Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Inferring FOLLOW Relationship from Repost Relationship between Users on Sina Weibo

Xu Hao[1], **Xiang Li**[1,2]

1. Adaptive Networks and Control Lab, Department of Electronic Engineering, and Research Center of Smart Networks and Systems, School of Information Science and Engineering, Fudan University, Shanghai 200433, China.
2. MOE Frontiers Center for Brain Science, Institutes of Brain Science, Fudan University, Shanghai 200433, China.
(e-mail: 18210720029@fudan.edu.cn, lix@fudan.edu.cn)

Abstract: In recent years, online social platforms such as sina weibo have been playing an increasingly important role in our lives. With the improvement of public awareness of privacy protection, however, it is increasingly difficult to obtain exhaustive direct data of *FOLLOW* relationship between users, which restricts the analysis and research on real online social networks. In the context of incomplete disclosure of information, we have to consider inferring from limited public data to more unknown information before further analysis and research with network topology. To overcome this obstacle, we try to recover the *FOLLOW* relationship by repost relationship. We collect repost data of six different bloggers on sina weibo to generate six independent networks, and propose an effective method to recover *FOLLOW* relationship between users. Our method is superior to other existing methods, and helps us draw a conclusion that two users establishing an indirect repost relationship with other two users, compared to other indirect relationship, are of greater possibility to maintain the *FOLLOW* relationship. At the same time, we also extend our method to greatly improve the recovering accuracy.

Keywords: Network inference, Sina weibo, data protection

1. INTRODUCTION

The last decades have witnessed considerable developments in utilizing complex network theory to analyze important issues as reported in Li et al. (2016). The examples of research content cover the disease transmission analysis and control by Castellano and Pastor-Satorras (2010), the discovery of the network community structure by Van Lierde et al. (2019), network synchronization by Wu and Li (2018) and so on. In recent years, with the rapid development of data tracking and collection technology, more complicated network model such as time-varying network which is studied by Liu et al. (2014b), Wang and Li (2019), multiplex network by Gomez et al. (2013), Ma et al. (2020) have gradually become the research hotspots.

Social network is the most high-profile branch in which one person is regarded as a node, and the relationship or interaction between people is regarded as a link generally. Though we have witnessed fruitful and exciting advances in the analysis of complicated, large-scale social networks, for example in Zhang and Li (2012), Liang et al. (2016), the relationship predicting or recovering between people in social networks is still an important and challenging research direction as reported in Clauset et al. (2008), Aiello et al. (2012), and a lot of efforts have been put into this direction in Lü and Zhou (2011), Zhang et al. (2013), Zhang et al. (2017).

As online social media platforms enter our lives, many scholars have proposed their methods to infer or predict user relationship aiming at online social networks rather than physical contact networks. De et al. (2013) proposed a novel framework, which integrates signals from node features, the existing local link neighborhood of a node pair, community-level link density, and global graph properties for link prediction. Liu et al. (2014a) predicted the links not only relying on the structural features of the network but also relying on of the interaction behaviors between agents. Zhang et al. (2017) studied the social link prediction of the target network aligned with multiple social networks concurrently. The prediction and inference on social relationships have many potential applications on online social platforms such as helping new users to find their friends or the content they are interested, as reported in Fouss et al. (2007), Dong et al. (2012). More importantly, the inference of social relationships provides new sights in social reality mining, and makes it feasible for further analysis especially on social network topology, when the direct relationship is hidden.

Though there are many popular online social platforms including twitter, facebook, Sina weibo and so on, our research focuses on one type of them—Sina weibo, which is widely used in China and has great influence. *FOLLOW* relationship is a basic unilateral relationship between users in Sina weibo. If user Y is user X's follower, then the weibo user X posts will be automatically distributed to the browsing area of user Y promptly. Obviously, on Sina weibo, the information spreads most commonly through *FOLLOW* relationship, and a complete *FOLLOW* relationship between users is significantly important to explore online social networks. In recent years, however, with the increasing awareness of privacy protection, many online social platforms do not disclose all the relationships of users, so does Sina weibo. For example, Sina weibo

only discloses users' latest 100 *FOLLOW* relationships, which means that ordinary users cannot directly obtain other users' all *FOLLOW* relationships. The insufficient data restricts deeper analysis and research on real online social networks, and in the current trends, there is no doubt more and more online social platforms will choose not to open all relationships. Therefore, in order to obtain network topology, the research on using other indirect information to recover the user's direct *FOLLOW* relationship is of important significance.

Despite of the fruitful advances on link inference mentioned above, recovering the *FOLLOW* relationship on Sina weibo is still a challenging task because the available information is quite rare. The previous methods all require part of the user's direct relationship, or even more information such as users' preference. However, due to the small amount of known information, the repost network is so sparse that the previous traditional link inference methods have a very poor performance on our data set (Specific simulation results can be seen in Sec. IV). To get higher accuracy, we propose a new method, the indirect repost relationship algorithm(IRRA), to cope with the situation of insufficient public data. We find that this method performs much better than the previous methods in our network consisting of repost relationship .

Our main contributions are as follows :

(1) In the context of incomplete public information, we propose a new perspective to recover the direct *FOLLOW* relationship based on the indirect repost relationship between Sina weibo users.
(2) We propose a new method, indirect repost relationship algorithm(IRRA), to solve the above problem. It is able to effectively infer users' *FOLLOW* relationship based on repost data, and outperforms other methods on our data set. It is also found that users who can establish an indirect relationship by two other users are more likely to maintain the *FOLLOW* relationship.
(3) Inspired by the existing methods, we consider the different types of users and extend our algorithm, which greatly improves the inference accuracy by complementing the near and far repost relationship.

2. DATA SET

2.1 Data Collection

Since it is the first time to propose that the *FOLLOW* relationship can be totally recovered from the perspective of reposting, we make an effort to collect a whole new set of data. We selected six bloggers on Sina weibo(see Table 1), tracked and collected all the repost information of their weibos during a period of time. The number of reposts on a single weibo ranges from 5 to 8,000. In dataset *A*, for example, the selected blogger recorded as blogger *A* has 601283 followers and 330 *FOLLOW*, and has totally published 1892 weibos before. During the collection period (from March 20, 2018 to April 25, 2018), 70 weibos were published, and all the reposts corresponding to these weibos were collected. It should be noted that we collect not only the users who directly repost the blogger *A*'s weibos, but also the users who repost other users' repost of *A*. In other words, if user *X* reposts a weibo from blogger *A*, and user *Y* reposts this weibo from *X*, then both user *X* and user *Y* are collected with the corresponding reposting time. And if other users repost this weibo from user *Y*, all of them are also collected.

In order to ensure that the experiment is feasible and accurate, the bloggers we selected meet the following two conditions. On the one hand, these bloggers have enough followers, and the followers are relatively active, so that the reposting number of their weibos is sufficient, and the reposts have an obvious hierarchy as shown in Fig. 1. For this reason, we choose the blogger focusing on the topics of current affairs. On the other hand, the number of reposts on each of their weibos is not so large to avoid that they become the top search topic and other disturbances introduced. In this way, we can get six basic data sets in total based on the repost information of the six bloggers. Table 1 shows the detailed information about the data collection. *Follow*, *Follower*, *Weibo* in the table refer to the number of *FOLLOW*, followers and weibos of the original bloggers we selected respectively. *Collection Period* shows the start and end dates between which we collected the data. *reposts_collected*, *weibo_collected*, means the total number of the reposts and weibos we collected.

2.2 Data Preprocessing

Before the recovering of network edge connection, the raw data needs to be pre-processed. The users we collected are divided into two categories: marginal users and core users. During data preprocessing, the marginal users are deleted.

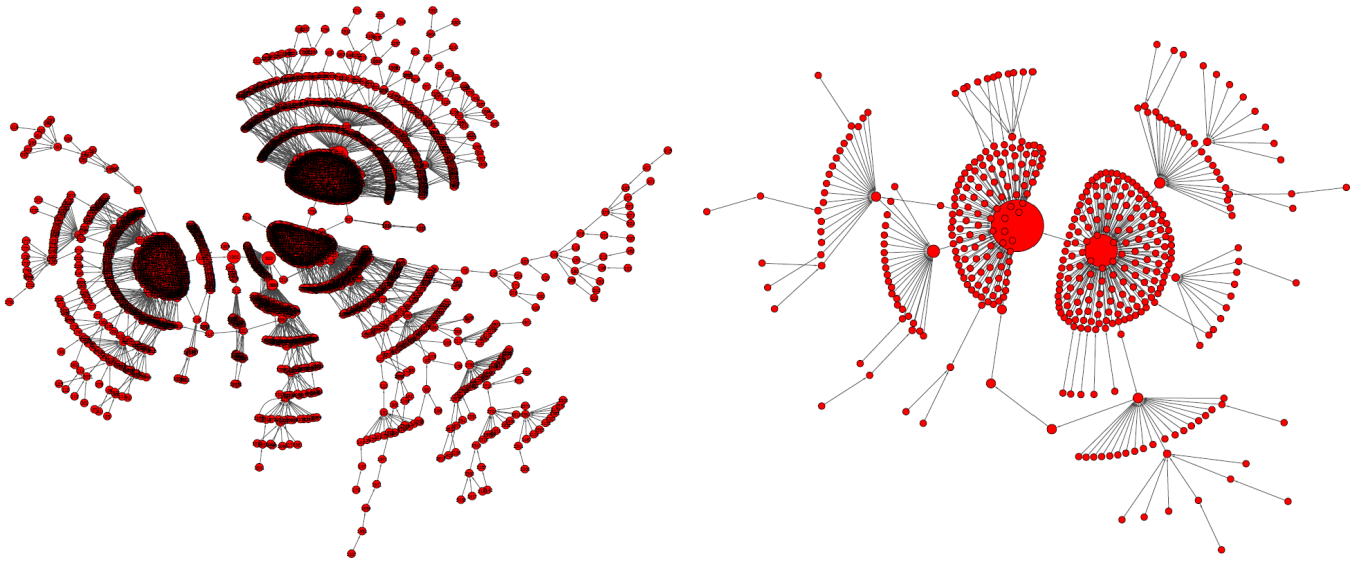
Here, the marginal users are defined as the users that meet one of the following two conditions (again, take data set *A* as an example):

(1) Among 70 weibos, the user only participates in reposting the weibo once, which is not reposted by any user. In other words, the out-degree of the node is 1, and the in-degree of the node is 0.
(2) Among 70 weibos, the user only participates in reposting the weibo once, and all the users who repost him are marginal users.

Marginal users meeting the first condition can be considered that they only participate in the repost by chance, and they are not active followers. They have little influence on the results of the following link recovering, because they're not shared by any two users, nor are upon any path that two users can reach each other. Marginal users meeting the second condition are only upon the paths between marginal users, while we only focus on the relationships between core users, so these users also have little influence on the recovering results in this paper. During the actual preprocessing, we first delete all the marginal users in the original data. After the first delete operation, new marginal users appear and we repeat the delete operation iteratively until no more new marginal users appear. After calculation, 78.5% of the users are located at the margin of the network and deleted. After deletion, the scale of the data set is greatly reduced and the subsequent computational complexity is also greatly reduced. What's more, the relationship between marginal users who participate in the repost by chance is avoided to be recovered. The above simplified processing is just a preprocessing on the collected raw data because of the noise in the raw data. In Sec. IV, we used the same network as input when we compare our performance with other algorithms.

Table 1. Detailed information about the data collection

Data Set	Follow	Follower	Weibo	Collection Period	reposts_collected	weibo_collected
A	330	601283	1892	2018.03.20-2018.04.25	47648	70
B	3767	984852	73394	2018.08.26-2018.09.01	79744	210
C	457	6991828	15029	2018.08.28-2018.09.16	45153	56
D	653	2525346	8421	2018.07.06-2018.09.04	12236	110
E	391	1354295	2249	2018.08.24-2018.09.20	72994	47
F	632	589908	8671	2018.09.16-2018.09.30	85184	28



(a) (b)

Fig. 1. (a), (b) are generated from the repost data in data set *A* of 2 weibos, respectively.

2.3 Network Generation

In this subsection, we generate six independent repost networks from the six collected data sets. We take each weibo user as a node, that is, user X is denoted as node x, and user Y is denoted as node y. If user Y reposts a weibo posted by user X, then we consider that there is an edge between node x and node y. Taking data set A as an example, the selected bloggers in this data set published the total of 70 weibos during the collection period. According to the above method, each weibo in 70 weibos of data set A can generate a network with clear structure and hierarchy. To show the original structure of information dissemination intuitively and clearly, we show the network generated from raw data in a weibo in Fig. 1. That is to say, we can get 70 networks like Fig. 1(a) and Fig. 1(b) from data set A. It can be clearly seen in these figures that how a weibo is reposted one by one. We actually use the information transmission process to recover the users' relationship.

In order to make a comprehensive use of the data, we integrate all the networks in one data set. It is not difficult to find that many users participate in the spreading process of different weibos, which means there are many nodes shown in different graphs at the same time, and the relations between 70 networks are set up in this way. We identify the same node in different networks as a unique node, and obtain a more complicated repost network rather than 70 simple networks. By the same method, six integrated networks can be obtained(See Table 2). In Table 2, *N_raw*, *E_raw*, *D_raw* refer to the number of nodes, the number of edges and the average degree of the networks generated from raw data, respectively. *N_simplified*, *E_simplified*, *D_simplified* refer to the number of nodes, the number of edges and the average degree of the networks generated from the preprocessed data, respectively.

3. METHOD

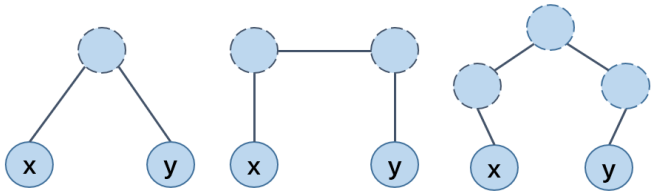
3.1 Indirect Repost Relationship Algorithm(IRRA)

First of all, we assign a similarity score S_{xy} to any two nodes x and y in the network. The larger the value S_{xy} is, the corresponding users X and Y is more likely to have a *FOLLOW* relationship. Then we further elaborate on the concept of indirect repost relationship. If node x can reach node y through n other indirect nodes in the networks, in which n is a constant natural number, we define there is indirect repost relationship between node x and node y. Given the value of n, we calculate S^n_{xy} for every node pair (x, y) in the network. Finally, we sort S^n_{xy} from high to low, then the higher the ranking of node pair is, the more likely there is a latent *FOLLOW* relationship between the corresponding users. Fig. 2(a) shows that node x can reach node y through one other indirect node, namely n=1. Fig. 2(b) shows that node x can reach node y through two other indirect nodes, namely n=2, and so on so forth. It's not difficult to imagine that node x is possible to reach y by different indirect nodes. x can reach y by more different combinations of indirect nodes means users X and Y share more indirect contacts, which also implies the possibility of latent *FOLLOW* relationship between them will be greater. As the result, we use the number of combinations of n indirect nodes as the similarity score, denoted as S^n_{xy}. When n=2, for example, S^2_{xy}=s if node x has s different combinations of 2 other nodes to reach node y (Here, s is a constant). In Fig. 3(a), it's

Table 2. Six graphs from six datasets

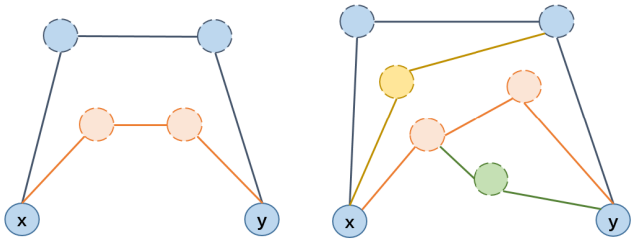
Data Set	Network	raw			simplified		
		N_raw	E_raw	D_raw	N_simplified	E_simplified	D_simplified
A	graph_1	29223	40951	2.8026	6287	9570	3.0434
B	graph_2	52447	58909	2.2464	7193	13356	3.7136
C	graph_3	33885	39710	2.3438	5641	11466	4.065
D	graph_4	7234	9995	2.7633	1248	2415	3.8702
E	graph_5	53462	65013	2.4321	8387	15783	3.7636
F	graph_6	67771	78898	2.3283	10805	21601	3.9983

easy to find out S^2_{xy}=2, and in Fig. 3(b), it's easy to count that S^2_{xy}=4.



(a) (b) (c)

Fig. 2. (a), (b), (c) give a typical example of indirect repost relationship between x and y when n= 1, 2, 3, respectively.



(a) (b)

Fig. 3. (a), (b) give a simple concrete example of S_{xy}=2 and S_{xy}=4 when n=2, respectively

3.2 Extension

Next, in order to further improve the accuracy of recovering, we take different n values into account at the same time inspired by the diverse node adaption algorithm (DNAA) proposed by Wang et al. (2017). The DNAA algorithm indicates that combining different link prediction algorithms to predict the links can effectively improve the accuracy, because different links may have different generation mechanisms even in the same network. In fact, it matches the appropriate algorithm to each node pair from a set of link prediction algorithms.

In this paper, considering that different users may prefer to establish indirect repost relationships by different number of nodes, we try to match the appropriate value of n for each node pair to improve the accuracy of inference. We denote the algorithm set as Ψ, and the IRRA algorithm with different values of n are placed in Ψ as the candidates. Hence Ψ=1,2,3 means the IRRA with n=1,2,3 are contained in set Ψ. It's obvious that when there are more values of n, the accuracy of the prediction is higher. However, the computational complexity of the algorithm also increases. After extensive simulations, it can be concluded that selecting 3 different values of n into Ψ is the best, taking both the computational complexity and the accuracy of the inference in to account. In addition, to consider all the indirect repost relationship comprehensively, the selected values of n are different in size and parity. As can be observed from the simulation results in the fourth part, the network recovering accuracy is greatly improved after extension.

4. EXPERIMENTS

4.1 Experiments Setup

In the simulation, seven classical link prediction methods are considered for comparison. They are Common Neighbors (CN), Jaccard Index(Jaccard), Salton Index (Salton), Resource Allocation Index (RA) as reported in Zhou et al. (2009), Adamic-Adar Index (AA) in Adamic and Adar (2003), Local Path (LP) in Lü et al. (2009), Katz Index (Katz) in Katz (1953).

At the beginning, the whole network is divided into training set E^T and test set E^P by 9:1. Then, all algorithms mentioned above are run on the training set E^T to get the possibility of latent link for each node pair in the network. Next, we verify the accuracy of the inference on the test set E^P. Finally, each algorithm is compared with the unified metric AUC.

The training set and the test set were redivided in each experiment, and each data result is obtained by taking the average after 20 independent repeated experiments.

4.2 Result Analysis

First of all, we examine the effect of n on the inference result. It can be intuitively seen from Fig. 4 that although the accuracy is also affected by different data sets, the highest accuracy is reached unanimously at n=2 in every data set. That is to say, users who are able to establish a repost relationship through two other users are more likely to have a *FOLLOW* relationship.

It is easy to understand why we cannot get better results when $n \geq 2$. Obviously, the larger n is, the farther the distance between x and y's indirect contact is. Therefore, when n is too large, the indirect relation between x and y is too far, which cannot reflect the real hidden relation between them. In the similar way, if node x and node y can be related by n other indirect nodes, we believe that the smaller n is, the more likely there is a link between them. In this way, the best recovering result should appear when n=1. Interestingly, the experimental results indicate the best performance at n=2.

The experiment is reliable although the result that achieves the best at n=2 is not completely consistent with our conventional thought. On the one hand, the data sets generated by six randomly selected bloggers here all conform to this conclusion. On the other hand, we also use the configuration model to generate six random networks with the same number of nodes,

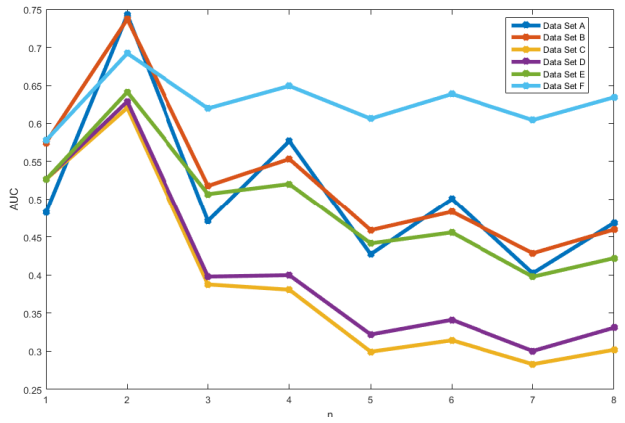


Fig. 4. Results with different values of n in every dataset. The horizontal axis represents n in our method, and the vertical axis represents the AUC result.

the same number of edges and the same degree sequences as real data sets. We find that these random networks do not have the best results at n=2 (take two of the data sets C,D for example, the results are shown in Fig. 5). This indicates that the result is caused by some inherent characteristics unique to the repost network. Therefore, it can be concluded that nodes can establish an indirect repost relationship by two other users are more likely to have the *FOLLOW* relationship.

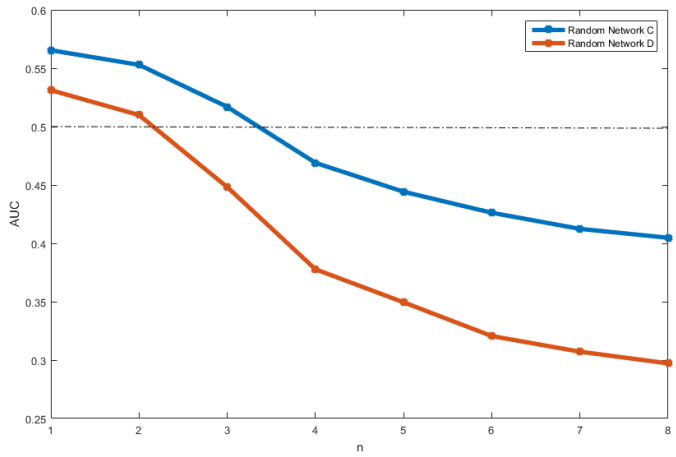


Fig. 5. Result with random networks corresponding to data sets C and D.

Table 3 compares the recovering accuracy when n equals 2 with those of other algorithms. The recovering accuracy of these algorithms are very low, and many of them are even less than 0.5, which is caused by the sparsity of the repost network structure as mentioned above. It's worth mentioning that the AUC of random guessing is 0.5 theoretically, and some algorithms do not work at all on this problem. It is obvious that our method of inferring on the indirect repost relationship is better than the classical link prediction algorithms.

In order to further improve the recovering accuracy of inferring the *FOLLOW* relationship from repost data, we consider different values of n at the same time, that is, to find a matching value of n for each pair of nodes. Table 4 shows that the accuracy of inference is the highest when Ψ=1,2,3,4,5,6,7. Actually the

Table 3. Comparison between our algorithm and existing algorithms

	A	*B*	*C*	*D*	*E*	*F*
CN	0.4826	0.5747	0.5266	0.5269	0.5264	0.5777
Jaccard	0.3684	0.5197	0.3691	0.3780	0.4970	0.5545
Salton	0.3645	0.5209	0.3751	0.3813	0.4951	0.5539
RA	0.5421	0.5956	0.5959	0.6154	0.5437	0.5918
AA	0.5496	0.5952	0.5949	0.6152	0.5451	0.5897
LP	0.6084	0.6268	0.5684	0.5900	0.5719	0.6573
Katz	0.5923	0.6188	0.5655	0.5904	0.5550	0.6399
IRRA	0.7425	0.7369	0.6206	0.6289	0.6921	0.6920

Table 4. Accuracy at different values of n

Ψ	*A*	*B*	*C*	*D*	*E*	*F*
all	0.8518	0.7841	0.8172	0.8273	0.7481	0.7889
2&3&6	0.8460	0.7776	0.8116	0.8259	0.7401	0.7786
2&3&7	0.8401	0.7747	0.8095	0.8271	0.7328	0.7757
3&4&6	0.8437	0.7743	0.8068	0.8177	0.7395	0.7459
3&4&7	0.8456	0.7763	0.8063	0.8210	0.7383	0.7746

result is not hard to understand, because we are thinking about the most comprehensive situation when Ψ=1,2,3,4,5,6,7.

However, due to the need to match the appropriate n value for each pair of nodes, when the value of n is more diverse, the computational complexity is also increased, especially that the number of nodes in our repost network is relatively large. In this case, we try to select several representative n values, and put them in Ψ to reduce the computational complexity, while the recovering accuracy is not significantly reduced. Although here we only show some representative combinations of n values, in fact, we traversed all the combinations during the experiment. Experiments show that when n takes three different values, the computational complexity can be greatly reduced, while the accuracy rate hardly decreases. It is important to note that not all combinations of three n values are able to have the same performance. Combinations shown in Table 4 is all cases in which the accuracy can be not significantly reduced, and other combinations will lead to a large decrease in accuracy. In terms of the results, this inference method greatly improves the accuracy of inferring *FOLLOW* relationships at a lower computational complexity, excellently solves the problem we are studying.

5. CONCLUSIONS

In the context of increasing attention to data protection and increasing difficulty of data acquisition, we try to infer more unknown information from the limited public information, and have proposed a novel perspective to recover *FOLLOW* relationship from repost information based on Sina weibo. We have developed a new algorithm based on indirect repost contacts, which noticeably outperforms other existing methods in the experiments. In addition, we find an interesting inherent characteristics in the repost relationship that the nodes have an indirect repost relationship through two reposts are more likely to have the *FOLLOW* relationship.

ACKNOWLEDGEMENTS

This work was partly supported by the National Natural Science Foundation of China (No. 71731004, No. 61425019, and No. 61751303).

REFERENCES

Adamic, L.A. and Adar, E. (2003). Friends and neighbors on the web. *Social networks*, 25(3), 211–230.

Aiello, L.M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., and Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web (TWEB)*, 6(2), 1–33.

Castellano, C. and Pastor-Satorras, R. (2010). Thresholds for epidemic spreading in networks. *Physical review letters*, 105(21), 218701.

Clauset, A., Moore, C., and Newman, M.E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191), 98–101.

De, A., Ganguly, N., and Chakrabarti, S. (2013). Discriminative link prediction using local links, node features and community structure. In *2013 IEEE 13th International Conference on Data Mining*, 1009–1018. IEEE.

Dong, Y., Tang, J., Wu, S., Tian, J., Chawla, N.V., Rao, J., and Cao, H. (2012). Link prediction and recommendation across heterogeneous social networks. In *2012 IEEE 12th International conference on data mining*, 181–190. IEEE.

Fouss, F., Pirotte, A., Renders, J.M., and Saerens, M. (2007). Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Transactions on knowledge and data engineering*, 19(3), 355–369.

Gomez, S., Diaz-Guilera, A., Gomez-Gardenes, J., Perez-Vicente, C.J., Moreno, Y., and Arenas, A. (2013). Diffusion dynamics on multiplex networks. *Physical review letters*, 110(2), 028701.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1), 39–43.

Li, X., Wang, J.B., and Li, C. (2016). Towards identifying epidemic processes with interplay between complex networks and human populations. In *2016 IEEE Conference on Norbert Wiener in the 21st Century (21CW)*, 1–5. IEEE.

Liang, D., Li, X., and Zhang, Y.Q. (2016). Identifying familiar strangers in human encounter networks. *EPL (Europhysics Letters)*, 116(1), 18006.

Liu, D., Wang, Y., Jia, Y., Li, J., and Yu, Z. (2014a). From strangers to neighbors: Link prediction in microblogs using social distance game. *Diffusion Networks and Cascade Analytics, WSDM*.

Liu, S., Perra, N., Karsai, M., and Vespignani, A. (2014b). Controlling contagion processes in activity driven networks. *Physical review letters*, 112(11), 118702.

Lü, L., Jin, C.H., and Zhou, T. (2009). Similarity index based on local paths for link prediction of complex networks. *Physical Review E*, 80(4), 046122.

Lü, L. and Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications*, 390(6), 1150–1170.

Ma, C., Chen, H.S., Li, X., Lai, Y.C., and Zhang, H.F. (2020). Data based reconstruction of duplex networks. *SIAM Journal on Applied Dynamical Systems*, 19(1), 124–150.

Van Lierde, H., Chen, G., and Chow, T.W. (2019). Scalable spectral clustering for overlapping community detection in large-scale networks. *IEEE Transactions on Knowledge and Data Engineering*.

Wang, H., Hu, W., Qiu, Z., and Du, B. (2017). Nodes' evolution diversity and link prediction in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 29(10), 2263–2274.

Wang, W. and Li, X. (2019). Temporal stable community in time-varying networks. *IEEE Transactions on Network Science and Engineering*.

Wu, J. and Li, X. (2018). Finite-time adaptive synchronization of drive-response two-layer networks. In *2018 IEEE Conference on Decision and Control (CDC)*, 1415–1420. IEEE.

Zhang, J., Chen, J., Zhi, S., Chang, Y., Philip, S.Y., and Han, J. (2017). Link prediction across aligned networks with sparse and low rank matrix estimation. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, 971–982. IEEE.

Zhang, J., Kong, X., and Philip, S.Y. (2013). Predicting social links for new users across aligned heterogeneous social networks. In *2013 IEEE 13th International Conference on Data Mining*, 1289–1294. IEEE.

Zhang, Y.Q. and Li, X. (2012). Characterizing large-scale population's indoor spatio-temporal interactive behaviors. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, 25–32.

Zhou, T., Lü, L., and Zhang, Y.C. (2009). Predicting missing links via local information. *The European Physical Journal B*, 71(4), 623–630.