

Relational Learning with Social Status Analysis

Liang Wu[†], Xia Hu[‡] and Huan Liu[†]

[†]Computer Science & Engineering, School of CIDSE, Arizona State University, Tempe, AZ, USA

[‡]Department of Computer Science and Engineering, Texas A&M University, College Station, TX, USA
{wuliang, huan.liu}@asu.edu, hu@cse.tamu.edu

ABSTRACT

Relational learning has been proposed to cope with the interdependency among linked instances in social network analysis, which often adopts network connectivity and social media content for prediction. A common assumption in existing relational learning methods is that data instances are equally important. The algorithms developed based on the assumption may be significantly affected by outlier data and thus less robust. In the meantime, it has been well established in social sciences that actors are naturally of different social status in a social network. Motivated by findings from social sciences, in this paper, we investigate whether social status analysis could facilitate relational learning. Particularly, we propose a novel framework RESA to model social status using the network structure. It extracts robust and intrinsic latent social dimensions for social actors, which are further exploited as features for supervised learning models. The proposed method is applicable for real-world relational learning problems where noise exists. Extensive experiments are conducted on datasets obtained from real-world social media platforms. Empirical results demonstrate the effectiveness of RESA and further experiments are conducted to help understand the effects of parameter settings to the proposed model and how local social status works.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database applications—*Data Mining*; J.4 [Social and Behavioral Sciences]: Sociology

Keywords

Social Dimensions, Social Media, Relational Learning

1. INTRODUCTION

Social media websites have become a popular open platform for people to connect with their friends, express opinions

and share feelings. Online user-generated data becomes a key data source for researchers to study scientific and commercial problems [37] which are related to social actors. To better understand online behaviors of social network users, categorization of social network users is fundamental for various tasks. Take stock price prediction [4] as an example, the sentiment dynamics of users in a specific group may indicate corresponding stock prices. Although some social media platforms enable users to categorize themselves by uploading tags, such information is often incomplete and unavailable in real applications, which necessitates the use of relational learning (RL).

RL has become an effective tool for classifying nodes in a network. They leverage network structures as well as attributes of social actors, which often result in improvement over models without considering network structures [11]. Given attributes and social links of social actors, and labels from training examples, RL methods view the prediction as a supervised learning problem over networked data, assuming links between nodes to be indicators of proximity over labels, i.e., social actors with links connected with each other are assumed to have similar labels [20, 21]. However, since social media content is often noisy, RL could be prone to poor performance due to the massive amount of information of low quality. For example, expressions such as ‘Cooooo!’ and ‘omg!’ are confusing, phenomena like ‘sarcasm’ even reverse the meaning of a word vector [28]. In addition, the proximity assumption of social links does not hold. Random social links are established due to reflexive reciprocity and link farming [12].

Meanwhile, social status has been intensively studied in profiling social actors [13]. As indicated by self-categorization theory, social status influences goals individuals pursue, and further affects the way they pursue [17] and the positive social identity they obtain. The social psychological finding motivates us to investigate the use of social status for relational learning. However, it is challenging to incorporate social status into relational learning. First, social status is difficult to estimate. Although various kinds of algorithms have been proposed to estimate social influence, such methods focus on estimating individual centrality in terms of certain network structure instead of social status. For example, experts in a professional area may have a high social status but fewer friends in a social network, who will be overlooked by existing influence modeling methods. Second, we are lack of an effective way to incorporate social status into RL. Existing work only considers links between

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WSDM’16, February 22–25, 2016, San Francisco, CA, USA.

© 2016 ACM. ISBN 978-1-4503-3716-8/16/02...\$15.00

DOI: <http://dx.doi.org/10.1145/2835776.2835782>

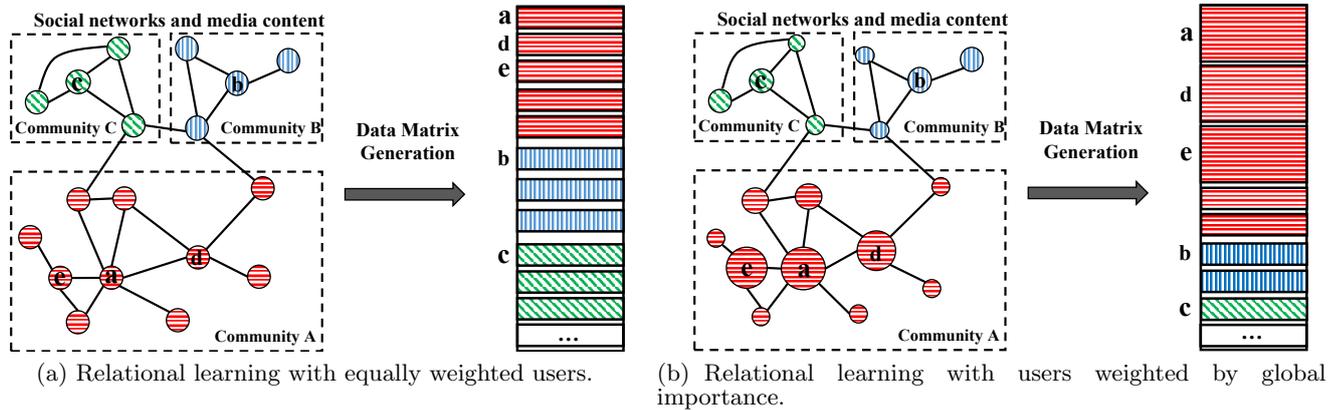


Figure 1: Illustrative examples of relational learning with different kinds of user weights.

nodes as an indicator of homophily, which fails to leverage them to reveal the social status.

In this paper, We propose a novel RElational learning framework based on Social status Analysis (RESA) to better predict labels of linked nodes through selecting influential agents of higher social status. Social status is used for constructing robust social dimensions based on social media content. Supervised models can then be built on the data in the latent social space.

The contribution of this work is threefold:

- We exploit social connections as an indicator of social importance to find influential individuals for RL. The proposed framework is resilient to noisy data through incorporating social status.
- A mathematical model is put forward to leverage social network structures. The model enables the automatic selection of information sources with higher social status and then generates robust social dimensions.
- Extensive experiments are conducted to test the effectiveness of the proposed model. It is demonstrated that network classification with the proposed social dimensions outperforms competitive baselines.

The rest of this paper is organized as follows. In Section 2, we briefly introduce the proposed framework. In Section 3, we investigate incorporating social status into RL. In Section 4, we show empirical evaluation with discussion. Related work is discussed in Section 5. Conclusion and possible future work are presented in Section 6.

2. RESA FRAMEWORK

Social media platforms enable users to label themselves through user profiling and content subscription. Since such labels are often unavailable and may be incomplete, it is of great significance to predict such labels of social actors with limited training information. Rather than merely focus on the network structure [22, 32], we focus on social media content. Since social content is idiosyncratic, informal and casual, we examine how to generate latent social dimensions in the presence of misinformation and noise by incorporating social status. In this section, we will introduce our motivation of the proposed framework and

formally formulate the problem and introduce intuitions and the general framework of RESA.

2.1 Motivation

Social actors are multidimensional, which indicates that social status should also be evaluated from different aspects. Existing social influence prediction algorithms often produce a scalar to represent individual status, which overlook individuals with high status but is not universally central. In this work, we reduce social status into two perspectives: global social status and local social status in a specific domain. Next, we describe details of the two aspects.

Figure 1 shows two illustrative examples for incorporating social status into relational learning, where the area of circles and heights of rows in the data matrix indicate the social status. Figure 1(a) illustrates why existing RL methods are prone to low performance without considering social status. Since nodes marked with letters are connected by more people, their information is supposed to be more useful than information generated by others. As shown by the data matrix, without considering it, information generated by important users like Barack Obama are treated the same with anybody else, though their posts are more important.

Figure 1(b) illustrates how social status works for RL. Through upweighting importance of nodes with greater centrality, information from nodes of higher social status are better focused. The problem is that the iterative mutual reinforcement between neighbors of existing centrality measures forces less popular communities to be overwhelmed by the globally influential groups, since communities are of different levels of activeness. As shown in Figure 1(b), although b and c may be less important than d and e , once node a is selected, b and c are more informative. Thus, b and c represent social actors with high social status but are of less centrality. For example, though Barack Obama and Pope Francis are more important globally, it is favorable to select social actors from diverse groups and backgrounds, such as athletes and professors, as well.

In this work, we aim to select social actors with high social status and upweight the content they generate during training. The content is extracted from their posts and social status depends on global centrality and local social influence. Next, we will introduce formulation and general framework of the problem.

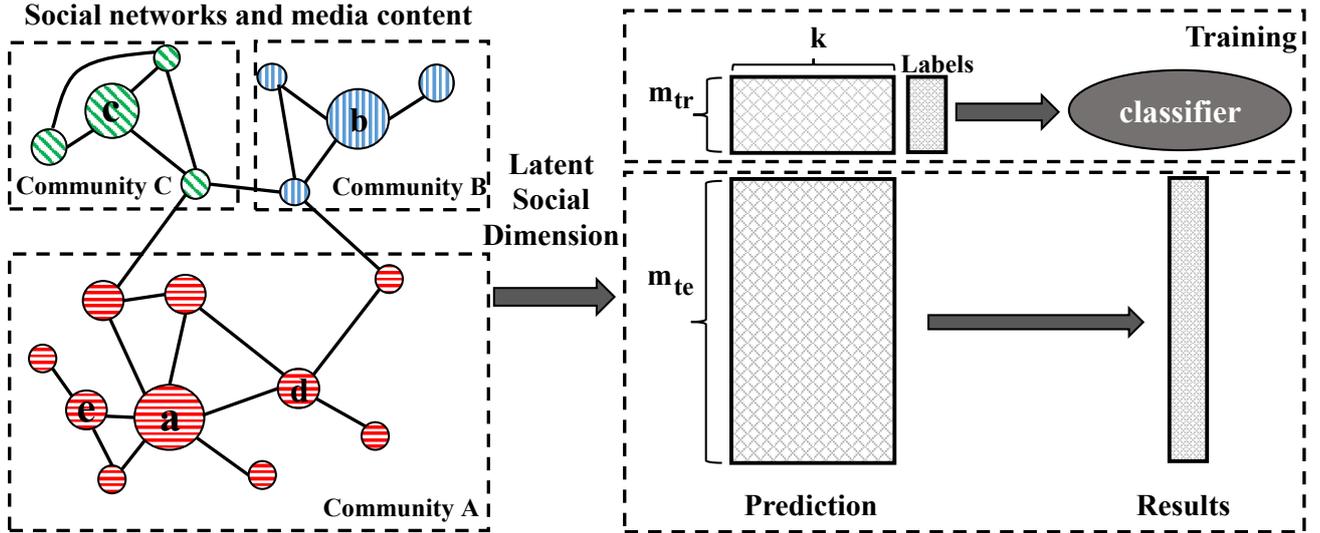


Figure 2: Relational Learning with Social Status Analysis.

2.2 Problem Formulation

Consider $\mathbf{V} \in \mathbb{R}^{n \times m}$ is the attribute matrix, where $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m)$ and each row is represented as \mathbf{v}_i , which is the attribute vector of user i . The columns may refer to user attributes depending on different websites. \mathbf{A} is the adjacency matrix of the underlying social network dataset with each entry \mathbf{A}_{ij} equaling to one if i follows j and equaling zero otherwise. A label set \mathbf{y} represents certain social tags and each user owns one or multiple labels. We aim to predict labels of a user given the network structure and his/her social media content. More formally, the problem is stated as follows:

Input

a user-attribute matrix \mathbf{V} , an adjacency matrix \mathbf{A} and the training information matrix $\mathbf{Y}^{tr} = \{0, 1\}^{m_{tr} \times t}$, where t is the number of labels and m_{tr} is the number of training instances.

Output

labels of test users, $\mathbf{Y}^{te} = \{0, 1\}^{m_{te} \times t}$, where m_{te} is the size of testing data.

The social labels can be obtained from different sources in different platforms. For example, Flickr users can join different groups and BlogCatalog users are able to subscribe and add tags for themselves. Such explicit group memberships and interest tags can then be extracted as labels. A user usually joins various groups and subscribes various interest tags.

2.3 General Framework

Latent social dimensions have been proposed to transform user links to intrinsic features for RL [32]. Since interests of an individual are often complex, such social dimensions enable user interests to be better represented. Different from existing RL methods which only regard social links as an indicator of homophily, we view them as an indicator of social status, which represents the weights of a user and its content.

For example, if a user has been followed by numerous people and has higher social influence in a social network, his posted content is more probable to be useful and of better quality. Thus, we first adopt linkage analysis algorithms to calculate how important a user is globally. Locally influential nodes are important for modeling structures of social networks. As mentioned in [1], local forceful nodes are important for preventing misinformation during the process of information aggregation. Thus, we further propose to construct social dimensions through leveraging the local social influence as well as global influence using social relations. The detailed procedure of our proposed work is illustrated in Figure 2, which consists of two steps: constructing latent dimensions and learning a relational classifier.

Dimension Extraction: First, we concentrate on extracting latent dimensions from social media content. Social actors are assumed to be equally weighted by existing methods. As social media content in real world is noisy, information from influential sources should be zoomed in. This inspires us to incorporate a weight into the extraction procedure. The social weights are derived from the network structure. Although global social status is easy to be derived from the network structure, local status is difficult to obtain. In order to highlight more local influencers from small groups and balance contribution from different communities, we introduce a intra-group exclusive sparsity regularizer. It stimulates representative nodes in small groups to be higher weighted. A toy example is depicted in Figure 2, where the area of each circle represents its importance. In our framework, the influencers with high local social status are better focused, meaning that Node b and c are of more importance than e and d .

Relational Classifier Construction: After the social dimensions are generated, social actors are mapped to the new space and represented by latent factors. Supervised learning algorithms, such as regression and classification, can then be employed upon them to enable network classification. The dimension can easily be extended through combining more features generated from other sources, like

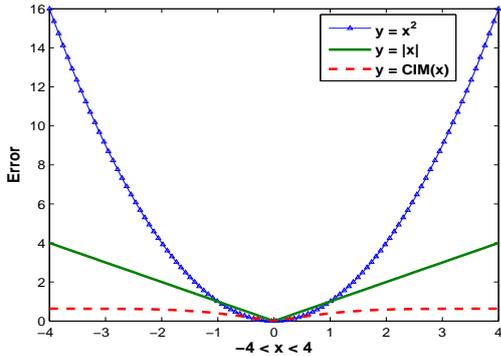


Figure 3: Curves of different cost functions in the range from -1 to 1.

network structure and user profiles. In this work, we only concentrate on the social media content. Since the classifier determines the identity label by learning upon the features, its prediction accuracy is an indicator telling whether the extracted dimensions are informative and descriptive. Due to the scalability and simplicity, we follow [32] and selected one-vs-rest linear SVM for discriminative learning.

3. RELATIONAL LEARNING FRAMEWORK WITH SOCIAL STATUS

In this section, we introduce how we induce latent social dimensions. The resultant latent social dimensions should be informative in the presence of noise. Since how to cope with noisy links has been discussed in previous work [32], we focus on reducing the negative impact of noise through selecting information sources with higher social status. The problem of social dimensions induced from user-generated content is to obtain an optimal low-rank representation $\mathbf{W} \in \mathbb{R}^{m \times k}$ ($k \ll n$), which best describes the social affiliations. As the social dimension represents the probability an instance is assigned to a cluster, we constrain it to be nonnegative.

3.1 Global Social Status Modeling for Relational Learning

Existing RL methods regard each instance equally weighted, which ignore that social actors are of different social status. Since the degree distribution of real social network subjects to power law and has a heavy tail [2], the equivalence assumption leads the model to be biased by less popular users. Unfortunately, they may even be dominated by the outliers in real world applications. Eq.(1) is a general form of dimension reduction algorithms, and it minimizes the sum of squared reconstruction loss:

$$\min_{\mathbf{W}, \mathbf{H}} \sum_i^m (\mathbf{v}_i - \mathbf{w}_i \mathbf{H}^T)(\mathbf{v}_i - \mathbf{w}_i \mathbf{H}^T)^T, \quad (1)$$

where \mathbf{V} is decomposed into two latent factor \mathbf{W} and \mathbf{H} . $\mathbf{H} \in \mathbb{R}^{n \times k}$ ($k \ll n$) describes associations between features and social affiliations, .

Figure 3 denotes the curve of different cost functions. As shown in the figure, the squared loss increases quadratically when error becomes large. So such models are optimal when

the noise $n \sim \mathcal{N}(\mu, \sigma^2)$ and it is small. However, when the noise becomes large, such algorithms are not resilient to outliers. In order to solve the problem, we propose to adopt the global social impact as weights of social actors to select more important information agents for constructing social dimensions. Given the social influence of each user, Eq.(1) can be reformulated as:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}} \quad & \sum_{i=1}^m \mathbf{c}_i (\mathbf{v}_i - \mathbf{w}_i \mathbf{H}^T)(\mathbf{v}_i - \mathbf{w}_i \mathbf{H}^T)^T \\ \text{subject to} \quad & \sum_{i=1}^m \mathbf{c}_i = 1, \end{aligned} \quad (2)$$

where $\mathbf{c} \in \mathbb{R}^m$ and \mathbf{c}_i is the social impact of user i , the values are positive and sum to 1 for all users. Eq.(2) is a row based normalization based on social impact. Influential nodes will be upweighted and users in the heavy tail will be avoided to dominate the calculation.

Since Eq.(2) is not jointly convex in terms of \mathbf{W} and \mathbf{H} , we solve the problem in an iterative manner. By fixing one variable each time, the objective is convex to the counter part and the optimal solution can be obtained. Motivated by the optimization method proposed in [8], we develop the update rules as follows:

While fixing \mathbf{H} , update \mathbf{W} :

$$\mathbf{W}_{ik} = \mathbf{W}_{ik} \frac{(\mathbf{C} \otimes \mathbf{V} \mathbf{H})_{ik}}{(\mathbf{C} \otimes (\mathbf{W} \mathbf{H}^T) \mathbf{H})_{ik}}, \quad (3)$$

where \mathbf{c} is the global social status and will be discussed in Section 3.4 and $\mathbf{C} = \mathbf{c} \mathbf{1}_n^T$. \otimes is Hadamard product and has higher operator precedence over matrix product.

While fixing \mathbf{W} , update \mathbf{H} :

$$\mathbf{H}_{jk} = \mathbf{H}_{jk} \frac{((\mathbf{C} \otimes \mathbf{V})^T \mathbf{W})_{jk}}{((\mathbf{C} \otimes (\mathbf{W} \mathbf{H}^T))^T \mathbf{W})_{jk}}. \quad (4)$$

3.2 Local Social Status Modeling for Relational Learning

By incorporating the global social status, if there exist some groups with significantly higher influence than the others, the resultant social dimensions will be biased by them. Different kinds of social affiliations are of different popularity. Globally influential groups may contribute redundant information, since users with similar social affiliations tend to have overlapped social attributes. As mentioned earlier, once the most influential social actor is selected, members in the same group become less informative. Our aim is to actively select more various nodes from different groups through upweighting locally influential information sources.

In this section, we use \mathbf{c} to represent individual weight for each user. Assume we have a static affiliation matrix $\mathbf{S} \in \mathbb{R}^{m \times g}$, where $\mathbf{S}_{ij} = 1$ means user i is a member of affiliation j and g is the number of all groups, we introduce the exclusive group LASSO penalty [16], $\mathcal{L}_{1,2}$ -norm, to achieve the intra-group sparsity for \mathbf{c} :

$$\mathcal{L}_{1,2}(\mathbf{c}) = \frac{1}{2} \sum_{j=1}^g \left(\sum_{i, \mathbf{S}_{ij}=1} |\mathbf{c}_i| \right)^2. \quad (5)$$

$\mathcal{L}_{1,2}$ -norm first sums up over intra-group variables, and imposes an \mathcal{L}_2 -norm to regularize the sum. Its minimization leads to intra-group sparsity. Concretely, it enforces more zero entries to take place in a group. The property

is desirable for our assumption since it tends to select locally representative influencers in different groups. By the definition of $\mathcal{L}_{1,2}$ -norm, Eq.(5) can be formulated as:

$$\mathcal{L}_{1,2}(\mathbf{c}) = \frac{1}{2} \sum_{i=1}^g (\mathbf{c}^T \mathbf{S}_{*i})^2 \quad (6)$$

$$= \frac{1}{2} \sum_{i=1}^g \mathbf{c}^T \mathbf{S}_{*i} \mathbf{S}_{*i}^T \mathbf{c} \quad (7)$$

$$= \frac{1}{2} \mathbf{c}^T \mathbf{S} \mathbf{S}^T \mathbf{c}, \quad (8)$$

where $\mathbf{S}_{*i} \in \mathbb{R}^{m \times 1}$ and denotes the membership of group i . By jointly considering the global and local social status, the objective can be reformulated as:

$$\min_{\mathbf{c}, \mathbf{W}, \mathbf{H}} \sum_i^m \mathbf{c}_i (\mathbf{v}_i - \mathbf{w}_i \mathbf{H}^T) (\mathbf{v}_i - \mathbf{w}_i \mathbf{H}^T)^T + \frac{1}{2} \mathbf{c}^T \mathbf{S} \mathbf{S}^T \mathbf{c}. \quad (9)$$

For simplicity, we replace $\mathbf{S} \mathbf{S}^T$ by $\mathbf{D} \in \mathbb{R}^{m \times m}$. Each entry \mathbf{D}_{ij} denotes the correlation between user i and j in terms of their overlapping affiliations. A large correlation \mathbf{D}_{ij} forces user i 's weight \mathbf{c}_i to be small if \mathbf{c}_j has a higher social status. Thus, intra-group sparsity actively selects the most influential people in one group to be upweighted. The resultant \mathbf{c} balances the contribution of different social affiliations and avoids being biased by only the top group and their redundant information.

3.3 Optimization

The optimization objective is not jointly convex to \mathbf{W} , \mathbf{H} and \mathbf{c} . We update them in an iterative manner.

When \mathbf{c} is fixed, the optimization has the same form with Eq.(2). Thus, \mathbf{W} and \mathbf{H} can be updated iteratively as denoted in Eq.(3) and Eq.(4).

When \mathbf{W} and \mathbf{H} are fixed, the objective function can be reformulated as:

$$\min_{\mathbf{c}} \sum_{i=1}^m \mathbf{c}_i t_i + \frac{1}{2} \mathbf{c}^T \mathbf{D} \mathbf{c} \quad (10)$$

$$\text{subject to } lb < \mathbf{c}_i < ub, \sum_{i=1}^m \mathbf{c}_i = 1,$$

where $t_i = (\mathbf{v}_i - \mathbf{w}_i \mathbf{H}^T) (\mathbf{v}_i - \mathbf{w}_i \mathbf{H}^T)^T$. Then optimization of Eq.(10) is transformed to a quadratic programming (QP) problem and the solution is similar to that introduced in [6]. Here we adopt two parameters, the upper bound ub and lower bound lb . The two bounds control the maximum and minimum weight of a node correspondingly.

The algorithm to extract social dimensions is summarized in Algorithm 1.

3.4 Discussion

In order to select both globally and locally influential social actors, RESA needs each user's global social influence and the social affiliations as input. In this section, we introduce how we measure the global influence and their affiliations based on network structures. We adopt the individual PageRank centrality to reveal the global social impact and describe a modularity maximization algorithm to discover the community memberships of different users.

The network structure of social network can be exploited to infer the authority of each user. Various algorithms

Algorithm 1 Social Dimension Extraction

Input: The user attribute matrix $\mathbf{V} \in \mathbb{R}^{m \times n}$,
the initial latent factors $\mathbf{W} \in \mathbb{R}^{m \times k}$ and $\mathbf{H} \in \mathbb{R}^{n \times k}$,
the social affiliation matrix $\mathbf{S} \in \mathbb{R}^{m \times k}$,
the vector of individual global influence $\mathbf{c} \in \mathbb{R}^m$,

Output: Social dimension \mathbf{W} , \mathbf{H} and \mathbf{c} .

Repeat

Calculate $\mathbf{C} = \mathbf{c} \mathbf{1}_n^T$

Update \mathbf{W} by $\mathbf{W}_{ik} = \mathbf{W}_{ik} \frac{(\mathbf{C} \otimes \mathbf{V} \mathbf{H})_{ik}}{(\mathbf{C} \otimes (\mathbf{W} \mathbf{H}^T) \mathbf{H})_{ik}}$;

Update \mathbf{H} by $\mathbf{H}_{jk} = \mathbf{H}_{jk} \frac{((\mathbf{C} \otimes \mathbf{V})^T \mathbf{W})_{jk}}{((\mathbf{C} \otimes (\mathbf{W} \mathbf{H}^T))^T \mathbf{W})_{jk}}$;

Update \mathbf{c} by Eq.(10);

until Convergence

have been proposed to predict the social influence in a social network [39]. Since we focus on leveraging the connectivity to build up robust social dimensions in this work, we directly adopt PageRank (PR) [27] to estimate the celebrity of each user. PR was proposed for web search to infer page importance based on the hyperlink structure and was proven to be useful for finding important people in a social network. Note that other centrality estimation methods can also be incorporated into our framework. In order to discover the social group affiliations, we adopt the modularity maximization approach [32]. The concept of modularity is introduced to measure the community structure of complex networks [7, 25]. It is to measure the deviation of interactions between users from a real social network to a uniform random graph. Here, we denote the social affiliation matrix as $\mathbf{S} \in \mathbb{R}^{m \times g}$. \mathbf{S} and the optimal PageRank scores $\mathbf{c} \in \mathbb{R}^{m \times g}$ are then be used by the proposed model in Section 3.2. Note that the proposed framework is independent to the estimation of global social impact and social affiliation and different measures can be used. To choose the best measure is not the main focus of this paper, which remains an important direction of future work.

4. EXPERIMENTS

In this section, we introduce experiment details to validate the effectiveness of the proposed framework. We first introduce our datasets and the methods used for comparison. Then we discuss the performance based on the two datasets. After investigating the sensitivity of RESA when the social dimensionality changes, we further verify how the social constraint works during training. Finally, we examine the effectiveness of RESA on the datasets. Empirical results of RL based on the social dimensions generated by other models are also presented.

4.1 Data Sets

In order to examine how the relational classifier behaves on real world social media websites, two datasets obtained from BlogCatalog¹ and Flickr² are adopted in our work. Both of them are publicly available and have been used previous work [32, 34].

Table 1 illustrates some statistics about the two datasets. The users are randomly sampled from the two websites, and various kinds of intrinsic attributes are extracted as features. We adopt the user interest tags in BlogCatalog and explicit

¹<http://www.blogcatalog.com/>

²<http://www.flickr.com/>

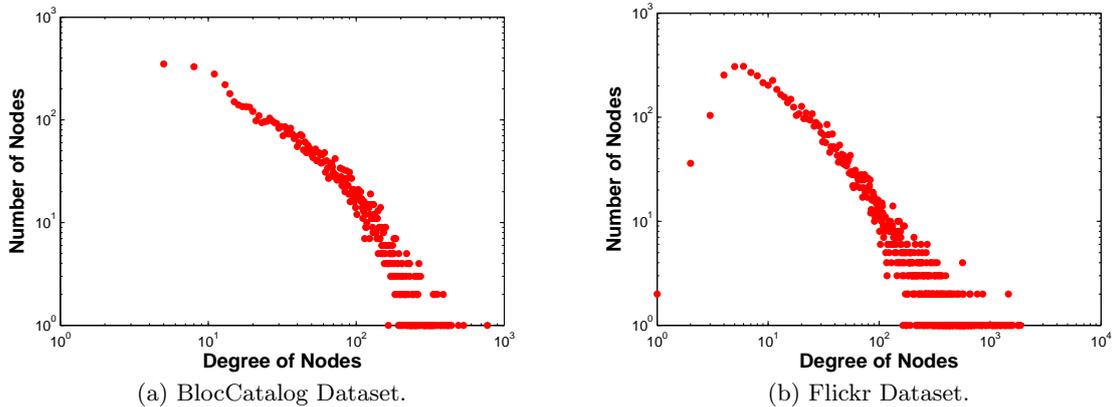


Figure 4: The degree distribution of nodes in two datasets.

Table 1: The statistics about employed datasets.

	# of Instances	# of Labels	# of features
BlogCatalog	5198	6	8189
Flickr	7575	9	12047

group memberships as identity labels. In order to better understand the distributions, we further explore the degree distribution of them in Figure 4. The x-axis and y-axis are both of log scale. As shown in the figure, the two networks are scale-free [3], meaning that only few nodes are of a high degree and a large portion of users have few connections.

4.2 Baseline Methods and Metrics

We select several representative methods for comparison. As we mainly focus on leveraging social status for building robust RL, more variants are not included. LFM only considers the latent structure of social media content and GNMF incorporates social relations as homophily indicators. CIMNMF is robust to large noise and GSIM is used for validating the global social status.

Latent Factor Model: Latent Factor Model (LFM) has been proven a useful tool for reducing dimensions in real applications. For social media analysis, since the feature often represents the tweeted words and user tags, a negative value is often meaningless and confusing, here we constrain all entries of factors to be nonnegative. Such constraint can reduce the model into Nonnegative Matrix Factorization. We employ the model proposed by Lee and Seung [29], and adopt the multiplicative update rule.

Graph Regularized NMF: To take advantage of the linkage information between users in social network data, another common assumption is users with social links should share a higher similarity in terms of content. In order to compare our model with such intuition, we adopt the Graph Regularized NMF (GNMF) model [5].

Robust NMF: Traditional dimension reduction algorithms suffer from noise due to the cost function they used. For example, one of the most common cost function, L_2 -norm, which is adopted for reducing dimensions by methods like NMF and Principal Component Analysis (PCA), assumes noise $n \sim \mathcal{N}(\mu, \sigma^2)$ and it is small. But in real applications, the noise are large and impulsive, and sometimes contain large corruptions. As shown in Figure 3, squared loss

increases quadratically when its argument becomes large. Various robust cost functions are introduced to solve this problem in areas of signal processing and computer vision. In this work, we adopt Correntropy Induced Metric (CIM) to compare our proposed model with such robust methods. Correntropy was proposed in information theoretic learning [19] to calculate the distance between two random variables, and it was further extended as CIM to process non-Gaussian noise and large corruptions in measuring distance of two vectors. In this work, the CIMNMF model [8] is adopted as a baseline method.

Global Social Impact Model: In order to further investigate the effectiveness of local social status, we also adopted the method introduced in Section 3.1 (GSIM) as a baseline, which only considers global social status.

4.3 Metrics

F_1 -measure: To test the prediction accuracy in terms of both precision and recall, we adopted the F_1 -measure to evaluate the performance.

Macro- F_1 : Since in our methods, there are multiple labels to be predicted. For each task t , F_1^t can be computed. In order to get the overall performance, we first adopt the Macro-averaged F_1 -measure as:

$$\text{Macro} - F_1 = \frac{1}{|\mathbf{T}|} \sum_{t \in \mathbf{T}} F_1^t, \quad (11)$$

where \mathbf{T} is the set of all identity labels and F_1^t is the F_1 -measure of task t .

Micro- F_1 : A possible problem of Macro- F_1 is, since the size of different labels varies, the task with fewer instances may be overemphasized. In order to cope with this problem, we adopted Micro-averaged F_1 -measure. First, we calculate the micro averaged precision and recall:

$$\text{Micro} - \text{precision} = \frac{\#TP}{\#TP + \#FP} \quad (12)$$

$$\text{Micro} - \text{recall} = \frac{\#TP}{\#TP + \#FN}, \quad (13)$$

where $\#TP$ is the number of true positives, $\#FP$ is the number of false positives and $\#FN$ is the number of false negatives. Then Micro- F_1 is the harmonic average of Micro-precision and Micro-recall.

Table 2: The F_1 -measure of different methods on BlogCatalog Data with varying training ratio.

	Training Ratio	10%	20%	30%	40%	50%	60%	70%	80%	90%
Micro- F_1 (%)	LFM	0.6977	0.7074	0.7223	0.7359	0.7425	0.7476	0.7562	0.7581	0.7669
	GNMF	0.7221	0.7352	0.7507	0.7530	0.7546	0.7550	0.7568	0.7627	0.7707
	CIMNMF	0.7170	0.7280	0.7382	0.7405	0.7476	0.7492	0.7507	0.7508	0.7518
	GSIM	0.6975	0.7143	0.7283	0.7328	0.7382	0.7441	0.7495	0.7499	0.7520
	RESA	0.7527	0.7566	0.7630	0.7677	0.7757	0.7760	0.7826	0.7883	0.8103
Macro- F_1 (%)	LFM	0.6958	0.7109	0.7206	0.7305	0.7363	0.7397	0.7532	0.7541	0.7632
	GNMF	0.7162	0.7308	0.7476	0.7508	0.7512	0.7537	0.7550	0.7577	0.7695
	CIMNMF	0.7116	0.7241	0.7349	0.7370	0.7406	0.7412	0.7464	0.7466	0.7481
	GSIM	0.6911	0.7093	0.7252	0.7280	0.7398	0.7406	0.7459	0.7468	0.7474
	RESA	0.7497	0.7525	0.7572	0.7637	0.7710	0.7779	0.7795	0.7809	0.8087

Table 3: The F_1 -measure of different methods on Flickr Data with varying training ratio.

	Training Ratio	10%	20%	30%	40%	50%	60%	70%	80%	90%
Micro- F_1 (%)	LFM	0.5443	0.5622	0.5705	0.5722	0.5746	0.5857	0.5868	0.5939	0.6065
	GNMF	0.5545	0.5588	0.5677	0.5735	0.5761	0.5767	0.5792	0.5811	0.5924
	CIMNMF	0.3587	0.4084	0.4556	0.4561	0.4808	0.4934	0.5043	0.5173	0.5381
	GSIM	0.5395	0.5584	0.5627	0.5640	0.5646	0.5652	0.5685	0.5709	0.5724
	RESA	0.5690	0.5770	0.5967	0.6086	0.6136	0.6169	0.6364	0.6394	0.6444
Macro- F_1 (%)	LFM	0.5306	0.5536	0.5673	0.5694	0.5687	0.5792	0.5823	0.5873	0.6053
	GNMF	0.5512	0.5521	0.5615	0.5667	0.5682	0.5703	0.5762	0.5761	0.5849
	CIMNMF	0.3394	0.3658	0.4126	0.4089	0.4494	0.4393	0.4764	0.4908	0.5262
	GSIM	0.5364	0.5500	0.5528	0.5539	0.5560	0.5537	0.5639	0.5608	0.5673
	RESA	0.5170	0.5293	0.5298	0.5366	0.5419	0.5774	0.5777	0.5894	0.6159

4.4 Experiments on BlogCatalog Data

The performance of different methods on BlogCatalog dataset with varying training ratio, from 10% to 90%, is illustrated in Table 2. For each training ratio, samples are randomly split into training and testing set. We repeat this process 10 times and report the average results. The highest performance under each setting is highlighted in bold face. In terms of both Micro- F_1 and Macro- F_1 , our proposed model RESA outperforms the baselines. As shown in the table, the performance of GNMF, which adopts the user graph regularization, is the closest to that of RESA. It implies that the homophily phenomena are observed on this dataset, meaning that users are likely to follow similar people. When less data is used for training, CIMNMF performs better than LFM. But it is outperformed by LFM when more data are used for training. The performance of CIMNMF is also worse than GNMF on both measures. This indicates that the robust methods are not directly applicable to social media analysis. GSIM has the lowest Micro- and Macro- F_1 among all five methods. It shows that using the static global social status leads the model to be biased by certain nodes, which finally resulted in an unfavorable result.

4.5 Experiments on Flickr Data

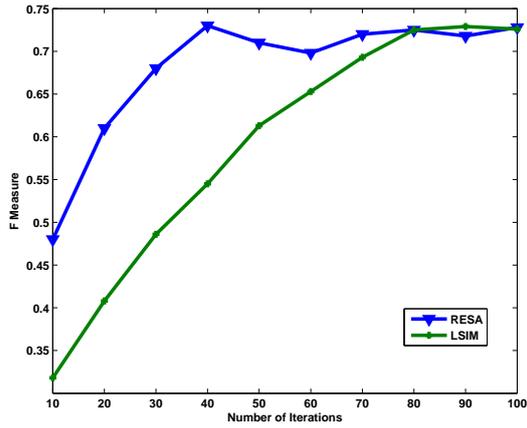
The performance of different methods on Flickr dataset with varying training ratio, from 10% to 90%, is illustrated in Table 3. In terms of Micro- F_1 , the proposed RESA outperforms other methods all the time. Compared with experimental results of BlogCatalog dataset, where GNMF is the runner-up and outperforms other baselines, LFM performs better than GNMF on Flickr dataset. It indicates that their performance varies with different settings and they are not stable enough. Similarly, GSIM does not perform very well, which proves the static and global social impact

leads to a biased representation, although introducing locally influential information sources can significantly improve the performance. But since we only use PageRank centrality in our work to induce the global influence, other centrality measures may have different performance. CIMNMF has the worst performance among all methods, and falls behind RESA about 20% to 60% under different settings. This large lag further proves that the CIMNMF method cannot directly be applied to social media dataset.

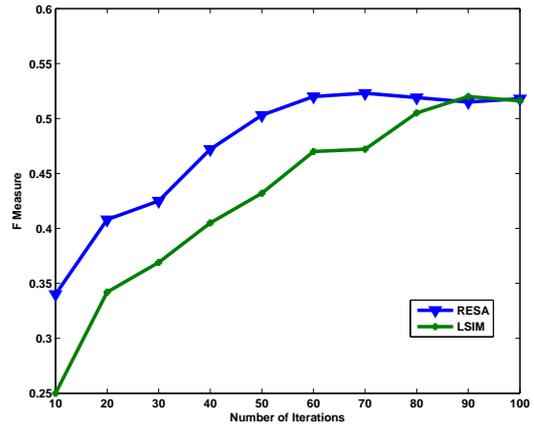
But the proposed RESA fails to perform the best under some settings when we use Macro- F_1 . As depicted in Table 3, though it performs the best under the setting of 80% and 90% training ratio, RESA is outperformed by LFM and GNMF when less portion is used for training. Since LFM is the runner up method under most settings of Micro- F_1 , the Macro- F_1 result indicates that LFM, which does not consider any graph structure, is a competitive baseline for generating social dimensions. By both considering the Macro- F_1 and Micro- F_1 , a conclusion can be made that RESA fails to identify certain tasks with fewer training instances. But when we increase the training ratio, RESA converges to and outperforms LFM.

4.6 Effectiveness of Social Status

In this section, we plan to test how social status works to facilitate RL. As discussed in Section 3.4, global social status is obtained separately from centrality measures. In Eq.9, global social status is used as prior for \mathbf{c} . To test how it contributes to RESA, we propose a variant of RESA, which initializes \mathbf{c} using random values. We named the model as Local Social Influence Model (LSIM). We vary number of training iterations and record results of two models, which are presented in Figure 5. As shown by results of BlogCatalog and Flickr datasets, RESA converges



(a) Micro- F_1 on BlogCatalog Data.



(b) Micro- F_1 on Flickr Data.

Figure 5: The Micro- F_1 measure of *RESA* and *LSIM* .

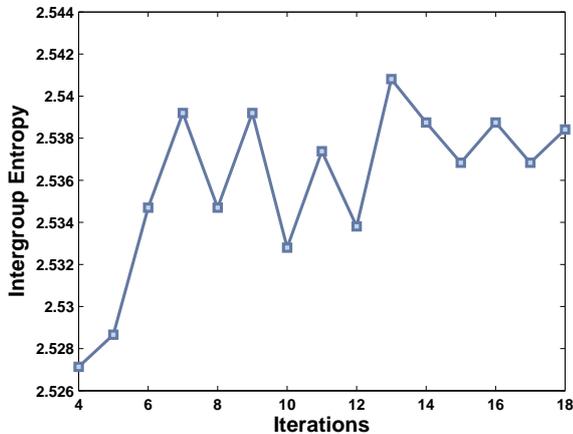


Figure 6: Changes of Inter-group Entropy.

more efficiently than LSIM, meaning that it takes less epochs to obtain the optimal solution. LSIM is able to gradually converge to RESA with more iterations. The result indicates that LSIM can converge to the optimal solution with more iterations and global social status helps the model to converge more efficiently.

Unlike global social impact, which is easily accessible, it is difficult to validate usefulness of local social status. In order to further investigate how the proposed model achieves to capture the local social influence, here we define the inter-group diversity:

$$P = S^T c. \quad (14)$$

The weight of each group can be calculated through collecting weight of instances from the group as in Eq.(14). We treat the group weights as their probabilities and calculate the entropy of the underlying distribution. Thus, the entropy measures the inter-group diversity of group weight distribution. Figure 6 denotes the changes of inter-group entropy. As shown in the figure, the entropy

increases with the number of iterations. At first it fluctuates around between 2.535 and 2.540, and finally converges after more iterations. The results indicate that the $L_{1,2}$ -norm is effective in increasing inter-group diversity and thus selecting influencers of high local social status.

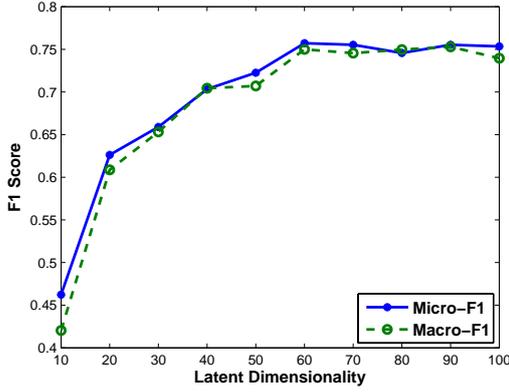
4.7 Sensitivity of Latent Dimensionality

In this section, we vary the dimensionality of social dimensions and observe the performance on both datasets. The dimensions are changed in the range of 10 to 100. The corresponding results on BlogCatalog and Flickr data are illustrated in Figure 7. As denoted by the figures, both Micro- and Macro- F_1 increase significantly when the dimensionality is small. The increasing ratio decreases when there are more dimensions and finally converges when k increases to 60 or more. When k increases to 100, the Macro- F_1 on BlogCatalog data and the Micro- F_1 on Flickr data even decrease. The results prove that user behaviors are idiocratic: when there are too few social dimensions, it fails to describe the idiocrasy. But a too large dimensionality is also meaningless, since the extra dimensions are not informative and may worsen the performance.

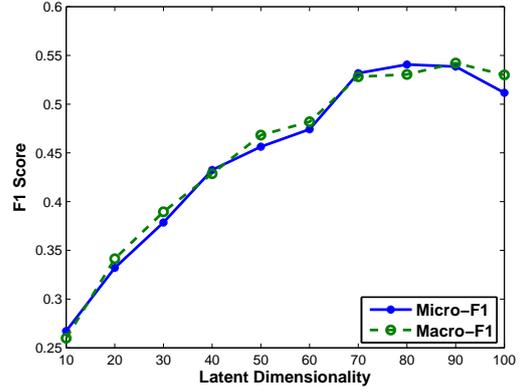
5. RELATED WORK

Relational learning refers to the learning over relational information between connected instances. Unlike traditional machine learning tasks, the connections made the classical i.i.d. assumption no longer applicable. In order to predict labels of linked instances, various methods have been proposed. Collective inference has been proposed to classify graph nodes based on the network structure [10, 15], where the graph structure of a node is often transformed into its features. The structural features are further fed to a supervised learning framework, which iteratively predicts class labels of unlabeled examples. Such methods are able to achieve a better accuracy through capturing autocorrelation between labels of neighbors.

Besides merely focusing on the relationship between one hub friends, other models are proposed to capture the long-distance autocorrelation [38]. They assume there exist hidden topics, where connections between users and



(a) F_1 -measure on BlogCatalog Data.



(b) F_1 -measure on Flickr Data.

Figure 7: The sensitivity in terms of different varying dimensionality.

social attributes of users are generated according to their group memberships. Such latent memberships are used for learning a supervised model to classify connected nodes. In [24], Neville and Jensen propose to replace the soft clustering with disjoint clusters, and use such cluster memberships as features to infer labels of social actors. The hard clustering manner reduces the computational cost, but it requires each actor to be assigned to one group. The assumption may not hold for real world social networks, since some nodes may not show a clear relationship to any groups. The strong assumption forces these nodes to be assigned randomly and thus bringing in noise.

Learning latent social dimensions based on social connections, attributes and social media content has attracted much attention from both RL [32] and other related classification tasks on social media instances. Since social content is often of high dimensionality and noisy, these methods solve the problem through learning a low rank representation social actors, which can be regarded as feature extraction. Another more direct way is to select discriminant features based on the latent social affiliations[30]. Since social connections and social media content are often noisy, methods in this stream which assume all nodes are equally weighted are not resilient to outliers. In order to infer user weights from social connections, PageRank [14] and topic propagation [31] have been adopted. A static social weight can reflect how globally influential a node is globally, but learning with such a weight may bias the model with influential groups. Nodes from the same group may introduce too much overlapped information, which is unfavorable. In our work, we use global influence as a prior and actively upweight local influencers from small groups, so as to achieve greater inter-group diversity.

Our work is also related to matrix factorization and sparse machine learning methods. Matrix factorization has been used in various applications such as recommendation [35] and distance metric learning [36] to reduce dimensionality. Sparse learning has been proposed to enforce the data representation to have more zeros. The structures of sparsity include group level [23], tree level [18], differences between features [33] and so on. Most sparse learning algorithms focus on learning a sparse representation for instances [9], while our work aims to select informative instances. Nie

et al. propose to select representative instances from the dataset using $L_{2,1}$ -norm [26]. The proposed model is different since we measure how informative a node is based on its social status and the social group it belongs to.

6. CONCLUSION AND FUTURE WORK

The availability of social relations enables RL algorithms to better model and predict data instances in a network environment, where the i.i.d. assumption no longer holds. In this paper, we investigate how to leverage the global and local social influence for RL. To capture global social context, we adopt PageRank centrality to weight the importance of different nodes. To capture the local social influence, we force individuals in a same group to compete with each other, thus obtaining the intra-group sparsity. The inter-group sparsity enable influential nodes in small communities to survive and reduce overlapped information from the top influential groups. With these solutions, we propose a framework RESA to integrate local and global social influence for constructing the latent social dimensions. Experimental results on real-world datasets show that the proposed method outperforms under most settings. We also conduct experiments to further understand how the local social status works.

There are several possible directions remaining to be investigated in the future. First, as new users may join the community and social media content and social relations of existing social actors may evolve overtime, it is challenging to efficiently update the model. In our paper, the exclusive group LASSO forces nodes to be either influencers or non-influential nodes. But in reality, a hierarchy may exist among social actors. Thus, we plan to explore whether different structures are hidden behind social status, and how other kinds of sparse learning methods can help to better model social influences. Third, we will investigate centrality measures other than PageRank on analyzing the social network structure.

7. ACKNOWLEDGEMENTS

This work was supported, in part, by the Office of Naval Research grants N000141410095 and N000141310835.

8. REFERENCES

- [1] D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi. Spread of (mis) information in social networks. *Games and Economic Behavior*, 70(2):194–227, 2010.
- [2] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.
- [3] A.-L. Barabási et al. Scale-free networks: a decade and beyond. *science*, 325(5939):412, 2009.
- [4] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 2011.
- [5] D. Cai, X. He, J. Han, and T. S. Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [6] J. Currie and D. I. Wilson. Opti: lowering the barrier between open source optimizers and the industrial matlab user. *Foundations of computer-aided process operations*, pages 8–11, 2012.
- [7] L. Danon, J. Duch, A. Arenas, and A. Díaz-guilera. Comparing community structure identification. *J. Stat. Mech*, 2005.
- [8] L. Du, X. Li, and Y. Shen. Robust nonnegative matrix factorization via half-quadratic minimization. In *ICDM*, pages 201–210, 2012.
- [9] L. Du and Y.-D. Shen. Unsupervised feature selection with adaptive structure learning. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 209–218. ACM.
- [10] H. Eldardiry and J. Neville. An analysis of how ensembles of collective classifiers improve predictions in graphs. In *CIKM*, pages 225–234, 2012.
- [11] L. Getoor and B. Taskar. *Introduction to statistical relational learning*. MIT press, 2007.
- [12] S. Ghosh and et al. Understanding and combating link farming in the twitter social network. In *Proceedings of the 21st international conference on World Wide Web*, pages 61–70. ACM, 2012.
- [13] M. A. Hogg and D. I. Terry. Social identity and self-categorization processes in organizational contexts. *Academy of management review*, 25(1):121–140, 2000.
- [14] X. Hu, J. Tang, H. Gao, and H. Liu. Actnet: Active learning for networked texts in microblogging. In *SDM*, pages 306–314. Citeseer, 2013.
- [15] D. Jensen, J. Neville, and B. Gallagher. Why collective inference improves relational classification. In *KDD*, pages 593–598. ACM, 2004.
- [16] D. Kong, R. Fujimaki, J. Liu, F. Nie, and C. Ding. Exclusive feature learning on arbitrary structures via $l_{1,2}$ -norm. In *NIPS*, pages 1655–1663. 2014.
- [17] N. Lin. Social networks and status attainment. *Annual review of sociology*, pages 467–487, 1999.
- [18] J. Liu and J. Ye. Moreau-Yosida regularization for grouped tree structure learning. In *NIPS*, pages 1459–1467, 2010.
- [19] W. Liu, P. P. Pokharel, and J. C. Principe. Correntropy: properties and applications in non-gaussian signal processing. *Signal Processing, IEEE Transactions on*, 55(11):5286–5298, 2007.
- [20] Q. Lu and L. Getoor. Link-based classification. In *ICML*, volume 3, pages 496–503, 2003.
- [21] S. A. Macskassy and F. Provost. A simple relational classifier. In *MRDM-2003*, page 64. Citeseer.
- [22] S. A. Macskassy and F. Provost. Classification in networked data: A toolkit and a univariate case study. *JMLR*, 8:935–983, 2007.
- [23] L. Meier, S. Van De Geer, and P. Bühlmann. The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53–71, 2008.
- [24] J. Neville and D. Jensen. Leveraging relational autocorrelation with latent group models. In *4th International Workshop on Multi-relational mining*, pages 49–55. ACM, 2005.
- [25] M. Newman. Modularity and community structure in networks. *PNAS*, 103(23):8577–8582, 2006.
- [26] F. Nie, H. Wang, H. Huang, and C. Ding. Early active learning via robust representation and structured sparsity. In *Proceedings of IJCAI*, pages 1572–1578. AAAI Press, 2013.
- [27] L. Page and et al. The pagerank citation ranking: Bringing order to the web. 1999.
- [28] A. Rajadesingan, R. Zafarani, and H. Liu. Sarcasm detection on twitter: A behavioral modeling approach. In *WSDM 2015*. ACM, 2015.
- [29] D. Seung and L. Lee. Algorithms for non-negative matrix factorization. In *NIPS*, volume 13, pages 556–562, 2011.
- [30] J. Tang and H. Liu. Unsupervised feature selection for linked social media data. In *SIGKDD*, 2012.
- [31] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *KDD*, pages 807–816. ACM, 2009.
- [32] L. Tang and H. Liu. Relational learning via latent social dimensions. In *ACM International Conference on Knowledge Discovery and Data Mining*, 2009.
- [33] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005.
- [34] X. Wang, L. Tang, H. Gao, and H. Liu. Discovering overlapping groups in social media. In *ICDM*, 2010.
- [35] L. Wu, A. Chin, G. Xu, L. Du, X. Wang, K. Meng, Y. Guo, and Y. Zhou. Who will follow your shop? exploiting multiple information sources in finding followers. In *DASFAA*, pages 401–415, 2013.
- [36] L. Wu, L. Du, B. Liu, G. Xu, Y. Ge, Y. Fu, J. Li, Y. Zhou, and H. Xiong. Heterogeneous metric learning with content-based regularization for software artifact retrieval. In *ICDM*, pages 610–619, 2014.
- [37] G. Xu. *Social Media Mining and Social Network Analysis: Emerging Research: Emerging Research*. IGI Global, 2013.
- [38] Z. Xu, V. Tresp, S. Yu, and K. Yu. Nonparametric relational learning for social network analysis. In *KDD 2008 Workshop on Social Network Mining and Analysis*, 2008.
- [39] R. Zafarani, A. A. Mohammad, and H. Liu. *Social Media Mining: An Introduction*. 2014.