed ties. As can be seen in the figure, for all the networks, the coverage of information starts a sharp drop at $f_r = 0.4$ when removing the weak ties first. In contrast, when removing the strong ties first, the information coverage remains beyond 0.9 until f_r reaches a high value, say 0.8. This experiment reveals the important bridge effect of the positive weak ties. They become the only channels for the information propagation to some remote local communities.

6.2 The relationship between clustering coefficients and weak ties

Here, we explore the relationship between the clustering coefficient and the number of weak ties. Assume we know the degrees of node i and its friends. Recall that the clustering coefficient of node i is

$$c_i = \frac{2|E_i|}{k_i(k_i-1)} \Rightarrow |E_i| = \frac{c_i k_i(k_i-1)}{2}, \tag{7}$$

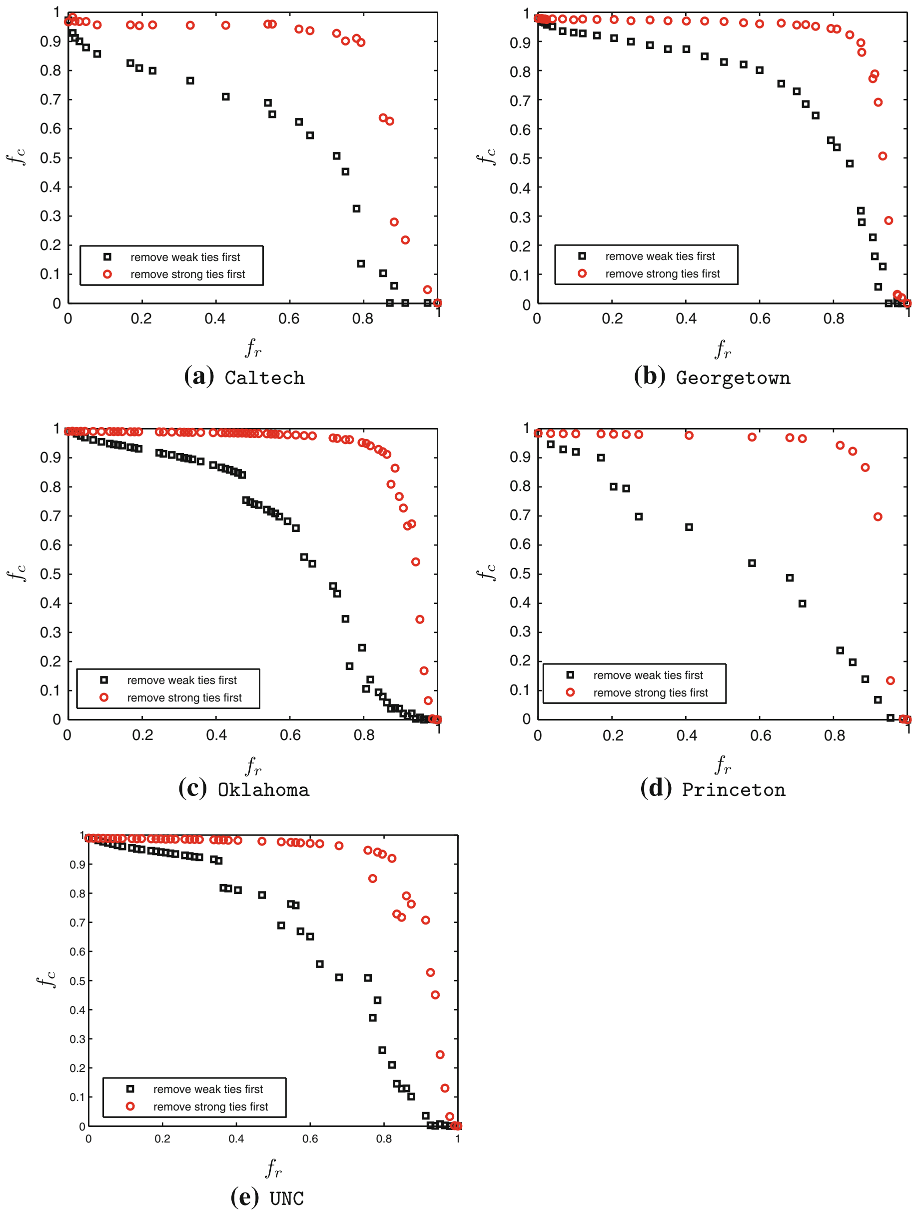


Fig. 7 The effects of removing weak/strong ties first. **a** `Caltech`; **b** `Georgetown`; **c** `Oklahoma`; **d** `Princeton`; **e** `UNC`

where E_i is the set of all ties existing between the neighbors of node i, and $k_i > 1$ is the degree of node i. Let

$$S_i = \sum_{m \in \{\text{neighbors of } i\}} k_m. \tag{8}$$

Then, we have the fact that built on S_i degrees, i's neighbors have altogether $2|E_i|$ common acquaintances with i. In other words, we can expect that one degree of i's neighbors leads to $2|E_i|/S_i$ common acquaintances with i. Now, assume node j is an arbitrary neighbor of node i. Then, the number of common acquaintances of i and j can be estimated by

$$c_{ij} \approx \frac{2|E_i|}{S_i} k_j = \frac{c_i k_i (k_i - 1) k_j}{S_i}. \tag{9}$$

Then, the strength of a tie can relate with its clustering coefficient as follows:

$$s_{ij} = \frac{c_{ij}}{k_i - 1 + k_j - 1 - c_{ij}} \approx \frac{1}{\frac{S_i (k_i + k_j - 2)}{c_i k_i (k_i - 1) k_j} - 1}. \tag{10}$$

By Eq. 10, when k_i, k_j and S_i are fixed, $c_i \uparrow \Rightarrow s_{ij} \uparrow$. This implies that, *when the averaged clustering coefficient of a network increases, the fraction of weak ties in that network will decrease*. This result, however, is solely based on the above theoretic reasoning. In what follows, we validate it through extensive empirical studies.

To this end, we should first find an authoritative online social network model, for which we can set the value of clustering coefficient in a broad range. However, to our best knowledge, it is still a very challenging work to find such a model. Nonetheless, it is widely accepted that online social networks have the mixed small-world and scale-free properties [1,25]. So we generate both the small-world networks and the scale-free networks using the Small-World model (SW) [35] and the Dorogovtsev and Mendes model (DM) [11], respectively. We assume: If the two types of networks have a similar result, then the result can be extended to the online social networks.

We denote the network generated by the DM model as $DM(N, m)$, where N is the size of the network and m is the number of links which will be established when a new node is added to the network. Also, we denote the network generated by the SW model as $SW(N, K, p)$, where $2K$ is the number of initial degrees of each node and p is the probability to rewire each link. As one notable merit, the DM (SW) model can generate networks with broad clustering coefficient values by tuning m (p). Specifically, the clustering coefficient of a DM (SW) network decreases when m (p) increases.

Now, let us look at the cumulative distribution functions (CDF) of tie strength in $DM(20000, m)$ and $SW(20000, 10, p)$, respectively. As shown in Fig.8, when m or p increases, i.e., the averaged clustering coefficient decreases, the CDF curve moves to the left, which indicates the increase of the fraction of weak ties. A similar effect can also be observed from the real-world data sets. As shown in Fig. 8c, the fraction of weak ties in `Caltech` is clearly much less than the ones in other networks for its highest clustering coefficient.

6.3 Understanding weak ties from a combined view

Based on the findings of the above two subsections, we can now explain the interesting Observations 3 and 4 raised in Sect. 5. In a nutshell, two key points are crucial for the explanations as follows:

First, although the negative weak ties tend to hinder the information from being further diffused, the positive weak ties have an important bridge effect which can facilitate the information propagation across various isolated clusters.

Recall the four networks, i.e., `Georgetown`, `Oklahoma`, `Princeton`, and `UNC`, with relatively small clustering coefficients in Table 1. According to Sect. 6.2, the numbers of weak ties tend to be large in these networks. In other words, these networks may contain

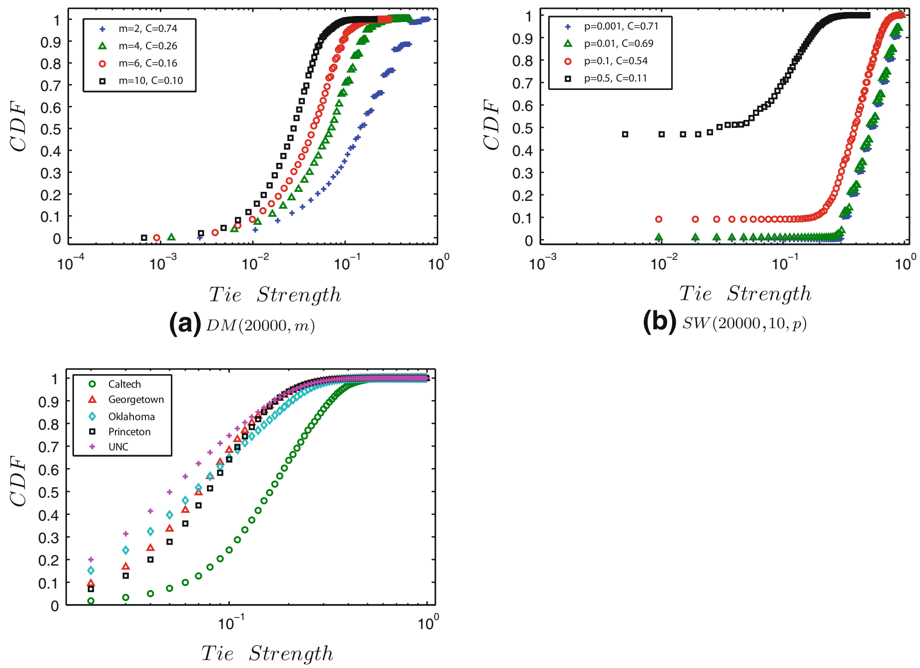


Fig. 8 The CDF of tie strength with different clustering coefficients. **a** $DM(20000, m)$; **b** $SW(20000, 10, p)$; **c** Real-world datasets

many negative weak ties, which will lead to the significantly worse performances of the weak-tie-first strategy. On the other hand, although the strong-tie-first strategy can bypass the negative weak ties, it is at the cost of missing the positive weak ties simultaneously, which will result in the local trapping of the information in some isolated clusters, and eventually make it hard to propagate the information further in the network. As a "compromise" between the weak-tie-first and strong-tie-first strategies, the random selection strategy has a higher probability in selecting the positive weak ties, which can help avoid the local trapping of the information. This explains why the random selection strategy beats the strong-tie-first strategy in all the cases.

Second, as the clustering coefficient goes up, the number of weak ties especially the negative weak ties will decrease, and thus the bridge effect of the positive weak ties will be more significant or even dominant in the networks.

Recall the `Caltech` network in Table 1 with a much higher clustering coefficient than the rest networks. According to Sect. 6.2, we can expect that the weak ties in `Caltech` have a more significant bridge effect, which indeed has led to the comparable performances of the weak-tie-first strategy and the strong-tie-first strategy in Fig. 1a. To better understand this, we also apply the $IP(-1, 0.01, ST)$ model for the five network data sets, and compare their information coverage. Figure 9 shows the results. As can be seen, when the weak-tie-first strategy is preferred, information diffusion in `Caltech` is indeed faster than information diffusion in other networks.

As an interesting corollary, for a network with an extremely high clustering coefficient, we can expect that the weak-tie-first strategy will be dominant, since the positive weak ties in

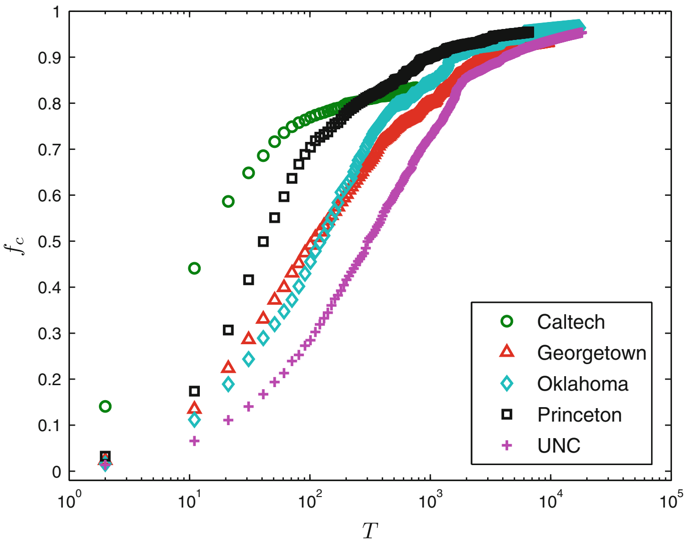


Fig. 9 A comparison of the weak-tie-first strategy on five social networks

such networks become the only channels that connect the isolated local clusters. To illustrate this, we again employ simulations on SW and DM models as follows.

Let $m = 2$ and $p = 0.001$. We then perform simulations on $DM(N, 2)$ and $SW(20000, K, 0.001)$, respectively. Figure 10 shows the results when $T = |V|$. As can be seen, selecting weak ties as preferred republishing paths always performs the best no matter what the size of the DM network is or what the density of the SW network is. These results indeed follow our expectations. Note that, compared with real OSNs, the averaged clustering coefficients of $DM(N, 2)$ and $SW(20000, K, 0.001)$ are much higher. For instance, we have $C_{DM(20000,2)} = 0.74$ and $C_{SW(20000,10,0.001)} = 0.71$.

7 Business insights

Nowadays, online social networking sites have established themselves as the most powerful media for reforming our social and knowledge sharing patterns in the new era. Indeed, according to the data released by Hitwise in March 2010, the independent pageview of Facebook in USA has exceeded the pageview of Google, which indicates that Facebook has become the most popular website in USA. As a result, the design of OSNSs and the issues regarding the information diffusion in OSNSs are getting more and more important from a business viewpoint. Our study in this paper can give some business insights as follows:

1. First, knowledge sharing websites within or outside the organizations should adopt the *push* mechanism of OSNSs to make the knowledge sharing faster and wider.
2. Second, pushing the information to the friends using a strong-tie-first strategy is a good choice to speed up the information propagation in OSNSs, and it can also lower the loads of the website in pushing the information to all the friends in peak time.
3. Third, the growing popularity of the online social networks does not mean that it is safe and reliable. As a simple and cost-efficient way, we can make the virus or the private

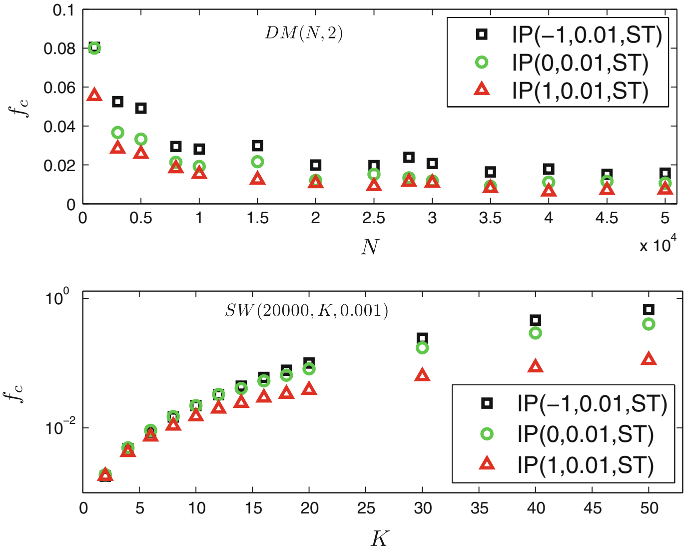


Fig. 10 Information diffusion in highly clustered social networks

information trapped in local communities by removing positive weak ties and stop them from diffusing further in the network.

8 Conclusion

In this paper, we investigated the impact of tie strength on information propagation in online social networks (OSNs). Specifically, we proposed an information diffusion model, which has flexibilities in controlling the preferences and the channels for information propagation. The model analysis revealed that: (1) The natural information propagation is very fast in OSNs using the information pushing mechanism; (2) Due to the bridge effect and the inverse correlation with clustering coefficients, weak ties play an important role for information diffusion in OSNs; (3) As an interesting extension, for the networks with very high clustering coefficients, selecting weak ties preferentially can speed up the information propagation. In the future work, we plan to extend the $IP(\alpha, \beta, w)$ model to an individual-oriented model which can characterize the differences of network users, e.g., on information strength β.

Acknowledgments This work was supported by the National Natural Science Foundation of China (NSFC) (No. 61170189), the fund of the State Key Laboratory of Software Development Environment (SKLSDE-2011ZX-02) and the Program for New Century Excellent Talents in University (NCET-07-0047). The second author was partially supported by NSFC (Nos. 70901002, 71171007, 70890080, 90924020) and the Doctoral Fund of Ministry of Education of China (No. 20091102120014).

References

1. Ahn YY, Han S, Kwak H, Moon S, Jeong H (2007) Analysis of topological characteristics of huge online social networking services. In: Proceedings of the 16th international conference on world wide web, WWW '07, pp 835–844

2. Benevenuto F, Rodrigues T, Cha M, Almeida V (2009) Characterizing user behavior in online social networks. In: Proceedings of the 9th ACM SIGCOMM conference on internet measurement conference, IMC '09, pp 49–62
3. Bonneau J, Anderson J, Anderson R, Stajano F (2009) Eight friends are enough: social graph approximation via public listings. In: Proceedings of the second ACM EuroSys workshop on social network systems, SNS '09, pp 13–18
4. Bonneau J, Anderson J, Danezis G (2009) Prying data out of a social network. In: Proceedings of the 2009 international conference on advances in social network analysis and mining, pp 249–254
5. Centola D, Eguíluz VM, Macy MW (2007) Cascade dynamics of complex propagation. Physica A Stat Mech Appl 374(1):449–456
6. Centola D, Macy M (2007) Complex contagions and the weakness of long ties. Am J Sociol 113(3): 702–734
7. Cha M, Mislove A, Gummadi KP (2009) A measurement-driven analysis of information propagation in the flickr social network. In: Proceedings of the 18th international conference on world wide web, WWW '09, pp 721–730
8. Chen W, Wang Y, Yang S (2009) Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '09, pp 199–208
9. Chen W, Yuan Y, Zhang L (2010) Scalable influence maximization in social networks under the linear threshold model. In: Proceedings of the 2010 IEEE international conference on data mining, ICDM '10, pp 88–97
10. De Choudhury M, Sundaram H, John A, Seligmann DD (2009) Social synchrony: predicting mimicry of user actions in online social media. In: Proceedings of the 2009 international conference on computational science and engineering, vol 04, pp 151–158
11. Dorogovtsev S, Mendes J (2002) Evolution of networks. Adv Phys 51:1079–1187
12. Eric S, Itamar R, Cameron AM, Thomas ML (2009) Gesundheit! modeling contagion through facebook news feed. In: 3rd international conference on weblogs and social media (ICWSM), pp 146–153
13. Facebook, http://www.facebook.com
14. Gao C, Liu J, Zhong N (2011) Network immunization and virus propagation in email networks: experimental evaluation and analysis. Knowl Inf Syst 27(2):253–279
15. Gilbert E, Karahalios K (2009) Predicting tie strength with social media. In: Proceedings of the 27th international conference on human factors in computing systems, CHI '09, pp 211–220
16. Granovetter MS (1973) The strength of weak ties. Am J Sociol 78(6):1360–1380
17. Granovetter MS (1978) Threshold models of collective behavior. Am J Sociol 83(6):1420–1443
18. Guo L, Tan E, Chen S, Zhang X, Zhao YE (2009) Analyzing patterns of user content generation in online social networks. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '09, pp 369–378
19. Goldenberg J, Libai B, Muller E (2001) Talk of the network: a complex systems look at the underlying process of word-of-mouth. Mark Lett 12(3):211–223
20. Kazumi S, Masahiro K, Hiroshi M (2009) Discovering influential nodes for sis models in social networks. Discov Sci 5808:302–316
21. Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining, KDD '03, pp 137–146
22. Kimura M, Saito K, Motoda H (2009) Blocking links to minimize contamination spread in a social network. ACM Trans Knowl Discov Data 3:9:1–9:23
23. Korolova A, Motwani R, Nabar SU, Xu Y (2008) Link privacy in social networks. In: Proceeding of the 17th ACM conference on Information and knowledge management, CIKM '08, pp 289–298
24. LinkedIn, http://www.linkedin.com
25. Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B (2007) Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM conference on internet measurement, IMC '07, pp 29–42
26. MySpace, http://www.myspace.com
27. Newman M, Park J (2003) Why social networks are different from other types of networks. Phys Rev E 68(3):036–122
28. Onnela JP, Saramaki J, Hyvonen J, Szabo G, Lazer D, Kaski K, Kertesz J, Barabási AL (2007) Structure and tie strengths in mobile communication networks. PNAS 104(18):7332–7336
29. Porter MA (2010) http://people.maths.ox.ac.uk/~porterm/data/facebook5.zip
30. Saito K, Kimura M, Ohara K, Motoda H (2011) Efficient discovery of influential nodes for sis models in social networks. Knowl Inf Syst. doi:10.1007/s10115-011-0396-2

31. Shi X, Zhu J, Cai R, Zhang L (2009) User grouping behavior in online forums. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '09, pp 777–786
32. Traud AL, Kelsic ED, Mucha PJ, Porter MA (2008) Comparing community structure to characteristics in online collegiate social networks. ArXiv e-prints, arXiv0809.0690
33. Twitter, http://www.twitter.com
34. Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in facebook. In: Proceedings of the 2nd ACM workshop on online social networks, WOSN '09, pp 37–42
35. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442
36. Yang J, Leskovec J (2010) Modeling information diffusion in implicit networks. In: Proceedings of the 2010 IEEE international conference on data mining, ICDM '10, pp 599–608
37. Ying X, Wu X (2011) On link privacy in randomizing social networks. Knowl Inf Syst 28(3):645–663
38. Zhao J, Wu J, Xu K (2010) Weak ties: subtle role of information diffusion in online social networks. Phys Rev E 82(1):016–105

Author Biographies

Jichang Zhao received a B.E. degree from Beihang University, Beijing, China, in 2008. He is currently a Ph.D. student at the Department of Computer Science, Beihang University, China. His research interests include robustness of complex networks, information diffusion in online social networks, and evolution of complex networks.

Junjie Wu received his Ph.D. degree in Management Science and Engineering from Tsinghua University, China, in 2008. He is currently an associate professor in Information Systems Department, School of Economics and Management, Beihang University, China. His general area of research is data mining and complex networks, with a special interest in solving the problems raised from the emerging data intensive applications. As a young scientist, he has published over 35 papers in refereed conference proceedings and journals, such as KDD, ICDM, DMKD, TKDE, and TSMCB. He has also been a reviewer for NSFC proposals, the leading academic journals and many international conferences in his area. He is the recipient of the National Excellent Doctoral Dissertation Award (2010), and the "star track" visiting scholar of MSRA. He is a member of AIS, ACM, and IEEE.

Xu Feng received a B.E. degree from Beihang University, Beijing, China, in 2009. He is currently a Ph.D. student at the Department of Computer Science and Technology, Beihang University, Beijing, China. His research interests include link mining in complex networks and social computing.

Hui Xiong is currently an Associate Professor and the Vice Chair of the Management Science and Information Systems department at Rutgers University. He received the B.E. degree from the University of Science and Technology of China (USTC), China, the M.S. degree from the National University of Singapore (NUS), Singapore, and the Ph.D. degree from the University of Minnesota (UMN), USA. His general area of research is data and knowledge engineering, with a focus on developing effective and efficient data analysis techniques for emerging data intensive applications. He has published over 90 technical papers in peer reviewed journals and conference proceedings. He is a co-editor of Clustering and Information Retrieval (Kluwer Academic Publishers, 2003) and a co-Editor-in-Chief of Encyclopedia of GIS (Springer, 2008). He is an Associate Editor of the Knowledge and Information Systems journal and has served regularly in the organization committees and the program committees of a number of international conferences and workshops. He is a senior member of the ACM and IEEE.

Ke Xu is a professor at Beihang University, China. He received his B.E., M.E., and Ph.D. degrees from Beihang University in 1993, 1996, and 2000, respectively. His research interests include algorithm and complexity, data mining, and complex networks.