Influence of Reciprocal links in Social Networks

YU-XIAO ZHU\textsuperscript{1,2}, XIAO-GUANG ZHANG\textsuperscript{3}, GUI-QUAN SUN\textsuperscript{3,4}, MING TANG\textsuperscript{1}, TAO ZHOU\textsuperscript{1} and ZI-KE ZHANG\textsuperscript{4} (a)

\textsuperscript{1} Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 610051, P. R. China
\textsuperscript{2} Beijing Computational Science Research Center, Beijing 100084, P. R. China
\textsuperscript{3} Department of Mathematics, North University of China, Taiyuan 030051, P. R. China
\textsuperscript{4} Institute of Information Economy, Hangzhou Normal University, Hangzhou 310018, P. R. China

PACS 89.20.Ff – Computer science and technology
PACS 89.75.Hc – Networks and genealogical trees
PACS 89.65.-a – Social and economic systems

Abstract. – In this Letter, we empirically study the influence of reciprocal links, in order to understand its role in affecting the structure and function of directed social networks. Experimental results on two representative datasets, Sina Weibo and Douban, demonstrate that the reciprocal links indeed play a more important role than non-reciprocal ones in both spreading information and maintaining the network robustness. In particular, the information spreading process can be significantly enhanced by considering the reciprocal effect. In addition, reciprocal links are largely responsible for the connectivity and efficiency of directed networks. This work may shed some light on the in-depth understanding and application of the reciprocal effect in directed online social networks.

Introduction. – Nowadays, the emergence of social networks and affiliated applications have triggered an increasing attention from various disciplines, ranging from studying the social interactions and spreading patterns in social sciences \cite{12} to uncovering the underlying structure and dynamics in mathematics and physics \cite{3,4}. Generally, social networks can be classified into two typical classes according to the edge properties: undirected and directed. Undirected social networks, such as Flickr \cite{5} and Okut \cite{5}, do not allow two users to be connected unless the relation is mutually confirmed, hence, they are normally regarded as equivalent individuals in graph theory. Comparatively, directed social networks, such as Twitter \cite{5} and Epinions \cite{5}, contain both unidirectional and bidirectional links, which consequently build up a so-called follower/followee structure \cite{6,7}. An online user is considered as a follower once s/he collects some other users as friends (followees), and puts close attention to them via automatically receiving their real-time information, as well as online activities \cite{9}. A considerable fraction of those followees would also give positive feedback and add some of their followers with similar interests as online neighbors. Subsequently, such intermediate directed structure property, namely reciprocity \cite{10}, facilitates a great deal of attention from the scientific community. Nowak and Sigmund discussed that the indirect reciprocity would help in building reputation systems, judging morality and eventually promote the cooperation level \cite{12} and benefit the evolution of natural selection \cite{11} in both social environment \cite{10,13} and supply networks \cite{15}. Pereira et al. experimentally discussed that negative reciprocity, because of lower cost and less effort, was somehow more favored than the positive reciprocity \cite{16}. Moreover, the power of reciprocity \cite{17} does not only play a vital role in social economic systems \cite{18,19} and human social organizations \cite{20,21}, but also has been found wide applications in characterizing the property \cite{22,23}, maintaining the structure \cite{24,25}, and uncovering the underlying function of directed social networks \cite{26,27,28,29}.

Typically, the simplest definition of reciprocity, $r$, can be quantified as the ratio of the number of bidirectional links, $L^{\leftrightarrow}$, to the total number of links $L$ \cite{28,29},

$$r = \frac{L^{\leftrightarrow}}{L}. \quad (1)$$

For the extreme cases, $r = 0$ represents an absolute directed network where all links are unidirectional, and $r = 1$ stands for a complete undirected network where
all links are reciprocal. Therefore, the value of \( r \) measures the probability that two nodes of a given link are mutually connected. However, Garlaschelli and Loffredo [22] argued that Eq. (1) failed to precisely describe the full network information. For example, the network density and self-loops can significantly affect the final measurement of mutual connections. Alternatively, they proposed a new measure of reciprocity considering the ordering of different networks according to their actual degree of reciprocity, denoted as

\[
\rho = \frac{\bar{L} - \bar{a}}{1 - \bar{a}} = \frac{r - \bar{a}}{1 - \bar{a}}, \quad (2)
\]

where \( \bar{a} = L/N(N - 1) \) measures the ratio of observed links to all possible directed links (namely link density). Based on this improved measure, Zlatić et al. [23] reported that the reciprocity of Wikipedia could be very similar to other directed networks, but having a stronger reciprocity than the networks of associations and dictionary terms, and smaller than that of World Wide Web. Besides that, they found that such a measure is quite stable for different scales of Wikipedia networks, hence is very important for describing the structure and evolution of wiki-based networks. Boguñá et al. [30] found that reciprocal connections played a crucial role in constructing the giant connected component and possibly affecting the Web navigability. Furthermore, Serrano et al. [31] provided an in-depth study of the effect of reciprocal links on degree-degree correlations and clustering. They found that reciprocal links indeed organized the local subgraphs of the World Wide Web network by forming start-like structures, as well as cliques and communities, which contained highly interconnected pages. What’s more, Gorka et al. [32] argued that the reciprocity was largely dependent on degree-degree correlation, which, consequently could partially reveal the underlying hierarchical structure of networks. Zlatić and Štefanić [33] discussed the influence of reciprocity on vertex degree distribution and degree correlations. They found that networks driven by reciprocal mechanisms are significantly different from static networks.

In this Letter, we shall provide a specific empirical study of the reciprocity influence on the structure and function of social networks. In particular, we apply a widely used epidemic spreading model [34][35] to observe the effect of reciprocity on information spreading. Numerical results show that reciprocal links can noticeably enhance the speed of information spreading. In addition, we show how reciprocal links affect the structure robustness as percolation catalysts in maintaining the global connectivity by investigating the avalanche of giant components, the network susceptibility and the network distance [37][38].

Data and Analysis. – In this Letter, we consider two representative directed social networks: (i) Sina Weibo [39]: the largest Chinese microblogging website, where a user (follower) can add others as his/her friends (followee) and automatically receive their posts and events. In addition, users can forward, comment or share their followees’ news on their own post walls; (ii) Douban [40]: the largest Chinese website for reviewing online movies, books, and music. Besides users’ generally proactive contribution, Douban also provides services via its recommendation mechanism, which can suggest items of users’ potential interests by mining their personalized preferences. Similar with Sina Weibo, users in Douban can also build follower-followee relationship with each other.

Table 1 summarizes the basic statistics of the two observed data sets. \( N = |V| \) and \( M = |E| \) are the total number of nodes and links, respectively. \( \rho \) is the network reciprocity denoted by Eq. (2), and \( S = M/N(N - 1) \) denotes the network sparsity.

| Data sets     | \( N \)     | \( M \)   | \( \rho \)   | \( S \)    |
|---------------|-------------|-----------|--------------|-----------|
| Sina Weibo    | 3,592       | 12,522    | 0.307        | 9.7 \times 10^{-4} |
| Douban        | 3,097       | 6,417     | 0.474        | 6.7 \times 10^{-4} |

Fig. 1: In-degree (left) and out-degree (right) distributions of the two observed data sets.

Consequently, such relationship can be represented by a directed network \( G(V, E) \), where \( V \) is the set of nodes and \( E \) is the set of edges. Each node represents a user, and one link from user \( i \) to user \( j \) indicates \( i \) is followed by \( j \), that is to say, \( i \) is the followee of \( j \), and \( j \) is one of \( i \)’s followees.

Table 1 summarizes the basic statistics of the observed datasets. In addition, Fig. 1 shows the out-degree distributions which power-law \( p(k_{\text{out}}) \propto k_{\text{out}}^{-\lambda} \) with exponents \( \lambda = 1.366 \) and 1.958, for Sina Weibo and Douban, respectively. This common feature suggests that most users are ordinary beings who have relative small number of followers and keep only a small fraction of celebrities. Comparatively, the in-degree distribution of the two datasets does
not exhibit the same phenomenon. The in-degree distribution of Douban still keep power-law shape with exponent 2.387, but Sina Weibo has a cut-off around $k_{in} = 20$. One possible reason is that Sina Weibo only allows a certain number of followers for each free account. It might also suggest the different mechanisms driving the growth of two sites: information diffusing automatically in microblogging system of Sina Weibo, comparing with the information filtering by recommendation-related technique in Douban. Similar difference between passive and automatic patterns was also empirically reported in bipartite and hypergraph networks \cite{11, 12}. In addition, we further investigate the average number of common follower and followees (see Table 2). Compared to non-reciprocal node pairs, reciprocal ones tend to have more common followers and followees, which is in accordance with previous work \cite{27}.

Table 2: Comparisons of the average number of common followees ($NCI$) and followers ($NCF$) for reciprocal and non-reciprocal node pairs, respectively.

|            | Sina Weibo | Douban |
|------------|------------|--------|
| Reciprocal | 0.498      | 0.451  |
| Non-reciprocal | 0.274  | 0.295  |

Methods and Results. – To better understand the influence of link reciprocity in social networks, in the following, we shall evaluate its effects on information spreading and network robustness from the perspectives of the network function and structure, respectively.

Effect on Information Spreading. Information spreading \cite{13} is one of the most important functions of social networks, where the information (messages, tweets, comments, etc.) can distribute at a remarkably fast speed through the whole online society via frequent interactions among users, although its structure is not designed on purpose for spreading news \cite{14}. Up to now, there is a considerable number of theoretical models to study information diffusion on social networks \cite{15, 19}. In this Letter, in order to understand the underlying mechanisms and possible factors that would result in the information outbreaks, we adopt the classic epidemic spreading model, Susceptive-Infected (SI) model \cite{34}, to evaluate the effect of reciprocal links in the two aforementioned social networks. The diffusion process is described as following,

- Initially, user $i$ publishes an information item, $I$, in the corresponding social network. $I$ could be about a piece of news, a photo, a comment, etc;
- All $i$’s followers will automatically receive $I$ according to the follower-followee directed network structure. Then an arbitrary fraction of those followers might notice $I$, and forward it on their own homepages if they find it interesting. We consider this forwarding willingness as the transmission probability, denoted by $p$;
- The above step will be repeated to the followers of $i$’s followers, and eventually diffuses to the all achievable network nodes.

Note that, the main difference between the directed SI (DSI) and classical SI model is that the link direction is taken into account. In the proposed DSI model, the information can only be transmitted from the followee to its own followers along with the direction of edges. Therefore, the final fraction of influenced nodes, $p_I$, is determined by such a structure. In order to observe the effects of reciprocal links on information diffusion, we quantify the influence according to an edge percolation process \cite{38, 31, 53}. Obviously, if one reciprocal link is more important than two separate non-reciprocal links, the information diffusion results will be affected significantly when we remove the same fraction of reciprocal and non-reciprocal links. Fig. 2 compares the information coverage of removing the two types of links. Compared with removing non-reciprocal links, $p_I$ decays much faster when we remove the same amount of reciprocal links. Analogously, it also can be seen from Fig. 3 that the diffusion speed is affected much remarkably when removing reciprocal links. Therefore, it demonstrates that reciprocal links indeed play a more important role in the information diffusion process on directed social networks.

Effect on Structural Robustness. In conventional complex network theory, it is widely agreed that the network function is largely influenced by its specific structure \cite{50}. Therefore, to give solid and comprehensive understanding of the aforementioned results, we adopt the a dynamical removing process to measure the effects of reciprocal links on maintaining the structural robustness of networks \cite{35}. For comparison, we apply three metrics to quantify the corresponding performance. (i) $R_{GSCC}$: the relative size of the strongly connected giant component. A sudden decline of $R_{GSCC}$ will be observed if the network disintegrates after deleting a certain fraction of edges; (ii) the network susceptibility ($\tilde{S}$): defined as

$$\tilde{S} = \sum_{s<s_{max}} \frac{n_s s^2}{N},$$

where $n_s$ is the number of components with $s$ nodes, $N$ is the size of the network, and the sum runs over all the components except the largest one ($s_{max}$). Note that, different with the definition in undirected networks, in Eq. (3), we only consider the strongly connected component in directed networks. Considering $\tilde{S}$ as the function of the fraction of removed edges $f$, usually, an obvious peak can be observed when the network disintegrates \cite{54, 55}; (iii) the average distance ($d$), calculated by

$$d = \frac{1}{N(N-1)} \sum_{<i,j>\in E, i \neq j} d_{i,j},$$

(4)
Fig. 2: (Color online) The fraction of influenced nodes as the function of the fraction of removed links \( f \). In each subgraph, the red and green curves correspond to removing reciprocal and non-reciprocal links, respectively. The experimental results are averaged over 30 independent realizations.

Fig. 3: (Color online) The fraction of influenced nodes as the function of observed time-step \( t \), where \( f \) is the fraction of removed links. The red and green curves correspond to removing reciprocal links and non-reciprocal links, respectively. The experimental results are averaged over 30 independent realizations.
where \( d_{<i,j>} \) is the distance from node \( i \) to \( j \). \( d_{<i,j>} \) is set to \( N \) when there is no directed path from node \( i \) to \( j \). Clearly, the smaller \( (d) \) is, the better connectivity and more efficient the network will be.

Fig. 4 and Fig. 5 show the corresponding results of the three examined matrices. In Fig. 4 it shows different dynamical patterns of removing reciprocal and nonreciprocal links, respectively. The size of strongly connected giant component \( (R_{GSCC}) \) decreases more sharply when removing reciprocal links than deleting non-reciprocal ones. Accordingly, the network susceptibility \( (\tilde{S}) \) result shows a percolation phenomenon when removing reciprocal links. Comparatively, this phenomenon is not observed when removing non-reciprocal links. In addition, Fig. 4 shows that the average network distance \( (\langle d \rangle) \) increases much faster when removing reciprocal links than deleting the nonreciprocal ones. In a word, different dynamical results indicate that reciprocal links play a more important role in both maintaining the connectivity and keeping the efficiency of directed networks than non-reciprocal links. It also strongly supports the results in the previous section that reciprocity can much promote the speed of information diffusion, as it takes a more significant responsibility for the robustness of directed networks.

![Fig. 4: (Color online) The fraction of giant component size \( (R_{GSCC}) \) and the susceptibility \( (\tilde{S}) \) as the function of the fraction of removed links \( f \) on the two observed datasets, (a) Sina Weibo and (b) Douban. In each subgraph, the red and green curves correspond to the results of reciprocal and non-reciprocal links, respectively. The experimental results are averaged over 30 independent realizations.](image)

![Fig. 5: (Color online) The average network distance \( (\langle d \rangle) \) as the function of removed links \( f \) on the two observed datasets, (left panel) Sina Weibo and (right panel) Douban. The red and green curves correspond to the results of reciprocal and non-reciprocal links, respectively. The experimental results are averaged over 30 independent realizations.](image)

**Conclusions and Discussion.** — In this Letter, we have studied the influence of reciprocal links of directed networks from two perspectives: (i) information spreading; (ii) structural robustness. Experimental results on two representative directed social networks, Sina Weibo and Douban, show that reciprocal links indeed play a more important role than non-reciprocal ones. In particular, the results of information spreading show that reciprocity can significantly enhance the spreading speed. In addition, the corresponding observations on the two examined datasets show that the reciprocity is also largely responsible for maintaining the connectivity and keeping the efficiency of directed networks, which suggests its significant impact in information spreading on networks.

The findings of this work may have a wide-range application in studying the role and influence of reciprocal links. Firstly, the topic of community detection has been well discussed [59], however, the progress on directed networks [57] is relatively slow. The main reason is that the modularity [58] of directed networks is rather difficult to be precisely defined. Secondly, most studies on epidemic spreading and information diffusion [59] focus on studying the corresponding dynamics on undirected networks, the in-depth theoretical understanding of the underlying spreading mechanism on directed networks [58] still remains to be solved. Finally, the area of information filtering [61] confronts a huge challenge as more and more directed social services are provided in the information era. The present work just provides a start point to see the preliminary effects of reciprocal links, a more comprehensive and in-depth understanding of reciprocity still need further efforts to discover.

***

This work was partially supported by the National Natural Science Foundation of China (Grant Nos. 11105024, 11105025, 1147015 and 11205040). ZKZ acknowledges the Zhejiang Provincial Natural Science Foundation of China (Grant Nos. LY12A05003 and LQ13F030015), the start-up foundation and Pandeng project of Hangzhou Normal University. ZYX acknowledges the Fundamental Research Funds for Central Universities (Grant No. A03008023401042).
REFERENCES

[1] Jackson M. O. and Watts A., J. Economic Theory, 106 (2002) 265.
[2] Kwak H., Lee C., Park H., and Moon S., Proc. 19th Intl. Conf. WWW, 74 (2006) 59.
[3] Castellano C., Fortunato S., and Loreto V., Rev. Mod. Phys., 81 (2009) 591.
[4] Lancichinetti A., Fortunato S., Kertész J., Nature, 446 (2007) 664.
[5] Websites of data can be accessed online via: http://www.flickr.com/, http://www.okut.com/, http://www.epinions.com/ and http://wikipedia.org/.
[6] Huberman B., Romero D. M. and Wu F., First Monday, 14 (2009).
[7] Gayo-Avello D., Brener D. J., Fernández-Fernández D., Fernández-Menéndez M. E. and García-Suárez R., EPL, 94 (2011) 38001.
[8] Grabowsicz P. A., Ramasco J. J., Moro E., Pujol J. M. and Eguiluz V. M., PLoS ONE, 7 (2012) e29358.
[9] Cimini G., Chen D.-B., Medo M., Lü L., Zhang Y.-C. and Zhou T., Phys. Rev. E, 85 (2012) 046108.
[10] Falk A. and Fischbacher U., Games and Economic Behavior, 54 (2006) 293.
[11] Nowak M. A. and Sigmund K., Nature, 437 (2005) 1291.
[12] Rong Z., Yang H.-X. and Wang W.-X., Phys. Rev. E, 82 (2010) 047101.
[13] West S. A. and Gardner A., J. Evolution. Biol., 20 (2007) 415.
[14] Rockenbach B. and Milinski M., Nature, 444 (2006) 718.
[15] Ge Z., Zhang Z.-K., Lü L., Zhou T. and Xi N., Physica A, 391 (2012) 647.
[16] Pereira P. T., Silva N. and Silva J. A., J. Econ. Behavior Organ., 59 (2006) 406.
[17] Diekmann A., J. Conflict Resolution, 48 (2004) 487.
[18] Berg J., Dickhaut J. and McCabe K., Games Econ. Behavior, 10 (1995) 122.
[19] Feigh E., Fischbacher U. and Gächter S., Human Nature, 13 (2002) 1.
[20] Milinski M., Semmann D. and Krambeck H. P., Roy. Soc. B-Biol. Sci., 269 (2002) 881.
[21] Feigh E. and Fischbacher U., Nature, 425 (2003) 785.
[22] Gal拉斯chelli D. and Loffredo M. I., Phys. Rev. Lett., 93 (2004) 26.
[23] Zlatić V., Bozicević M., Štefanić H. and Domazet M., Phys. Rev. E, 74 (2006) 016115.
[24] Zhou T., Medo M., Cimini G., Zhang Z.-K. and Zhang Y.-C., PLoS ONE, 6 (2011) e20648.
[25] Lü L., Zhang Y.-C., Yeung C. H. and Zhou T., PLoS ONE, 6 (2011) e21202.
[26] Medo M., Zhang Y.-C. and Zhou T., EPL, 80 (2009) 68003.
[27] Cui A.-X., Zhang Z.-K., Tang M., Hui P. M. and Fu Y., PLoS ONE, 7 (2012) e50702.
[28] Newman M. E. J., Forrest S. and Balthrop J., Phys. Rev. E, 66 (2002) 035101(R).
[29] Serrano M. A. and Boguña M., Phys. Rev. E, 68 (2003) 015101(R).
[30] Boguña M. and Serrano M. A., Phys. Rev. E, 72, (2005) 016106.