IJICTR International Journal of Information & Communication Technology Research ITRC

Volume 8- Number 1- Winter 2016 (9-14)

Ensemble Link Predictor for Heterogeneous Complex Networks

Hadi Shakibian
Faculty of Electrical
and Computer Engineering
Tarbiat Modares University
Tehran, Iran
h.shakibian@modares.ac.ir

Nasrollah Moghadam Charkari
Faculty of Electrical
and Computer Engineering
Tarbiat Modares University
Tehran, Iran
moghadam@modares.ac.ir

Saeed Jalili
Faculty of Electrical
and Computer Engineering
Tarbiat Modares University
Tehran, Iran
sjalili@modares.ac.ir

Received: November 26, 2015- Accepted: March 2, 2016

Abstract— **Link prediction in complex networks aims to explore similarities between node pairs. Recently, link prediction has been considered in the presence of network heterogeneity which makes the majority of the homogeneous link prediction approaches infeasible. A meta-structure, known as meta-path, has been proposed to explore such networks. Generating good meta-paths and selecting the best of them introduce some new challenges to link prediction problem. In this paper, a new ensemble-based link prediction approach is proposed in heterogeneous complex networks. This approach consists of three steps: (*i*) a set of meta-paths are selected such that each of them represents a different semantic between the target node pairs; (*ii*) a feature vector is extracted for each node pair using each meta-path; (*iii*) an ensemble of learners would be established on different feature sets. The final link predictor is obtained after the ensemble aggregation. The results on DBLP network show that the proposed approach has more accurate predictions than a single meta-path based link predictor.**

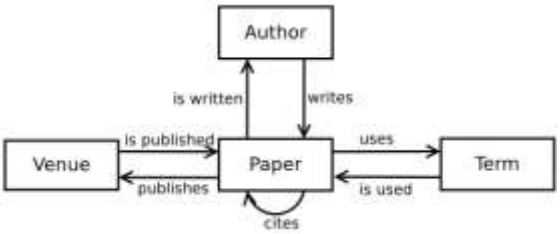
Keywords-Heterogeneous Complex Networks; Link Prediction; Meta-path

I. INTRODUCTION

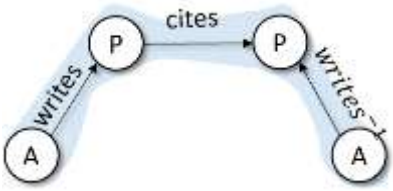
Many real-world systems are modeled as complex networks in order to exploit the dependencies between a group of linked entities. When entities are appeared in different types, the system can be modeled as a heterogeneous complex network (HCN). Among the well-known HCNs, Heterogeneous Information Networks (HINs) have been well studied in recent years which have many real-world examples including bibliographic, Flickr, Twitter, and healthcare networks [1].

Mining interesting links in complex networks is a crucial issue. The aim of this problem, known as link prediction, is to exploit the potential relationships between any pair of entities which has many interesting applications in real networks such as: extracting potential interactions between two proteins in PPI networks [2], friendship recommendation in online social networks, detecting spam emails [3], detecting selfish nodes in mobile social networks [4], and collaboration and citation prediction in bibliographic networks [5].

Network sparsity, network size, and extracting similarity features are three main challenges in link prediction problem. Most of the proposed link predictors assume the network is homogeneous. These



(a)



(b)

Fig. 1. (a) Network schema of DBLP network [24]; (b) an example of a meta-path A-P-P-A.

approaches are divided into three categories [6]. In the first category, local similarity features are extracted from the network. These measures are generally based on the neighborhood set *N(x)* or degree connectivity *D(x)* of each individual node *x*. Although, they are computationally efficient, their accuracy is poor. In the second category, global similarity measures are extracted which are mainly based on the shortest path between the node pairs. These measures provide better accuracy but their computational complexities are high. In the third category, a model is assumed for the existing network and the likelihood of the model is maximized by a probabilistic technique. It has been shown that the probabilistic link prediction methods become slow when the network grows in size. Moreover, it depends on a prior assumption about the network. A good survey of link predictions methods has been provided in [6].

In HCNs, different types of entities and relationships add some new challenges for predicting links. In this regard, the concept of meta-path has been proposed in order to explain the relation of the pair nodes in HCNs. A meta-path is sequence of nodes and edges types that follows a network schema. In Fig. 1 the network schema of DBLP network and a typical meta-path are shown.

In recent years, a number of studies have been worked on the concept of meta-paths. In [7], [8], [9], and [10], new meta-path based similarity measures have been proposed denoted as *Path Count*, *Path Constraint Random Walk*, *PathSim*, and *HeteSim*, respectively. Using these measures, the prediction of the relationships between two node types can be derived. For example, in DBLP network [11], they can be used for predicting the co-authorships or future citations between a pair of authors or papers, respectively.

In [12] a recommender system has been proposed that considers different recommendation factors via different meta-paths. In [13] meta-path based Heterogeneous Collective Link Prediction (HCC) has been proposed that effectively assigns entity class to one type of objects in the network. The proposed HCC model can exploit a large number of different types of dependencies among objects simultaneously [13].

In [14], the problem of collective prediction of multiple types of links in heterogeneous information networks has been studied. The authors introduce the linkage homophily principle and design a relatedness measure, called RM, between different types of objects to compute the existence probability of a link. Also, an iterative framework for heterogeneous collective link prediction, called as HCLP, has been proposed to predict multiple types of links collectively by exploiting diverse and complex linkage information in heterogeneous information networks.

Automatic discovery of meta-paths in large scale HCNs has been studied in [15]. In the proposed method, the users are asked to provide example pairs of nodes that exhibit high proximity. Then, generating meta-path that best explain the given node pairs are investigated by a greedy algorithm.

Associating a feature vector and a binary {0, 1} label to each node pair, link prediction problem can be considered as a two-class learning problem [16]. In [17], a supervised link predictor, denoted as MLI (Multi-network Link Identifier), has been proposed in partially aligned social networks. The authors define a set of meta-social-paths in the network and extract some topological features per meta-path. Integrating the features, the link predictor is achieved after optimizing a probabilistic learning method in which one optimization function is assigned to each aligned network. Transferring information between the aligned networks in MLI method leads to achieve the higher accuracy.

All the proposed meta-path based link predictors deal with three main challenges: defining and generating valid/useful meta-paths, selecting the best meta-paths, and extracting meta-path based similarity measures. In this paper, an ensemble link predictor is proposed in three steps. Firstly, a set of meta-paths is selected that covers different semantics between the node pairs. Next, some topological features are extracted based on each meta-path for each node pair. Then, by using the feature sets on a learning algorithm, an ensemble of link predictors would be obtained. Finally, the result which will be aggregated to compute the final predictor. The main contributions of this paper are as follows:

- Learning an ensemble of link predictors and combining them leads a better coverage on the dependency relations between the node pairs. This reduces the negative impact of the meta-path selection task by decreasing the probability of missing high accurate meta-paths.

- In order t deal with the network sparsity and consequently imbalanced data problem, Twin SVM is employed which inherently has been proposed for imbalanced data sets.
- To the best of our knowledge, this is the first meta-path based supervised framework for predicting links in HCNs.

The paper is organized as follows. In section 2, the problem is formulated. The proposed ensemble link predictor is introduced in section 3 and is evaluated in section 4. Finally the last section is concluding remarks.

II. Problem Definition

In this section, basic concepts used in this paper and the problem of heterogeneous link prediction are formally stated.

A. Terminology Definitions

Definition 1. *Heterogeneous Complex Network: A complex network is heterogeneous if it contains different types of nodes and links. A heterogeneous complex network can be represented as a graph* $G=<\Gamma, \mathrm{E}>$, *in which* $\Gamma = \{V_1, \ldots, V_m\}$ *and* $\mathrm{E} = \{E_1, \ldots, E_n\}$ *denote to the union of different node sets and heterogeneous link sets, respectively.*

The network used in this paper is DBLP. It is a heterogeneous information network in which four types of nodes are exist: Authors *(A)*, Papers *(P)*, Venues *(V)*, *and* Terms *(T)*. In the proposed approach, the nodes of type *T* have been omitted and therefore $\Gamma = \{A, P, V\}$.

Definition 2. *Network Schema: The network schema is a meta-structure defined over* $G=<\Gamma, \mathrm{E}>$. *It is constructed using an object type mapping as* $\varphi : \Gamma \rightarrow \Omega$ *and a link mapping as* $\psi : \mathrm{E} \rightarrow R$ *where* Ω *and* R *denote the node and link type sets, respectively. Therefore, a network schema can be represented as* $T_G = <\Omega, R>$.

In Fig. 1, Ω contains four node types as *A*, *P*, *V*, and *T* and *R* includes any link type between these types.

Definition 3. *Meta Path: A meta-path P is defined over the network schema* $T_G = <\Omega, R>$. *It refers to a composite relation between any pair of nodes with types* Ω_i *and* Ω_j, *and has the following form:*

$$\Omega_i \xrightarrow{R_i} \Omega_{i+1} \xrightarrow{R_{i+1}} \ldots \xrightarrow{R_{j-1}} \Omega_j$$

A meta-path is called symmetric (asymmetric) if the corresponding composite relation is symmetric (asymmetric). Also, the length of a meta-path is defined as the number of links in the corresponding composite relation.

Definition 4. *Path Instance: Given a meta-path P in* $T_G = <\Omega, R>$, *a path* $p = (p_1, \ldots, p_L)$ *is called a path instance of P, if* $\varphi(p_i) = \Omega_i$ *for* $\forall p_i \in p$. *In other words, a path instance of a meta-path P is any path p that follows the meta-structure of P.*

In HCNs, it has been shown that the most important topological features of meta-paths are computed by counting the number of path instances between any pair of nodes.

B. Heterogeneous Link Prediction

Suppose a heterogeneous complex network is given as $G=<\Gamma, \mathrm{E}>$. Also, assume that the similarity measures of any node pair $(x, y) \in \mathrm{E} \cup \mathrm{E}'$ is defined in a feature vector *F* where $\mathrm{E}' = \mathrm{M} \setminus \mathrm{E}$ and M is the set of all possible links of type $(\varphi(x), \varphi(y))$. Now, the aim of link prediction is to obtain a score function $f : F \rightarrow R$ that assigns a real score from *R* to a given feature vector from *F*.

If the score of each link is chosen from the set of {-1, +1}, the link prediction problem can be considered as a two-class classification problem. In this case, a set of example of node pairs is firstly selected. Then, the label of each node pair is set to +1 or -1 according to the existence or non-existence of that link. Now, a learning algorithm can be applied on the feature space *F* and class labels {-1, +1}. The focus of this paper is on the supervised link prediction. As stated in definition 1, DBLP network has been chosen for the experiments and the goal is to predicting the *coauthorships*.

III. Proposed Approach

In this section, the proposed approach is described in two steps. In the first step, constructing an ensemble of link predictors based on the selected meta-paths is introduced. In the second step, the method of ensemble aggregation is discussed.

A. Constructing an ensemble of link predictors

Meta-Paths. From the network schema of DBLP shown in Fig. 1, the nodes of type *T* have been omitted. Instead, a set of terms is extracted for each venue. Then, a pearson score is computed between the venue nodes according to their associated terms. Dosing this, two venues will have a relation if their pearson score is equal or greather than 0.75.

From the constructed network, three meta-paths are selected as follows:

$$P_1 = Author \xrightarrow{writes} Paper \xrightarrow{publishedIn} Venue \xleftarrow{publishedIn^{-1}} Paper \xleftarrow{writen^{-1}} Author$$

$$P_2 = Author \xrightarrow{writes} Paper \xrightarrow{cites} Paper \xleftarrow{writes^{-1}} Author$$

$$P_3 = Author \xrightarrow{writes} Paper \xrightarrow{publishedIn} Venue \xrightarrow{HasSameScope} Venue \xleftarrow{publishedIn^{-1}} Paper \xleftarrow{writen^{-1}} Author$$

where -1 denotes to the inverse relation. Each meta-path shows a different semantic and describes the authors dependencies from a different point of view. For example, the first meta-path measures the similarity of two authors based on their published papers by a single publisher. Although, the above meta-paths have been selected practically, it is possible to use more meta-paths and increase the ensemble size.

Feature Extraction. In order to exploit the future co-authorships between an author pair, three topological features are extracted for each defined meta-path.

- *Path Count (PC)*. Using definition 4, given a meta-path P, $PC_P(A_i, A_j)$ is computed by counting the number of path instances between two authors A_i and A_j following P. Simply, the larger the number of paths satisfy a meta-path, the more similar the author pair is [15].

- *Normalized Path Count (NPC)*. This measure normalizes the PC measure by the degree of each author as follows:

$$NPC_P(A_i, A_j) = \frac{PC_P(A_i, A_j) + PC_{P^{-1}}(A_i, A_j)}{PC_P(A_i,.) + PC_P(., A_j)} \quad (1)$$

where $PC_P(A_i,.)$ is the number of path instances starting from A_i following P, and $PC_P(., A_j)$ is the number of path instances ending to A_j following P. This measure penalizes the nodes that have very high degree.

- *Random Walk (RW)*. This measure defines the similarity of two authors A_i and A_j as a random walk starting at A_i, following meta-path P, and ending at A_j. $RW_P(A_i, A_j)$ is computed as follows:

$$RW_P(A_i, A_j) = \frac{PC_P(A_i, A_j)}{PC_P(A_i,.)} \quad (2)$$

Ensemble Construction. Each meta-path can be used as an independent link predictor. The major drawback of using a single meta-path as a predictor is that it might not make accurate predictions for all node pairs. The idea of ensemble link predictor, proposed in this paper, is to keep all the discovered meta-path in an ensemble.

Suppose a set of training author pairs is given. In order to construct the ensemble, three topological features, formulated as in above, are extracted based on each meta-path for the given training pairs. So, three training sets are created. Now, a learning algorithm can be applied on each training set to obtain a link predictor. To do this, Twin Support Vector Machine (TSVM) algorithm is used.

Inherently, TSVM has been known as an imbalanced-data based classifier in the literature. Moreover, it is 4 times faster than simple SVM [18]. Recalling the main challenges of link prediction problem from section 1, TSVM might obtain promising results.

Suppose the positive and negative training author pairs are given in two matrices $[A^+]_{n\times k}$ and $[A^-]_{m\times k}$, respectively. The number of dimensions k is 3 in our work. The linear TSVM finds two non-parallel hyperplanes in R^k as:

$$x^T w_1 + b_1 = 0 \quad (3)$$

$$x^T w_2 + b_2 = 0 \quad (4)$$

such that each hyperplane is proximal to data points of one class and as far as possible from the data points of the other class [23]. In Eq. (3) and (4) $w_1, w_2 \in R^k$ and $b_1, b_2 \in R$. An author pair will have collaboration in future if its feature vector in R^k space be closets to the first hyperplane in Eq. (3). The linear TSVM solves two quadratic problems, one for each class, as Eq. (5) and (6):

$$\min_{w_1,b_1} \frac{1}{2}\left\|A^+ w_1 + e_1 b_1\right\| + c_1 e_2{}^T \lambda_2 \quad (5)$$
$$s.t. -(A^- w_1 + e_2 b_1) + \lambda_2 \geq e_2, \lambda_2 \geq 0$$

$$\min_{w_2,b_3} \frac{1}{2}\left\|A^- w_2 + e_2 b_2\right\| + c_2 e_1{}^T \lambda_1 \quad (6)$$
$$s.t. -(A^+ w_2 + e_1 b_2) + \lambda_1 \geq e_1, \lambda_1 \geq 0$$

where $c_1, c_2 > 0$ are penalty factors, e_1 and e_2 are two vectors of 1s, and λ_1 and λ_2 are two slack variable vectors. If the feature vector of an author pair is represented by x, its label is determined by:

$$\arg\min_i \frac{|x^T w_i + b_i|}{\|w_i\|} \quad (7)$$

which yields +1 or -1.

Ensemble Aggregation. The output of the link predictors in the ensemble should be aggregated after training. There are a number of aggregation techniques including majority voting and LSE-based weighted voting as two linear combination methods, and double-layer hierarchical SVM as a nonlinear technique [19]. In our experimental results, the majority voting has been used as follows.

Let $f_l (l = 1,..., K)$ be the decision function of the l-th link predictor in the ensemble. If N^+ and N^- denote to the number of link predictors that vote respectively to class +1 and -1, a given test author pair x is assigned to class +1 if $N^+ \geq N^-$, and -1 in otherwise. The steps of the proposed approach have been shown in Fig. 2.

IV. EXPERIMENTS

A. Data Set

DBLP [11] is a heterogeneous information network which is frequently used in the study of heterogeneous complex networks [20]. The used dataset in this paper [21] contains 1,712,433 authors, 2,092,356 papers, 4,258,615 co-authorships, and 8,024,869 citations. Following [25], a sub-graph of this network has been selected in which there are 5 papers for each author. Accordingly, the selected sub-graph contains 35832 authors, 170035 papers, 6225 co-authorships, 70480 citations, and 35443 publishers with 141161 same-scope relations. In order to create the training and test sets for co authorships prediction,

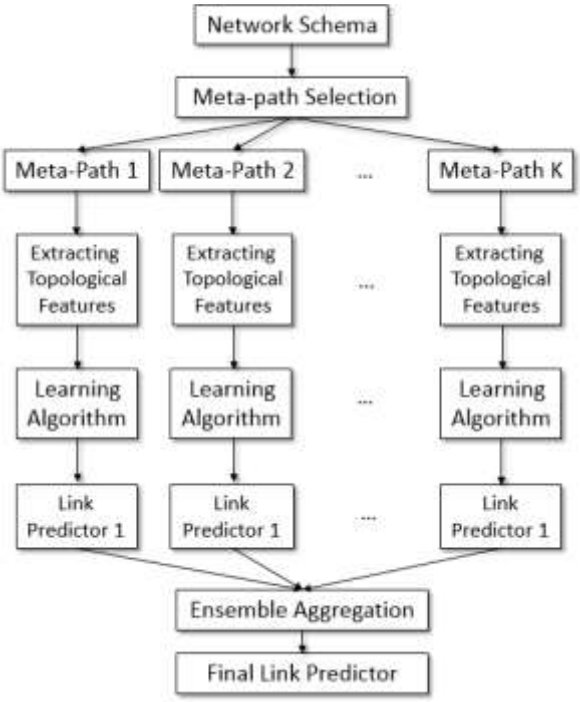


Fig. 2. The steps of the proposed ensemble link predictor. As it is shown, the size of the ensemble equals to the number discovered/selected meta-path from the network schema.

the network has been considered in two time periods. The sub-network with papers published in [2000, 2010] has been used for training and the sub-network including papers published in [2011, 2014] has been used for the test. From the first sub-network, 4353 positive samples and 43530 negative samples have been selected randomly. So, the imbalance ratio is 0.1. Similarly, 1650 positive samples and 16500 negative samples have been chosen from the second sub-network.

B. *Prediction Accuracy.*

The prediction accuracy of the proposed ensemble link predictor has been shown in Table 1. As it is shown the proposed approach obtains an accuracy better than each single link predictor as well as the PathSim [9]. Its accuracy is %3.92 better than PathSim and %11.92, %7.62, and %17.26 better than the first, second, and third meta-paths respectively. Thus, it can be inferred an ensemble of link predictors works better that a single meta-path based predictor Also, three meta-paths have been ranked according to their accuracies. As it is shown in Table 1, the weights of the three meta-paths are computed as 0.33, 0.36, and 0.29 respectively. This shows that the prediction accuracy is decreased when the length or the number of different object types in the meta-path increases. The negative impact of the meta-path length has also been discussed in [22].

C. *Computational Complexity*

In feature extraction step, the time complexity is proportional to the structure of the longest meta-path,

Table 1. The prediction accuracy of the proposed ensemble link predictor (ELP), PathSim [9], and each meta-path as a single link predictor. MP1, MP2, and MP3 denote to three meta-paths introduced in section 3. Also, the weight of each meta-path has been computed based on their prediction accuracies.

Method	MP1	MP2	MP3	PathSim [9]	ELP
Accuracy	66.46	70.76	61.12	74.16	**78.38**
Weight	0.33	0.36	0.29	-	-

i.e. A-P-V-V-P-A. In this regard, the required time to extract *PC* for the *j*-th author pair is $O(k_{12}^{j} \times k_{23}^{j} \times 1 \times k_{45}^{j} \times k_{56}^{j})$ where $k_{i,i+1}$ refers to the number of relations between the object types *i* and *i*+1 in the meta-path. Thus, the time complexity for all *n* example pairs is $O(\sum_{j=1}^{n} k_{12}^{j} \times k_{23}^{j} \times 1 \times k_{45}^{j} \times k_{56}^{j})$ which can be simplified to $O(nk_{12}^{a} \times k_{23}^{a} \times 1 \times k_{45}^{a} \times k_{56}^{a})$ where $k_{i,i+1}^{a}$ denotes to the average degree between the object types of *i* and *i*+1 in the meta-path. The time complexity of two other features is in the order of computing *PC*. In the learning step, the time complexity is proportional to standard SVM algorithm. It has been known that the complexity of SVM is no more than $O(|m+n|)^3$ where *m*+*n* is the number of training author pairs [18]. Accordingly, the complexity of TSVM is in the order of $O(\frac{1}{4}(|m+n|)^3)$, i.e. 4 times faster than SVM.

V. Conclusion

In this paper, a new ensemble-based link prediction approach has been proposed. The idea of the proposed approach is to define different semantics in a number of meta-paths and learn an independent link predictor based on each meta-path. Keeping different meta-paths brings a promising opportunity to obtain a better coverage on the pairing nodes. The experimental results on DBLP as a large-scale complex network, show that the proposed approach obtains higher accuracy compared to each single link predictor. Moreover, thanks to fast running time of TSVM, the size of the network and training sets have not been found as an obstacle. Another advantage of using TSVM is to overcome the network sparsity and imbalanced data issues in link predition problem.

However, the prediction accuracy of the proposed approach should be improved further. In this regard, more meta-paths as well as the other topological features can be used. Also, applying the proposed supervised framework on the weighted network and extracting weighted meta-paths are placed in our future work.

References

[1] Sun, Yue, and Jinguang Han. "Meta-path-based search and mining in heterogeneous information networks." *Tsinghua Science and Technology* 18, no. 4 (2013).

[2] Luo, Jiawei, and Shiyu Liang. "Prioritization of potential candidate disease genes by topological similarity of protein–protein interaction network and phenotype data." *Journal of biomedical informatics* 53 (2015): 229-236.

[3] Wu, Fangzhao, Jinyun Shu, Yongfeng Huang, and Zhigang Yuan. "Social Spammer and Spam Message Co-Detection in Microblogging with Social Context Regularization." In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 1601-1610. ACM, 2015.

[4] Ciobanu, Radu Ioan, Ciprian Dobre, Valentin Cristea, Florin Pop, and Fatos Xhafa. "SPRINT-SELF: Social-Based Routing and Selfish Node Detection in Opportunistic Networks." *Mobile Information Systems* 2015 (2015).

[5] Jawed, Mujtaba, Mehmet Kaya, and Reda Alhajj. "Time Frame based Link Prediction in Directed Citation Networks." In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pp. 1162-1168. ACM, 2015.

[6] Lü, Linyuan, and Tao Zhou. "Link prediction in complex networks: A survey." *Physica A: Statistical Mechanics and its Applications* 390, no. 6 (2011): 1150-1170.

[7] Y. Sun, R. Barber, M. Gupt, C. Aggarwal, and J. Han. Co-author relationship prediction in heterogeneous bibliographic networks. In ASONAM, 2011.

[8] N. Lao and W. W. Cohen. Relational retrieval using a combination of path-constrained random walks. Mach. Learn., 81(1), 2010.

[9] Sun, Yizhou, Jiawei Han, Xifeng Yan, Philip S. Yu, and Tianyi Wu. "Pathsim: Meta path-based top-k similarity search in heterogeneous information networks." *VLDB'11* (2011).

[10] C. Shi, X. Kong, P. Yu, S. Xie, and B. Wu. Relevance search in heterogeneous networks. EDBT, 2012.

[11] M. Ley. DBLP: some lessons learned. PVLDB, 2(2), 2009.

[12] Yu, X. Ren, Y. Sun, B. Sturt, U. Khandelwal, Q. Gu, B. Norick, and J. Han, HeteRec: Entity recommendation in heterogeneous information networks with implicit user feedback, in Proc. of 2013 ACM Int. Conf. Series on Recommendation Systems (RecSys'13), Hong Kong

[13] China, Oct. 2013. Kong, Xiangnan, Philip S. Yu, Ying Ding, and David J. Wild. "Meta path-based collective classification in heterogeneous information networks." In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pp. 1567-1571. ACM, 2012.

[14] Cao, Bokai, Xiangnan Kong, and Philip S. Yu. "Collective Prediction of Multiple Types of Links in Heterogeneous Information Networks." In *Data Mining (ICDM), 2014 IEEE International Conference on*, pp. 50-59. IEEE, 2014.

[15] Meng, Changping, Reynold Cheng, Silviu Maniu, Pierre Senellart, and Wangda Zhang. "Discovering meta-paths in large heterogeneous information networks." In *Proceedings of the 24th International Conference on World Wide Web*, pp. 754-764. International World Wide Web Conferences Steering Committee, 2015.

[16] Wang, Peng, BaoWen Xu, YuRong Wu, and XiaoYu Zhou. "Link prediction in social networks: the state-of-the-art." *Science China Information Sciences* 58, no. 1 (2015): 1-38.

[17] Zhang, Jiawei, Philip S. Yu, and Zhi-Hua Zhou. "Meta-path based multi-network collective link prediction." In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1286-1295. ACM, 2014.

[18] Khemchandani, R., and Suresh Chandra. "Twin support vector machines for pattern classification." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29, no. 5 (2007): 905-910.

[19] Kim, Hyun-Chul, Shaoning Pang, Hong-Mo Je, Daijin Kim, and Sung Yang Bang. "Constructing support vector machine ensemble." *Pattern recognition* 36, no. 12 (2003): 2757-2767.

[20] Gimenes, Gabriel P., Hugo Gualdron, Thiago R. Raddo, and Jose F. Rodrigues. "Supervised-learning link recommendation in the DBLP co-authoring network." In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*, pp. 563-568. IEEE, 2014.

[21] *http://arnetminer.org/lab-datasets/aminerdataset*

[22] Sun, Yizhou, and Jiawei Han. "Mining heterogeneous information networks: a structural analysis approach." *ACM SIGKDD Explorations Newsletter* 14, no. 2 (2013): 20-28.

[23] Shao, Yuan-Hai, Wei-Jie Chen, Zhen Wang, Chun-Na Li, and Nai-Yang Deng. "Weighted linear loss twin support vector machine for large-scale classification." *Knowledge-Based Systems* 73 (2015): 276-288.

[24] Yu, Xiao, Quanquan Gu, Mianwei Zhou, and Jiawei Han. "Citation Prediction in Heterogeneous Bibliographic Networks." In *SDM*, vol. 12, pp. 1119-1130. 2012.

[25] Yu, Qi, Chao Long, Yanhua Lv, Hongfang Shao, Peifeng He, and Zhiguang Duan. "Predicting Co-Author Relationship in Medical Co-Authorship Networks." (2014): e101214.

Hadi Shakibian received his M.Sc in Computer Engineering from Tarbiat Modares Univeristy, Tehran, in 2011. Currently he is a PhD candidate in computer engineering at Tarbiat Modares University, Tehran. His main research interests include complex networks and machine learning.

Nasrollah Moghadam Charkari received his B.Sc. degree in Computer Engineering from Shahid Beheshti University, Tehran, Iran, in 1987, and his M.Sc. and Ph.D. degrees in Computer Engineering and Information System Engineering from Yamanashi University , Japan, in 1993 and 1996, respectively. His research interests include robot vision and image analysis, image mining, complex networks, parallel algorithms and processing, information technology, data hiding and water marking.

Saeed Jalili received P.hD. in Computer Science from Bradford University, Bradford, in 1991. In 1992 he joined the School of Electrical and Computer Engineering at Tarbiat Modares University, Tehran, Iran. He is currently an associate professor at Tarbiate Modares University and his main research interests are bioinformatics, security protocol verification, network security, machine learning and software runtime verification.