A new link prediction method for improving security in social networks

Kasra Majbouri Yazdi[1], Adel Majbouri Yazdi[2], Saeid Khodayi[3], Jiaojiao Jiang[4], and Saeed Saedy[5]
1. School of Information Technology, Deakin University.
2. Department of Computing, Kharazmi University.
3. Faculty of Computer & Electrical Engineering. Qazvin Islamic Azad University.
4. School of Information Technology, Deakin University.
5. Faculty of Engineering, Khavaran Higher Education Institute.

Abstract
Networks can represent a wide range of complex systems such as social, biological and technological systems. In such complex environments, there are many challenges and problems that can be studied and considered. One of the most important issues in such systems that has attracted a lot of interests in recent years, is link prediction. Many studies have been accomplished on link prediction over the last few years, but the existing approaches are not satisfactory in processing topological information as they have high time complexity. Many researches in traditional methods assume that endpoint influence represented by endpoint degree, prefers to facilitate the connection between big-degree endpoints. In this paper, we propose a new link prediction approach using Louvain community detection algorithm [25] for clustering nodes in different groups and then estimating future links by precise analyzing of the nodes relationships. The experimental results demonstrate that our proposed methods outperform the base methods. The comparison analysis with main stream baselines on 10 benchmark datasets shows that the results have been effectively improved on link prediction accuracy.

Key words:
Link prediction, community detection, Social networks, social network security

1. Introduction

Complex networks have got a significant potential in modeling and analyzing a wide range of complex systems such as social, biological and information systems. In other words, complex networks are used as a great tool in simulating and studying complicated systems in different fields of science [1].

Many social, biological, and information systems can be naturally described as networks where vertices represent entities (individuals or organizations) and links denote relations or interactions between vertices [2]. Networks or graphs are a powerful representation that has been employed in different tasks of machine learning (ML) and data mining (DM). This growing interest in use of graph can be justified by the expressiveness of this representation and its applications include: supervised learning, unsupervised learning [3], and semi-supervised learning, to cite just a few.

One of the most important topics of network analysis that has attracted increasing attention in recent years is link prediction. Link prediction aims to estimate the likelihood of the future existence of a link between two disconnected vertices in a network, based on the observed available links. Prediction, or forecasting, is a statement of an event uncertainty which is often but not always, based upon experience or knowledge. Link prediction has many applications in various sciences. For example, in biology, accurate prediction of protein–protein interaction has great value to hugely reduce the experimental costs. In some researches, link prediction algorithms were used in partially labeled networks to predict protein functions and also predicting research areas of scientific publication. Also link prediction can be used in social networks to recommend new relationships.

Social networks are social structures that consist of vertexes which are connected via several specific dependencies. One way of improving social network security is considering the existing nodes and their connections and predicting future links. Then recommending the future links to users in the network, can help them in detecting correct friendship relationships and decrease possibility of misusing and fraud/scam in social networks. This improvement on calculating future links, provide a safer community that its users are more match and reliable.

Researches on link prediction have got an important role on considering problems of network evolving mechanisms [4]. Qianming Zhang and Tao Zhou et al. [5] considered the problem of choosing the communication structure of networked systems. They showed that a small percentage of all possible communication links suffices to get short paths from the leader to all followers. This research derives an analytical expression for the expected path length in the network with the additional links and shows

Manuscript received May 5, 2018
Manuscript revised May 20, 2018

that a small number of additional links considerably shortens the path length from the leader to all followers.

Furthermore, recent findings on considering multiple evolution mechanisms of multiple complex networks shows that evolution of most networks is affected by both popularity and clustering.

Majority of learning-based link prediction methods that have been proposed so far can be divided into two categories: featured-based classification and probabilistic graph. In feature-based classification, it is crucial to define and extract a set of appropriate features from social networks. Li and Chen [6] presented a graph of kernel-based learning method and used features such as age, education level, book title, keywords and introduction to predict user-item link in a bipartite network. Scellato et al. [7] studied the problem of designing a link prediction system for online location-based social networks using social, placing and global features. In this work, they have gathered extensive data of periodic snapshots for Gowalla, which is one of the services, to capture its temporal evolution. They studied the link prediction space and their findings show that about 30% of new links are added among "place-friends", i.e., among users who visit the same places. The VCP metric can be seen as a kind of special feature which describes local topology information [8]. Al Hasan et al. [9] studied link prediction as a supervised learning task. They identified a set of features that are key to the performance under the supervised learning setup. Those features are very easy to compute and at the same time effective in solving the link prediction issues.

In this paper, we design an efficient incremental dynamic algorithm that can predict the relationships between users according to evolution of social network structure. In our proposed method, we used Louvain community detection algorithm [25] to improve understanding of the complex relationships.

The rest of the paper is organized as follows. Section 2, will discuss previous works and findings. Our proposed method is presented in section 3. The experimental results and discussion of algorithms are presented in Section 4. Finally, in Section 5, we draw a conclusion and perspectives of development of this work.

2. Related Work

Most of proposed link prediction approaches are based on network structure information. Also, most of them use similarity based metrics for forecasting connections which is highly depended on similarity proximity between two vertices. These methods are under a main assumption that the likelihood of the existence of a link between two nodes is proportional to the similarity between them [10]. The Common Neighbors (CN) metric [11] is a classic similarity-based link prediction index. It also tends to take specie attributes of the nodes into account rather than purely relaying on the graph structure. Some other indexes were proposed based on CN metric that considers the effects caused by different nodes, such as AdamicAdar (AA) [12] that weakens the contribution of large degree nodes and the Resource Allocation (RA) index [13]. Also, there are some other similarity-based indexes presented which depends on path information and random walk [14].

Analyzing the maximum likelihood of nodes on anticipating connections is another technique [15, 16], which includes Hierarchical Structure Model (HSM) and Stochastic Block Model (SBM) [15]. In this method, a general mathematical and computational framework is presented to deal with data reliability problem in complex networks. In particular, it allows to reliably identify both missing and fake interactions in noisy network observations.

HSM take advantage of hierarchical structure information to predict the missing links in the partly known networks with high precision. SBM can not only predict the missing links but also correct the spurious links, such as the spurious links in the protein–protein interaction network. In general, the accuracy of likelihood analysis methods is higher than similarity-based methods in most networks [17]. In addition, there are some other link prediction approaches that consider different criteria, such as probability model [18], machine learning [19] etc. Link prediction have been also studied on weighted [20], directed [21] and bipartite networks in addition to undirected and un-weighted networks.

In similarity-based approaches, similar vertices share the same relations (links). When the similarity between nodes is only based on the structure of the network, it is called structural similarity. Structural similarity measures can be classified in different ways such as local or global information [22].

Liben-Nowell and Kleinberg [11] and Zhou et al. [13] tried to systematically compare a number of structural similarity measures on real networks. According to their research results, global measures can provide higher accuracy, but its time complexity is very high and usually infeasible in large-scale networks, while local measures are generally faster but less accurate [23].

3. Proposed method

In this section, we introduce our method of link prediction which is based on community structure analysis. Community detection framework describes network from both macroscopic and microscopic structure. It means that our proposed method fully uses global structure information of the network when mapping the network into a hyperbolic environment. After mapping the network into a hyperbolic environment, the missing links are predicted according to the hyperbolic coordinates of the nodes. Despite of the existing mapping methods, our approach takes full advantage of the community structure and its hierarchical organization to map the network into the hyperbolic space. The community structure is an inherent property of complex networks which depicts the local aggregation characteristic and the heterogeneous distribution of the edges.

Link prediction in social network refers to anticipating creation of future links between unconnected nodes based on social metrics and general graph structure information. This prediction may have three conditions. First, predicting the links that exist in reality, but not in online network relationships. Second, predicting the links that do not exist in reality but will be generated in the future, and finally detecting wrong links which caused by erroneous online data or other reasons [35]. The distance between two neighboring nodes is smaller than unconnected nodes and if we consider each dimension as a feature of a node, the closer the two nodes are, the smaller difference they have. In a practical sense, it is obvious that two people with similar personalities are more likely to be friends and in a certain point, the assumption is consistent with the actual situation.

Communities are usually composed of nodes that are similar to each other in structure or in function. Also in real networks, they usually have hierarchical organization. It means they are composed of smaller communities, which in turn contain smaller communities. It is interesting that the structurally similar and closely connected nodes are usually aggregated in the same sector in the hyperbolic space. It accords with the community structure that is a cluster of similar nodes [36].

In comparison to previous works that use local feature of a network in finding similarities and neighborhood of nodes, we consider also the longer paths between neighbors to improve the current approaches. Unlike other approaches that take the maximum path between the source node and its neighbors, we suppose a node can connect to another node through other paths with various length.

In this section, we introduce the criteria for expressing similarities of nodes in a graph that we borrow from [37]. Let's assume Vi and Vj are two nodes and Sim=(Vi,Vj) is a function that estimate the similarities between them. Whatever the Sim is bigger, the possibility of their friendship relationship is higher. Let's say two users in an Online Social Network (OSN) want to have a connection, but the shortest path between them is blocked by a reluctant broker. If there is another path; even if it is longer or less usable; it is still possible that these two nodes use that path. In general, nodes can use all paths between them to connect instead of just using the shortest path. Thus, we expand the available approach in using shortest path by considering other possible paths between them. Therefore, two nodes that can connect to each other via many other unique paths, have a higher possibility of knowing each other proportional to the length of the paths that connect them together. For example, in fig 1, if we only consider paths with length of 2, then U4 and U7 have equal possibility to be recommended as a new friend of U1. However, if paths with length of 3 are considered as well, then U4 have higher likelihood to be recommended as U1's friend.

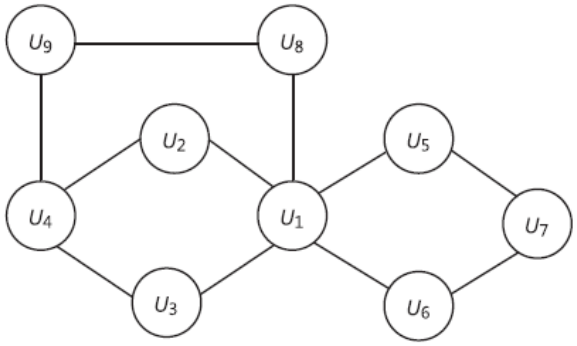

Fig. 1 Link prediction possibilities for U1

As mentioned above we use CCLP criteria [37] to calculate the similarities between nodes in a graph. These criteria use clustering factor in determining the similarities and it is as follows:

$$CCLP(x,y) = \sum_{z \in \Gamma(x) \bigcap \Gamma(y)} CC_z$$

$$CC_z = \frac{t_z}{k_z(k_z-1)/2}$$

tz denotes triangle paths (path with length of 3) that starts from z and also ends at z. kz denotes the degree of node k.

As it can be seen in the equation, whatever the similarities of nodes will be higher, CCLP is higher and whatever the similarities of nodes are smaller, the CCLP will be smaller.

The next step after calculating the similarities of nodes in a graph is the process of clustering. The purpose of clustering is to divide initial nodes into different clusters based on their similarities. Therefore, the nodes in a same cluster have more similarities together. Also features of different clusters are less similar together. Majority of proposed clustering approaches have some defects as follows [1]:

- The number of clusters should be determined before running the clustering algorithms in most clustering methods.
- Data distribution is one of the most important criteria in clustering nodes, but in most clustering approaches it is not considered properly. Considering the extent of data distribution can considerably improve the performance of the clustering algorithms.
- In most clustering methods, all features of nodes are considered equally during the process of clustering so it provides the same effects on clustering features. However, in real situation, some nodes may have more effects than other nodes based on their features.

The community structure is one of the most important patterns in network. Since detecting communities in a network can significantly improve understanding of the complex relations, many researches have been done over the last few years [38]. Classical clustering approach of k-means, is one of the most accurate methods in detecting communities. However, it is quite sensitive to the initial centroids or seeds [38, 39]. The goal of community detection is to cluster the similar vertices into one community and separate to others. Since the vertices in the same community share similar properties, the communities in the network provides deep understanding of complicated relations. Community detection also has many other applications in complex networks such as classification in social dimensions, finding influential bloggers and recommendation system. We use Louvain method [25] to detect communities in this work. Louvain algorithm tries to detect clusters in a network by maximizing a modularity function. This method is a simple, efficient and easy-to-implement method that can detects communities in very large networks in short computing time. It is based on the two simple steps. In the first step, each node is assigned to a community chosen in order to maximize the network modularity; the gain derived from moving a node into a community can simply be calculated as follows:

$$\Delta Q = \frac{\sum c + k_i^C}{2m} - \left(\frac{\sum c' + k_i}{2m}\right)^2 - \left[\frac{\sum c}{2m} - \left(\frac{\sum c'}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)\right]$$

Where $\sum c$ is the sum of the weights of the edges inside C, $\sum c'$ is the sum of the weights of the edges incident to nodes in C, ki is the sum of the weights of the edges incident to node i, kc i is the sum of the weights of the edges from i to nodes in c, m is the sum of the weights of all the edges in the network. The second step, makes a new network consisting of nodes in those detected communities. Then this process iterates until network modularity gain a significant improvement. Louvain method have been implemented in several network analysis software such as NetworkX [26] and Gephi [27]. In the given graph, we consider the relations of nodes and edges as interactions among them. Also, a k-dimension latent space is assumed so that each node has its own coordinates, then we can infer nodes specific location in latent space according to the interaction in reality. Since the position of the latent space is implicit, we need to infer the position of the node by an indirect method.

Our proposed method which is based on latent space model, is very accurate in theory, however its time complexity is relatively high. To moderate this complexity, we propose another algorithm, called community link prediction, which is based on improved resource allocation algorithm. The main idea of resource allocation is based on the community node metrics. That is to say, the more neighbors the two nodes have, the more correlation they have, and they are more likely to generate connections by introducing their neighbors. However common neighbors have two fatal disadvantages. On one hand, all nodes in the network will have the same effect on their neighbors by default, although, it does not accord with the reality. For example, a star as a popular user, has got a lot of people's attention in Weibo. If two strangers both have such a neighbor, his contribution to their similarity is actually very small. On the other hand, it does not consider the impact of the degree of two unconnected nodes. In real world, on the basis of same number of common neighbors, the persons with less friends would be easier to acquaint with each other. Thus, the smaller degree nodes would have more possibility to establish a connection. Thereby, we measure the impact of common neighbors and consider the degree of the unconnected nodes.

Input: Adjacency Matrix $W_1,..,W_t$, dimension of latent space k
Output: W_{t+1}'s prediction value Y_{t+1}, latent $Z_1,...Z_t$
Calculate community detection for all nodes
for τ in [1,t] do
—for u in V do
——If $W_\tau(u)$ has not changed then
———$Z_\tau(u) = Z_{\tau-1}(u)$
——else
—end
end
Calculate $Z_\tau(u)$ Affected node set S = ΔV_t
while Z t is convergence do
—for u in S do
——update Z_t (u)
—update (S)
return $Y_{t+1} = f([Z_1,..,Z_t])$ and $[Z_1,..,Z_t]$

Alg 1: Link production based on community structure

4. Experimental results

In this section, we discuss the experimental results of our proposed asymmetric link clustering (ALC) based link prediction methods on 8 network datasets collected from various fields. Also, the performance of our work is compared with three other neighborhood-based link prediction methods. Please note in most related literatures, the results are presented in solving globalized link prediction problem, but here we demonstrate the performance of our methods in predicting both globalized and personalized top-L latent links.

4.1 Datasets

The experimental datasets that we used in this work are as follows:

(1) US Air97 (USAir) [28]. USAir is the aviation network of USA.

(2) Yeast PPI (Yeast) [29]. Yeast is the protein interaction network which is for proteins and their interaction as node and edge respectively.

(3) Food Web of Florida ecosystem (FW) [30]. FW is the network of food chain in the rainy season of Florida. The prediction relationship is described by the carbon exchange relation in this network.

(4) Power [31]. Power is the American West electrical network. The nodes represent generator, transformer and substation, while the edges represent the high voltage transmission line between nodes.

(5) NetScience (NS). NS is the scientist cooperation network in which the node represents scientists and the edges are the cooperation relationships between scientists.

(6) C.elegans (CE) [31]. CE is a neural network of the worm Caenorhabditis elegans.

(7) E-mail network (Email) [32]. Email indicates the communication network which uses the email in University Rovira i Virgili (URV) in Tarragona, Spain.

(8) Jazz [33]. Jazz denotes the network including the collaboration between jazz musicians.

(9) Political blogs (PB). PB is the network constructed by the web pages of US political blogs. The edge represents the hyperlink of the web pages.

(10) Slavko. Slavko is the network involving the Facebook friendship of Slavko Žitnik.

The basic topological features of these networks are shown in Table 1 where |V| and |E| represent the numbers of nodes and links respectively. <k> is the average degree, <d> denotes the shortest average distance, indicates the clustering coefficient, r expresses the assortativity coefficient and H is the degree heterogeneity.

Table 1: The basic topology features of benchmark networks

Network	$\|V\|$	$\|E\|$	$\langle d \rangle$
USAir	332	2 126	2.73
Yeast	2364	10 904	5.16
FW	128	2 075	1.77
Power	4941	6 594	15.87
NS	1461	2 742	5.82
CE	453	2 025	2.66
Email	1133	5 451	3.61
Jazz	198	2 742	2.24
PB	1222	16 717	2.51
Slavko	334	2 218	3.05

4.2 Metric

To evaluate our method, AUC [34] is used as a standard metric to determine the performance of our work. AUC is the probability that the score of a link randomly chosen in the testing set is higher than that of a nonexistent link. (It is selection of a random link from the EP simulation continuously). In this method, each time a link is selected randomly from the EP testing set, and also another link is selected from the nonexistence link set; U\E. If the score of the selected link from EP is higher than the one selected from U \E, 1 point is given. If the score is equal, 0.5 point is given. After n times independent comparisons, if the links selected from the testing set have higher scores for n' times and also have equal score for n" the AUC is presented as:

$$Auc = \frac{n' + 0.5n''}{n}$$

If all the scores are generated randomly, AUC will approximately be equal to 0.5 obviously. The AUC greater than 0.5 measures the extent to which the index is more accurate than the randomly selected method. Furthermore, we use accuracy metric (Eq. (2)), which is predicting the correct ratio in the top-k prediction edge results. This metric produces top-k prediction results in which have edges belonging to the simulation (test set). Accuracy is defined by the following proportion:

$$Auccracy = \frac{m}{n}$$

The larger accuracy means higher performance for the prediction algorithm.

4.3 Results and discussions

In order to compare and evaluate all aforementioned prediction algorithms in terms of accuracy, we employ two verification modes; AUC and Accuracy. The experimental results are shown in Figs. 1 to 8. According to the results, the overall prediction of our algorithm on NC is better than other datasets. The reason is that Email dataset is large enough to give more structural information to the algorithm in comparison to Yeast dataset. In full-graph mode, it only infers test edges by static graph instead of dynamic information, therefore, the base paper is not as good as the traditional proposed method and common neighbor index algorithms in both AUC and accuracy metric. As the results show, our proposed method has more accuracy than other methods. For instance, the proposed methods obtained 84% accuracy in predicting correct links in CE dataset while the base paper [2] obtained 81%.

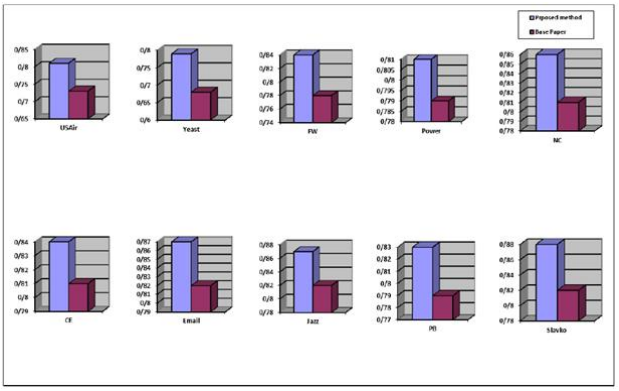

Fig. 2 AUC comparison results on different datasets

Also in another experiment, the performance of our method was evaluated in terms of Accuracy metric. Table 2, represents the comparison of link prediction accuracy of our approach against the base method.

As Table 2 depicts, the performance of our methods is higher than the base method for all datasets. For example, for the Yeast dataset, the prediction accuracy of our method is 8 percent bigger than the base method.

Table 2: Comparison of link prediction accuracy

Dataset	Proposed method Accuracy	Base paper
USAir	87.14	84.32
Yeast	89.82	81.68
FW	86.81	80.09
Power	82.45	76.91
NS	83.12	76.37
CE	79.91	74.12
Email	85.35	81.72
Jazz	81.78	76.83
PB	85.72	79.81
Slavko	84.29	80.06

Furthermore, in another experiment the performance of our proposed method and the base paper is compared in terms of time complexity. Table 3 shows the run time duration of link prediction for both methods.

Table 3: Run time duration of the proposed method with the base method in milliseconds

Dataset	time of Proposed method	time of Base paper
USAir	112	162
Yeast	354	391
FW	281	306
Power	561	697
NS	119	162
CE	276	286
Email	312	345
Jazz	352	398
PB	4361	4981
Slavko	471	508

As shown in table 3, in this experiment also, the proposed method has a better result in predicting the links in a shorter period of time.

5. Conclusions

Link prediction is an important interesting topic that many studies have been accomplished around it in recent years. However, the existing methods are unsatisfactory for processing topological information. Also, they have high time complexity. This paper presents a novel approach based on community structure. It deals with the network in overall perspective and takes advantage of the community structure and its hierarchical organization to map the network into hyperbolic space and then predicts missing links. The performance of our proposed method is compared to the well-known and state-of-the-art available

methods in the different aspects of classification accuracy. The results show that our method obtains a better classification accuracy. Furthermore, the results indicate that the execution time of the proposed method is comparable to those of the multivariate feature selection methods.

References

[1] Zhao, Y.-D., et al., Coarse cluster enhancing collaborative recommendation for social network systems. Physica A: Statistical Mechanics and its Applications, 2017. 483(Supplement C): p. 209-218.

[2] Zhang, Z., et al., Efficient incremental dynamic link prediction algorithms in social network. Knowledge-Based Systems, 2017. 132(Supplement C): p. 226-235.

[3] Pacheco, F., et al., Attribute clustering using rough set theory for feature selection in fault severity classification of rotating machinery. Expert Systems with Applications, 2017. 71: p. 69-86.

[4] Charu Aggarwal and Huan Liu, Negative Link Prediction in Social Media Jiliang Tang, Shiyu Chang, 2015.

[5] Lunze, J., Structural properties of networked systems with random communication links. Automatica, 2017. 80: p. 300-304.

[6] X. Li , H. Chen , Recommendation as link prediction in bipartite graphs: a graph kernel-based machine learning approach, Decis. Support Syst. 54 (2) (2013) 880–890.

[7] S. Scellato , A. Noulas , C. Mascolo , Exploiting place features in link prediction on location-based social networks, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2011, pp. 1046–1054.

[8] R.N. Lichtenwalter , N.V. Chawla, Vertex collocation profiles: subgraph count- ing for link analysis and prediction, in: Proceedings of the 21st International Conference on World Wide Web, ACM, 2012, pp. 1019–1028.

[9] M. Al Hasan , V. Chaoji , S. Salem , M. Zaki , Link prediction using supervised learning, SDM06: Workshop on Link Analysis, Counter-Terrorism and Security, 2006 .

[10] D. Lin, An information-theoretic definition of similarity, in: Proceedings of the Fifteenth International Conference on Machine Learning, Morgan Kaufmann Publishers Inc., 1998, pp. 296–304.

[11] D. Liben-Nowell, J. Kleinberg, The link-prediction problem for social networks, J. Amer. Soc. Inform. Sci. Technol. 58 (7) (2007) 1019–1031.

[12] L.A. Adamic, E. Adar, Friends and neighbors on the web, Social Networks 25 (3) (2003) 211–230.

[13] T. Zhou, L. Lü, Y. Zhang, Predicting missing links via local information, Eur. Phys. J. B 71 (4) (2009) 623–630.

[14] E.A. Leicht, P. Holme, M.E.J. Newman, Vertex similarity in networks, Phys. Rev. E 73 (2) (2006) 026120.

[15] R. Guimerà, M. Sales-Pardo, Missing and spurious interactions and the reconstruction of complex networks, Proc. Natl. Acad. Sci. USA 106 (52) (2009) 22073–22078.

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

[17] L. Lü, T. Zhou, Link Prediction, High Education Press, 2013.

[18] H. Kashima, N. Abe, A parameterized probabilistic model of network evolution for supervised link prediction, in: Sixth International Conference on Data Mining, 2006. ICDM'06, IEEE, 2006, pp. 340–349.

[19] B. Chen, L. Chen, A link prediction algorithm based on ant colony optimization, Appl. Intell. 41 (3) (2014) 694–708.

[20] J. Zhao, L. Miao, J. Yang, H. Fang, et al., Prediction of links and weights in networks by reliable routes, Sci. Rep. 5 (2015) 12261.

[21] Q.M. Zhang, L. Lü, W.Q. Wang, Y.X. Zhu, T. Zhou, Potential theory for directed networks, PLoS One 8 (2) (2013) e55437.

[22] Chuan Shi, et al., A link clustering based overlapping community detection algorithm. Data & Knowledge Engineering, 2013. 87: p. Pages 394–404.

[23] Yakun Li , et al., Efficient community detection with additive constrains on large networks. Knowledge-Based Systems, 2013. 52: p. Pages 268–278.

[24] Yawen Jiang , Caiyan Jia, and J. Yu, An efficient community detection method based on rank centrality. Physica A, 2013. 392(9): p. Pages 2182–2194.

[25] V. Blondel, et al., Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008. 10008: p. pp. 1–12.

[26] A.A. Hagberg, D.A. Schult, and P.J. Swart, Exploring network structure, dynamics, and function using NetworkX. in: Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, CA, USA, 2008: p. pp. 11–15.

[27] Bastian, M., et al.,, Gephi: an open source software for exploring and manipulating networks. In: International AAAI Conference on Weblogs and Social Media, North America,, 2009.

[28] V. Batagelj, A. Mrvar, Pajek-program for large network analysis, Connections 21 (2) (1998) 47–57.

[29] D. Bu, Y. Zhao, L. Cai, et al., Topological structure analysis of the protein-protein interaction network in budding yeast, Nucleic Acids Res. 31 (9) (2003) 2443–2450.

[30] J. Bascompte, Food web cohesion, Ecology 85 (2) (2004) 352–358.

[31] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks, Nature 393 (6684) (1998) 440–442.

[32] R. Guimera, L. Danon, A. Diaz-Guilera, et al., Self-similar community structure in a network of human interactions, Phys. Rev. E 68 (6) (2003) 065103.

[33] P.M. Gleiser, L. Danon, Community structure in jazz, Adv. Complex Syst. 6 (04) (2003) 565–573.

[34] J.A. Hanley, B.J. McNeil, A method of comparing the areas under receiver operating characteristic curves derived from the same cases, Radiology 148 (3) (1983) 839–843.

[35] Shikhar Sharma and Anurag Singh. An efficient method for link prediction in weighted multiplex networks. Computational Social Networks, 3(1):7, Nov 2016.

[36] Yiwen Cui, Lingling Zhang, Quandong Wang, Peng Chen, and Chunyu Xie. Heterogeneous network linkage-weight based link prediction in bipartite graph for personalized

recommendation. Procedia Computer Science, 91:953 – 958, 2016. Promoting Business Analytics and Quantitative Management of Technology: 4th International Conference on Information Technology and Quantitative Management (ITQM 2016).

[37] Zhihao Wu, Youfang Lin, Jing Wang, and Steve Gregory. Efficient link prediction with node clustering coefficient. Physica A: Statistical Mechanics and its Applications, 452, 10 2015.

[38] Messaoudi Imane and Kamel Nadjet. Hybrid bat algorithm for overlapping community detection. IFAC-PapersOnLine, 49(12):1454 – 1459, 2016. 8th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2016.

[39] Parham Moradi and Mehrdad Rostami. A graph theoretic approach for unsupervised feature selection. Eng. Appl. Artif. Intell., 44(C):33–45, September 2015.