

Link Prediction via Ranking Metric Dual-Level Attention Network Learning

Zhou Zhao¹, Ben Gao², Vicent W. Zheng³, Deng Cai², Xiaofei He² and Yueting Zhuang¹

¹College of Computer Science, Zhejiang University

²State Key Lab of CAD&CG, Zhejiang University

³Advanced Digital Sciences Center, Singapore

{zhaozhou, gaoben, yzhuang}@zju.edu.cn, vincent.zheng@adsc.com.sg, {dengcai, xiaofeihe}@gmail.com

Abstract

Link prediction is a challenging problem for complex network analysis, arising in many disciplines such as social networks and telecommunication networks. Currently, many existing approaches estimate the proximity of the link endpoints from the local neighborhood around them for link prediction, which suffer from the localized view of network connections. In this paper, we consider the problem of link prediction from the viewpoint of learning path-based proximity ranking metric embedding. We propose a novel proximity ranking metric attention network learning framework by jointly exploiting both node-level and path-level attention proximity of the endpoints to their betweenness paths for learning the discriminative feature representation for link prediction. We then develop the path-based dual-level attentional learning method with multi-step reasoning process for proximity ranking metric embedding. The extensive experiments on two large-scale datasets show that our method achieves better performance than other state-of-the-art solutions to the problem.

1 Introduction

Link prediction is an important problem for complex network analysis [Liben-Nowell and Kleinberg, 2007], which has a large number of applications in many disciplines like network ranking [Zhao *et al.*, 2015] and network clustering [Zhao *et al.*, 2016a]. The essential problem of link prediction is to infer the missed or new relationships of the network based on the currently observed connections, which has attracted considerable attention recently [Katz, 1953; Cao *et al.*, 2014].

Many previous approaches for the problem of link prediction are neighborhood-based methods [Menon and Elkan, 2011; Wang *et al.*, 2007; 2016], which learn the proximity model for the link endpoints from the local neighborhood around them, and then infer their closeness for link prediction. Although the methods arising from these approaches could be simple and efficient, they still suffer from the localized view of network connection [Martínez *et al.*, 2016]. To tackle the localized view problem, the path-based approaches [Katz, 1953; Cao *et al.*, 2014; Zhang *et al.*, 2014]

were proposed for learning the proximity of link endpoints based on their betweenness paths, and it can be shown that such approaches are more effective [Martínez *et al.*, 2016]. Although the path-based prediction methods have achieved promising performance, most of them still suffer from the insufficiency of discriminative path feature representation [Sun *et al.*, 2013].

Currently, the existing path-based link prediction methods [Cao *et al.*, 2014; Zhang *et al.*, 2014; Liu *et al.*, 2014] learn the semantic path representation based on the hand-crafted meta feature (e.g., bag-of-words). However, the rich content information in networking nodes is not fully utilized on the feature representation in existing methods, which is critical for learning discriminative path representation [Miller *et al.*, 2009]. Recently, various embedding methods are proposed to learn the semantic of similar words [Mikolov *et al.*, 2013] and encode the content information into low-dimensional continuous embedding space. Since the node contents are always the sequential data with variant length, it is natural to employ the deep recurrent neural networks [Hochreiter and Schmidhuber, 1997] to learn the semantic representation of the nodes. On the other hand, learning the effective joint representation of betweenness paths is also a challenging issue for path-based link prediction. The shorter the path, the higher chance that the path influences the link formation between its endpoints [Katz, 1953]. Furthermore, the importance of the betweenness paths for link formation varies according to the specific endpoints, and a number of betweenness paths are redundant and irrelevant for link formation [Sun *et al.*, 2013]. Thus, learning the discriminative joint representation of the betweenness paths is critical for path-based link prediction.

In this paper, we consider the problem of link prediction from the viewpoint of learning path-based proximity ranking metric embedding. We then develop the ranking metric dual-level attention networks to learn the discriminative joint representation of the betweenness paths of the endpoints for link prediction. The main contributions of this paper are illustrated as follows:

- Unlike the previous studies, we study the problem of link prediction from the viewpoint of path-based proximity ranking. We propose the dual attention networks to learn the discriminative joint representation from the betweenness paths of the endpoints for learning proxim-

ity ranking metric embedding.

- We incorporate the multi-step reasoning framework for our proposed attention networks to enable the progressive joint representation learning from the endpoint content information and their betweenness paths to further improve the performance of link prediction.
- We conduct extensive experiments on two large-scale datasets to show that our method achieves better performance than other state-of-the-art solutions to the problem.

The rest of this paper is organized as follows. In Section 2, we introduce the problem of link prediction from the viewpoint of path-based proximity ranking metric learning framework and propose the path-based ranking metric dual-level attention networks to learn the discriminative joint representation of the endpoints for link prediction. A variety of experimental results are presented in Section 3. We provide a brief review of the related work about link prediction in Section 4. Finally, we provide some concluding remarks in Section 5.

2 Link Prediction via Ranking Metric Dual-Level Attention Network Learning

In this section, we first present the problem of link prediction from the viewpoint of learning proximity ranking metric embedding, and then introduce the path-based ranking metric dual-level attention network learning framework.

2.1 The Problem

Before presenting the problem, we first introduce some basic notions and terminologies. Since the contents of nodes in networking sites are always the sequential data with variant length, we then encode them into the fixed length feature representation using recurrent neural networks. Given a set of nodes $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ of size n , we take the last hidden layer of recurrent neural networks as their semantic embeddings by $V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ where \mathbf{v}_i is the embedding vector for the latent representation of the i -th node. We then denote the link set of nodes in networks by $E = \{e_1, e_2, \dots, e_m\}$ of size m , where $e_i = (\mathbf{x}_j, \mathbf{x}_k)$ is the i -th edge.

We denote the set of betweenness paths for the s -th node and the t -th node with length l by $P_{(s,t)}^{(l)}$, which is derived from set E . Given the path length threshold L , we consider that the path set $P_{(s,t)}$ is formed by $P_{(s,t)} = P_{(s,t)}^{(2)} \cup P_{(s,t)}^{(3)} \dots \cup P_{(s,t)}^{(L-1)}$, which is the set union of $P_{(s,t)}^{(l)}$. We denote the collection of path sets by P . We then consider that the predictive function $f_w(P_{(s,t)}) = \mathbf{w}^T h(P_{(s,t)})$ quantifies the likelihood of the link formation between the s -th node and the t -node based on path set $P_{(s,t)}$. The \mathbf{w} is the parameter vector and $h(P_{(s,t)})$ is the embedding model for path set $P_{(s,t)}$. Inspired by negative edge sampling [Tang *et al.*,], we introduce the relative ranking metric to learn the discriminative model for link prediction, which is in the form of triplet constraints. Given edge $e = (\mathbf{x}_i, \mathbf{x}_j)$, we randomly sample another node $\mathbf{x}_k \in X$ to obtain the negative link $e' = (\mathbf{x}_i, \mathbf{x}_k)$

(i.e., $e' \notin E$). We then denote the triplet constraint by the order tuple (i, j, k) , meaning that “the likelihood of link formation between the pair of the i -th node and the j -th node is higher than that of the pair of the i -th node and the k -th node”. Let $T = \{(i, j, k)\}$ be the set of triplet constraints derived from the negative sampling process based on set E in networks. More formally, we aim to learn the ranking metric function that for any $(i, j, k) \in T$, the inequality holds:

$$f_w(P_{(i,j)}) > f_w(P_{(i,k)}) \iff \mathbf{w}^T h(P_{(i,j)}) > \mathbf{w}^T h(P_{(i,k)}).$$

Using the notations above, we define the problem of link prediction from the viewpoint of path-based proximity ranking metric learning as follows. Given the network graph $G = (X, E)$ with node set X and edge set E , our goal is to learn the ranking metric predictive function $f_w(P_{(s,t)})$ for link prediction. The endpoints $e = (\mathbf{x}_s, \mathbf{x}_t)$ ($e \notin E$) with the highest value according to $f_w(P_{(s,t)})$ is then returned.

2.2 Path-based Proximity Ranking Metric Embedding with Recurrent Neural Networks

In this section, we propose the hierarchical path-based proximity ranking metric learning method for link prediction in networking sites in Figures 1(a) and 1(b). We first present the dual-level attention learning method and then introduce the multi-step reasoning framework for path-based ranking metric embedding.

We employ the efficient path enumeration algorithm [Wang *et al.*, 2007] to collect the set $P_{(s,t)}$ of betweenness paths for the link endpoints \mathbf{x}_s and \mathbf{x}_t in network. We then choose the proper embedding method for path representation. Given a path of nodes $\mathbf{p}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ik})$ ($\mathbf{p}_i \in P_{(s,t)}$), we learn the node-level embedding of \mathbf{x}_j by pre-trained LSTM model [Hochreiter and Schmidhuber, 1997] as \mathbf{v}_{ij} and then use the sequence $(\mathbf{v}_{i1}, \mathbf{v}_{i2}, \dots, \mathbf{v}_{ik})$ as the input of the bidirectional LSTM model for path representation. The bidirectional LSTM model consists of a forward LSTM model and a backward LSTM model. The backward LSTM model has the same network structure with the forward one while its input sequence is reversed. We denote the hidden state of the j -th node in path \mathbf{p}_i as \mathbf{h}_{ij}^f by forward LSTM model, and its hidden state by backward LSTM model as \mathbf{h}_{ij}^b . Thus, the hidden state of the j -th node from the bidirectional LSTM layer is denoted by $\mathbf{h}_{ij} = [\mathbf{h}_{ij}^f, \mathbf{h}_{ij}^b]$, and then the representation of path \mathbf{p}_i is given by $\mathbf{p}_i = (\mathbf{h}_{i1}, \mathbf{h}_{i2}, \dots, \mathbf{h}_{ik})$.

Inspired by attention mechanism [Xu *et al.*, 2015], we propose the dual-level attention networks that jointly explore the discriminative nodes and critical paths according to the paired nodes \mathbf{x}_s and \mathbf{x}_t for link prediction. The joint representation of the paired nodes \mathbf{x}_s and \mathbf{x}_t is given by the element-wise product of their embeddings from the pre-trained LSTM model as $\mathbf{z}_{(s,t)}^{(0)} = \mathbf{v}_s \otimes \mathbf{v}_t$. We denote the vector $\mathbf{z}_{(s,t)}^{(0)}$ as the joint representation for paired nodes and \otimes is the element-wise product operator. We first introduce the node-level attention network to explore the discriminative node embeddings from the path representation. Given the joint representation of paired nodes $\mathbf{z}_{(s,t)}^{(0)}$ and path representation $\mathbf{p}_i = (\mathbf{h}_{i1}, \mathbf{h}_{i2}, \dots, \mathbf{h}_{ik})$, the node-level attention score of

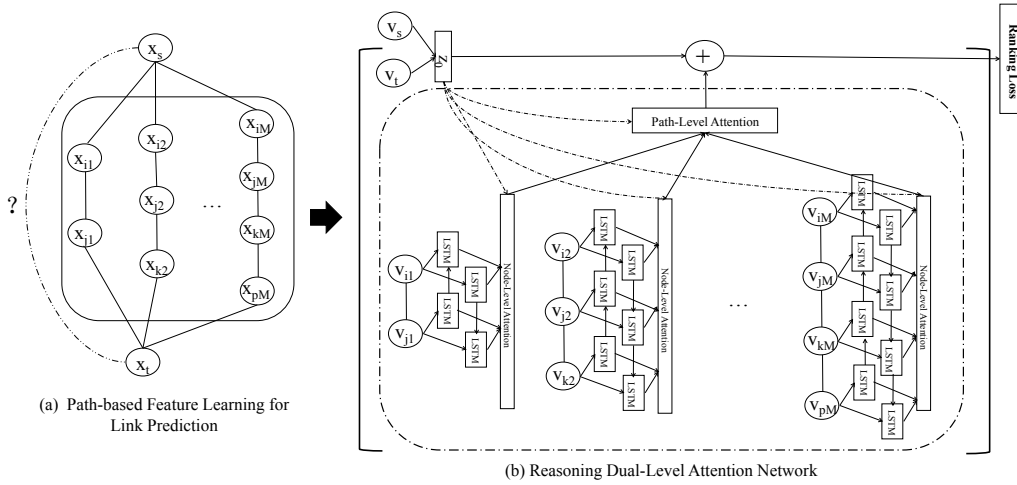


Figure 1: The Overview of Path-based Proximity Ranking Metric Dual-Level Attention Learning (PRML). (a) The set $P_{(s,t)}$ of the betweenness paths for paired nodes \mathbf{x}_s and \mathbf{x}_t is collected by path enumeration algorithm. (b) The dual-level attention network $h(P_{(s,t)})$ learns the discriminative joint representation of the betweenness paths for paired nodes \mathbf{x}_s and \mathbf{x}_t . (c) The path-based proximity ranking metric function $f_w(\cdot)$ is then trained with triplet constraints T and joint representation of their betweenness paths.

node \mathbf{x}_{ij} is given by:

$$s_{ij}^{(n)}(\mathbf{z}_{(s,t)}^{(0)}) = \mathbf{h}^{(n)} \tanh(\mathbf{W}_{zs}^{(n)} \mathbf{z}_{(s,t)}^{(0)} + \mathbf{W}_{hs} \mathbf{h}_{ij} + \mathbf{b}_n), \quad (1)$$

where $\mathbf{W}_{zs}^{(n)}$, \mathbf{W}_{hs} are parameter matrices, $\mathbf{h}^{(n)}$ is parameter vector and \mathbf{b}_n is the bias vector. The \mathbf{h}_{ij} is the hidden state of the j -th node from the bidirectional layer. For each node \mathbf{x}_{ij} , its activation by softmax function is given by $\alpha_{ij}(\mathbf{z}_{(s,t)}^{(0)}) = \frac{\exp(s_{ij}^{(n)}(\mathbf{z}_{(s,t)}^{(0)}))}{\sum_j \exp(s_{ij}^{(n)}(\mathbf{z}_{(s,t)}^{(0)}))}$, which is the normalization of node-level attention scores. Therefore, the latent representation of path \mathbf{p}_i with attention mechanism is denoted by $\mathbf{y}_i = \sum_j \alpha_{ij}(\mathbf{z}_{(s,t)}^{(0)}) \mathbf{h}_{ij}$. We then propose the path-level attention network to explore the critical paths according to paired nodes representation $\mathbf{z}_{(s,t)}^{(0)}$. Given the representation of paired nodes $\mathbf{z}_{(s,t)}^{(0)}$ and the path set $P_{(s,t)}$, the path-level attention score of path $\mathbf{p}_i \in P_{(s,t)}$ is given by:

$$s_i^{(p)}(\mathbf{z}_{(s,t)}^{(0)}) = \mathbf{h}^{(p)} \tanh(\mathbf{W}_{zs}^{(p)} \mathbf{z}_{(s,t)}^{(0)} + \mathbf{W}_{ys} \mathbf{y}_i + \mathbf{b}_p), \quad (2)$$

where $\mathbf{W}_{zs}^{(p)}$, \mathbf{W}_{ys} are parameter matrices, $\mathbf{h}^{(p)}$ is parameter vector and \mathbf{b}_p is the bias vector. The vector \mathbf{y}_i is the latent representation of path \mathbf{p}_i with node-level attention mechanism. We note that the shorter the path, the higher chance that the path influences the link formation between two nodes. Unlike the node-level attention, we introduce the damping factor that penalizes the contribution of longer path exponentially for path-level attention score regularization. For each path \mathbf{p}_i , its activation by softmax function is given by $\beta_i(\mathbf{z}_{(s,t)}^{(0)}) = \frac{\exp(s_i^{(p)}(\mathbf{z}_{(s,t)}^{(0)})) \cdot \exp(-\gamma l_i)}{\sum_{\mathbf{p}_i \in P_{(s,t)}} \exp(s_i^{(p)}(\mathbf{z}_{(s,t)}^{(0)})) \cdot \exp(-\gamma l_i)}$, which is the normalization of path-level attention scores with path length awareness. The $\exp(-\gamma l_i)$ is the damping factor for path \mathbf{p}_i ,

where l_i is the path length and γ is the coefficient. Therefore, the latent representation of path set $P_{(s,t)}$ with attention mechanism of paired nodes representation $\mathbf{z}_{(s,t)}^{(0)}$ is given by $g_{\mathbf{z}_{(s,t)}^{(0)}}(P_{(s,t)}) = \sum_{\mathbf{p}_i \in P_{(s,t)}} \beta_i(\mathbf{z}_{(s,t)}^{(0)}) \sum_j \alpha_{ij}(\mathbf{z}_{(s,t)}^{(0)}) \mathbf{h}_{ij}$, where $\beta_i(\mathbf{z}_{(s,t)}^{(0)})$ is the path-level activation coefficient and $\alpha_{ij}(\mathbf{z}_{(s,t)}^{(0)})$ is the node-level activation coefficient.

We then incorporate the multi-step reasoning framework [Sukhbaatar *et al.*, 2015] for the proposed dual-level attention network to further improve the performance of link prediction. Given dual-level attention network $g_{\mathbf{z}}(\cdot)$ and path set $P_{(s,t)}$, the reasoning dual-level attention network learning is given by

$$\mathbf{z}_{(s,t)}^{(k)} = \mathbf{z}_{(s,t)}^{(k-1)} + g_{\mathbf{z}_{(s,t)}^{(k-1)}}(P_{(s,t)}), \quad (3)$$

$$\mathbf{z}_{(s,t)}^{(0)} = \mathbf{v}_s \otimes \mathbf{v}_t, \quad (4)$$

which is recursively updated. The latent representation of path set $P_{(s,t)}$ is returned after the K -th update, given by $\mathbf{h}(P_{(s,t)}) = \mathbf{z}_{(s,t)}^{(K)}$. The learning process of reasoning dual-level attention network is illustrated in Figure 1(b).

2.3 Proximity Ranking Metric Learning

In this section, we present the details of our path-based proximity ranking metric learning method and summarize the main training process in Algorithm 1.

We first start with negative edge sampling for set E to construct the triplet constraints T for ranking metric learning. For each triplet constraint $t = (i, j, k)$, we design its ranking loss function by

$$\mathcal{L}(t) = \max(0, C + f_w(P_{(i,k)}) - f_w(P_{(i,j)})), \quad (5)$$

Algorithm 1 Path-based Proximity Ranking Metric Learning

Input: Network Graph $G = (X, E)$

Output: Ranking metric predictive function $f_w(\cdot)$

```

1: for  $e = (\mathbf{x}_i, \mathbf{x}_j) \in E$  do
2:    $\mathbf{x}_k \leftarrow \text{NegativeSampling}(\mathbf{x}_i, \mathbf{x}_j)$ 
3:    $P_{(i,j)} \leftarrow \text{PathEnum}(\mathbf{x}_i, \mathbf{x}_j)$ 
4:    $P_{(i,k)} \leftarrow \text{PathEnum}(\mathbf{x}_i, \mathbf{x}_k)$ 
5:    $T \leftarrow T \cup (i, j, k)$ 
6: end for
7: for  $t = (i, j, k) \in T$  do
8:   Accumulate ranking loss  $\mathcal{L}(t)$  by Equation (5)
9: end for
10: Accumulate the total loss by  $\mathcal{L} = \sum_{t \in T} \mathcal{L}(t)$ .
11: Update predictive function  $f_w(\cdot)$  by SGD.

```

where the hyper-parameter C ($0 < C < 1$) controls the margin in the loss function and $f_w(P_{(i,k)})$, $f_w(P_{(i,j)})$ are the embedding functions of set $P_{(i,k)}$ and $P_{(i,j)}$ by reasoning dual-level attention network, respectively. We denote all the model coefficients including neural network parameters and the result function parameter by Θ . Therefore, the objective function in our learning process is given by:

$$\min_{\Theta} \mathcal{L}(\Theta) = \sum_{t \in T} \mathcal{L}(t) + \lambda \|\Theta\|^2,$$

where λ is the trade-off parameter between the training loss and regularization. To optimize the objective, we employ the stochastic gradient descent (SGD) with the diagonal variant of AdaGrad in [Kingma and Ba, 2014].

3 Experiments

In this section, we conduct several experiments on the co-authorship network DBLP and the collaborative development network Github, to show the effectiveness of our approach for the problem of link prediction in heterogeneous networks.

3.1 Experimental Setting

We present the data preparation procedure for obtaining the datasets from both networks. For DBLP dataset, we choose the faculty in research institutions as the user node, and the paper with the number of citations greater than 3 as the item node. The DBLP dataset contains 15,728 user nodes and 97,523 item nodes. The number of observed links between user node and item node is 118,245, and the number of observed links among item nodes is 247,081. For Github dataset, we choose the developer that participates in more than 20 projects as user node, and the project developed by more than 500 users as item node. The Github dataset is composed of 2,407 user nodes and 6,490 item nodes. The number of observed links between user node and item node is 19,153 and the number of observed links among user nodes is 13,153. Thus, the average degree in DBLP data is 3.2 and average degree in Github data is 3.6. We sort the network links based on their established timestamps. We use the first 70% links as training set, 10% ones as the validation set and

Table 1: Experimental results using Github dataset.

Method	MAP@10	NDCG@10	AUC
FactRank	0.5824	0.5831	0.6292
Katz	0.5918	0.6323	0.6439
LPGM	0.5239	0.6803	0.6815
MMMF	0.6439	0.6688	0.7157
HNE	0.7419	0.7183	0.7141
SDNE	0.7430	0.7280	0.7252
PRML ₁	0.7442	0.7318	0.7730
PRML ₂	0.7487	0.7394	0.7823

Table 2: Experimental results using DBLP dataset.

Method	MAP@10	NDCG@10	AUC
FactRank	0.5021	0.5927	0.6513
Katz	0.5312	0.6306	0.6915
LPGM	0.5422	0.6278	0.7023
MMMF	0.6542	0.6927	0.7481
HNE	0.7649	0.7610	0.7805
SDNE	0.7663	0.7626	0.7882
PRML ₁	0.7721	0.7826	0.7804
PRML ₂	0.7802	0.7923	0.7915

the remaining 20% links as the testing set. So the training, validation and test data do not have overlap. The dataset will be provided later.

3.2 Performance Comparisons

We evaluate the performance of our proposed method based on three widely-used ranking evaluation criteria for the problem of link prediction, i.e., mean average precision (MAP) [Chang *et al.*, 2015; Wang *et al.*, 2016], normalized discounted cumulative gain (nDCG) and area under the curve (AUC) [Menon and Elkan, 2011; Wang *et al.*, 2007]. We compare our proposed method with other six state-of-the-art methods for the problem of network link prediction as follows:

- **FactRank** method [Menon and Elkan, 2011] employs the low-rank approximation based ranking function for link prediction form the partial observed network adjacent matrix.
- **Katz** method [Katz, 1953] estimates the likelihood of the unknown links based on the set of betweenness path of their endpoints.
- **LPGM** method [Wang *et al.*, 2007] is the local probabilistic graphical model based on the co-occurrence probability of two nodes for link prediction.
- **MMMF** method [Zhu *et al.*, 2016] is the max-margin matrix factorization model for predicting the missing links in the partially observed network.
- **HNE** method [Chang *et al.*, 2015] learns the heterogeneous network embedding of both user node and item node based on the network content and its observed structure for link prediction.
- **SDNE** method [Wang *et al.*, 2016] learns the vertex embedding of heterogeneous network based on first-order

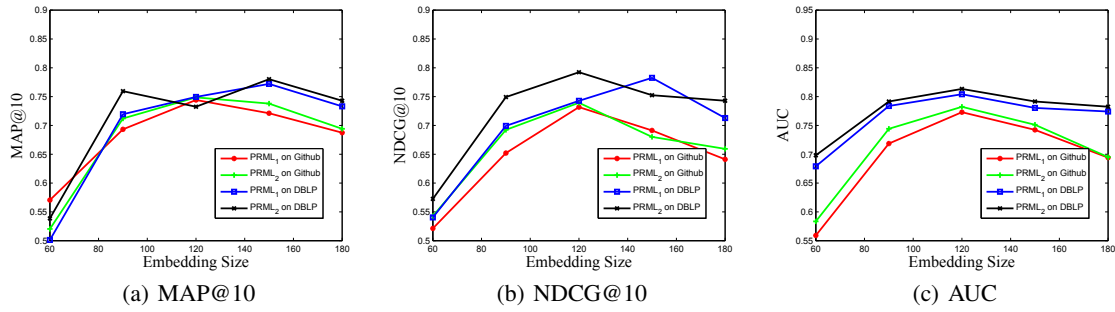


Figure 2: Effect of embedding dimension on MAP@10, NDCG@10 and AUC.

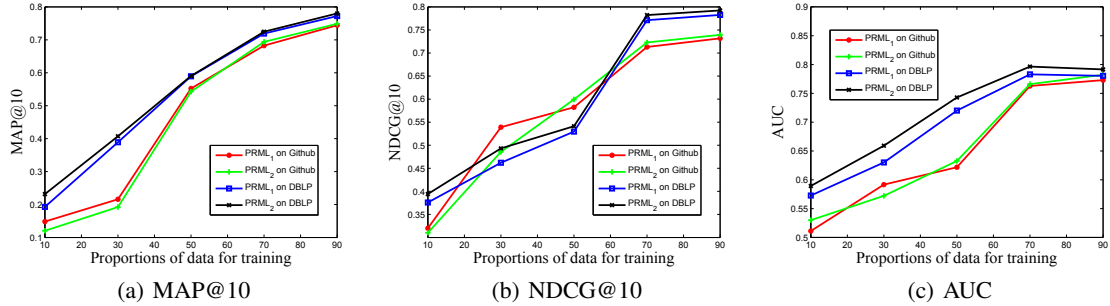


Figure 3: Effect of proportion of data for training on MAP@10, NDCG@10 and AUC.

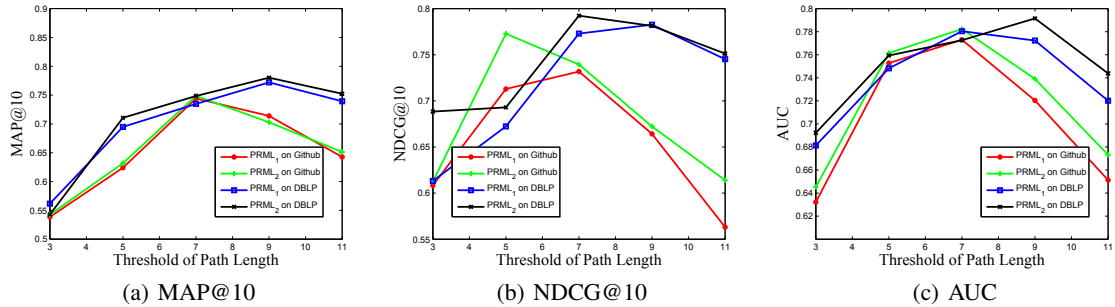


Figure 4: Effect of threshold of path length on MAP@10, NDCG@10 and AUC.

and second-order network proximity for link prediction.

Among them, method Katz is based on set of betweenness path for the link endpoints, methods FactRank is ranking model based on low-rank approximation, methods LPGM and MMMF are probabilistic model, and methods HNE and SDNE are network embedding model for link prediction. Unlike the previous studies, our method PRML learns the ranking metric predictive function based on path set for link prediction. We then denote our PRML with one reasoning step by PRML₍₁₎ and our method with two steps by PRML₍₂₎, respectively.

Tables 1 and 2 show the experimental results of the methods on MAP@10, NDCG@10 and AUC using Github and DBLP datasets, respectively. The hyperparameters and parameters which achieve the best performance on the validation set are chosen to conduct the testing evaluation. We report the average value of all the methods on three evaluation

criteria. The experiments reveal two-fold interesting points:

- The path-based learning method Katz achieves better performance than the neighborhood-based factorization method FactRank, which suggests that the path-based information is critical for link prediction.
- The deep network embedding methods HNE and SDNE outperform the other baseline algorithms. This suggests that the performance of link prediction can be improved with discriminative node feature representation.
- In all the cases, our PRML method achieves the best performance. This fact shows that the proximity ranking metric learning framework that exploits both discriminative joint representation of betweenness paths and multi-step reasoning can further improve the performance of link prediction.

In our approach, there are three essential parameters, which

are the dimension of node embedding, the proportion of training data used for model learning and the threshold of path length using both validation datasets. We plot the performance of $PRML_{(1)}$ and $PRML_{(2)}$ using Github dataset with red and green lines, while illustrating the performance of $PRML_{(1)}$ and $PRML_{(2)}$ using DBLP dataset using blue and black lines. We then investigate the effect of these parameters of our method by varying the dimension of node embedding from 60 to 180 in Figures 2(a), 2(b) and 2(c), the proportion of training data for model learning from 10% to 90% in Figures 3(a), 3(b) and 3(c), and the threshold of path length from 3 to 11 in Figures 4(a), 4(b) and 4(c), respectively. We observe that our method achieves the best performance on Github dataset when the dimension of node embedding is set to 120 and the threshold of path length is set to 7. We also report that the best performance of our method can be achieved on DBLP dataset when the dimension of node embedding is set to 150 and the threshold of path length is set to 9. This suggests that by leveraging the proximity sequence of nodes with reasoning dual-level attention network, we can further improve the performance of link prediction.

We train the proposed method on machines with Linux OS, Intel(R) Core i7-5930K 3.50GHz and two GTX TITAN X graphic cards. The running time of training $PRML_{(1)}$ and $PRML_{(2)}$ on Github dataset are 8 hours and 14 hours. The time cost of learning $PRML_{(1)}$ and $PRML_{(2)}$ on DBLP dataset are 11 hours and 19 hours. The running time for the training process validates the efficiency of our proposed method.

4 Related Work

In this section, we briefly review some related work on link prediction and network embedding.

The problem of link prediction is to estimate the proximity of the link endpoints based on the partially observed links and the node contents in network, which has attracted considerable attention recently [Menon and Elkan, 2011; Wang *et al.*, 2007; Zhu *et al.*, 2016]. The early link prediction approaches are to compute the similarity between each pair of network nodes for determining the proximity, which are based on the number of shared common features [Lin, 1998] and localized neighborhood features [Lü *et al.*, 2009]. To leverage the correlated link features, several matrix factorization models [Menon and Elkan, 2011; Cao *et al.*, 2014; Agrawal *et al.*, 2013; Leskovec *et al.*, 2010] have been proposed to make prediction from the observed network links. Menon *et al.* [Menon and Elkan, 2011] employ the supervised matrix factorization method for solving the link prediction problem. Cao *et al.* [Cao *et al.*, 2014] study the problem of collective prediction of multiple types of links in heterogeneous information networks. Agrawal *et al.* [Agrawal *et al.*, 2013] propose a matrix factorization based method with trace norm for link sign prediction. Zhao *et al.* [Zhao *et al.*, 2016b] propose the user preference learning method for predicting the online link between the user and the item. Leskovec *et al.* [Leskovec *et al.*, 2010] study the link sign prediction problem in online social networks. On the other hand, various probabilistic approaches [Wang *et al.*, 2007;

Miller *et al.*, 2009] have also been developed for learning a link probability distribution model from the observed network. Unlike previous studies, we formulate the problem of link prediction from the viewpoint of path-based proximity ranking metric learning with recurrent neural networks.

Recently, deep learning models show great potential for learning effective representation for network mining applications [Perozzi *et al.*, 2014; Chang *et al.*, 2015], which exploit the network structure for node embedding. Perozzi *et al.* [Perozzi *et al.*, 2014; Tang *et al.*,] propose the local network structure embedding method. Chang *et al.* [Chang *et al.*, 2015] propose the embedding method for heterogeneous networks. Yang *et al.* [Yang *et al.*,] develop the learning method for attributed networks. Liu *et al.* [Liu *et al.*, 2017] propose the proximity embedding for the semantic search on heterogeneous graph. Fang *et al.* [Fang *et al.*, 2016] develop the metagraph-based learning for semantic proximity graph search. However, the objective of proximity ranking metric learning in our problem is different from these deep learning models. Thus, these methods may not be directly suitable for our problem.

5 Conclusion

In this paper, we consider the problem of link prediction from the viewpoint of learning path-based proximity ranking metric embedding. We propose a novel ranking metric network learning framework by jointly exploiting both node-level and path-level attentional proximity of the endpoints to their betweenness paths in order to learn the discriminative joint representation for link prediction. We then develop the path-based dual-level attention learning method with recurrent neural networks for proximity ranking metric embedding. We next incorporate the multi-step reasoning framework for the proposed dual-level attention networks to further improve the performance of link prediction. The extensive experiments on two large-scale datasets show that our method achieves better performance than other state-of-the-art solutions to the problem.

Acknowledgements

This work was supported by National Basic Research Program of China (973 Program) under Grant 2013CB336500, and National Natural Science Foundation of China under Grant 61602405, Fundamental Research Funds for the Central Universities 2016QNA5015 and the China Knowledge Centre for Engineering Sciences and Technology. The Project is also Supported by the Key Laboratory of Advanced Information Science and Network Technology of Beijing (XDXX1603). We thank the support from the National Natural Science Foundation of China (No. 61502418) and the research grant for Human-centered Cyber-physical Systems Programme at Advanced Digital Sciences Center from Singapore Agency for Science, Technology and Research (A*STAR).

References

- [Agrawal *et al.*, 2013] Priyanka Agrawal, Vikas K Garg, and Ramasuri Narayanam. Link label prediction in signed social networks. In *IJCAI*, 2013.
- [Cao *et al.*, 2014] Bokai Cao, Xiangnan Kong, and S Yu Philip. Collective prediction of multiple types of links in heterogeneous information networks. In *ICDM*, pages 50–59. IEEE, 2014.
- [Chang *et al.*, 2015] Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C Aggarwal, and Thomas S Huang. Heterogeneous network embedding via deep architectures. In *SIGKDD*, pages 119–128. ACM, 2015.
- [Fang *et al.*, 2016] Yuan Fang, Wenqing Lin, Vincent W Zheng, Min Wu, Kevin Chen-Chuan Chang, and Xiao-Li Li. Semantic proximity search on graphs with metagraph-based learning. In *ICDE*, pages 277–288. IEEE, 2016.
- [Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [Katz, 1953] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.
- [Kingma and Ba, 2014] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [Leskovec *et al.*, 2010] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Predicting positive and negative links in online social networks. In *WWW*, pages 641–650. ACM, 2010.
- [Liben-Nowell and Kleinberg, 2007] David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *Journal of the American society for information science and technology*, 58(7):1019–1031, 2007.
- [Lin, 1998] Dekang Lin. An information-theoretic definition of similarity. In *ICML*, volume 98, pages 296–304. Cite-seer, 1998.
- [Liu *et al.*, 2014] Xiaozhong Liu, Yingying Yu, Chun Guo, and Yizhou Sun. Meta-path-based ranking with pseudo relevance feedback on heterogeneous graph for citation recommendation. In *CIKM*, pages 121–130. ACM, 2014.
- [Liu *et al.*, 2017] Zemin Liu, Vincent W Zheng, Zhou Zhao, Fanwei Zhu, Kevin Chen-Chuan Chang, Minghui Wu, and Jing Ying. Semantic proximity search on heterogeneous graph by proximity embedding. 2017.
- [Lü *et al.*, 2009] Linyuan Lü, Ci-Hang Jin, and Tao Zhou. Similarity index based on local paths for link prediction of complex networks. *Physical Review E*, 80(4):046122, 2009.
- [Martínez *et al.*, 2016] Víctor Martínez, Fernando Berzal, and Juan-Carlos Cubero. A survey of link prediction in complex networks. *ACM Computing Surveys (CSUR)*, 49(4):69, 2016.
- [Menon and Elkan, 2011] Aditya Krishna Menon and Charles Elkan. Link prediction via matrix factorization. In *ECML/PKDD*, pages 437–452. Springer, 2011.
- [Mikolov *et al.*, 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *NIPS*, pages 3111–3119, 2013.
- [Miller *et al.*, 2009] Kurt Miller, Michael I Jordan, and Thomas L Griffiths. Nonparametric latent feature models for link prediction. In *NIPS*, pages 1276–1284, 2009.
- [Perozzi *et al.*, 2014] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *SIGKDD*, pages 701–710. ACM, 2014.
- [Sukhbaatar *et al.*, 2015] Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. End-to-end memory networks. In *NIPS*, pages 2440–2448, 2015.
- [Sun *et al.*, 2013] Yizhou Sun, Brandon Norick, Jiawei Han, Xifeng Yan, Philip S Yu, and Xiao Yu. Pathselclus: Integrating meta-path selection with user-guided object clustering in heterogeneous information networks. *TKDD*, 7(3):11, 2013.
- [Tang *et al.*,] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *WWW*.
- [Wang *et al.*, 2007] Chao Wang, Venu Satuluri, and Srinivasan Parthasarathy. Local probabilistic models for link prediction. In *ICDM*, pages 322–331. IEEE, 2007.
- [Wang *et al.*, 2016] Daixin Wang, Peng Cui, and Wenwu Zhu. Structural deep network embedding. In *SIGKDD*, pages 1225–1234. ACM, 2016.
- [Xu *et al.*, 2015] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron C Courville, Ruslan Salakhutdinov, Richard S Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *ICML*, volume 14, pages 77–81, 2015.
- [Yang *et al.*,] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y Chang. Network representation learning with rich text information.
- [Zhang *et al.*, 2014] Jiawei Zhang, Philip S Yu, and Zhi-Hua Zhou. Meta-path based multi-network collective link prediction. In *SIGKDD*, pages 1286–1295. ACM, 2014.
- [Zhao *et al.*, 2015] Zhou Zhao, Lijun Zhang, Xiaofei He, and Wilfred Ng. Expert finding for question answering via graph regularized matrix completion. *TKDE*, 27(4):993–1004, 2015.
- [Zhao *et al.*, 2016a] Zhou Zhao, Xiaofei He, Deng Cai, Lijun Zhang, Wilfred Ng, and Yueting Zhuang. Graph regularized feature selection with data reconstruction. *TKDE*, 28(3):689–700, 2016.
- [Zhao *et al.*, 2016b] Zhou Zhao, Hanqing Lu, Deng Cai, Xiaofei He, and Yueting Zhuang. User preference learning for online social recommendation. *TKDE*, 28(9):2522–2534, 2016.
- [Zhu *et al.*, 2016] Jun Zhu, Jiaming Song, and Bei Chen. Max-margin nonparametric latent feature models for link prediction. *PAMI*, 2016.