

Identifying missing and spurious connections via the bi-directional diffusion on bipartite networks

Peng Zhang^a, An Zeng^{b,c,*}, Ying Fan^{c,*}

^a School of Science, Beijing University of Posts and Telecommunications, Beijing, 100876, PR China

^b Department of Physics, University of Fribourg, Chemin du Musee 3, CH-1700 Fribourg, Switzerland

^c School of Systems Science, Beijing Normal University, Beijing, 100875, PR China

Link prediction and spurious link detection in complex networks have attracted increasing attention from both physical and computer science communities, due to their wide applications in many real systems. Related previous works mainly focus on monopartite networks while these problems in bipartite networks are not yet systematically addressed. Containing two different kinds of nodes, bipartite networks are essentially different from monopartite networks, especially in node similarity calculation: the similarity between nodes of different kinds (called inter-similarity) is not well defined. In this letter, we employ the local diffusion processes to measure the inter-similarity in bipartite networks. We find that the inter-similarity is asymmetric if the diffusion is applied in different directions. Accordingly, we propose a bi-directional hybrid diffusion method which is shown to achieve higher accuracy than the existing diffusion methods in identifying missing and spurious links in bipartite networks.

Link prediction and spurious link detection problems are long-standing challenges in network study [1]. They have applications in many fields such as chemistry, biology, sociology and computer science. In many biological networks, such as food webs, protein-protein interaction networks and metabolic networks, whether a link between two nodes exists must be demonstrated by field or laboratorial experiments, which is usually very costly [2]. In addition, the data in constructing biological and social networks may contain inaccurate information, resulting in spurious links [3]. Correcting the network connections can be very expensive if it is done by laboratorial experiments. To solve this problem, many network-based methods have been developed and they are shown to have high accuracy in identifying missing and spurious links [4].

So far, most related works focus on monopartite networks in which only one type of nodes exists [5]. However, many systems coupled by two different building blocks should be modeled by bipartite networks. For example, the e-commercial systems consisting of online users and items [6], the scientific collaboration system consisting of authors and papers [7], family name inheritance system consisting of babies and names [8] are naturally described by such networks. The link prediction and spurious link detection problems are not yet well addressed in bipartite networks. One close problem is the so-called "network-based

recommendation" in which network structure information is used to generate recommendation list to online users [9–11]. This field aims at predicting the future links for each individual. However, which links will most likely to emerge at system level still remains unknown.

Compared to the link prediction, the spurious link detection is more difficult. Though many link prediction methods are claimed to be applicable to spurious link detection, they may lead to serious distortion of networks' structural and dynamical properties, as shown in Ref. [12]. Therefore, there is not yet a well-accepted method for this task. The spurious link detection in bipartite network is even newer and less touched, to the best of our knowledge. One closely related topic is the design of the reputation system in online user-object rating networks through which the malicious spammers can be identified [13].

Most link prediction and spurious link detection are based on calculating the similarity between nodes [14]. Most similarity measures in monopartite networks can be easily applied to calculate the intra-similarity (i.e. the similarity between nodes of the same kind) in bipartite networks. However, the inter-similarity (i.e. the similarity between nodes of different kinds) in bipartite networks is not yet well-defined [15,16]. In this letter, we employ a well-known hybrid diffusion process from recommender system to calculate inter-similarities in bipartite networks. The hybrid diffusion process was originally designed to be initialized from only one type of nodes [17]. Interestingly, we find that the obtained

* Corresponding authors.

E-mail addresses: an.zeng@unifr.ch (A. Zeng), yfan@bnu.edu.cn (Y. Fan).

inter-similarity will be entirely different if the direction of the diffusion process is reversed. A similar phenomenon has been found when coarse graining bipartite networks [18]. We therefore design a bi-directional hybrid diffusion method and find that it can achieve a much higher accuracy than the existing diffusion methods. We investigate the bi-direction hybrid diffusion in many different conditions and it is shown that the diffusion information from both directions are equally important in all the considered conditions.

1. Method

A bipartite network consists of two different types of nodes. Without loss of any generality, here we use the online commercial system as an example. In such systems, the two kinds of nodes are respectively users and items. The bipartite network with N users and M objects can be represented by an adjacency matrix A , where the element $a_{i\alpha}$ equals 1 if user i has collected item α , and 0 otherwise. (In this letter, item nodes will be labeled by Greek letters and user nodes will be identified by Latin letters.)

We first briefly introduce the hybrid diffusion method in Ref. [17]. The hybrid diffusion method is a combination of two independent algorithms called Mass diffusion (MD for short) [19] and Heat conduction (HC for short) [20]. For a user node i , the MD algorithm starts by assigning one unit of resource to each item collected by i , and redistributes the resource through the user-item network. We denote by the vector \mathbf{f} the initial resources on items, where the α -th component f_α^i is the resource possessed by object α . The three-step diffusion aims at estimating the inter-similarity between i and α , namely the likelihood of existing a link between these two nodes. In the first step, the resource reaches the objects selected by the target user. In the second step, the resource reaches the users who selected the same objects as the target user, and these users are usually refereed as relevant users. In the final step, the resource reaches the objects selected by the relevant users. These objects are more likely to be selected by the target user in the future. In real networks there can be many objects that reached by the diffusion but they are not with the same likelihood to be selected by the target user, the amount of resource on the object is the estimation of the likelihood. The inter-similarity between user i and items is obtained by setting the elements in \mathbf{f} to be $f_\alpha^i = a_{i\alpha}$, and redistributing them by $\tilde{\mathbf{f}} = W\mathbf{f}$, where

$$W_{\alpha\beta} = \frac{1}{k_\beta} \sum_{j=1}^N \frac{a_{j\alpha}a_{j\beta}}{k_j}, \quad (1)$$

is the diffusion matrix, with $k_\beta = \sum_{l=1}^N a_{l\beta}$ and $k_j = \sum_{\gamma=1}^M a_{j\gamma}$ denoting the degree of object β and user j respectively [19]. Physically, the diffusion is equivalent to a three-step random walk starting with k_i units of resources on the target user i .

The HC algorithm works similarly to the MD algorithm, but instead follows a conductive process represented by

$$W_{\alpha\beta} = \frac{1}{k_\alpha} \sum_{j=1}^N \frac{a_{j\alpha}a_{j\beta}}{k_j}. \quad (2)$$

Physically, the recommendation scores can be interpreted as the temperature of an item, which is the average temperature of its nearest neighborhood, i.e. its connected users. The higher the temperature of an item, the higher its similarity to the target user node [20]. The MD and HC methods are illustrated in Fig. 1(a).

The hybrid algorithm of MD and HC was proposed in [17], with the diffusion matrix given by

$$W_{\alpha\beta} = \frac{1}{k_\alpha^{1-\lambda}k_\beta^\lambda} \sum_{j=1}^N \frac{a_{j\alpha}a_{j\beta}}{k_j}, \quad (3)$$

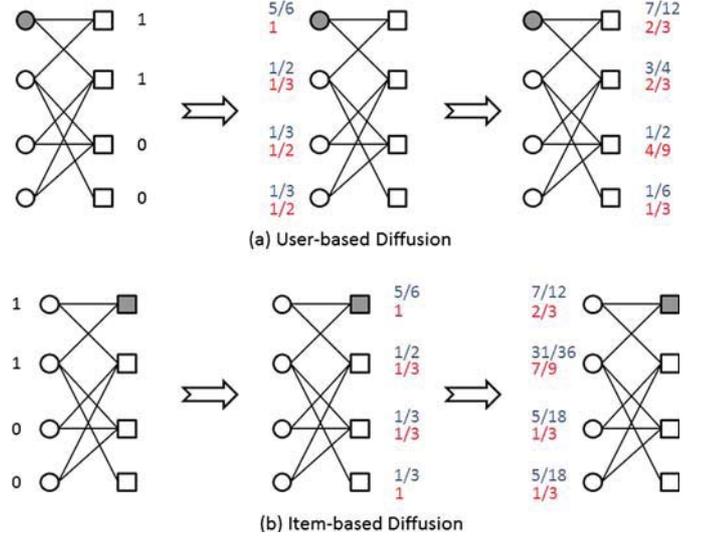


Fig. 1. (Color online.) The illustration of the (a) user-based and (b) item-based diffusion processes. The blue number next to a node is the resource the node received in the mass diffusion process. The red number is the resource the node received in the heat conduction process. The final resource on nodes can be used to estimate the inter-similarity between them and the target node.

where the parameter λ adjusts the relative weight between the two algorithms. When λ increases from 0 to 1, the hybrid algorithm changes gradually from HC to MD. By tuning the parameter λ , the inter-similarity between user nodes and item nodes will be more accurately measured.

The above diffusion methods are defined to be only initialized from the user side, while the diffusion from the item side is always ignored. Here, we consider the hybrid diffusion from the item side. We first set the initial resource in the user side as $f_i^\alpha = a_{i\alpha}$ and redistribute it in the bipartite network via

$$W_{ij} = \frac{1}{k_i^{1-\lambda}k_j^\lambda} \sum_{\beta=1}^M \frac{a_{j\alpha}a_{j\beta}}{k_\beta}. \quad (4)$$

Again, the hybrid algorithm is HC when $\lambda = 0$ and MD when $\lambda = 1$. This process is illustrated in Fig. 1(b).

In this letter, we denoted the diffusion initialized from the user side as user-based hybrid diffusion and the obtained similarity matrix as $\tilde{\mathbf{f}}^{\text{UHD}}$. The diffusion initialized from the item side is denoted as the item-based hybrid diffusion and the similarity matrix is written as $\tilde{\mathbf{f}}^{\text{IHD}}$. One can immediately see from Fig. 1 that $\tilde{f}_{i\alpha}^{\text{UHD}} \neq \tilde{f}_{\alpha i}^{\text{IHD}}$. To make use of the information of diffusion from both directions, we design a bi-direction hybrid diffusion (BHD for short) method in which the final inter-similarity matrix is calculated as

$$\tilde{f}_{i\alpha}^{\text{BHD}} = \theta \frac{\tilde{f}_{i\alpha}^{\text{UHD}}}{\sum_j \sum_\beta \tilde{f}_{j\beta}^{\text{UHD}}} + (1 - \theta) \frac{\tilde{f}_{\alpha i}^{\text{IHD}}}{\sum_j \sum_\beta \tilde{f}_{j\beta}^{\text{IHD}}}, \quad (5)$$

where θ is a tunable parameter which controls the weight between the user-based diffusion and item-based diffusion.

2. Data and metric

2.1. Data

To test the performance of the BHD method, we make use of two benchmark data sets: RYM and Econophys. The RYM data was obtained by downloading publicly available data from the music ratings Web site *RateYourMusic.com*. In RYM, users can rate the music from 1 to 10 (i.e., the worst to the best). We consider only the

rating higher than 5 as a link here. We use a random subset of this data with 986 users, 1197 items and 26681 links [17]. The Econophys data consists of a set of papers in the field of econophysics that were published between April 1995 and September 2010 in 78 scientific journals and an e-print server (i.e. arXiv.org) [21]. The network has 1608 distinct authors, 1204 papers and 19061 links. In these two networks, link prediction is to predict future ratings or citations. Both networks may also have spurious links since people may give ratings or cite papers carelessly or some malicious behavior exists. Therefore, it is naturally to test our method in these bipartite networks.

In the link prediction problem, each data is randomly divided into two parts: the training set contains 90% of the links (E^T) and the remaining 10% of links constitutes the probe set (E^P). The algorithm will run on E^T while E^P will be used to estimate the prediction accuracy. In the spurious link detection problem, the original data is called true network. 10% spurious links (E^S) are randomly added to the true network, which generate the observed network (E^O). Algorithms will be run on E^O , and E^S will be used to evaluate the detection accuracy.

2.2. Metric

In order to evaluate the accuracy of different algorithms, we adopt the *ranking score* (RS) index [22]. In the link prediction problem, RS measures whether the links in the probe set have highest rank among all nonexistent links. For each link $i\alpha$ in the probe set, its inter-similarity rank is denoted as $R_{i\alpha}$ among all L nonexistent links. Then RS can be calculated as

$$RS = \frac{1}{|E^P|} \sum_{i\alpha \in E^P} \frac{R_{i\alpha}}{L}. \quad (6)$$

According to the definition, a well-performed link prediction algorithm should have a small RS .

Likewise, we use the rank of spurious link to evaluate the accuracy of algorithms in the spurious link detection problem. For each link $i\alpha$ in the spurious link set, its rank based on inter-similarity (\tilde{r}) is denoted as $R_{i\alpha}$ among all L observed links. A good algorithm will rank the spurious links behind the true links, thus these spurious links are supposed to have large rank. In order to be consistent with the ranking score in link prediction problem, we calculate RS as

$$RS = 1 - \frac{1}{|E^S|} \sum_{i\alpha \in E^S} \frac{R_{i\alpha}}{L}. \quad (7)$$

In this way, it is still the smaller RS the better.

3. Results

We first investigate the performance of the BHD method under different parameters λ and θ . The heatmaps of RS in link prediction and spurious link detection are shown in Fig. 2. In each panel, we use the dashed line to mark the region where the BHD method outperforms the original hybrid method (with optimal λ when $\theta = 1$) in RS . One can see that in both data sets, the region is large. Moreover, the region mainly locates in the region with $0 < \theta < 1$, which indicates that a well-performed algorithm should combine the diffusion information initialized from both user and item sides. When applied to link prediction (see Fig. 2(a) and (b)), the BHD method will outperform the hybrid method by 8.2% in RYM and 15.0% in Econophysics. In the case of spurious link detection, the BHD method can achieve 22.1% higher RS than the hybrid method in RYM and 14.4% higher RS in Econophysics. The results imply that the BHD has bigger advantage in spurious link detection.

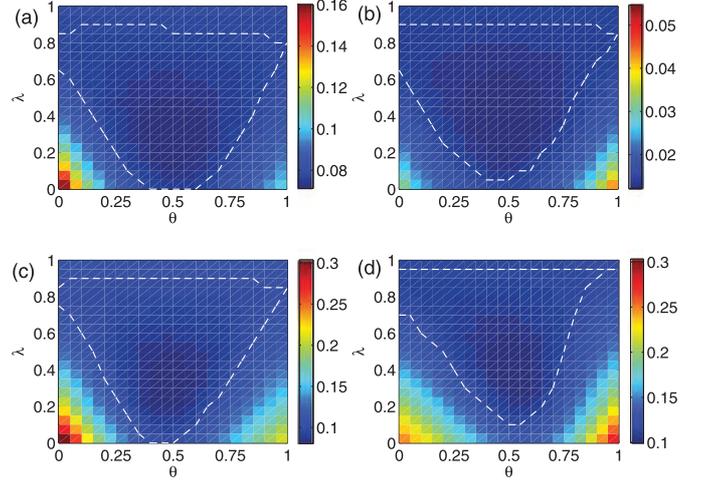


Fig. 2. (Color online.) Ranking score RS of the BHD method in the parameter space (θ, λ) in link prediction for (a) RYM and (b) Econophysics. RS of the BHD method in the parameter space (θ, λ) in spurious link detection for (c) RYM and (d) Econophysics. In each panel, the dashed line marks the region where RS is better than the best RS value achievable given $\theta = 1$.

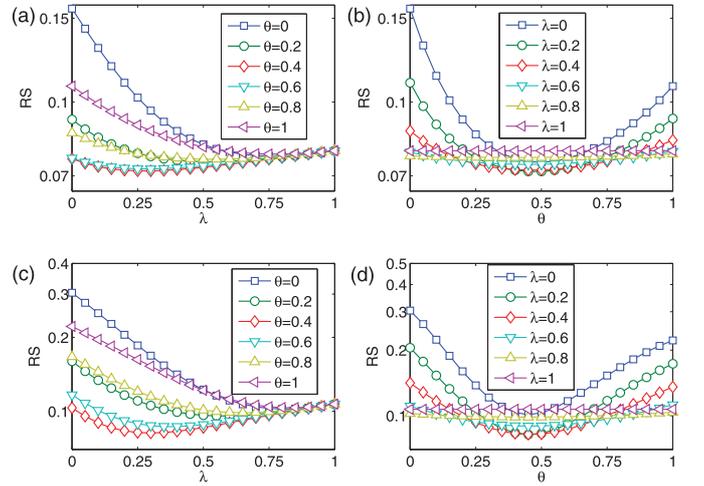


Fig. 3. (Color online.) (a) RS vs λ and (b) RS vs θ in link prediction. (c) RS vs λ and (d) RS vs θ in spurious link detection. The network in this figure is RYM. Note that the y-axis in this figure is in log scale.

In order to better understand the results in Fig. 2, we pick several θ and plot RS as a function of λ in Fig. 3(a), (c). One can see that there is an optimal RS^* when tuning λ , which is consistent with the results in [17]. We can also see that the optimal RS^* when $\theta < 1$ can be smaller than RS^* when $\theta = 1$, which confirms that the BHD can outperform the original hybrid method (where only user-based diffusion is considered). For completeness, we also show RS as a function of θ for several λ in Fig. 3(b), (d). Clearly, there is also an optimal RS^* when tuning θ . The optimal parameter θ is around 0.5, which indicates that the information from user-based diffusion and item-based diffusion are equally important. The optimal parameters for link prediction are $\theta^* = 0.5$ and $\lambda^* = 0.25$. For spurious link detection, the optimal parameters are $\theta^* = 0.5$ and $\lambda^* = 0.3$. The results in Fig. 3 are based on RYM data, we remark here that the results based on Econophysics data are similar.

In link prediction, it is generally difficult to predict the missing link of the nodes with small degree. This is known as the “cold-start” problem [23]. In [17], it has already been shown that the item cold-start problem can be well addressed in the hybrid method by tuning the parameter λ . More specifically, the

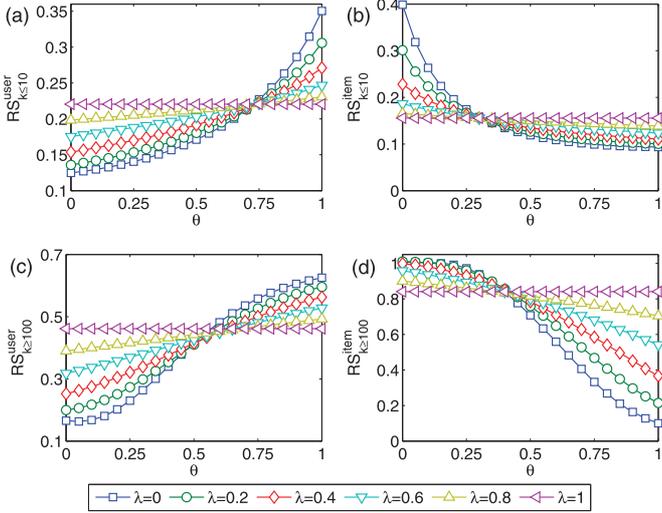


Fig. 4. (Color online.) (a) $RS_{k \leq 10}^{\text{user}}$ vs θ and (b) $RS_{k \leq 10}^{\text{item}}$ vs θ in link prediction. (c) $RS_{k \geq 100}^{\text{user}}$ vs θ and (d) $RS_{k \geq 100}^{\text{item}}$ vs θ in spurious link detection. The network in this figure is RYM.

prediction accuracy for small degree items can be largely improved when $\lambda < 1$. However, the user cold-start problem (i.e., predicting links for small degree users) cannot be solved by tuning λ . Actually, addressing the user cold-start is also of great importance from the commercial point of view. In real online systems, web sites are competing for users. In order to attract more users, web sites should provide new/inactive users with more accurate recommendations, which means that they have to more accurately predict the future links for small degree users. The cold-start problems have their counterparts in spurious link detection. Different from the link prediction case, the spurious links connecting to the large degree nodes are usually more difficult to detect. Therefore, it is important to improve the detection accuracy of spurious links for the nodes with large degree.

We argue here that the user cold-start problem can be solved by the BHD method. In order to show this, we pick all the users with degree smaller than 10 and calculate the average ranking score of all their links in the probe set as $RS_{k \leq 10}^{\text{user}}$. For completeness, we also calculate the average ranking score of the probe set links connecting to the items with degree smaller than 10 as $RS_{k \leq 10}^{\text{item}}$. In this way, we can study $RS_{k \leq 10}^{\text{user}}$ and $RS_{k \leq 10}^{\text{item}}$ as a function of θ for different λ in Fig. 4(a), (b). One can see that when $\theta = 1$ (i.e. the original hybrid algorithm), $RS_{k \leq 10}^{\text{item}}$ can be indeed reduced by tuning λ . However, $RS_{k \leq 10}^{\text{user}}$ will be significantly enlarged, which indicates that the user cold-start problem becomes even more serious. On the other hand, this problem cannot be well addressed also when $\theta = 0$ ($RS_{k \leq 10}^{\text{user}}$ is small but $RS_{k \leq 10}^{\text{item}}$ is too large). In BHD method where θ is around 0.5, one can see that both $RS_{k \leq 10}^{\text{user}}$ and $RS_{k \leq 10}^{\text{item}}$ can be improved by tuning λ . In the spurious link detection, we mainly consider the spurious links connecting to the nodes with degree larger than 100 in Fig. 4(c) and (d) (i.e., we focus on $RS_{k \geq 100}^{\text{user}}$ and $RS_{k \geq 100}^{\text{item}}$). Similarly to the link prediction case, both $RS_{k \geq 100}^{\text{user}}$ and $RS_{k \geq 100}^{\text{item}}$ can be improved when θ is around 0.5.

We further investigate the effects of data sparsity on the performance of the BHD method. In link prediction, we select a fraction $1 - p$ (p ranging from 0.1 to 0.7 with interval 0.1) links from the whole data set as the training set; the fraction p of the links form the probe set. Clearly, a higher p indicates a sparser known network (i.e., less available information). We report the minimum RS^* of both the BHD method and the hybrid method in Fig. 5(a) and (b). Note that RS^* of BHD is obtained by the double seek in

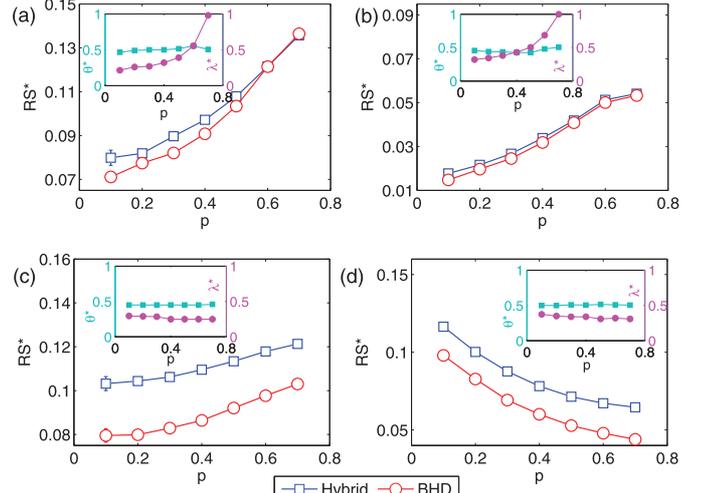


Fig. 5. (Color online.) The optimal ranking score RS^* of BHD and hybrid methods vs p in link prediction for (a) RYM and (b) Econophysics. The optimal ranking score RS^* of BHD and hybrid methods vs p in spurious link detection for (c) RYM and (d) Econophysics. The insets are the optimal θ and λ in BHD method as a function of p . The error bars are obtained based on 10 independent realizations. Some error bars are invisible since they are smaller than the size of the markers.

the parameter space (θ, λ) . Obviously, the BHD method has a lower RS^* than the hybrid method in both data sets. The advantage of the BHD method becomes smaller when p is large. This is because the available information in this case is very limited, all algorithms will have low detecting accuracy and their performance becomes almost the same. In spurious link detection, we add a fraction p of spurious links to the real network to create the observed network. The higher p is, the more difficult for the algorithms to detect spurious links. RS^* against p is shown in Fig. 5(c) and (d). In RYM, RS^* increases with p . However, the BHD method can still have a lower RS^* than the hybrid method under different p in both data sets. In Econophysics data, the RS^* decreases with p . This is because the Econophysics network is very sparse, a small number of spurious links will increase the connectivity and improve the accuracy of the spurious link detection. However, note that RS^* becomes worse when p is too large (e.g. $RS^* = 0.1365$ when $p = 5$). This phenomenon is very similar to a recent finding where the link prediction is improved by adding some virtual links [23].

The insets in Fig. 5 show the optimal θ in the BHD method. One can see that θ^* is rather stable around 0.5 in different p , which indicates that both user-based diffusion and item-based diffusion are useful under different conditions. In the insets of Fig. 5, the optimal λ in the BHD method is reported as well. One can see from Fig. 5(a), (b) that λ^* approaches 1 as the training set gets sparse, which is consistent with a recent finding [24]. This phenomenon is due to the fact that mass diffusion ($\lambda = 1$) works better than the heat conduction ($\lambda = 0$) in sparse data. In Fig. 5(c), (d), the λ^* is rather stable under different p , indicating that the optimal parameters are not sensitive to the fraction of spurious links in the networks.

4. Conclusion

Link prediction and spurious link detection in complex networks have been intensively studied recently. Most of related works are focused on monopartite networks. In this letter, we investigate the problem of missing and spurious link identification in bipartite networks. We propose a bi-directional hybrid diffusion method in which inter-similarity between nodes is estimated by the diffusion resource and the diffusion initialized from both types of nodes is considered. We find that this method can achieve

higher accuracy than the existing methods in identifying missing and spurious links in bipartite networks. Moreover, our method is very effective in solving the cold-start problems in information filtering.

In the literature, there are other important examples of spurious link detection approaches for bipartite networks (e.g. stochastic block model [25]), which, however, designed for networks with ratings. Moreover, the method is based on global algorithms that can be prohibitive to use for large-scale systems. Instead, our method would be easily applicable for large networks since it is based on local diffusion. We also remark that developing the spurious link detection approaches has many applications. For instance, a related method has been applied to predict future conflicts between team-members in social network [26].

Many real networks such as online social networks are modeled by directed networks [27]. So far, the link prediction in directed network is still a challenge [28–30]. It is already pointed out that the directed networks are similar to bipartite networks and in some cases the problems in directed networks can be mapped to bipartite networks [31]. Therefore, we believe that the bi-directional hybrid diffusion method in this letter can be easily extended to directed networks, and we expect our method to have high accuracy as well in directed networks. Moreover, by considering link senders as user-like nodes and link receivers as object-like ones, some link prediction methods for directed networks could be extended to bipartite networks [28–30]. However, such an adaptation may lead to the prediction of some self-loops. A systematic study of adapting methods between directed and bipartite networks will be an important subject for future investigation.

Acknowledgements

This work is supported by the Swiss National Science Foundation (Grant No. 200020-143272). Y. Fan acknowledges the National Natural Science Foundation of China under Grant No. 61174150, NCET-09-0228, the Doctoral Fund of the Ministry of Education

(20110003110027) and the major project of National Social Science Fund (12&ZD217).

References

- [1] L. Lu, T. Zhou, *Physica A* 390 (2011) 1150.
- [2] H. Yu, et al., *Science* 322 (2008) 104.
- [3] C. von Mering, R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Field, P. Bork, *Nature* 417 (2002) 399.
- [4] R. Guimera, M. Sales-Pardo, *Proc. Natl. Acad. Sci. USA* 106 (2009) 22073.
- [5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, *Phys. Rep.* 424 (2006) 175.
- [6] M.-S. Shang, L. Lü, W. Zeng, T. Zhou, Y.-C. Zhang, *Europhys. Lett.* 88 (2009) 68006.
- [7] F. Radicchi, S. Fortunato, B. Markines, A. Vespignani, *Phys. Rev. E* 80 (2009) 056103.
- [8] T.S. Evans, A.D.K. Plato, *Phys. Rev. E* 75 (2007) 056101.
- [9] L.-Y. Lü, M. Medo, C.H. Yeung, Y.C. Zhang, Z.K. Zhang, T. Zhou, *Phys. Rep.* 519 (2012) 1.
- [10] A. Zeng, C.H. Yeung, M.S. Shang, Y.C. Zhang, *Europhys. Lett.* 97 (2012) 18005.
- [11] L.-G. Liu, K. Shi, Q. Guo, *Phys. Rev. E* 85 (2012) 016118.
- [12] A. Zeng, G. Cimini, *Phys. Rev. E* 85 (2012) 036101.
- [13] Y.-K. Yu, Y.-C. Zhang, P. Laureti, L. Moret, *Physica A* 371 (2006) 732.
- [14] T. Zhou, L. Lu, Y.-C. Zhang, *Eur. Phys. J. B* 71 (2009) 623.
- [15] C.-J. Zhang, A. Zeng, *Physica A* 391 (2012) 1822.
- [16] P. Zhang, J. Wang, X. Li, M. Li, Z. Di, Y. Fan, *Physica A* 387 (2008) 6869.
- [17] T. Zhou, Z. Kucsik, J.G. Liu, M. Medo, J.R. Wakeling, Y.C. Zhang, *Proc. Natl. Acad. Sci. USA* 107 (2010) 4511.
- [18] Y. Wang, A. Zeng, Z. Di, Y. Fan, *Chaos* 23 (2013) 013104.
- [19] T. Zhou, J. Ren, M. Medo, Y.C. Zhang, *Phys. Rev. E* 76 (2007) 046115.
- [20] Y.-C. Zhang, M. Blattner, Y.K. Yu, *Phys. Rev. Lett.* 99 (2007) 154301.
- [21] Y.-B. Zhou, L. Lu, M. Li, *New J. Phys.* 14 (2012) 033033.
- [22] G. Adomavicius, A. Tuzhilin, *IEEE Trans. Knowl. Data Eng.* 17 (2005) 734.
- [23] F. Zhang, A. Zeng, *Europhys. Lett.* 100 (2012) 58005.
- [24] A. Zeng, A. Vidmer, M. Medo, Y.-C. Zhang, *Europhys. Lett.* 105 (2014) 58002.
- [25] R. Guimera, A. Llorente, E. Moro, M. Sale-Pardo, *PLoS ONE* 7 (2012) e44620.
- [26] N. Rovira-Asenjo, T. Gumi, M. Sale-Pardo, R. Guimera, *Sci. Rep.* 3 (2013) 1999.
- [27] B. Ball, M.E.J. Newman, *Netw. Sci.* 1 (2013) 16.
- [28] Q.-M. Zhang, L. Lu, W.Q. Wang, Y.X. Zhu, T. Zhou, *PLoS ONE* 8 (2012) e55437.
- [29] M. Kim, J. Leskovec, in: *Proceedings of 11th SIAM International Conference on Data Mining*, 2011, pp. 47–58.
- [30] J. Sanz, E. Cozzo, Y. Moreno, *J. Stat. Mech.* (2013) P12008.
- [31] W. Zhan, Z. Zhang, J. Guan, S. Zhou, *Phys. Rev. E* 83 (2011) 066120.