The following publication L. Xu, X. Wei, J. Cao and P. S. Yu, "Multiple Social Role Embedding," 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Tokyo, Japan, 2017, pp. 581-589 is available at https://doi.org/10.1109/DSAA.2017.23.

Multiple Social Role Embedding

Linchuan Xu*, Xiaokai Wei†, Jiannong Cao* and Philip S. Yu‡
*The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
Email: 13901626r@connect.polyu.hk, csjcao@comp.polyu.edu.hk
†Facebook Inc., 1 Hacker Way, Menlo Park, CA, USA
Email: weixiaokai@gmail.com
‡University of Illinois at Chicago, Chicago, Illinois, USA
Email: psyu@uic.edu

***Abstract*—Network embedding has been increasingly employed in networked data mining applications as it is effective to learn node embeddings that encode the network structure. Existing network models usually learn a single embedding for each node. In practice, a person may interact with others in different roles, such as interacting with schoolmates as a student, and with colleagues as an employee. Obviously, different roles exhibit different characteristics or features. Hence, only learning a single embedding responsible for all roles is not appropriate. In this paper, we thus introduce a concept of multiple social role (MSR) into social network embedding for the first time. The MSR models multiple roles people play in society, such as student and employee. To make the embedding more versatile, we thus propose a multiple social role embedding (MSRE) model to preserve both the network structure and social roles. Empirical evaluation on various real-world social networks demonstrates advantages of the proposed MSRE over the state-of-the-art embedding models in link prediction and multi-label classification.**

***Index Terms*—network embedding; social networks; data mining;**

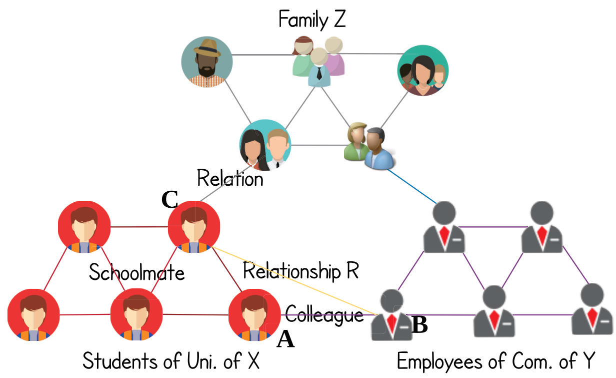

Fig. 1: An example of network with multiple social roles, where straight lines indicate interactions between people, and each color denotes a kind of relationship.

I. Introduction

Social network analysis has been attracting increasing interests as social networks, such as Facebook and Twitter, are integral parts of daily life. Moreover, social network analysis is beneficial in various aspects, from information diffusion [3], community detection [6] [26], recommendation [10] to link prediction [12] [27]. Among the information we can utilize from networks, interactions of vertices are the foundations on which all these applications are based.

Recently, network embedding has been used to assist analyzing social networks as it is effective to learn low-dimensional latent features that encode interactions[17] [21] [9] [30] [31]. The major reason for turning vertex interactions into latent features is that interactions are discrete and usually spare with respect to the entire network, e.g., the adjacent matrix of the network is usually spare. Hence, vertex interactions are not suitable features for tuple-based machine learning models, such as SVM and Logistic Regression. The basic idea of embedding is to preserve the network structure by presenting vertices which are close in the original space to be close in the latent space. Existing embedding methods are effective, but they share a common problem of not distinguishing the different roles of the same vertex.

Fig. 1 explains why different roles should be distinguished. In practice, a person may have multiple roles simultaneously, such as student and employee. For example, person A has a role of student of university X, and employee in company Y. In the theory of social role taking [11] [19], normal people would adopt and act out a particular role in each interaction with reference to the social environment. In other words, only a single role participates in each interaction with the other roles being inactive. For example, person A interacts with schoolmate C most likely as a student while A interacts with colleague B most likely as an employee. Obviously, different roles exhibit different characteristics as students are concerned mostly about studies while employees are concerned mostly about work. Hence, each role should be separately preserved and appropriate roles should be adopted in each interaction. Conversely, if all the roles are mixed up in a single embedding, multiple roles would be active in each interaction, which violates the theory of social role taking.

In this paper, we thus propose a multiple social role embedding (MSRE) model to preserve the network structure and social roles simultaneously. In contrast to existing methods where each node has a single global embedding, MSRE model assigns multiple embeddings to each vertex. As a result, each embedding can account for the unique nature of the corresponding role and therefore finer grained information could be preserved in the embeddings. Hence, role-specific embeddings can lead to better performance in various tasks of network analysis, such as visualization, link prediction and classification.

The major challenge for learning role-specific embeddings is how to infer the social roles. One key insight we rely on is the following: roles produce relationships, e.g., role *student* produces relationship *schoolmate* and relationships formed by the same role tend to share certain common characteristics, such as similar interests or background [7]. We introduce the concept of *role representative* embedding, which aims to capture such common information shared by this role. Then we introduce a mixture gating function in which each node is assigned to a role with a certain probability, based on their *role affinity* to each role representative. With this soft role assignment, the involvement of each node over different social roles can be reflected by different probabilities.

Hence, each link is formed by the interaction between different roles of two nodes. To infer multiple role embedding for network nodes, we need to perform a joint inference on the role representative and role-specific embedding, by maximizing the likelihood of links generated from these roles.

The contributions of this paper are summarized as follows:

1. We point out the limitation of learning single global embedding and propose to learn multiple role-specific embeddings. To our best knowledge, this is the first attempt to consider multiple social roles while embedding networks.
2. We propose a multiple social role embedding (MSRE) model to learn role-specific embeddings for network vertices via joint social role and embedding inference.
3. We present comprehensive evidence on five real-world social networks to show that the proposed MSRE model outperforms three recent network embedding models in link prediction and multi-label classification.

The rest of the paper is organized as follows. Section 2 presents related work. Section 3 presents the proposed MSRE model and the optimization algorithm to solve the model. In section 4, we evaluate the proposed MSRE on five real-world social networks against three recent network embedding models. In section 5, we conclude and introduce our future work.

II. Related Work

The *major* category of related work should be network embedding to learn latent representations for network vertices. Various embedding methods [4] [5] have been proposed before, but they are not designed for social networks since they just construct affinity networks using feature vectors of independent data points.

A recent social network embedding model is DeepWalk [17], which presents sequences of vertices obtained from random walks to be close in the latent space. This embedding principle is based on the connection between a sequence of nodes and a sequence of words in natural languages, i.e., both node frequency and word frequency in random sequences follow the power law. And presenting a sequence of words to be close has been demonstrated effective to learn word embeddings [14]. Afterwards, TADW [32] extends DeepWalk to embed both the network structure and node content.

Another recent model is LINE [21] for large-scale network embedding. LINE presents nodes with either first-order or second-order interactions to be close. Although MSRE also presents nodes with first-order interactions to be close like LINE, MSRE assigns multiple role-specific embeddings to each node.

The more recent one is node2vec [9], which defines a flexible notion of neighborhood instead of a rigid one in DeepWalk and LINE so as to exploit more diverse interaction information. However, none of these methods consider multiple social roles, which can be observed in practice.

A *Minor* category of related work should be studies inferring multiple social roles in social networks, such as mixed membership stochastic block models [1] [28], and multi-agent based simulation models [16]. This category of related work is considered as minor because the proposed MSRE is a network embedding model. Moreover, as a comparison, the MSRE model not only estimates the role probabilities, but also learns role representations, which can be used in subsequent data mining tasks, such as link prediction and classification.

III. Methodology

A. Overview

The proposed MSRE deals with social networks denoted as $G(V, E)$, where V is a set of vertices, E is a set of weighted or unweighted, directed or undirected edges. The MSRE starts with learning a global embedding by preserving the network structure without considering social roles for each vertex. The global embedding is learned to facilitate the discovery of common characteristics among nodes. After discovering common characteristics shared by certain nodes, a social role is defined and a corresponding role representative is selected, and every vertex sharing the common characteristics is assigned with a corresponding role as discussed in the introduction. This inference process acts as the pre-training of the MSRE. Afterwards, the MSRE learns role-specific embeddings via jointly embedding the structure and inferring roles. The role inference has to be further performed because the role assignment is just performed without considering multiple roles, and role-specific embeddings may lead to the change of role representatives.

B. Global Pre-training

Global pre-training is to learn a global embedding for each vertex by presenting vertices connected by edges to be close in the latent space. Closeness is formally defined as follows:

DEFINITION 1. *The* ***closeness*** *between two vertices in a certain latent space is defined as the probability of an edge between them, where the probability is quantified as follows:*

$$p(\boldsymbol{v}_i, \boldsymbol{v}_j) = \frac{1}{1 + exp\{-\boldsymbol{v}_i^\top \boldsymbol{v}_j\}}, \tag{1}$$

where $\boldsymbol{v}_i \in \mathbb{R}^D$ and $\boldsymbol{v}_j \in \mathbb{R}^D$ are column vectors of global embeddings for vertices i and j, respectively, and D is the dimension of the latent space.

Accordingly, pairs of vertices connected by edges are expected to have large probabilities and conversely, pairs of vertices not connected with small ones. To cast the structure preserving mechanism to an optimization problem, both small probabilities of pairs of vertices with an edge and large ones of pairs of vertices without an edge should be penalized. A simple yet effective function to perform the former penalty is the negative natural logarithmic function denoted as follows:

$$l(\boldsymbol{v}_i, \boldsymbol{v}_j) = -w_{ij} \times \log(p(\boldsymbol{v}_i, \boldsymbol{v}_j)), \tag{2}$$

where $w_{ij} \in \mathbb{R}$ is the weight of the edge reflecting the relationship strength. This function is chosen because it is convex, and thus it enables us to solve the optimization by gradient-based algorithms as well as obtain the global optimal. Correspondingly, the penalty for large probabilities of pairs of vertices without an edge is quantified as follows:

$$l(\boldsymbol{v}_h, \boldsymbol{v}_k) = -\log(1 - p(\boldsymbol{v}_h, \boldsymbol{v}_k)), \tag{3}$$

where $\boldsymbol{v}_h$ and $\boldsymbol{v}_k$ is pair of vertices without an edge.

The overall loss function is thus quantified as follows:

$$\begin{aligned} L(\boldsymbol{V}) = &- \sum_{(i,j)\in E} w_{ij} \log(p(\boldsymbol{v}_i, \boldsymbol{v}_j)) - \sum_{(h,k)\notin E} \log(1 - p(\boldsymbol{v}_h, \boldsymbol{v}_k)) \\ &+ \lambda ||\boldsymbol{V}||_F^2, \end{aligned} \tag{4}$$

where $\boldsymbol{V}$ is a collection of vertex embeddings, $\lambda \in \mathbb{R}$, and $||\cdot||_F^2$ is F_2-norm used as regularization.

The derivative for minimizing the loss function can obtained by differentiating $L(\boldsymbol{V})$ with respect to $\boldsymbol{v}_i$, we get:

$$\begin{aligned} \frac{\partial L(\boldsymbol{V})}{\partial \boldsymbol{v}_i} = &- \sum_{(i,j)\in E} \left[\frac{w_{ij} \exp\{-\boldsymbol{v}_i^\top \boldsymbol{v}_j\}}{1 + \exp\{-\boldsymbol{v}_i^\top \boldsymbol{v}_j\}} \times \boldsymbol{v}_j \right] \\ &+ \sum_{(i,k)\notin E} \left[\frac{\boldsymbol{v}_k}{1 + \exp\{-\boldsymbol{v}_i^\top \boldsymbol{v}_k\}} \right] + 2\lambda(\boldsymbol{v}_i), \end{aligned} \tag{5}$$

Minimizing Eq. (4) produces a global embedding for each vertex, which facilitates the discovery of common characteristics shared by nodes with a certain role so as to perform role inference. Specifically, certain nodes sharing common characteristics would be close in the embedding space. Before presenting the role inference of each node, a concept of role representative is introduced and defined as follows:

DEFINITION 2. ***Role representative*** *of a particular social role refers to the typical characteristics of the role, which are common characteristics shared by all nodes with the role.*

The role representative is useful because if a certain vertex has similar characteristics to a particular role representative, it is high likely that the vertex has the social role as well. We define a concept of role affinity to measure the similarity between a vertex and a role representative as follows:

DEFINITION 3. ***Role affinity*** *of a certain vertex to a particular role representative refers to the similarity between the vertex and the role representative.*

According to the definition, the larger the role affinity of a vertex to a particular role representative is, the larger likelihood that the vertex has such a role. Since a vertex can have a role affinity to every role representative, we adopt a multi-class logistic gating function [8] to quantify the role affinity as follows:

$$\pi_a(\boldsymbol{v}_i) = \frac{\exp\{\boldsymbol{c}_a^\top \boldsymbol{v}_i + k_a\}}{\sum_{t=1}^{A} \exp\{\boldsymbol{c}_t^\top \boldsymbol{v}_i + k_t\}}, \tag{6}$$

where $\pi_a(\boldsymbol{v}_i)$ is the role affinity of vertex i to the role representative of role a, $\boldsymbol{v}_i$ is the global embedding for vertex i, $\boldsymbol{c}_a \in \mathbb{R}^D$ is the column vector of the role representative, A is the total number of roles, and $k \in \mathbb{R}$ with different subscriptions are bias terms. The multi-class logistic gating function is adopted because it normalizes the role affinities of a vertex to all role representatives in the scale from 0.0 to 1.0. Accordingly, we adopt the role affinity as the probability with which a role is assigned to a vertex.

To conclude, the outputs of role inference in the global pre-training process are that (1) the common characteristics shared by a certain group of nodes is defined as a particular role representative, (2) the role affinity of each vertex to a particular role representative is defined as the probability that the vertex has the particular role, and (3) the role-specific embeddings of each vertex are initialized as the role representatives.

C. Joint Role Inference and Embedding Learning

With roles assigned, the probability that there exists an interaction between two vertices i and j with role-specific embeddings is reformulated by the following Eq. (1) as follows:

$$p(i,j) = \sum_a \sum_b \pi_a(\boldsymbol{v}_i) \pi_b(\boldsymbol{v}_j) p(\boldsymbol{v}_i^a, \boldsymbol{v}_j^b), \tag{7}$$

where $\pi_a(\boldsymbol{v}_i)$ and $\pi_b(\boldsymbol{v}_j)$ are role affinities, $\boldsymbol{v}_i^a \in \mathbb{R}^D$ is a role-specific embedding of vertex i, $\boldsymbol{v}_j^b \in \mathbb{R}^D$ is a role-specific embedding of vertex j, and $p(\boldsymbol{v}_i^a, \boldsymbol{v}_j^b)$ is quantified in Eq. (1).

The intuition behind the weighted probability using role-specific embeddings, i.e., $\pi_a(\boldsymbol{v}_i)\pi_b(\boldsymbol{v}_j)p(\boldsymbol{v}_i^a, \boldsymbol{v}_j^b)$, is the theory of social role taking where each person would act out a particular social role with reference to the social environment. The summation over all combinations of roles is needed because two persons are likely to interact in different environments with different roles. For example, in the environment of universities, both two persons are likely to act out as students while in the environment of job fairs, one of them may act out a student and the other act out an employer. Moreover, the summation makes the range of probability of an edge scale from 0% to 100% again.

The loss for preserving the network structure and social roles is quantified by following Eq. (4) as follows: $L(\boldsymbol{C}, \boldsymbol{V}) =$

$$
\begin{aligned}
&- \sum_{(i,j)\in E} \left[w_{ij} \log \sum_{a} \sum_{b} \pi_a(\boldsymbol{v}_i)\pi_b(\boldsymbol{v}_j) p(\boldsymbol{v}_i^a, \boldsymbol{v}_j^b) \right] \\
&- \sum_{(h,k)\notin E} \left[\log(1 - \sum_{a} \sum_{b} \pi_a(\boldsymbol{v}_h)\pi_b(\boldsymbol{v}_k) p(\boldsymbol{v}_h^a, \boldsymbol{v}_k^b)) \right] \\
&+ \lambda \sum_{a=1}^{A} \sum_{n=1}^{N} ||\boldsymbol{v}_n^a||_a^2 + \beta \sum_{a=1}^{A} ||\boldsymbol{c}_a||_a^2,
\end{aligned}
\tag{8}
$$

where $\beta \in \mathbb{R}$ is the regularization coefficient, N is the number of vertices, $\boldsymbol{C}$ is a collection of role representatives, and $||\cdot||$ is ℓ_2 norm used as regularization.

Recalling that role representatives are only pre-trained on the global embedding of each vertex, we thus need to perform joint role inference and role-specific embedding learning. In other words, the loss function needs to be minimized with respect to both $\boldsymbol{C}$ and $\boldsymbol{V}$. The loss function is not jointly convex with respect to and $\boldsymbol{C}$ and $\boldsymbol{V}$. We thus solve one variable at a time with the either variable fixed.

The derivative w.r.t $\boldsymbol{v}_i^a$ is calculated as follows: $\frac{\partial L(\boldsymbol{C},\boldsymbol{V})}{\partial v_i^a} =$

$$
\begin{aligned}
&- \sum_{(i,j)\in E} \left[\frac{w_{ij}\pi_a(\boldsymbol{v}_i)}{p(i,j)} \sum_{b}^{A} \pi_b(\boldsymbol{v}_j) p^2(\boldsymbol{v}_i^a, \boldsymbol{v}_j^b) \exp\{-(\boldsymbol{v}_i^a)^\top \boldsymbol{v}_j^b\} \boldsymbol{v}_j^b \right] \\
&- \sum_{(i,k)\notin E} \left[\frac{\pi_a(\boldsymbol{v}_i)}{p(i,k)-1} \sum_{b}^{A} \pi_b(\boldsymbol{v}_k) p^2(\boldsymbol{v}_i^a, \boldsymbol{v}_k^b) \exp\{-(\boldsymbol{v}_i^a)^\top \boldsymbol{v}_k^b\} \boldsymbol{v}_k^b \right] \\
&+ 2\lambda(\boldsymbol{v}_i^a),
\end{aligned}
\tag{9}
$$

The derivative w.r.t $\boldsymbol{c}_a$ is calculated as follows: $\frac{\partial L(\boldsymbol{C},\boldsymbol{V})}{\partial \boldsymbol{c}_a} =$

$$
\begin{aligned}
&- \sum_{(i,j)\in E} \frac{w_{ij}}{p_{ij}} \times \left\{ \sum_{b\neq a} \Big[\frac{\mathrm{d}\pi_a(\boldsymbol{v}_i)}{\mathrm{d}\boldsymbol{v}_a} \pi_b(\boldsymbol{v}_j) + \frac{\mathrm{d}\pi_a(\boldsymbol{v}_j)}{\mathrm{d}\boldsymbol{v}_a} \pi_b(\boldsymbol{v}_i) \Big] p_{ij}^{ab} \right. \\
&\left. + \frac{\mathrm{d}\left[\pi_a(\boldsymbol{v}_i)\pi_a(\boldsymbol{v}_j)\right]}{\mathrm{d}\boldsymbol{v}_a} p_{ij}^{aa} \right\} - \sum_{(h,k)\notin E} \frac{1}{(p_{hk}-1)} \\
&\times \left\{ \sum_{b\neq a} \Big[\frac{\mathrm{d}\pi_a(\boldsymbol{v}_h)}{\mathrm{d}\boldsymbol{v}_a} \times \pi_b(\boldsymbol{v}_k) + \frac{\mathrm{d}\pi_a(\boldsymbol{v}_k)}{\mathrm{d}\boldsymbol{v}_a} \pi_b(\boldsymbol{v}_h) \Big] p_{hk}^{ab} \right. \\
&\left. + \frac{\mathrm{d}\left[\pi_a(\boldsymbol{v}_h)\pi_a(\boldsymbol{v}_k)\right]}{\mathrm{d}\boldsymbol{v}_a} p_{hk}^{aa} \right\} + 2\beta(\boldsymbol{c}_a),
\end{aligned}
\tag{10}
$$

where $p_{ij}^{ab} = p(\boldsymbol{v}_i^a, \boldsymbol{v}_j^b)$ and $p_{hk}^{ab} = p(\boldsymbol{v}_h^a, \boldsymbol{v}_k^b)$.

D. The Optimization Algorithm for MSRE

We employ the block-coordinate descent algorithm [20] to solve the optimization problem of multiple social role embedding. The implementation of the block-coordinate descent algorithm in MSRE is to solve the minimization problem with respect to multiple role embeddings by fixing role representatives. Afterwards, it is to perform the optimization with respect to role representatives while fixing multiple role embeddings. This process is iteratively executed until convergence. The pesudo-codes of the algorithm are presented in Algorithm 1.

Input : $G(V, E)$, D, λ, β, and negative ratio
Output: Multiple role embeddings

Perform global pre-training and role assignment;
while *(not converge)* **do**
    Fix $\boldsymbol{C}$, solve $\boldsymbol{V}$ with gradient descent;
    Fix $\boldsymbol{V}$, solve $\boldsymbol{C}$ with gradient descent;
return $V = \{\boldsymbol{v}_1^1, ..., \boldsymbol{v}_1^a, ..., \boldsymbol{v}_i^1, ..., \boldsymbol{v}_i^a, ...\}$

Algorithm 1: The optimization algorithm

For each variable, i.e., role-specific embeddings or role representatives, the problem can be solved by gradient-based algorithms, such as gradient descent and L-BFGS. The updating rules for gradient descent are as follows: The updating rules are as follows:

$$
(\boldsymbol{v}_i^a)^{p+1} = (\boldsymbol{v}_i^a)^p - d^{p+1} \times \frac{\partial L(\boldsymbol{C},\boldsymbol{V})}{\partial v_i^a}, \tag{11}
$$

and

$$
\boldsymbol{c}_a^{p+1} = \boldsymbol{c}_a^p - h^{p+1} \times \frac{\partial L(\boldsymbol{C},\boldsymbol{V})}{\partial \boldsymbol{c}_a}, \tag{12}
$$

where p denotes the p-th iteration, $d \in \mathbb{R}$ and $h \in \mathbb{R}$ are descent rates. For all the decent rates, we implement the backtracking line search [2] to learn an appropriate one for each of the iteration.

For the total loss on pairs of vertices without an edge, it may be inefficient to be summarized on total pairs of them when dealing with a large-scale network. To compromise, the number of pairs of vertices without an edge is sampled to several times larger than that of pairs with an edge, which is referred to as negative ratio. Since networks, especially large-scale networks, are usually spare, the negative ratio can significantly save the computational costs. The MSRE model works well with the negative ratio as demonstrated in the experiments.

Referring to the computation of derivatives of role embeddings and role representatives, the complexity of the MSRE algorithm is $O(|E|ADi)$, where D is the dimension of the latent space, A is the number of roles, and i is the number of iterations. Moreover, the convergence can be guaranteed based on the general proof of convergence for block-wise coordinate descent [23]. In our experiments, we observe that Algorithm 1 converges very fast in terms of the outer iterations, which is presented in the evaluation section.

IV. Empirical Evaluation

A. Experiment settings

We evaluate the latent features learned by the proposed MSRE model against state-of-the-art latent feature learning methods, DeepWalk [17], LINE [21], and node2vec [9]. For the implementation, we set the embedding dimension as 128, as used in all the baselines, negative ratio as 5, which is used in LINE, commonly used settings for backtracking line search, all the coefficients for ℓ_2 regularization terms as 1, and 0.001 as the relative loss difference to determine that the decent algorithm

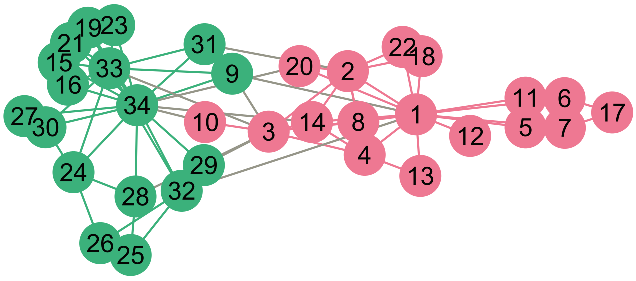

Fig. 2: Network layout of Karate Club friendship network.

converges. For the number of roles, we set it as the ground-truth number of roles for each network for simplification. For role representatives, we initialize them as centroids of clusters formed by node embeddings. Other methods such as cross-validation, or more sophisticated clustering methods [29] can be employed to learn an appropriate number of roles when the ground truth is unknown.

B. Case Study

This section presents a case study to show that learning multiple role-specific embeddings is meaningful in an intuitive way. Specifically, we visualize the network layout of Karate Club [33], a network of friendships between 34 members of a karate club. There are two social roles among the members because members in the left-hand side have few interactions with members in the right-hand side, and vice versa, which is illustrated in Fig. 2. We project the embeddings learned for these members into two-dimensional vectors by the t-SNE [24] tool, and then present the visualizations in Fig. 3. The network layout in the figure Global embedding preserves well the network structure. All the baselines show similar results, but they are omitted due to limited space. Additionally, two role representatives (RP) are plotted as "*" over the figures. Each role representative are surrounded by nodes having the role, which demonstrates that the role representatives are accurately estimated.

However, some information about the data points with cross-role edges may be lost. On the one hand, typical characteristics of roles may be lost as a result of locating at the middle area between the two role representatives, which means they have small role affinities to both role representatives, such as vertex 29 and 10. The location of vertex 29 is because it has balanced links to nodes with different roles. On the other hand, a role may be totally lost as a result of locating far away from the corresponding role representative, such as vertex 34. The location of vertex 34 results from imbalanced distribution of links to nodes with different roles.

These two problems can be solved by learning two role-specific embeddings, as illustrated in the right-hand sub-figure. Almost all of them have two distinct locations close to different role representatives, such as vertex 29 and 34. This demonstrates that each embedding preserves role-specific characteristics. Besides the more proper locations, the multiple social role embedding works well with the theory of social role taking [19] where normal people would adopt and act out a particular social role in every interaction with reference to the social environment. Specifically, when dealing with interactions, the MSRE model would take into account the role information and adopt an appropriate role-specific embedding. The effectiveness of this mechanism is demonstrated in following two practical applications, link prediction and multi-label classification.

C. Datasets

We introduce another four real-world social networks, which are Facebook social circles [13], DBLP co-authorship network [22], BlogCatalog friendship network [25], and Youtube friendship network [15]. Users in *Facebook social circles* have features like school and work type, which denotes that users have the roles of student and employee. Hence, these social networks are the real-world instances of the proposed multiple social network as illustrated in Fig. 1. We choose network 107 as the experiment network.

We construct a *DBLP co-authorship network* of researchers published papers in four research fields, Database, Data Mining, Machine Learning, and Information Retrieval during the 11-years period from 1999 to 2009. Each field corresponds to a social role. From each field, we select several popular conferences, such as SIGMOD, VLDB, and ICDE for Database, KDD, ICDM, SDM, and PAKDD for Data Mining, ICML, NIPS, AAAI, IJCAI and ECML for Machine Learning, SIGIR, WWW, ECIR, and WSDM for Information Retrieval. We filter out those authors without co-authorships with others.

For the *BlogCatalog friendship network*, we select users who post blogs under four popular categories, which are Art, Entertainment, Development and Growth, and Travel. Their friends that have not posted related blogs are filtered out. After this filtering, some users may have no friends. A second filtering is performed to filter those without friends.

For the *Youtube friendship network*, we select users that hold the group membership of any of five major groups, which are 23, 30, 81, 82 and 367 indicated in the the dataset. Similarly to the preprocessing on BlogCatalog dataset, two filters are performed to filter out users without related friends. Statistics of all these five social networks are presented in Table 1.

D. Link Prediction

The link prediction problem [12] refers to inferring new interactions between network vertices by using information of a certain snapshot of network. We deploy two scenarios of link prediction for evaluation of the proposed MSRE, which are future link prediction and missing link prediction. The future link prediction problem is to infer future interactions by employing the past network information while the missing link prediction problem is to infer missing interactions by employing existing ones. The DBLP co-authorship network is the only one that contains time information for the future link prediction scenario.

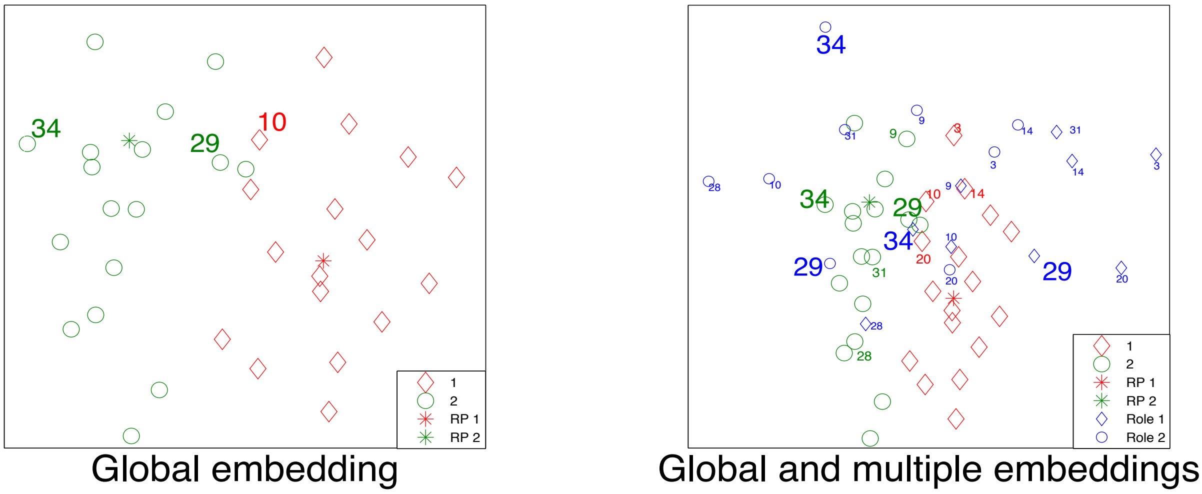

Fig. 3: Visualization of embeddings. In Global and multiple embeddings, two role embeddings of 9 vertices with cross-role edges are plotted, which are marked with vertex numbers. RP stands for the term role representative.

Dataset	Karate Club	Youtube	DBLP	BlogCatalog	Facebook
# Roles	2	5	4	4	9
# Vertices	34	8916	8230	6177	481
# Edges	78	33802	27659	36075	10066
Ave. degree	6.73	7.58	6.51	11.68	41.85

TABLE I: Statistics of five real-world social networks.

AUC	During 2010	2011-2012	2010-2012
DeepWalk	77.54(1.51)	75.62(1.32)	76.15(0.89)
LINE(1st)	65.04(1.94)	61.70(1.78)	62.00(1.26)
LINE(2nd)	79.85(1.19)	78.07(1.51)	77.42(1.10)
node2vec	78.66(0.88)	78.11(1.25)	77.39(1.19)
MSRE	**80.42(0.98)**	**78.75(1.21)**	**78.26(1.01)**

TABLE II: AUC score(standard error) on future link prediction.

	Vikas Sindhwani, Tao Li	Yannis Theodoridis, Salvatore Orlando
DeepWalk	0.58	0.49
LINE(1st)	0.43	0.38
node2vec	0.59	0.50
MSRE	**0.68**	**0.54**

TABLE III: Probabilities of positive co-authorships.

In future link prediction, we employ the DBLP co-authorship network presented in Table 1 as the past information, and infer future interactions between existing authors in three different time periods, which are during 2010, from 2011 to 2012, and from 2010 to 2012. Besides the ground-truth future interactions (positive) during these three periods, the same number of pairs of vertices without interactions (negative) are randomly generated for measuring the capability of detecting negative interactions. The inference of new interactions is made by computing the "similarity" between two vertices, which is the commonly used method to infer new interactions [12]. The "similarity" in embedding models is the closeness defined in Definition 1.

AUC score is used as the evaluation metric, and the performance of averaged scores of five repetitions of each task is presented in Table 2, where the numbers have been multiplied by 100% here and in the rest of the paper. It shows that the MSRE outperforms all the baselines. Other link prediction methods such as using Common Neighbors and Adamic/Adar to measure the similarities are omitted in the comparison because they significantly underperform node2vec as suggested in its paper [9].

For missing link prediction, we adopt five-fold cross validation as the evaluation method. Other experiment settings are the same as those in the future link prediction. Performance AUC score is presented in Table 4. Similarly, the proposed MSRE model outperforms all the baselines.

To explore the reason behind superior performance of the MSRE, we examine how it make predictions comparing with baselines for the link prediction task during 2010. We present two representatives of positive co-authorships in Table 3. In the first co-authorship, author Vikas Sindhwani at Google is an expert of Machine Learning publishing papers mostly in Machine Learning fields while the author Tao Li at Florida International University is an expert of Data Mining. Accordingly, the global embedding of the author Vikas Sindhwani would most likely be close to the role representative of Machine Learning and the author Tao Li close to the role representative of Data Mining. And the MSRE assigns the role affinity of Vikas Sindhwani to Machine Learning as 0.82, and that of Tao Li to Data Mining

AUC	Youtube	Facebook	DBLP	BlogCatalog
DeepWalk	80.23	86.01	94.76	71.63
LINE(1st)	87.68	85.85	93.82	83.56
LINE(2nd)	86.66	88.52	94.72	85.12
node2vec	85.28	87.14	94.01	85.98
MSRE	**91.05**	**90.32**	**94.82**	**87.70**

TABLE IV: Performance on missing link prediction.

Micro-F1	Youtube	Facebook	DBLP	BlogCatalog
DeepWalk	71.52	89.38	72.45	80.46
LINE(1st)	72.89	86.08	68.26	80.51
LINE(2nd)	77.42	87.23	65.46	79.58
node2vec	76.12	89.22	72.25	80.86
MSRE	**81.66**	**90.62**	**73.48**	**81.22**
Macro-F1	Youtube	Facebook	DBLP	BlogCatalog
DeepWalk	70.61	89.38	71.12	75.52
LINE(1st)	72.83	87.05	67.41	76.12
LINE(2nd)	77.09	88.02	64.51	73.62
node2vec	75.60	89.57	71.82	80.63
MSRE	**80.46**	**90.90**	**73.42**	**77.02**

TABLE V: Performance on multi-label classification.

0.79. Hence, it is not likely for them to coauthor a paper, which is the prediction of baselines. The probability given by LINE(2nd) is omitted since it is not that interpretable given that it produces probabilities about 99% for any test links, including negative links. In fact, they did coauthor a paper in SDM 2010. As we examine the published papers of both authors, we find that the author Vikas Sindhwani did published papers in Data Mining fields, but only a few compared with papers in Machine Learning fields. Similarly, the author Tao Li had published few papers in Machine Learning fields. Hence, they both actually have double distinct social roles. The MSRE model can capture this knowledge, and thus gives a larger estimate on the future co-authorship. The case with the second co-authorship is similar and the co-authorship occurred in ICDE 2010.

E. Multi-label Classification

In multi-label classification, multiple labels are assigned to each instance. In this setting, communities or groups are the labels for each vertex. The features by the MSRE model is the direct concatenation of all the role embeddings plus the role probabilities. The motivation behind this concatenation is that each role embedding acts as a hyper-feature of the vertex just like each social role is an attribute of a person. The binary-relevance based SVM implemented in Meka [18] is used as the classification tool. Performance of five-fold cross validation is presented in Table 5. Here we make similar observations that the MSRE outperforms all the baselines. The reason behind the superior performance can be explained by referring to the case study, i.e., multiple role characteristics can be well preserved in the MSRE while certain roles may be lost by only learning a global embedding in all the baselines . To this point, it is concluded that the concept of multiple social roles originating from daily life is applicable to broadly multiple communities or groups that are observed in DBLP, Blog and Youtube networks.

F. Convergence Analysis

This section studies the convergence of the proposed alternating algorithm as indicated in Algorithm 1. Specifically, we study the performance of the algorithm on applications with respect to the number of outer iterations. We only present the performance on the link prediction task for DBLP future link prediction and Youtube missing link prediction in Fig. 4 because other experiments show similar results. It shows that the algorithm converges very fast and can usually converge to stable performance after about 10 iterations.

G. Parameter Sensitivity

The performance of the MSRE w.r.t different number of roles in link prediction for the Youtube network is presented in Fig. 5, and similar results are observed on the other datasets omitted due to space limitation. It shows the performance of the MSRE improves as the number of roles gets larger when it is smaller than the ground-truth number. This is expected because the number determines how well distinct social roles are preserved. If the number is considerably smaller than the ground truth, such as 1, a large portion of social roles have to be mixed up in a single embedding. Accordingly, each interaction would involve multiple social roles, which violates the theory of social role taking. The performance is maximized when the number of roles is around the ground truth, but it drops when the number is larger than the ground truth. Accordingly, it is not always the case that more roles bring better performance, especially when the number is larger than the ground truth. As a result, an appropriate number should be adopted, which can be learned by cross validation or advanced clustering models that can automatically learn one, such as [29].

Besides, the performance with respect to the dimension of embeddings is studied in Fig. 5. It is worthy noting the actual dimension is (# dimensions + 1) $\times$ # roles in the MSRE model as a result of concatenation of all role embeddings. The figure

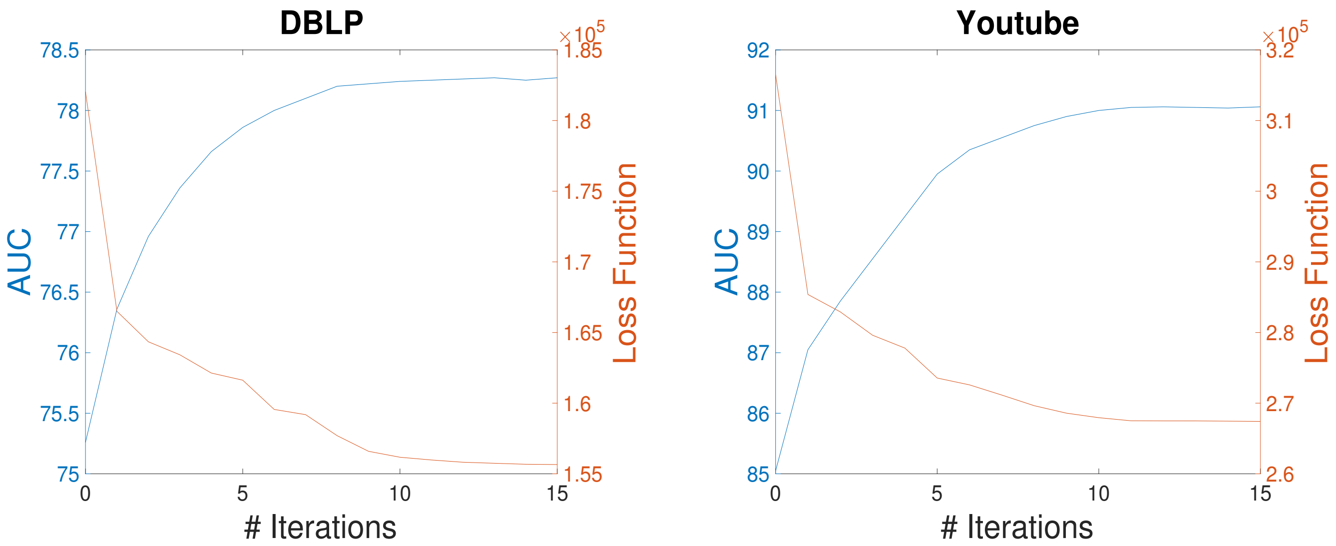

Fig. 4: Convergence analysis

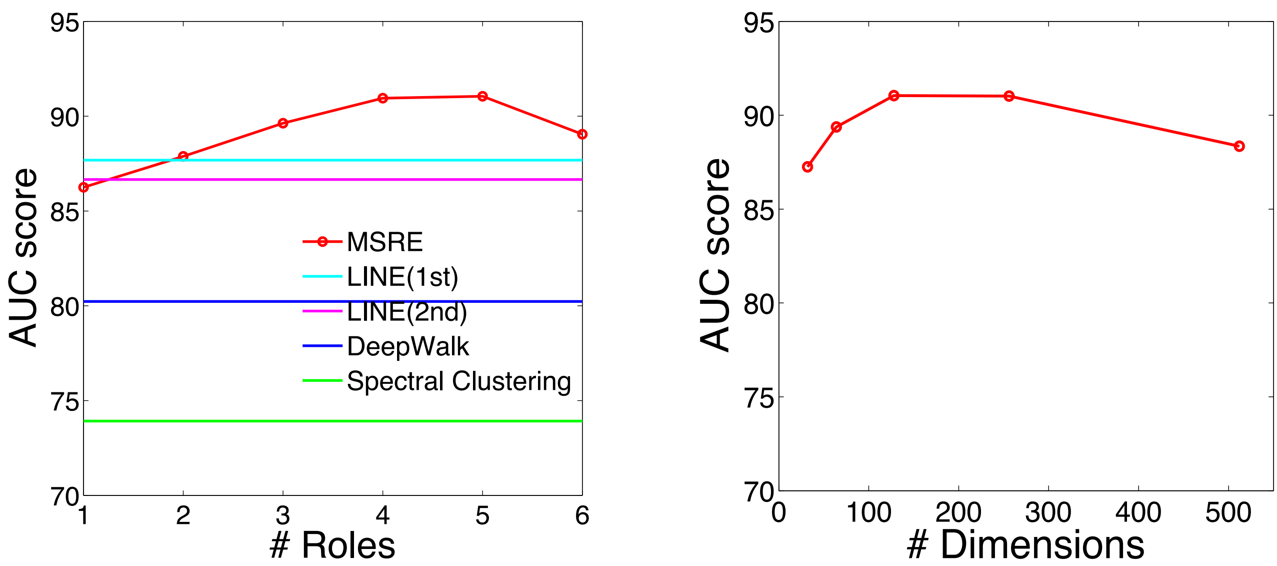

Fig. 5: Parameter sensitivity

suggests the embedding dimension should not be too small (e.g., 32) and too large (e.g., 512).

V. Conclusion and Future work

In this paper, we propose the multiple social role embedding (MSRE) to preserve the network structure and social roles simultaneously. Multiple role embeddings are useful because the unique characteristics of each role can be preserved. Each unique role embedding, in turn, is responsible for a particular type of social interaction, which is suggested by the theory of social role taking. Multiple role embeddings are demonstrated as more effective features in link prediction and multi-label classification. We plan to enable MSRE to perform online learning to handle evolving networks in the future.

Acknowledgment

The work described in this paper was partially supported by the funding for Project of Strategic Importance provided by The Hong Kong Polytechnic University (Project Code : 1-ZE26), the funding for demonstration project on large data provided by The Hong Kong Polytechnic University (project account code : 9A5V), and NSF through grants IIS-1526499, and CNS-1626432, and NSFC 61672313.

References

[1] Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9(Sep):1981–2014, 2008.

[2] Larry Armijo. Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of mathematics*, 16(1):1–3, 1966.

[3] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, pages 519–528. ACM, 2012.

[4] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*, volume 14, pages 585–591, 2001.

[5] Trevor F Cox and Michael AA Cox. *Multidimensional scaling*. CRC press, 2000.

[6] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3):75–174, 2010.

[7] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[8] Mehmet Gönen and Ethem Alpaydin. Localized multiple kernel learning. In *Proceedings of the 25th international conference on Machine learning*, pages 352–359. ACM, 2008.

[9] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[10] Mohsen Jamali and Martin Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 135–142. ACM, 2010.

[11] David W Johnson. Cooperativeness and social perspective taking. *Journal of Personality and Social Psychology*, 31(2):241, 1975.

[12] 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.

[13] Julian J McAuley and Jure Leskovec. Learning to discover social circles in ego networks. In *NIPS*, volume 2012, pages 548–56, 2012.

[14] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[15] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and Analysis of Online Social Networks. In *Proceedings of the 5th ACM/Usenix Internet Measurement Conference (IMC'07)*, San Diego, CA, October 2007.

[16] Davide Nunes and Luis Antunes. Achieving consensus with segregation in multiple social contexts. In *Advances in Computational Social Science*, pages 85–103. Springer, 2014.

[17] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710. ACM, 2014.

[18] Jesse Read, Peter Reutemann, Bernhard Pfahringer, and Geoff Holmes. MEKA: A multi-label/multi-target extension to Weka. *Journal of Machine Learning Research*, 17(21):1–5, 2016.

[19] David Shaffer. *Social and personality development*. Nelson Education, 2008.

[20] David Sontag, Amir Globerson, and Tommi Jaakkola. Introduction to dual decomposition for inference. *Optimization for Machine Learning*, 1:219–254, 2011.

[21] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1067–1077. International World Wide Web Conferences Steering Committee, 2015.

[22] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 990–998. ACM, 2008.

[23] Paul Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of optimization theory and applications*, 109(3):475–494, 2001.

[24] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.

[25] Xufei Wang, Lei Tang, Huiji Gao, and Huan Liu. Discovering overlapping groups in social media. In *the 10th IEEE International Conference on Data Mining series (ICDM2010)*, Sydney, Australia, December 14 - 17 2010.

[26] Xiaokai Wei, Bokai Cao, and Philip S. Yu. Unsupervised feature selection on networks: A generative view. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.*, pages 2215–2221, 2016.

[27] Xiaokai Wei, Linchuan Xu, Bokai Cao, and Philip S. Yu. Cross view link prediction by learning noise-resilient representation consensus. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 1611–1619, 2017.

[28] Eric P Xing, Wenjie Fu, Le Song, et al. A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, 4(2):535–566, 2010.

[29] Lei Xu, Adam Krzyżak, and Erkki Oja. Rival penalized competitive learning for clustering analysis, rbf net, and curve detection. *Neural Networks, IEEE Transactions on*, 4(4):636–649, 1993.

[30] Linchuan Xu, Xiaokai Wei, Jiannong Cao, and Philip S. Yu. Embedding identity and interest for social networks. In *Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, April 3-7, 2017*, pages 859–860, 2017.

[31] Linchuan Xu, Xiaokai Wei, Jiannong Cao, and Philip S. Yu. Embedding of embedding (EOE): joint embedding for coupled heterogeneous networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM 2017, Cambridge, United Kingdom, February 6-10, 2017*, pages 741–749, 2017.

[32] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y Chang. Network representation learning with rich text information. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina*, pages 2111–2117, 2015.

[33] Wayne W Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, pages 452–473, 1977.