TELELINK: Link Prediction in Social Network Based on Multiplex Cohesive Structures

Di Jin[1], Mengdi Wang[2], and Yu-Ru Lin[3]

[1] School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2] Intelligent System Program, University of Pittsburgh, Pittsburgh, PA 15260, USA
[3] School of Information Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

Abstract. Given a network where the same set of nodes have multiple types of relationships, how do we efficiently predict potential links in the future (e.g., interactions between social actors), and how do we predict links using information from other relationships? These problems have been widely studied recently, most of the existing methods either aggregate multiple types of relationships into a single network or consider them separately and ignore the correlations across relationships, leading to information loss. In this work, we present TELELINK, a general link prediction model that works for networks with single and multiple relationships. TELELINK predicts potential links based on community detection and improves link prediction by bringing in a cohesive structure across multiple networks constructed by different relationships or node attributes. To further improve the prediction performance, we extend TELELINK to a semi-supervised scheme, incorporating partially labeled information. Our extensive experiments show that TELELINK outperforms existing methods in predicting new links. Specifically, among the various datasets that we study, TELELINK achieves a precision improvement by up to 110% compared to the baselines.

1 Introduction and Background

Recent years have witnessed a surge of interest for understanding and characterizing the properties of *social networks*, where nodes represent people or other entities embedded in a social context and links denote relationships or interactions, such as friendship, collaboration, or influence between entities. An important problem in this context is *link prediction*, which is to predict links that will appear in the network during the interval from time T to a later time $T + 1$, given a snapshot of a network at time T [1] or before. Link prediction is useful in various areas. In social networks, link prediction algorithms can be used to predict relationships among individuals such as friendship, partnership and their future behaviors such as communications and collaborations [2]. In biological networks such as protein-protein interaction networks where over 99% links are unknown [3], accurately predicting possible links could sharply reduce the experimental cost.

A typical framework of link prediction algorithm is based on "similarity", where each pair of nodes is given a similarity measure, and node pairs with high similarity scores are assumed more likely to be connected [4]. However, most of the similarity measures mainly consider the network topological structure (e.g., the number of common neighbors [5], or the length of the shortest path between nodes [6]). These kinds of methods ignore the information provided by the community structure. A community is a densely connected group of nodes while sparsely connected to other groups.

Community structure has been proved to be critical for link prediction [7]. For example, in the friendship network, a community could be a group of people in the same school, company or club and new links (i.e. friendships) are more likely to form within the group. [8] shows that group membership information can enhance the accuracy of link prediction. However, their method does not scale well to very large networks in practice.

The real world networks are always multi-relational, where links have different meanings and bring challenges to link prediction. For example, in a Twitter network, the links could be different interactions such as replying or mentioning. The problem is: how can one predict possible links representing a particular relationship using the information provided by other relationships? Two typical strategies have been employed: pre-fusion and post-fusion. Pre-fusion aggregates multiple types of relationships into a single link while post-fusion separately studies each type of links independently and ignores the correlations across types. But both approaches result in a loss of information.

Another issue in link prediction is how to develop a method that combines topological information and node attributes. Existing work has employed supervised classifiers [9], which is trained to discriminate between positive links (i.e. links that form) and negative links (i.e. links that do not form) by using multiple sources of information as features. Those methods suffer from the imbalance problem: in real networks, the number of positive links is significantly less than that of negative links. To overcome the imbalance problem, typical strategies are under-sampling and over-sampling [9], which lead to issues including overfitting.

Present work We present TELELINK, a general link prediction model to address the above challenges. We consider multiple layers of networks, one is target layer – the particular type of links to be predicted (e.g. Twitter follower-followee network) and others are *auxiliary* layers, which can be constructed based on other relations between nodes (e.g. reply or retweet between Twitter users) or nodes attributes (e.g. geographical information in users' profile). TELELINK predicts potential links using a probabilistic similarity measure between nodes defined by the path information of multilayer community structure revealed by *Multiplex Infomap* [10], a random-walk based community detection approach. We further extend TELELINK into a semi-supervised learning scheme to improve link prediction performance, using both networks at T and partially labeled information in the networks at time $T + 1$ in multiplex Infomap framework.

Contributions Our main contributions include:

1. **New link prediction approach**: We extend Infomap to address several challenges in link prediction problem, including incorporating community structure, combining multiple relationships and combining topological and node attributes.
2. **Prediction performance**: We conduct extensive experiments on two different datasets. The proposed methods achieve best prediction accuracy in new link prediction, compared with existing link prediction methods.
3. **Extended analysis on dynamic networks**: We conduct experiments on real-world social networks over time and present the influence of time intervals on prediction.

2 Related Work

Single-relational link prediction methods The seminal work of Liben-Nowell and Kleinberg [1] is the first comprehensive study on link prediction methods based on similarity measures derived from graph topology structure. Empirical results of comparison between random predictors and a variety of measures including *Jaccard's coefficient* [5], *Adamic/Adar* [11] , *Katz Index* [6] and *Rooted PageRank* [12] demonstrate the usefulness of topological information. However, the weakness of these methods is that they only consider a single (topological) feature. In addition to topological information, we often have the knowledge of attributes or covariance for the nodes. Intuitively, performance is expected to be enhanced by using this extra information. A classical approach is to use supervised classifiers [2] unitizing different sources of information, including topological information and node attributes as features. [9] suggests that placing classification algorithms in an ensemble framework can benefit by reducing variance, especially for unstable algorithms like decision trees. These methods have to use down-sampling or over-sampling strategies to overcome imbalance, which lead to a loss of information or over-fitting. [13] proposes an algorithm based on *Supervised Random Walks*, which uses node attributes to guide the random walker, but it has high computational cost and does not scale well in practice.

Most existing link prediction methods do not consider community information, which is proved to be useful to link prediction. The community in a network is a densely connected group of nodes while sparsely connected to other groups. In [7], experiments on both synthetic and real-world networks unveil how the community structure affects the performance of link prediction methods: with increasing number of communities, the performance of link prediction could be improved remarkably. Recently, [14] also shows that supplement the similarity-based measures with community information could improve the accuracy of link prediction methods.

Multi-relational link prediction methods The above link prediction techniques only consider homogeneous links with the same semantic meaning, while in reality networks comprise multiple types of links or interactions among nodes. Only a few studies address link prediction problems in these heterogeneous networks. In [15], an unsupervised method extending the Admic/Adar measure and a supervised method in multi-relational networks are proposed. [16] develops machine learning approaches based on graphical models to infer new links across heterogeneous networks assuming that people will form relationships in different networks with similar principles. However, none of the multi-relational link prediction methods consider the information of multi-relational communities, which is potentially useful in improving the performance of link prediction algorithm.

TELELINK addresses the limitations of prior works in three perspectives. First, TELELINK combines the topological information and nodes attributes by extracting structures from multiple layers of networks. Second, TELELINK handles link prediction across multi-relational networks with auxiliary layers based on multiple relationships (e.g. reply or retweet of Twitter users). Last, to make use of the community information, TELELINK defines "similarity" between nodes based on the path information of multilayer community structure calculated through random walk on multiple networks.

3 Proposed Method

First we give the definitions of two fundamental concepts in our proposed method:

Definition 1 (layer) *A layer of the social network is defined as a specific relationship. A physical node n belongs to at least one layer of the social network.*

Definition 2 (link) *A link from node i to node j, l_{ij}, is defined as the behavior starting from node i to j.* For example, in the retweet layer of the Twitter network, l_{ij} denotes user i retweets user j's tweet.

The challenges discussed in previous sections can be summarized into two research questions in social networks: a) How do we express similarity through communities, thus predicting possible new links in a single layer? b) How do we combine the attributes and connectivity from other layers so as to improve the prediction accuracy? We present our solutions to these questions in the next two sections. The symbols used in this paper are defined in Table 1.

Symbol	Definition
$\mathcal{L}_{\mathcal{C}}$	a collection of all links between communities
$L_{\mathcal{C}_i\mathcal{C}_j}$	a directed link starting from community i to community j
l_{ij}	a directed link starting from node i to node j
$\mathcal{C}, \mathcal{C}_i$	a set of all communities in the network; the ith community in the set
$\mathcal{C}^{(i)}$	the community that node i belongs to
n	a physical node in the network, $n \in 1, 2, \ldots, N$
v_i	the probability that the random walker reaches node i
$w^{\alpha}_{i,j}$	the out-flow volume from node i to j in layer α
N, M	total number of nodes and total number of communities of the network
$\mathbf{M}$	a partition of the network with minimum Huffman coding length
α, β	specific layer of the network, represented in Greek letters
ΔT	length of time interval for community detection

Table 1: Table of symbols

3.1 TELELINK

TELELINK is based on MapEquation[17], a flow-based network partitioning algorithm. MapEquation characterizes the behavior of the network through the system-wide flow of information and considers community detection as solving a coding problem.

These flows are described using Huffman coding, following the Random Walk algorithm to identify communities. Groups of nodes among which information flows frequently are described as well-connected communities; the paths between communities are also captured. [17] gives the community partition $\mathbf{M}$ by solving Equ. 1 to obtain the minimized expected coding length of Random Walk paths:

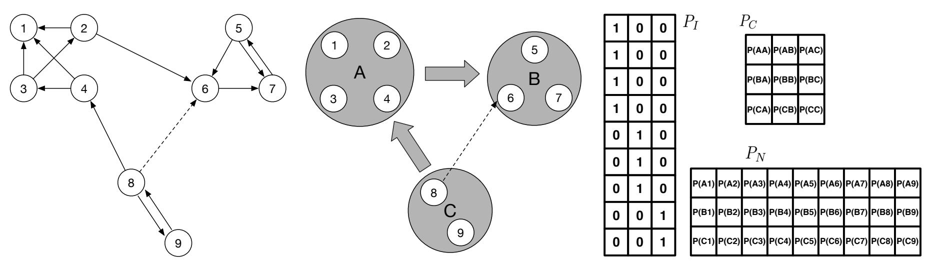
$$\min L(\mathbf{M}) = q_{\curvearrowright} H(\mathcal{L}) + \sum_{i=1}^{M} p^{i}_{\circlearrowright} H(\mathcal{P}^{i}) \,, \tag{1}$$

where $L(\mathbf{M})$ denotes the Huffman coding description of random walker's path; $q_{\curvearrowright}$ is the probability that the random walk switches communities on any given step, and $p^{i}_{\circlearrowright}$ describes the fraction of within-community movements that occur in community i, $H(.)$ is the entropy of the community.

Given the partition $\mathbf{M}$, the network $\langle \mathcal{V}, \mathcal{E} \rangle$ could be described as $\langle \mathcal{C}_{\mathcal{V}}, \mathcal{L}_{\mathcal{C}} \rangle$. $\mathcal{C}_{\mathcal{V}}$ describes the set of communities that compose the network: $\mathcal{C}_{\mathcal{V}} = \{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$.

$\mathcal{C}_i$ is the ith community. $\mathcal{L}_\mathcal{C}$ is the collection of links between communities: $\mathcal{L}_\mathcal{C} = \{L_{\mathcal{C}_1\mathcal{C}_1}, L_{\mathcal{C}_1\mathcal{C}_2}, \ldots, L_{\mathcal{C}_M\mathcal{C}_M}\}$. The links between communities can be interpreted as the flow volumes of all nodes reaching from the source community to the destination community. $P(L_{\mathcal{C}_i\mathcal{C}_j})$ denotes the transition probability from community $\mathcal{C}_i$ to $\mathcal{C}_j$.

Now the problem can be formulated as: given $P(\mathcal{C}_p)$ where $p \in \{1, \ldots, M\}$ and $P(L_{\mathcal{C}_p\mathcal{C}_q})$ where $p, q \in \{1, \ldots, M\}$, how do we compute the probability of a specific link l_{ij} between node i and j? We estimate the probability of l_{ij} through the product of a) the transition probability from the community of node i to the community of node j, and b) the conditional probability of the random walker reaching node j inside its community. An example is illustrated in Fig. 1a. To predict the link in dashed line, the random walker has to follow each possible path from node 8 to 6, which could be expensive in real-world networks. TELELINK calculates the probability based on the transition flows between communities (bold arrow): from community C to community B through community A, and the probability to reach node 6 inside community B, as illustrated in Fig. 1b.



(a) A synthetic network **(b) Community structure** **(c) Three matrices**

Fig. 1: Single layer link prediction through Infomap

Considering nodes across the whole network, we could factorize the probability between node i and node j into three terms, as illustrated by Equ. 2:

$$P(l_{ij}) = \underbrace{\mathbb{1}_{\mathcal{C}^{(i)}}}_{P_I} \times \underbrace{P(L_{\mathcal{C}^{(i)}\mathcal{C}^{(j)}})}_{P_C} \times \underbrace{\frac{v_j}{\sum_{k\in\mathcal{C}^{(j)}}^{|\mathcal{C}^{(j)}|} v_k}}_{P_N}, \tag{2}$$

where $\mathcal{C}^{(i)}$ denotes the community that node i belongs to. The first term represents the community affiliation of node i: $\mathbb{1}_{\mathcal{C}_i}$ is the indicator matrix determining which community node i uniquely belongs to, so each row of this matrix has only one entry with value of 1; the second term describes transition probability from the community of node i to the community of node j, and the last term is the flow probability to node j normalized by the flows reaching all nodes inside its community. P_I, P_C and P_N correspond to the indicator matrix, transition matrix, and flow matrix of the example shown in Fig. 1c.

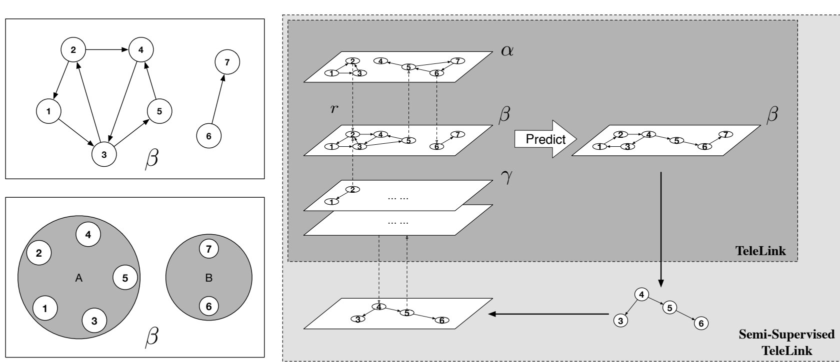
TELELINK computes the transition probability from the source community $\mathcal{C}^{(i)}$ to destination node j as the approximation to the probability of link l_{ij} used for prediction. The reason is that communities are partitioned using Random Walk algorithm, so information flows faster and more easily among nodes within the same community than those between different communities. If node i in community $\mathcal{C}^{(i)}$ has link to node j in community $\mathcal{C}^{(j)}$, i is also likely to reach other nodes in community j. In addition, by grouping nodes as a community, we don't have to compute every possible link from the source node to the destination node (which is what Random Walk algorithm does). As a result, TELELINK achieves precision with the guarantee of efficiency.

3.2 TELELINK in multiplex networks

Multiplex Infomap[10] extends MapEquation to multilayer networks in two ways: 1) inter-layer dynamics described in communities and 2) overlapping communities identification. The first generalization resembles PageRank algorithm to compute the flow volume between layers by introducing the relax rate r. With probability r, the random walker "teleports" to another network layer (thus the method is named TELELINK); otherwise, it stays at the same layer. Equ. 3 defines the transition probability between layer α and β given r:

$$P_{ij}^{\alpha\beta}(r) = (1-r)\delta_{\alpha\beta}\frac{w_{ij}^{\beta}}{s_i^{\beta}} + r\frac{w_{ij}^{\beta}}{S_i}\,, \tag{3}$$

where δ is the indicator function seeing if the random walker stays at the same layer; w_{ij}^{β} denotes the out-flow volume from node i to j in layer β; $s_i^{\beta} = \Sigma_j w_{ij}^{\beta}$ and $S_i = \Sigma_\beta s_i^{\beta}$. Similar to Infomap, the interaction between layers is denoted through communities, and overlapping communities are not allowed. With information from multiple layers of networks, TELELINK could predict links that cannot be predicted in single layer networks. Consider the example illustrated in Fig. 2a where network layer β consists of two disjoint communities A and B. Due to the isolation between community A and B, the random walk algorithm could not predict links such as l_{56}. However in Fig. 2b, TELELINK overcomes this limitation through flows between layers α and β with a "teleporting" relax rate r.



(a) Community structures of β

(b) TELELINK and semi-supervised TELELINK in multiplex networks

Fig. 2: Single layer link prediction through Infomap

3.3 Semi-supervised TELELINK

Semi-supervised TELELINK extends the above method by employing partially labeled data. When performing link prediction on real-world networks, historical data could be the double-edged sword: on the one hand, historical data provides valuable structural information about the network; on the other hand, it could bring in noise, impeding the prediction precision especially to rapidly changing networks. Semi-supervised TELELINK overcomes this problem by randomly selecting small parts of the structural information from the target network as an extra auxiliary layer in community detection. An example is illustrated in Fig. 2b where we use the multiplex network information at T to predict possible links in layer β at time $T + 1$. In a specific attempt, a small part of the β network at time $T + 1$ is selected: the connectivity between node 3, 4, 5 and 6. It then serves as an extra auxiliary layer for community detection at the stage of T. In this example, such information is the key to the flows between nodes in two disjoint communities and provides additional information to guide proper community assignment that could help predict future links.

4 Experiments

To demonstrate that TELELINK can be applied in different contexts, we perform experiments on two datasets to address three perspectives of TELELINK: 1) How well does TELELINK perform, comparing to existing methods? 2) How are information flows from other layers influencing the prediction in the specific layer? 3) Given different time intervals, how would the prediction results be different?

4.1 Datasets

Primary School This dataset contains 125,773 contact records among 236 students in a primary school in Lyon, France during two days in Oct. 2009[18,19]. We build a weighted undirected contact network with students as nodes, contacts between two students as links and contact frequency as weights. There are two features of students: class and gender, which are used as two unweighted undirected auxiliary networks. The link between two students is built if they are in the same class or of the same gender.

Twitter We use the Twitter 2012 election dataset[20], which contains approximately 48.7 million politically active users and approximately 0.2 billion tweets during 8 weeks starting from Sept. 2012. The dataset is divided into 8 sub-datasets according to post time (one per week). In each sub-dataset, one weighted directed network is established for the mention, reply and retweet respectively. In the mention (or reply, retweet) network, a directed link with weight k is built from user A to user B if user A mentions (or replies, retweets) user B exactly k times. Given the three networks in each week, we aim to predict new links of a specific network in week $T + 1$ from the networks during the period $[T - \Delta T + 1, T]$, where $\Delta T \in \{1, 2, \ldots, 7\}$. We also ignore inactive users (with two or fewer links) because for users only posting one or two tweets during eight weeks, their impacts on the analysis is trivial. The pre-processed dataset contains 24 networks with identical 2,073 nodes. The numbers of links in week 1 for the mention, reply and retweet are 14,248, 8,707 and 8,944. As the numbers of links increase from week 1 to week 8, the numbers of new links added every week for the mention, reply and retweet are 1,990, 652 and 1,775 on average.

4.2 Experiment Setup

Baseline We consider 2 classic approaches: *Jaccard's coefficient* and *Adamic/Adar*.

Our Methods We apply our methods on the above datasets in three scenarios:

1. **TELELINK** TELELINK is applied to the target single layer network.
2. **TELELINK in multiplex networks (different** r**)** TELELINK is applied to the multiplex network, flows moving between layers with different relax rate r.
3. **Semi-supervised TELELINK** TELELINK is applied to the multiplex network, flows moving between layers with $r = 0.15$. $1/3$ nodes of the target layer network are selected for training, and the result is evaluated with 3-fold cross validation.

Evaluation Metrics The experiments are designed to predict new links appearing in $T + 1$ from networks during $[T - \varDelta T + 1, T]$, $\varDelta T \in \{1, 2, \ldots, 7\}$. We consider the prediction precision at Top k nodes according to the probability to which the source node is not already connected, i.e., how many of top k nodes suggested by our algorithm during $T + 1$ actually receive links not exist in $[T - \varDelta T + 1, T]$. We set k equal to the total number of links in $T+1$. In addition, we measure the improvement over baselines.

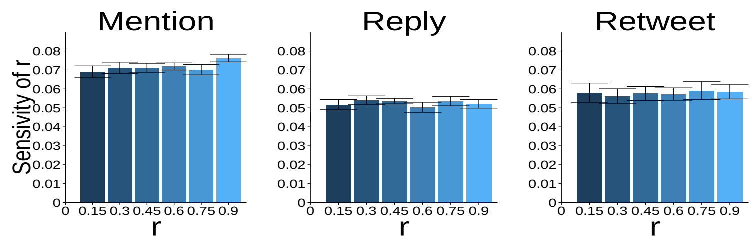
4.3 Results

Interaction Sensitivity We first explore the sensitivity of TELELINK to the interaction between layers in the multiplex networks, which is controlled by relax rate r. Higher values of r indicate that the random walker is more likely to "teleport" to other layers while lower values of r indicate more isolation between layers. The result is shown in Fig. 3a. The relatively flat pattern in each of the diagrams indicates that when we are considering information from different relationships, the interaction between them does not lead to the significant difference in prediction precision. Therefore in the following experiments, we set $r = 0.5$ and use it for evaluation.

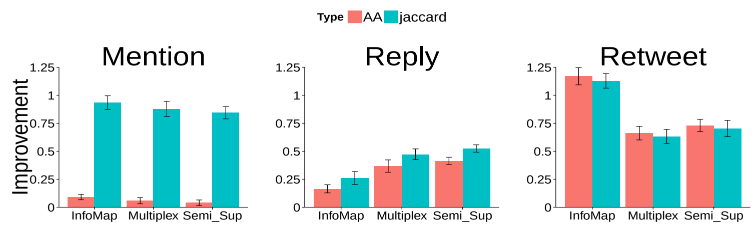
Method	InfoMap	Multiplex	Multiplex(i)	Semi-Supervised	Jaccard	AA
Precision	0.191	0.198	0.193	**0.205**	0.103	0.079

Table 2: Table of elementary school prediction precision

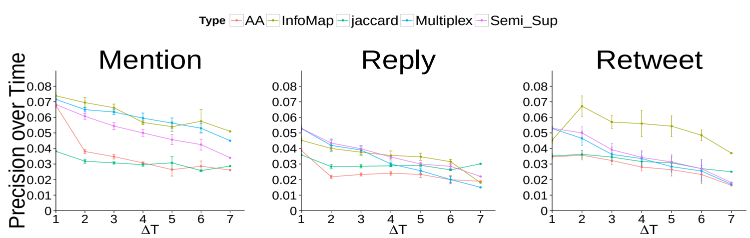
Prediction precision We measure the prediction precision on two datasets. For the primary school dataset, there are two intervals so $\varDelta T = 1$. The prediction result is illustrated in Table 2. In the Twitter dataset, we also set $\varDelta T = 1$ for consistency. Since there are prediction results for seven weeks, we only show the precision improvement of TELELINK over baselines in Fig. 3b. The results show that TELELINK under all settings outperform the baselines. The mention relationship is a loose way of communication between users, so TELELINK provides a relatively small improvement over the baseline AA. However, for relationships such as reply and retweet where interactions between users are stronger, TELELINK gives significant improvement over both baselines. In the reply relationship, TELELINK achieves $40\% \sim 50\%$ improvement through semi-supervised learning with auxiliary layers; Multiplex TELELINK also makes great improvements, indicating the relationship of mention and retweet is helpful when we are predicting replies between users. In the retweet relationship, TELELINK performs on single layer achieves approximately 110% improvement over both baselines, which implies that for this relationship, adding other types of interactions limits or even impedes the prediction performance.



(a) Sensitivity



(b) Improvement



(c) Precision over time interval

Fig. 3: Precision performance on (from left to right) mention, reply and retweet networks

Perdiction performance with dynamic networks In this section, we focus on the impacts of different time intervals (ΔT) used for community detection to prediction precision. We apply TELELINK to the Twitter dataset partitioned into different time intervals, ranging from 1 to 7. The result is shown in Fig. 3c. Intuitively, with longer time intervals to perform community detection, more information about the network would be included, rendering higher prediction precision. However, based on the overall result curves, the prediction precision drops as ΔT increases. This means that with long-time intervals, longer-term historical data becomes outdated, which impedes the prediction. One interesting observation is the increase in retweet curve from $\Delta T = 1$ to $\Delta T = 2$. This indicates that although too much long-term historical data could be harmful, with appropriate portion it could actually be beneficial to the prediction. This also explains the relatively "flat" pattern from $\Delta T = 1$ to $\Delta T = 2$ in other curves.

5 Conclusions

In this paper, we study the problem of link prediction in multiplex networks. We propose TELELINK to address this problem, which provides a novel link prediction approach through community detection. TELELINK detects community structure across different types of relationships of the network to incorporate the topological and attribute information. Experiments on two multiplex network datasets show that the impact of information from auxiliary layers on prediction varies with different relationships, and TELELINK improves the prediction precision in all scenarios. In addition, the learned interaction sensitivity and the prediction performance on dynamic networks provide a better understanding of information flows and community detection in social networks.

References

1. David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *JASIST*, 58(7):1019–1031, 2007.
2. Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki. Link prediction using supervised learning. In *SDM'06: Workshop*, 2006.
3. Michael PH Stumpf, Thorne, et al. Estimating the size of the human interactome. *PNAS*, 105(19):6959–6964, 2008.
4. Linyuan Lü and Tao Zhou. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6):1150–1170, 2011.
5. Paul Jaccard. *Distribution de la Flore Alpine: dans le Bassin des dranses et dans quelques régions voisines*. Rouge, 1901.
6. Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.
7. Xu Feng, JC Zhao, and Ke Xu. Link prediction in complex networks: a clustering perspective. *The European Physical Journal B*, 85(1):1–9, 2012.
8. Aaron Clauset, Cristopher Moore, and Mark EJ Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008.
9. Ryan N Lichtenwalter, Jake T Lussier, and Nitesh V Chawla. New perspectives and methods in link prediction. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 243–252. ACM, 2010.
10. Manlio De Domenico, Andrea Lancichinetti, Alex Arenas, and Martin Rosvall. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Physical Review X*, 5(1):011027, 2015.
11. Lada A Adamic and Eytan Adar. Friends and neighbors on the web. *Social networks*, 25(3):211–230, 2003.
12. Sergey Brin and Lawrence Page. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18):3825–3833, 2012.
13. Lars Backstrom et al. Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the fourth ACM WSDM*, pages 635–644. ACM, 2011.
14. Sucheta Soundarajan and John Hopcroft. Using community info. to improve the precision of link prediction methods. In *Proceedings of the WWW*, pages 607–608. ACM, 2012.
15. Darcy Davis, Ryan Lichtenwalter, and Nitesh V Chawla. Supervised methods for multi-relational link prediction. *Social network analysis and mining*, 3(2):127–141, 2013.
16. Yizhou Sun et al. Co-author relationship prediction in heterogeneous bibliographic networks. In *ASONAM, 2011 International Conference on*, pages 121–128. IEEE, 2011.
17. Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *PNAS*, 105(4):1118–1123, 2008.
18. Juliette Stehlé et al. High-resolution measurements of face-to-face contact patterns in a primary school. 2011.
19. Valerio Gemmetto, Alain Barrat, and Ciro Cattuto. Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC infectious diseases*, 14(1):695, 2014.
20. Yu-Ru Lin, Brian Keegan, Drew Margolin, and David Lazer. Rising tides or rising stars?: Dynamics of shared attention on twitter during media events. 2014.