2014 IEEE International Conference on Data Mining

Social Role Identification via Dual Uncertainty Minimization Regularization

Yu Cheng, Ankit Agrawal, Alok Choudhary
EECS Department
Northwestern University

Huan Liu
Computer Science and Engineering
Arizona State University

Tao Zhang
Department of Automation
Tsinghua University

***Abstract*—In this paper, we study a challenging problem of inferring individuals' role and statuses in a professional social network, which is of central importance in workforce optimization and human capital management. Realizing the natural setting of social nodes associated with dual view information, i.e., the local node characteristics and the global network influence, we present a novel model that explores graph regularization techniques and integrates such information to achieve improved prediction performance. In particular, our prediction model is built upon the graph transductive learning framework that encodes an uncertainty regularization term in the conventional empirical risk minimization principle. Through taking advantage of the information from both the local profile and the global network characteristics, the final inference of the role or statues achieves minimum an empirical loss on the labeled set, as well as a minimum uncertainty on the unlabeled social nodes. We perform extensive empirical study using real-wold data and compare with representative peer approaches. The experimental results on three real social network data sets show that the proposed model greatly outperforms a number of baseline models and is able to effectively infer in a wide range of scenarios.**

***Keywords*-Social Role Identification; Graph Regularization; Dual Uncertainty Minimization**

I. Introduction

Being hinged with different social roles and statuses, people tend to behave fairly differently in social networks. For example, users in some social blogospheres can behave as content contributors, information sharers, or information receivers, while in some professional social media like Linkedin, users experience the professional network in various roles, such as engineer, sales-person, recruiter etc [1]. Finally, in a collaborated workforce network (e.g, a enterprise social network), the users can be largely categoried based on their job functions and job titles. Note that studying and understanding social roles in a network can gain actionable insights for organizations and companies since such learned knowledge can be used to perform optimal resource allocation and human resource management. In addition, understanding social roles also help improve the user experiences since the identified social roles can help to build effective social network applications in recommendation systems, question answering systems, and advertising systems [2], [3].

The problem of identifying social role in social networks has attracted increasing amount of attention and many techniques are presented in recent years [1], [4]. Some previous works studied this social role prediction problem in a specific network or with some strong assumptions of the data. For example, researchers have developed methods for identifying social roles in Wikipedia [5] and email networks [6], [7]. Recently some work has been proposed on social network inference problems in different contexts. Henderson *et al.* [4] proposed a role discovery framework on networks. It is unsupervised and essentially a clustering approach (using matrix factorization). A semi-supervised semantic role labeling method is proposed in [8]. In a very recent study of identifying social roles in online social networks [1], the authors explored five social principles and concepts that represent a variety of network characteristics and quantify their relations with social roles and statuses. However, it remains as an open challenge to maximize the usage of all available information such as social profiles as local node characteristics and the network influence as global information. Especially in the scenario of online social networks, data are collected from diverse domains or obtained from various feature extractors and exhibit heterogeneous properties. Hence, it motivates us to design a principled way to leverage such types of information to carry out the prediction of social role and status.

Motivated by multi-view learning idea [9], [10], here we propose a graph regularization based learning framework that integrates heterogeneous information. The goal is to jointly optimizes all the functions to exploit views of node characteristics as well as graph influence, and improve the prediction performance. In particular, through encoding the prediction uncertainty into the conventional empirical risk minimization principle, we propose an optimal prediction with minimum empirical loss, as well as minimum uncertainty given the observed node characteristics and network information. Different from existing co-regularization method [11], [12], or co-training framework [13], which solely enforces prediction agreement across different views, the uncertainty regularization encourages to archive predictions with high confidence while exploiting the complementariness from dual-view data; thus is more feasible and robust for realistic scenario, where either view information might be problematic due to some reason. The extensive empirical studies using three real social network data sets clearly show that the proposed model consistently

1550-4786/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDM.2014.31

outperforms a number of competing models with significant performance margins for a wide range of social role prediction scenarios. Our proposed dual uncertainty minimization framework can be easily generalized to analyze other applications with the consideration of heterogenous factors from multiple view information.

II. Related Work

A. Social Roles Mining

In real social network, due to the missing, outdated and non-standard data issues, the task of social roles mining is quite challenging. In summary, most of the existing methods can be categorized in the following three groups: 1) rules based methods to identify roles from textual information; 2) unsupervised learning methods such as clustering social nodes with matrix factorization techniques; and 3) semi-supervised learning methods using node properties or link information. For example, editors in Wikipedia have been studied in [5]. Email users of Palins email network have been analyzed in [7]. In [14], the authors study user attribute inference in university social networks by applying community detection. Recently some work has been proposed on social network inference problems in different contexts. Henderson *et al.* [4] proposed a role discovery framework on networks. Zhao *et al.* [1] explored five social principles and concepts that represent a variety of network characteristics and quantify their relations with social roles and statuses. We extract the same categories of the features used in [1] as the node features in our experiments.

B. Graph Regularization

Single Graph Regularization: In graph based transductive learning setting, we have *iid* (independent and identically distributed) labeled samples $\mathbf{X}^L = \{\mathbf{x}_i, y_i\}_{i=1}^{l}$ and unlabeled samples $\mathbf{X}^U = \{\mathbf{x}_i\}_{i=l+1}^{l+u}$ ($l+u=N$). The objective is to infer the missing labels $\mathbf{Y}^U = \{\hat{y}_i\}_{i=l+1}^{l+u}$ corresponding to the unlabeled data $\mathbf{X}^U$. We can construct a similarity graph $\mathcal{G}$ from $\mathbf{X} = \{\mathbf{X}^L, \mathbf{X}^U\}$, where the vertices represent data samples and the weighted edges $\mathbf{W} = \{w_{i,j}\}_{i,j=1}^{n}$ denote similarities between samples. Let $f(\mathbf{x}_i)$ be the classifier, a classical regularization framework solves the following minimization problem:

$$\begin{aligned} f^* &= \arg\min_f \mathcal{T}(f;\mathcal{G}) \\ &= \arg\min_f \sum_{i=1}^{l} \mathcal{V}(f(\mathbf{x}_i), y_i)) + \lambda \mathcal{L}_{\mathcal{G}}(f) \end{aligned} \quad (1)$$

where $\mathcal{V}$ is some loss function, λ is parameter to balance the empirical fitness and function smoothness, $\mathcal{L}_{\mathcal{G}}(f)$ is the single-view graph regularizer. Denoting the degree matrix $D = \mathrm{diag}(\mathrm{d}_1, \mathrm{d}_2, ..., \mathrm{d}_\mathrm{n}), \mathrm{d}_\mathrm{i} = \sum_\mathrm{j} \mathrm{w}_\mathrm{ij}$ and the unnormalized graph Laplacian [15] by $\mathbf{L} = \mathbf{D} - \mathbf{W}$, a single-view graph regularizer $\mathcal{L}_{\mathcal{G}}(f)$ over the function $\mathbf{f}$ can be written as:

$$\begin{aligned} \mathcal{L}_{\mathcal{G}}(f) &= \sum_{i,j} w_{ij} ||f(\mathbf{x}_i) - f(\mathbf{x}_j)||^2 \\ &= \mathrm{trace}(\mathbf{f}^\mathrm{T}(\mathbf{D}-\mathbf{W})\mathbf{f}) = \mathrm{trace}(\mathbf{f}^\mathrm{T}\mathbf{L}\mathbf{f}) \end{aligned} \quad (2)$$

where $\mathbf{f} = [f(\mathbf{x}_1), f(\mathbf{x}_2), ..., f(\mathbf{x}_n)]$ is a column vector. Based on the aforementioned graph Laplacian based regularization framework, various semi-supervised learning algorithms are proposed, including the local and global consistency formulation [16], Gaussian fields and Harmonic function formulation [17], transductive support vector machines [18], and the manifold regularization formulation [19].

Graph Co-Regularization: In the two-view semi-supervised learning setting, each sample $\mathbf{x} = (\mathbf{x}^a, \mathbf{x}^b)$ is seen in two views with $\mathbf{x}^a \in \mathbf{X}^a$ and $\mathbf{x}^b \in \mathbf{X}^b$. The setup and the algorithms can also be generalized to more than two views. For simplicity, we assume the case with two graphs $\mathbf{G}^a = \{\mathbf{W}^a, \mathbf{D}^a, \mathbf{L}^a\}$ and $\mathbf{G}^b = \{\mathbf{W}^b, \mathbf{D}^b, \mathbf{L}^b\}$. The goal is to learn the prediction function pair $f = (f^a, f^b)$ associated with $\mathbf{G}^a$ and $\mathbf{G}^b$, in a joint form. Considering function smoothness over single graphs and the agreement between two views, the graph co-regularization is formed as:

$$\begin{aligned} \mathcal{L}_{\mathcal{G}^a,\mathcal{G}^b}(f^a, f^b) &= \lambda_a ||\mathbf{f}||^2_{\mathcal{G}^a} + \lambda_b ||\mathbf{f}||^2_{\mathcal{G}^b} + \gamma \mathcal{S}(\mathbf{f}^a, \mathbf{f}^b) \\ &= \lambda_a ||\mathbf{f}||^2_{\mathcal{G}^a} + \lambda_b ||\mathbf{f}||^2_{\mathcal{G}^b} + \gamma ||\mathbf{f}^a - \mathbf{f}^b||^2 \\ &= \mathrm{trace}(\lambda_\mathrm{a}(\mathbf{f}^\mathrm{a})^\mathrm{T}\mathbf{L}^\mathrm{a}\mathbf{f}^\mathrm{a} + \lambda_\mathrm{b}(\mathbf{f}^\mathrm{b})^\mathrm{T}\mathbf{L}^\mathrm{b}\mathbf{f}^\mathrm{b} \\ &+ \gamma(\mathbf{f}^a - \mathbf{f}^b)^\mathrm{T}(\mathbf{f}^a - \mathbf{f}^b)) \end{aligned} \quad (3)$$

where $||\mathbf{f}||^2_{\mathcal{G}^a}$ and $||\mathbf{f}||^2_{\mathcal{G}^b}$ are regularization terms over individual views, and $\mathcal{S}(\mathbf{f}^a, \mathbf{f}^b)$ is the term interpreted as a requirement of smoothness over the graph for the pair $f = (f^a, f^b)$. The core part of graph co-regularization is to utilize the agreement among learners trained on different representations of the same problem to improve the overall performance. However, the "consensus" assumption of multiple view could be questionable since each data graph might contribute different and complementary prediction power. And for the extreme case, the view agreement will significantly weaken the prediction performance when either of the view data is not reliable.

Combined Graph Laplacians: Except for Graph Co-Regularization, another way to construct a multi-view regularizer is Combined Graph Laplacians [20], [21], which applies to convex combinations of graph kernels parameterized by a compact set. Given two view representations $\mathbf{G}^a$ and $\mathbf{G}^b$, the convex combination of the graph Laplacians is defined as $K(\lambda) = \lambda\mathbf{L}^a + (1-\lambda)\mathbf{L}^b$ ($\lambda > 0$). Then the learning problem is formulated as:

$$(f^*, K(\lambda)^*) = \arg\min_{f,\lambda} \sum_{i=1}^{l} \mathcal{V}(f(\mathbf{x}_i), y_i)) + \gamma ||\mathbf{f}||_{K(\lambda)} \quad (4)$$

which is a joint minimization problem over f and $K(\lambda)$, and can be solved via an alternative optimizer iteratively proposed in [22].

III. The Proposed Methods

A. Preliminaries

We first introduce some notations and definitions that we will use throughout the rest of the paper. Assume we have a partially labeled social network $G = (\mathbf{V}^L, \mathbf{V}^U, \mathbf{E}, \mathbf{X}^a, \mathbf{X}^b)$, where $\mathbf{V}^L$ is the set of labeled users with social roles/statuses and $\mathbf{V}^U$ is the set of unlabeled users in the social network. We note that the set of all users in the network $\mathbf{V} = \{v_i\}_{i=1}^N = \mathbf{V}^L \cup \mathbf{V}^U$ and $\mathbf{V}^L \cap \mathbf{V}^U = \emptyset$. $\mathbf{E}$ represents the set of all edges in the network. $\mathbf{X}^a$ and $\mathbf{X}^b$ are are two-view data representations of V, where $\mathbf{X}^a$ is node view from users and $\mathbf{X}^b$ is the link view, separately. Let y_{v_i} be the label for user v_i, $\mathbf{x}_i^a$ be a vector of node attributes and $\mathbf{x}_i^b$ be the link view representation. Suppose we have the set of labels to be $C = \{1, ..., c\}$, which contains C different roles. The task is to infer labels of users with unknown social roles: $y_{v_i} \in C$ where $v_i \in \mathbf{V}^U$.

We formulate this problem as the graph-based transductive learning problem. That is, given the labeled nodes $\mathbf{V}^L$ and unlabeled nodes $\mathbf{V}^U$ with dual view representation $\mathbf{X}^a$ and $\mathbf{X}^b$, we aim to learn the pair function $f = (f^a, f^b)$ with a graph regularization framework, then the label of $\mathbf{V}^U$ can be predicted by combining $\mathbf{f}^a$ and $\mathbf{f}^b$.

Dual View Representation: For each user v_i, the representation from the node view $\mathbf{x}_i^a$ is pretty straightforward: either a BOW (bag-of-words) of profile information, or a combination of social factors [1], such as Local Clustering Coefficient, Degree Centrality, Average Neighbor Degree , Embeddedness and Number of Communities. There are various of methods for the representation of $\mathbf{x}_i^b$, such as direct link, demonstrating friend/follower/following information in social network. $\mathbf{x}_i^b$ is represented as N-dimension vector, with the j-th element $x_{i,j}^b$ denoting link connectivity of user v_j. Then we use a very natural way called neighbor node overlapping to measure link connectivity by Jaccard similarity [23].

B. Dual View Uncertainty Regularizer

In the co-regularization framework, the term $\mathcal{S}(\mathbf{f}^a, \mathbf{f}^b)$ measures of smoothness to enforce the right complexity for the dual view prediction. In this work, instead of the smoothness term in co-regularization, we propose a new term called uncertainty regularizer, which measures the uncertainty and complementariness of the prediction from different views. The objective is to enforce minimum uncertainty for the prediction from dual views instead of purely enforcing agreement between different views.

Given the two view data $\mathbf{X}^a = \{\mathbf{x}_i^a\}_{i=1}^{u+l}$ and $\mathbf{X}^b = \{\mathbf{x}_i^b\}_{i=1}^{u+l}$, the predictions over the two views are $f^a(\mathbf{x}_i^a)$ and $f^b(\mathbf{x}_i^b)$. For the predictions over a data point $\mathbf{x}_i$, let $p_{ik} = P(f^a(\mathbf{x}_i^a) = k)$ be the probability of the prediction of $\mathbf{x}_i$ of view A, and $q_{ik} = P(f^b(\mathbf{x}_i^b) = k)$ be the probability of the prediction of $\mathbf{x}_i$ of view B, where $k \in C = \{1, ..., c\}$ is the class label. The uncertainty of the prediction measured by entropy is computed as:

$$H(p_i) = -\sum_{k=1}^{C} p_{ik} \log p_{ik}, \qquad H(q_i) = -\sum_{k=1}^{C} q_{ik} \log q_{ik} \quad (5)$$

where $p_i = \sum_{k=1}^{C} p_{ik}$ and $q_i = \sum_{k=1}^{C} q_{ik}$. And the uncertainty over all the data points is expressed by:

$$\mathcal{U}(\mathbf{X}^a, \mathbf{X}^b) = \sum_{i=1}^{N} H(p_i) + H(q_i) \quad (6)$$

Note that the entropy from each single view presents different confidence for individual data points. And Eq. 6 measures the prediction uncertainty over the two views.

To embed the uncertainty measure into the graph regularization, let us rewrite the uncertainty formulation with f^a and f^b as:

$$\begin{aligned} \mathcal{U}(f^a, f^b) &= -\sum_{i=1}^{N} H(f^a(\mathbf{x}_i^a)) + H(f^b(\mathbf{x}_i^b)) \\ &= \sum_{i=1}^{N} f^a(\mathbf{x}_i^a)^{\mathrm{T}} \log(f^a(\mathbf{x}_i^a)) + f^b(\mathbf{x}_i^b)^{\mathrm{T}} \log(f^b(\mathbf{x}_i^b)) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{C} f_{ik}^a \log(f_{ik}^a) + f_{ik}^b \log(f_{ik}^b) \end{aligned} \quad (7)$$

where $f_{ik}^a = f^a(\mathbf{x}_i^a)|_{i=k}$ and $f_{ik}^b = f^b(\mathbf{x}_i^b)|_{i=k}$ represent the prediction values of $\mathbf{x}_i^a$ and $\mathbf{x}_i^b$, separately. Apparently, the uncertainty regularizer is monotonically related to the uncertainty value in Eq. 5. In general, we tend to derive the prediction with less uncertainty from at least one of the views and want to minimize Eq. 7.

Final Objective Combining the objective function $\mathcal{T}(f^a; \mathcal{G}_a)$ and $\mathcal{T}(f^b; \mathcal{G}_b)$ from each individual view (as defined in co-regularization), and the dual-view uncertainty regularization proposed, we can suggests the following objective function:

$$\begin{aligned} (f^{(a)*}, f^{(b)*}) &= \arg\min_{f^a, f^b} \mathcal{T}(f^a; \mathcal{G}_a) + \mathcal{T}(f^b; \mathcal{G}_b) + \gamma \mathcal{U}(f^a, f^b) \\ &= \sum_{i=1}^{l} \mathcal{V}(f^a(\mathbf{x}_i^a), y_i)) + \lambda_a \mathcal{L}_{\mathcal{G}_a}(f^a) \\ &+ \sum_{i=1}^{l} \mathcal{V}(f^b(\mathbf{x}_i^b), y_i)) + \lambda_b \mathcal{L}_{\mathcal{G}_b}(f^b) \\ &+ \gamma \mathcal{U}(f^a, f^b) \end{aligned} \quad (8)$$

where the empirical loss function over labeled data is

formulated as a squared loss form:

$$\begin{aligned}\mathcal{V}(f^a(\mathbf{x}_i^a), y_i)) &= \sum_{i=1}^{l} ||f^a(\mathbf{x}_i^a) - y_i|| \\ &= \mathrm{trace}((\mathbf{f}^a - \mathbf{y})^{\mathrm{T}}(\mathbf{f}^a - \mathbf{y})) \quad (9)\end{aligned}$$

where $\mathbf{y}$ is the label indicator over the l samples. And the graph regularizer $\mathcal{L}_{\mathcal{G}}(f^a)$ is defined as:

$$\mathcal{L}_{\mathcal{G}}(f^a) = \mathrm{trace}(\mathbf{f}^{\mathrm{a}\mathrm{T}}(\mathbf{D}^{\mathrm{a}} - \mathbf{W}^{\mathrm{a}})\mathbf{f}^{\mathrm{a}}) = \mathrm{trace}(\mathbf{f}^{\mathrm{a}\mathrm{T}}\mathbf{L}^{\mathrm{a}}\mathbf{f}^{\mathrm{a}}) \quad (10)$$

Then the pair prediction function $(f^{(a)*}, f^{(b)*})$ can be obtained by minimizing the objective function.

C. Optimization and Solution

We can derive the optimal solution for Eq. 8. Let's denote the objective function as $\mathcal{O}$ and take the partial derivatives over f^a and f^b:

$$\begin{aligned}\frac{\partial \mathcal{O}}{\partial \mathbf{f}^a} &= \mathbf{f}^a - \mathbf{y} + \lambda_a \mathbf{L}^a \mathbf{f}^a + \gamma \frac{\partial \mathcal{U}(\mathbf{f}^a, \mathbf{f}^b)}{\partial \mathbf{f}^a} \\ &= \mathbf{F}^a + \gamma \frac{\partial \mathcal{U}(\mathbf{f}^a, \mathbf{f}^b)}{\partial \mathbf{f}^a} \quad (11)\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathcal{O}}{\partial \mathbf{f}^b} &= \mathbf{f}^b - \mathbf{y} + \lambda_b \mathbf{L}^b \mathbf{f}^b + \mu \frac{\partial \mathcal{U}(\mathbf{f}^a, \mathbf{f}^b)}{\partial \mathbf{f}^b} \\ &= \mathbf{F}^b + \mu \frac{\partial \mathcal{U}(\mathbf{f}^a, \mathbf{f}^b)}{\partial \mathbf{f}^b} \quad (12)\end{aligned}$$

where for both equations, the first parts are expressed with matrix format as: $\mathbf{F}^a = \mathbf{f}^a - \mathbf{y} + \lambda_a \mathbf{L}^a \mathbf{f}^a$ and $\mathbf{F}^b = \mathbf{f}^b - \mathbf{y} + \lambda_b \mathbf{L}^b \mathbf{f}^b$, separately. However, the second part, need to be computed at the elements level:

$$\frac{\partial \mathcal{U}(f^a, f^b)}{\partial f_{ik}^a} = 1 + \log f_{ik}^a, \qquad \frac{\partial \mathcal{U}(f^a, f^b)}{\partial f_{ik}^b} = 1 + \log f_{ik}^b$$

Hence, given the computed partial derivatives, the update of $\mathbf{f}^a$ and $\mathbf{f}^b$ can be iteratively computed using the gradient descent as

$$\mathbf{f}^a|^{t+1} = \mathbf{f}^a|^t - \alpha(\frac{\partial \mathcal{O}}{\partial \mathbf{f}^a}), \qquad \mathbf{f}^b|^{t+1} = \mathbf{f}^b|^t - \beta(\frac{\partial \mathcal{O}}{\partial \mathbf{f}^b}) \quad (13)$$

where α and β are the learning step lengths. In particular, at single element level, f_{ik}^a and f_{ik}^b can be updated iteratively using the gradient descent

$$\begin{aligned}f_{ik}^a|^{t+1} &= f_{ik}^a|^t - \alpha F_{ik}^a|^t - \alpha\gamma(1 + \log f_{ik}^a) \\ f_{ik}^b|^{t+1} &= f_{ik}^b|^t - \beta F_{ik}^b|^t - \beta\gamma(1 + \log f_{ik}^b) \quad (14)\end{aligned}$$

where $f_{ik}^a|^t$ is the updated values of f_{ik}^a at time stamp t, and F_{ik}^a and F_{ik}^b are the i-th row and k-th column element of $\mathbf{F}^a$ and $\mathbf{F}^b$. The iterative updates can be terminated if there change for consecutive calculations are small enough for the two vectors $||\mathbf{f}^a|^{t+1} - \mathbf{f}^a|^t|| \leq \epsilon$ and $||\mathbf{f}^b|^{t+1} - \mathbf{f}^b|^t|| \leq \epsilon$ with a threshold constant ϵ. Finally, after both $\mathbf{f}^a$ and $\mathbf{f}^b$ achieve the convergence, the final prediction can be simply obtained by $\hat{\mathbf{Y}} = \frac{1}{2}(\mathbf{f}^a + \mathbf{f}^b)$, ($\hat{\mathbf{Y}}$ is the labels of all samples), similar to that used in co-regularization.

IV. Evaluation

A. Datasets

The first dataset we used is the email data of IBM employees, which was collected by IBM's internal system to analyze IBM workforce situation. We take a subset of the data starting from Nov 2011. After pre-processed, this email dataset is used to build a communication network with 3,547 nodes representing the authors of the emails, and 15,780 distinct weighted edges indicating the email communication between the nodes. Each node represents an employee associated with profile information and email content. The social roles of those social nodes in the email communication network consist of *department head*, *manager*, *secretary*, researcher, and *programmer*. The second dataset is from the professional social network Linkedin. We specifically identify a subsect of $25,265$ users in the finance industry. In total, there are about 2.5 millions connections and the average node degree is 100.8. Five social roles are identified as: *Finance*, *Sales*, *IT*, *Support and Operation*, which are corresponding to their job functions and titles. The third dataset used in our experiments is the IMDB dataset. We take a subset with the movies taken from the year 2001 to 2005. The three major social roles are identified as *actor/actress*, *director*, and *producer*. These social roles are treated as network nodes that are connected if they collaborate in a same movie. Finally, we construct a network with a total of $20,402$ nodes connected by $4,903,605$ links.

B. Experimental Settings

Baselines: In order to demonstrate the effectiveness of the proposed model, we compare the proposed DUMR method against a number of baselines. Since our method utilizes both node features and link information, intuitive baseline are those methods using a single view information, i.e., Homophily A with only the link view and Homophily B with only node view. For both single view based methods, we apply the popular graph regularization framework proposed in [18]. In addition, we compared with Social Roles and Statuses Inference Model (SRS), the method proposed in [1]. SRS first forms a factorized probabilistic model to learn a joint distribution over node and edge functions. Then the final prediction is achieved though an iterative algorithm that maximizes the marginal probabilities. Finally, we also compare with the co-regularization (Co-Reg) learning method in [11] and a combined regularization framework (Com-Reg), as described in [21]. For fair comparison, we use the same network with the uniform settings like the link weighting scheme [24], [25].

C. Results

In our experiments, we split the data into two subsets, a labeled set and an unlabeled set for testing. Specifically, we vary the fraction of the labeled subset from 30% to 80% to evaluate the prediction performance under different

Table I: The performance comparison on the IBM Email dataset.

Method	Homophily A	Homophily B	SRS	Co-Reg	Com-Reg	DUMR
80%	0.5924(±0.016)	0.6072(±0.016)	0.6945(±0.017)	0.6512(±0.017)	0.6656(±0.015)	**0.7556**(±0.016)
70%	0.5565(±0.019)	0.5737(±0.018)	0.6521(±0.020)	0.6085(±0.019)	0.6301(±0.018)	**0.7322**(±0.018)
60%	0.4828(±0.022)	0.5241(±0.021)	0.6202(±0.025)	0.5968(±0.023)	0.6168(±0.021)	**0.7168**(±0.021)
50%	0.4501(±0.029)	0.4832(±0.028)	0.5998(±0.031)	0.5404(±0.029)	0.5802(±0.026)	**0.6705**(±0.027)
40%	0.3803(±0.037)	0.4422(±0.036)	0.5313(±0.039)	0.4703(±0.036)	0.5187(±0.032)	**0.6032**(±0.035)
30%	0.3328(±0.044)	0.3855(±0.043)	0.4752(±0.048)	0.4419(±0.045)	0.4632(±0.041)	**0.5481**(±0.043)

Table II: The performance comparison on the LinkedIn dataset.

Method	Homophily A	Homophily B	SRS	Co-Reg	Com-Reg	DUMR
80%	0.6524(±0.022)	0.6672(±0.019)	0.7645(±0.020)	0.7512(±0.012)	0.7156(±0.017)	**0.8056**(±0.015)
70%	0.5965(±0.028)	0.6034(±0.024)	0.7221(±0.024)	0.7085(±0.017)	0.6601(±0.021)	**0.7501**(±0.018)
60%	0.5428(±0.033)	0.5524(±0.027)	0.6803(±0.026)	0.6649(±0.02)	0.6275(±0.025)	**0.7368**(±0.021)
50%	0.5001(±0.038)	0.5232(±0.033)	0.6598(±0.032)	0.6404(±0.025)	0.5902(±0.034)	**0.7005**(±0.026)
40%	0.4703(±0.046)	0.4922(±0.037)	0.5913(±0.040)	0.6003(±0.031)	0.5687(±0.043)	**0.6432**(±0.035)
30%	0.4028(±0.057)	0.4355(±0.045)	0.5452(±0.047)	0.5419(±0.039)	0.5032(±0.055)	**0.6081**(±0.042)

Table III: The performance comparison on the IMDB dataset.

Method	Homophily A	Homophily B	SRS	Co-Reg	Com-Reg	DUMR
80%	0.5724(±0.014)	0.6072(±0.013)	0.6545(±0.015)	0.6412(±0.012)	0.6556(±0.013)	**0.6956**(±0.011)
70%	0.5265(±0.017)	0.5434(±0.016)	0.6301(±0.019)	0.6085(±0.014)	0.6312(±0.015)	**0.6756**(±0.013)
60%	0.4628(±0.022)	0.4841(±0.020)	0.6168(±0.025)	0.5968(±0.017)	0.6044(±0.019)	**0.6301**(±0.017)
50%	0.4301(±0.028)	0.4632(±0.026)	0.5502(±0.030)	0.5404(±0.022)	0.5768(±0.024)	**0.6168**(±0.020)
40%	0.3703(±0.034)	0.4322(±0.033)	0.5187(±0.034)	0.4703(±0.028)	0.5404(±0.029)	**0.5802**(±0.026)
30%	0.3228(±0.041)	0.3855(±0.039)	0.4632(±0.041)	0.4419(±0.031)	0.4703(±0.036)	**0.5487**(±0.031)

amount of labeled data. Table I, II, and III give the average prediction accuracies after 30 runs for all the compared methods on the three tested datasets respectively. In all the testing experiments, two-fold cross-validation is used. From these reported results, we can easily see that the proposed DUMR approach consistently outperforms the compared methods with clear performance gains. In addition, the SRS approach achieves the second best performance with significantly higher accuracies than the other four baselines. Note that both the SRS method and the DUMR method are specifically designed models for social role prediction, which utilize sophisticated techniques to integrate multi-view information from social networks to perform accurate prediction. However, the proposed DUMR method forms a semi-supervised prediction model to maximize the usage of all available data to achieve more robust predictions. Since the Co-Reg and the Com-Reg approaches are general multi-view learning methods, they employ simple fusion strategies to combine different data sources that achieve limited performance gain. However, it is worth mentioning that the four methods using both link and node information always generate better prediction than the two Homophily methods, which confirms that both the local characteristics of nodes and the influence of the network present complementary information for better prediction.

D. Running time comparisons

We measure the efficiency of our implementation on three experimental datasets. Our Python implementation runs on a dual-core 3.3 GHz machine with 16G memory. We compare it with the running time of other baseline methods: Homophily A, Homophily B, SRS, Co-Reg, Com-Reg. In each run, 40% of the data are labeled and the methods outperform all the other methods. Table IV summarizes the results, measured as average running time over 30 runs. On Email, LinkedIn and IMDB datasets, the DUMP approach requires 1594.9, 2414.2 and 2714.5 seconds to output the prediction separately. Although the complexity of our algorithm is higher than single-view based methods, Homophily A, Homophily B and SRS, our method achieves comparable running time with the three basic methods. Compared with the dual-view methods, our proposed method achieves the same efficiency. The Com-Reg methods, on the contrast, takes even more time on each dataset.

Table IV: Running time (seconds) comparison between single-view algorithms and dual-view algorithms on all three datasets.

Datasets	Email	LinkedIn	IMDB
Single-view			
Homophily A	1,245.3	2,046.8	2,312.3
Homophily B	1,138.7	1,982.5	2,156.4
SRS	1,086.7	1,779.1	1,812.3
Dual-view			
Co-Reg	1,525.5	2,321.3	2,516.9
Com-Reg	1,711.8	2,950.6	3,100.6
DUMR	1,594.9	2,414.2	2,714.5

V. ACKNOWLEDGEMENT

This work is supported in part by the following grants: NSF awards CCF-1409601, IIS-1343639, ACI-1144061, and

CCF-1029166; DOE awards DESC0007456; AFOSR award FA9550-12-1-0458; NIST award 70NANB14H012.

REFERENCES

[1] Y. Zhao, G. Wang, P. S. Yu, S. Liu, and S. Zhang, "Inferring social roles and statuses in social networks," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '13. New York, NY, USA: ACM, 2013, pp. 695–703.

[2] Y. Cheng, Y. Xie, Z. Chen, A. Agrawal, A. Choudhary, and S. Guo, "Jobminer: A real-time system for mining job-related patterns from social media," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '13, 2013, pp. 1450–1453.

[3] A. Agrawal, Z. Chen, Y. Cheng, A. N. Choudhary, H. Liu, M. M. A. Patwary, Y. Xie, and K. Zhang, "Graphical modeling of macro behavioral targeting in social networks," in *Proceedings of the 13th SIAM International Conference on Data Mining, May 2-4, 2013. Austin, Texas, USA.*, 2013, pp. 740–748.

[4] K. Henderson, B. Gallagher, T. Eliassi-Rad, H. Tong, S. Basu, L. Akoglu, D. Koutra, C. Faloutsos, and L. Li, "Rolx: Structural role extraction & mining in large graphs," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '12. New York, NY, USA: ACM, 2012, pp. 1231–1239.

[5] H. T. Welser, D. Cosley, G. Kossinets, A. Lin, F. Dokshin, G. Gay, and M. Smith, "Finding social roles in wikipedia," in *Proceedings of the 2011 iConference*, ser. iConference '11. New York, NY, USA: ACM, 2011, pp. 122–129.

[6] A. Leuski, "Email is a stage: discovering people roles from email archives." in *SIGIR*. ACM, 2004, pp. 502–503.

[7] X. Hu and H. Liu, "Social status and role analysis of palin's email network," in *Proceedings of the 21st International Conference Companion on World Wide Web*, ser. WWW '12 Companion. New York, NY, USA: ACM, 2012, pp. 531–532.

[8] H. Fürstenau and M. Lapata, "Semi-supervised semantic role labeling via structural alignment," *Comput. Linguist.*, vol. 38, no. 1, pp. 135–171, Mar. 2012.

[9] N. Chen, J. Zhu, and E. P. Xing, "Predictive subspace learning for multi-view data: a large margin approach," in *Advances in Neural Information Processing Systems 23*, J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, Eds., 2010, pp. 361–369.

[10] S. Bickel and T. Scheffer, "Multi-view clustering," in *Proceedings of the Fourth IEEE International Conference on Data Mining*, ser. ICDM '04, 2004, pp. 19–26.

[11] V. Sindhwani and P. Niyogi, "A co-regularized approach to semi-supervised learning with multiple views," in *Proceedings of the ICML Workshop on Learning with Multiple Views*, 2005.

[12] G. Li, K. Chang, and S. C. H. Hoi, "Multiview semi-supervised learning with consensus," *IEEE Trans. on Knowl. and Data Eng.*, vol. 24, no. 11, pp. 2040–2051, Nov. 2012.

[13] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, ser. COLT' 98, 1998, pp. 92–100.

[14] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: Inferring user profiles in online social networks," in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, ser. WSDM '10. New York, NY, USA: ACM, 2010, pp. 251–260.

[15] F. R. K. Chung, *Spectral Graph Theory*. American Mathematical Society, 1997.

[16] T. Joachims, "Transductive inference for text classification using support vector machines," in *Proceedings of the Sixteenth International Conference on Machine Learning*, ser. ICML '99, San Francisco, CA, USA, 1999, pp. 200–209.

[17] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *IN ICML*, 2003, pp. 912–919.

[18] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schlkopf, "Learning with local and global consistency," in *Advances in Neural Information Processing Systems 16*. MIT Press, 2004, pp. 321–328.

[19] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Dec. 2006.

[20] C. A. Micchelli and M. Pontil, "Learning the kernel function via regularization," *J. Mach. Learn. Res.*, vol. 6, pp. 1099–1125, Dec. 2005.

[21] A. Argyriou, M. Herbster, and M. Pontil, "Combining graph laplacians for semi-supervised learning," in *Advances in Neural Information Processing Systems 18*. MIT Press, 2005, pp. 67–74.

[22] A. Argyriou, C. A. Micchelli, and M. Pontil, "Learning convex combinations of continuously parameterized basic kernels," in *COLT*, 2005, pp. 338–352.

[23] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining, (First Edition)*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.

[24] B. Qian, X. Wang, F. Wang, H. Li, J. Ye, and I. Davidson, "Active learning from relative queries," in *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, 2013.

[25] Y. Cheng, Q. Fan, S. Pankanti, and A. N. Choudhary, "Temporal sequence modeling for video event detection," in *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, 2014, pp. 2235–2242.