Perceiving Group Themes from Collective Social and Behavioral Information

Peng Cui, Tianyang Zhang
Department of Computer Science
Tsinghua University, Beijing, China

Fei Wang
Department of Computer Science
University of Connecticut, USA

Peng He
Department of Social Network Operation
Tencent Technology, Shenzhen, China

Abstract

Collective social and behavioral information commonly exists in nature. There is a widespread intuitive sense that the characteristics of these social and behavioral information are to some extend related to the themes (or semantics) of the activities or targets. In this paper, we explicitly validate the interplay of collective social behavioral information and group themes using a large scale real dataset of online groups, and demonstrate the possibility of perceiving group themes from collective social and behavioral information. We propose a REgularized miXEd Regression (REXER) model based on matrix factorization to infer hierarchical semantics (including both group category and group labels) from collective social and behavioral information of group members. We extensively evaluate the proposed method in a large scale real online group dataset. For the prediction of group themes, the proposed REXER achieves satisfactory performances in various criterions. More specifically, we can predict the category of a group (among 6 categories) purely based on the collective social and behavioral information of the group with the Precision@1 to be 55.16% , without any assistance from group labels or conversation contents. We also show, perhaps counterintuitively, that the collective social and behavioral information is more reliable than the titles and labels of groups for inferring the group categories.

Introduction

Collective social and behavioral information commonly exists in nature, e.g. groups of birds, insects and fishes coordinate their moves and speeds so that they can move together. Humans, as a typical social species, show a variety of social and behavioral information. The synchronization of applause in concerts, the formation of consensus in social groups, the common social attributes of the audience of a certain movie are easily recognizable examples. There is a widespread intuitive sense that the characteristics of these social and behavioral information are to some extend related to the themes (or semantics) of the activities or targets. For example, the style of applause synchronization in piano concerts should be quite different from that in rock concerts. But it has been difficult to evaluate this question quantitatively since it requires a setting where various social and behavioral information happen in a shared environment. In recent years, social networks, especially the online social groups, digitally and continuously record the collective behaviors and social profiles of users, as well as the themes of social groups in an unprecedented level. This provides precious opportunities to investigate the link between the collective social and behavioral information of group members and the group themes. The interplay between semantics and human behaviors is of significant interest to both academia to understand human collective behaviors and industry to serve customers more intelligently.

Social groups and the collective behaviors of their members have been studied in sociology for many years (Shaw 1971)(Barsade 2002), and most of them focus on the offline social groups and collective behaviors themselves and hardly investigate their relations with semantics. In recent years, it arouses considerable research interests in computer science to study social groups and collective behaviors in a large scale and with finer resolution. (Backstrom et al. 2006)(Kairam, Wang, and Leskovec 2012) studied the dynamic mechanisms of online groups, including group formation, growth, evolution and demise. In (Danescu-Niculescu-Mizil et al. 2013), the effects of users' susceptibility to linguistic change on their lifecycles in groups was investigated. (Qiu, Zhu, and Jiang 2013) exploited user behaviors and group information to regularize the topic categorization of online documents. Also, (Sachan et al. 2012) used content (semantics) and user interaction behaviors to assist community (group) discovery in social networks. These works either study the dynamics of social group themselves, or using the social and behavior information in social groups to assist community detection or topic categorization. In contrast, the goal of this paper is to demonstrate the possibility of perceiving group themes directly from collective social and behavioral information.

Online social groups, which are established, owned and maintained by group owners, is a feature in many social network platforms to provide interest- and niche-specific networks within the larger and more diverse global social networks. Normally, a group owner should select a category from a predefined category list for the group under establishment, name it with a group title and label it with several words to represent the theme of the group. Then other users can apply to or freely join the group to create, post, comment or read contents in the group. The revealing of the

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

links between semantics and collective social and behavioral information can at least benefit the identification of group themes with the assistance of social and behavioral information. However, the group themes can be represented in different granularity levels. For example, the theme of the FIFA World Cup group can be represented by ”interest related - sports related - football related - FIFA World CUP”. In order to infer the rich hierarchical semantics from social behavioral information, it is required to study the learning model that can incorporate multiple types of relations, including semantic relations across different semantic levels and the relations between semantics and social behavioral information.

In this paper, we extract a set of effective social and behavioral features for group theme prediction. We further propose a REgularized miXEd Regression (REXER) method based on matrix factorization to predict group themes with hierarchical semantics (including both group categories and label sets) from collective social behavioral information of group members, where flexible regularizers are imposed to alleviate the problems brought by sparse and noisy data. We validate the discoveries and evaluate the prediction performances of REXER in a large scale real online group dataset, which is collected from an MSN-style instant message platform in China. We have in total 50,000 online groups with 5,549,570 group members, the social profiles of these group members and their behavior information for on month. The experimental results show that the proposed REXER achieves satisfactory performances in group theme prediction. More specifically, we can predict the category of a group (among 6 categories) purely based on the collective social and behavioral information of the group with the Precision@1 to be 55.16% , without any assistance from group labels or conversation contents. We also show, perhaps counterintuitively, that the collective social and behavioral information is more reliable than the titles and labels of groups for inferring the group categories.

Data and Features

In this section, we will introduce the characteristics of the dataset and introduce the social and behavioral features for group theme prediction.

Data Description

The online group dataset is collected from the real social network platform QQ, an MSN-style instant messenger in China with more than 500 million users. According to our statistics, there are around 100 million online groups that are generated and maintained by users. In this paper, we randomly select 50 thousand online groups with 5,549,570 unique anonymized users. For each user, we have their profile information (e.g. sexuality, age, geo-location etc.), and their behaviors (e.g. their participating behaviors in group conversations, and the timestamps of these behaviors) for one month. Besides, we know the friendship relations between group members. (Note that online groups are overlays on the global social network. Any pair of members in one group may or may not have friendship relation in the social network.) In total, we have 91,600,486 user participating behaviors, and 69,272,454 social relations.

In the dataset, there are 6 categories for these groups (as listed in Table 1), and each group belongs to one category. These categories can be used as the themes of groups. When a group is established, a title and a short description will be assigned to the group, from which we can extract labels to represent more detailed themes than categories. Thus, we use standard LDA (Latent Dirichlet Allocation) model to categorize them into 30 label sets after removing the high-frequency and low-frequency labels,, and use these label sets to represent the detailed theme of groups. For abbreviation, we only list the descriptions and representative labels for 14 label sets in Table 2, and the following visualizations are based on these 14 label sets. Note that most label sets are interpretable, except some ones mixing several topics together, e.g. the No. 12 label set.

No.	Categories
1	Friends
2	Housing
3	Game
4	People
5	Education
6	Industry

Table 1: Description of categories for online groups.

Social Features

We suppose that most groups in a given theme have similar distributions on some collective (i.e. group level) social features. Here we first extract the following collective social features.

- **Friendship Relational Density (FriDty)**. It is calculated by the ratio of the number of group member pairs that have friendship relation to the number of all possible group member pairs. Intuitively, a group with high friendship relational density is more relation driven rather than topic driven.
- **Sex Ratio (SexRto)**. It is calculated by the ratio of the number of male members to the total number of members.
- **Average Age (AvgAge)**. It is calculated in a normal way except that the outliers in user profiles are filtered, such as the age of 0 and 99 years.
- **Variance of Age (VarAge)**. Here we use standard deviation to represent the variance of age.
- **Geo-Affinity (GeoAff)**. We first find out the geographical area (province-level in our case) that include the largest number of group members, and calculate the ratio of the number of group members in that area to the total number of group members.

These features are in different scales. In order to make these feature values comparable, we normalize all these features into the rage of $[0, 1]$.

No.	Label Set Description	Representative Words
1	Mobile and Telecom Agent	Mobile, Recharge Card, Telecom, 3G, Mobile Phone
2	Business	Corporation, Taobao, Product, E-Commerce, Marketing
3	Online Game	Game, Community, Entertainment, Forum, Hero
4	Middle School	Class, Friendship, Middle School, Teacher, Love
5	University	University, Student Union, College, Sports, Community
6	Stock Market	Stock, Investment, Gold, Security, Buy and Sell
7	Organizations	Organization, Techniques, Business, Government Officer, Experience
8	Living Areas	House Owner, Photo, Living Area, Property Management, Group Buy
9	Blogs	Family, Blog, Accountant, Network, Software
10	House	Building, Living Area, Garden, Building Level, Buy House
11	Middle School Alumni	Middle School, Memory, Address List, Old School, Emotion
12	Mixed	Game, Classmate, Community, Engineering, Mechanic
13	Engineering	Engineering, Training, Cost of Manufacturing, Construction, Channel
14	Education	Children, Parents, Teacher, English, Training

Table 2: Description of label sets for online groups.

Behavioral Features

We suppose that most groups in a given theme have similar distributions on some collective behavioral features. Thus we first extract the following collective behavioral features.

- **Mobile Conversation Ratio (MobRto)**. It represents the proportion of conversation participating behaviors that are generated through mobile phones.
- **No-response Conversation Ratio (NoReRto)**. It represents the proportion of conversations that do not get any responses from other group members, where no-response conversation is determined by a waiting duration of three hours.
- **Night Conversation Ratio (NgtRto)**. It represents the proportion of conversations that are generated between 8pm to 11pm each day.
- **Degree of Conversational Dominance (DegDom)**. We first figure out three most active users that generate most conversations, and calculate the proportion of conversations that are generated by these 3 active users to represent the degree of conversational dominance.

We normalize all these features into the rage of $[0, 1]$.

Group Theme Prediction

In this section, we will present the method of perceiving group themes from social and behavioral information. First we introduce some symbols and notations that will be used through out the paper.

Notations and Problem Statement

As stated in the introduction, the problem we focus on is to infer the themes of online groups from the collective social and behavioral information of group members. In this paper, the group themes is represented in both category level (coarser semantic granularity) and label level (finer semantic granularity). Thus, the problem is how to infer the category and labels for a group given the features extracted from the social and behavioral information in the group.

Suppose we have N groups. Each group belongs to one of the M categories, and corresponds to some of the K labels. Then we have the Group-Category matrix $\mathbf{M} \in \mathbb{R}^{N\times M}$, and the Group-Label matrix $\mathbf{Y} \in \mathbb{R}^{N\times K}$. As the matrix $\mathbf{M}$ and $\mathbf{Y}$ are incomplete and noisy, we aim to complete them based on social and behavioral information. Here we represent each group by P-dimensional social and behavioral features by concatenating social and behavioral features into vectors, then we have the group feature matrix $\mathbf{G} \in \mathbb{R}^{N\times P}$. We assume these social and behavioral features are correlated with group labels, thus we use $\mathbf{A} \in \mathbb{R}^{P\times K}$ to denote the Feature-Label matrix. Meanwhile, as label is a more fine-grained way to represent group themes in contrast with group categories, we assume that there should be a mapping between labels and categories, and use $\mathbf{R} \in \mathbb{R}^{M\times K}$ to denote the Category-Label matrix.

In this way, the problem of perceiving group themes from collective social and behavioral information is transformed into the matrix completion problem on $\mathbf{Y}$ and $\mathbf{M}$.

Problem Formulation

As stated above, group themes can be represented by both categories and labels. We assume that the social and behavioral features are correlated with labels rather than categories. As the labels are in finer granularity than category, the correlation between social behavioral features and group categories can be transited by labels. Thus, given the observed Group-Label matrix $\mathbf{Y}$, one objective is to find an optimal Feature-Label matrix $\mathbf{A}$ to approximate $\mathbf{Y}$ by minimizing $||\mathbf{Y} - \mathbf{GA}||_F^2$. In order to leverage both labeled groups and unlabeled groups, we impose a constraint on the loss in a similar way as semi-supervised learning.

$$\mathcal{J}_1 = ||\mathbf{\Omega}(\mathbf{Y} - \mathbf{GA})||_F^2 \tag{1}$$

where $\mathbf{\Omega}$ is a diagonal matrix, such that

$$\Omega_{ii} = \begin{cases} 1, & \text{if the label of the } i\text{-th group is available} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

The second objective is to find an optimal Category-Label matrix $\mathbf{R}$ to approximate the Group-Category matrix $\mathbf{M}$ by minimizing

$$\mathcal{J}_2 = ||\mathbf{\Lambda}(\mathbf{M} - \mathbf{GAR}^T)||_F^2 \tag{3}$$

where $\mathbf{\Lambda}$ is a diagonal matrix, such that

$$\Lambda_{ii} = \begin{cases} 1, & \text{if the category of the } i\text{-th group is available} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

As $\mathbf{Y}$ and $\mathbf{M}$ are very sparse, the learned model is quite easy to be overfitting. In order to further improve the prediction performance, we further impose two regularizers on the main objectives under the following assumptions:

- Any two groups that have similar group members, which can be measured by the number of common members or the number of similar members in profiles, should have similar labels.
- The mapping between labels and categories should be sparse, i.e. each category should have only a small number of representative labels.

For the first assumption, we have the third objective:

$$\mathcal{J}_3 = ||\mathbf{W} - \mathbf{A}^T\mathbf{G}^T\mathbf{GA}||_F^2 \tag{5}$$

where $\mathbf{W} \in \mathbb{R}^{N\times N}$ is the group similarity matrix.

For the second assumption, we simply use $L-1$ norm to regularize the Category-Label mapping matrix $\mathbf{R}$.

$$\mathcal{J}_4 = ||\mathbf{R}||_1 \tag{6}$$

By combining $\mathcal{J}_1$ to $\mathcal{J}_4$ together, we propose the REgularized miXed Regression (REXER) model, and we can get $\mathbf{A}$ and $\mathbf{R}$ by minimizing:

$$\begin{aligned}\mathcal{J} = \quad & ||\mathbf{\Omega}(\mathbf{Y} - \mathbf{GA})||_F^2 + \lambda_1||\mathbf{\Lambda}(\mathbf{M} - \mathbf{GAR}^T)||_F^2 \\ & +\lambda_2||\mathbf{W} - \mathbf{A}^T\mathbf{G}^T\mathbf{GA}||_F^2 + \sigma||\mathbf{R}||_1\end{aligned} \tag{7}$$

Optimization Algorithm

Before deriving any details, we define the following matrices

$$\widetilde{\mathbf{Y}} = \mathbf{\Omega Y},\ \ \widetilde{\mathbf{G}} = \mathbf{\Omega G},\ \ \widehat{\mathbf{M}} = \mathbf{\Lambda M},\ \ \widehat{\mathbf{G}} = \mathbf{\Lambda G} \tag{8}$$

Then

$$\begin{aligned}\min_{\mathbf{A},\mathbf{R}} \quad & ||\widetilde{\mathbf{Y}} - \widetilde{\mathbf{G}}\mathbf{A}||_F^2 + \lambda_1||\widehat{\mathbf{M}} - \widehat{\mathbf{G}}\mathbf{AR}^T||_F^2 \\ & +\lambda_2||\mathbf{W} - \mathbf{GAA}^T\mathbf{G}^T||_F^2 + \sigma||\mathbf{R}||_1\end{aligned} \tag{9}$$

There are only two group of variables, $\mathbf{A}$ and $\mathbf{R}$. We adopt an *Block Coordinate Descent* (BCD) strategy to solve it. Each time we fix one group of variable and solve the other. When fixing $\mathbf{A}$, the problem of solving $\mathbf{R}$ is to minimize

$$\mathcal{J}_{\mathbf{R}} = \lambda_1||\widehat{\mathbf{M}} - \widehat{\mathbf{G}}\mathbf{AR}^T||_F^2 + \sigma||\mathbf{R}||_1 \tag{10}$$

which is a standard ℓ_1 norm regularized least squares problem and can be solved with any LASSO solver.

Fixing $\mathbf{R}$ solving $\mathbf{A}$ is much more complicated, we need to minimize

$$\mathcal{J}_{\mathbf{A}} = ||\widetilde{\mathbf{Y}} - \widetilde{\mathbf{G}}\mathbf{A}||_F^2 + \lambda_1||\widehat{\mathbf{M}} - \widehat{\mathbf{G}}\mathbf{AR}^T||_F^2 + \lambda_2||\mathbf{W} - \mathbf{A}^T\mathbf{K}_G\mathbf{A}||_F^2 \tag{11}$$

where the first term

$$\mathcal{J}_{\mathbf{A}}^1 = ||\widetilde{\mathbf{Y}} - \widetilde{\mathbf{G}}\mathbf{A}||_F^2 = tr(\widetilde{\mathbf{Y}}^T\widetilde{\mathbf{Y}}) - 2tr(\widetilde{\mathbf{Y}}^T\widetilde{\mathbf{G}}\mathbf{A}) + tr(\mathbf{A}^T\widetilde{\mathbf{K}}_G\mathbf{A}) \tag{12}$$

the second term

$$\begin{aligned}\mathcal{J}_{\mathbf{A}}^2 = ||\widehat{\mathbf{M}} - \widehat{\mathbf{G}}\mathbf{AR}^T||_F^2 \quad = \quad & tr(\widehat{\mathbf{M}}^T\widehat{\mathbf{M}}) - 2tr(\mathbf{R}\widehat{\mathbf{M}}^T\widehat{\mathbf{G}}\mathbf{A}) \\ & +tr(\mathbf{RA}^T\widehat{\mathbf{K}}_G\mathbf{AR}^T)\end{aligned} \tag{13}$$

the third term

$$\begin{aligned}&\mathcal{J}_{\mathbf{A}}^3 = ||\mathbf{W} - \mathbf{GAA}^T\mathbf{G}^T||_F^2 \\ &= tr(\mathbf{W}^T\mathbf{W}) - 2(\mathbf{A}^T\mathbf{G}^T\mathbf{WGA}) \\ &+tr(\mathbf{A}^T\mathbf{G}^T\mathbf{GAA}^T\mathbf{G}^T\mathbf{GA})\end{aligned} \tag{14}$$

The partial gradient of those three term with respect to $\mathbf{A}$ are

$$\nabla_{\mathbf{A}}\mathcal{J}_{\mathbf{A}}^1 = -2\widetilde{\mathbf{G}}^T\widetilde{\mathbf{Y}} + 2\widetilde{\mathbf{K}}_G\mathbf{A} \tag{15}$$

$$\nabla_{\mathbf{A}}\mathcal{J}_{\mathbf{A}}^2 = -2\widehat{\mathbf{G}}^T\widehat{\mathbf{M}}\mathbf{R}^T + 2\widehat{\mathbf{K}}_G\mathbf{ARR}^T \tag{16}$$

$$\nabla_{\mathbf{A}}\mathcal{J}_{\mathbf{A}}^3 = -4tr(\mathbf{G}^T\mathbf{WGA}) + 4(\mathbf{K}_G\mathbf{AA}^T\mathbf{K}_G\mathbf{A}) \tag{17}$$

Therefore

$$\nabla_{\mathbf{A}}\mathcal{J}_A = \nabla_{\mathbf{A}}\mathcal{J}_{\mathbf{A}}^1 + \lambda_1\nabla_{\mathbf{A}}\mathcal{J}_{\mathbf{A}}^2 + \nabla_{\mathbf{A}}\mathcal{J}_{\mathbf{A}}^2 \tag{18}$$

Then we can adopt gradient descent to solve $\mathbf{A}$. The whole algorithm is summarized in Algorithm 1.

Algorithm 1 REgurlarized miXEd Regression (REXER)

```
Require: Tradeoff parameters $\lambda_1 > 0, \lambda_2 > 0, \sigma > 0$, Group-Category matrix $\mathbf{M}$, Group-Label matrix $\mathbf{Y}$, Group-Feature matrix $\mathbf{G}$, group similarity matrix $\mathbf{W}$, step size $0 < \alpha_{\mathbf{A}} < 1$.
Initialize Label-Feature matrix $\mathbf{A}^{(0)}$ and Category-Label matrix $\mathbf{R}^{(0)}$
Calculate the current value of $\mathcal{J}^{(0)} \leftarrow \mathcal{J}(\mathbf{A}^{(0)}, \mathbf{R}^{(0)})$ with Equation (7)
Initialize the iteration variable $k \leftarrow 0$
repeat
    $k \leftarrow k+1$
    Update $\mathbf{R}^{(k)}$ by solving $\mathcal{J}_{\mathbf{R}}(\mathbf{A}^{(k-1)})$ in Equation (10) with standard LASSO solver
    Initialize the iteration variable $l \leftarrow 0$
    Calculate the current value of $\mathcal{J}_{\mathbf{A}}^{(0)}(\mathbf{A}^{(k-1)}, \mathbf{R}^{(k)})$ using Equation (11)
    repeat
        $l \leftarrow l+1$
        Calculate $\nabla_{\mathbf{A}}\mathcal{J}_A^{(l-1)}$ using Equation 18.
        Update $\bar{\mathbf{A}}^{(l)} \leftarrow \mathbf{A}^{(l-1)} - \alpha_{\mathbf{A}}\nabla_{\mathbf{A}}\mathcal{J}_A^{(l-1)}$
        Calculate $\mathcal{J}_{\mathbf{A}}^l(\bar{\mathbf{A}}^{(l)}, \mathbf{R}^{(k)})$ using Equation (11)
    until $\mathcal{J}_{\mathbf{A}}^l(\bar{\mathbf{A}}^{(l)}, \mathbf{R}^{(k)})$ converged
    Update $\mathbf{A}^{(k)} \leftarrow \bar{\mathbf{A}}^{(l)}$
    Calculate $\mathcal{J}^{(k)} \leftarrow \mathcal{J}(\mathbf{A}^{(k)}, \mathbf{R}^{(k)})$
until $\mathcal{J}^{(k)}$ converged
Output: $\mathbf{A} = \mathbf{A}^{(k)}, \mathbf{R} = \mathbf{R}^{(k)}$
```

We analyze the model complexity for each BCD iteration, which includes two parts:

- **Fixing A, solving R**. This is standard LASSO problem and the complexity is $O(KNM \min(N, M))$
- **Fixing R, solving A**. This step require Gradient Descent (GD) iterations, each GD step evaluating the gradient takes $O(4PNK + NMK + P^2N + PN^2 + KM^2 + 4P^2K + PK^2)$ time, evaluating the objective function loss takes $O(K^2P)$ time. Out of the GD iterations, evaluating $\mathbf{K}_G$ takes $O(NP^2)$ time.

Experiments

In this section, we will present the empirical study results on the REXER method for group theme prediction.

Experimental Settings

Training and Testing. In this paper, the problem of group theme prediction is transformed into category prediction and label set prediction. In order to evaluate the prediction performance of REXER, we randomly select 20% groups and hide their corresponding category and label set information from $\mathbf{Y}$ and $\mathbf{M}$. After learning $\mathbf{A}$ and $\mathbf{R}$ from the remaining 80% entries, we reconstruct $\mathbf{Y}$ and $\mathbf{M}$ and calculate the loss on the hidden entries. For all the following experiments, we conduct 20-folds testing and report average results with standard deviation.

Groundtruth. As each group belongs to only one category, the groundtruth can be naturally derived from the data. For the label sets, as we derive them by applying LDA on the labels of groups, we directly use the topic relevance score of a group over label sets (i.e. categorized topics in LDA) as the groundtruth on label sets for the group. In order to filter the trivial topic distributions, we set a threshold on the topic relevance score, and the scores that are below the threshold will be set to 0.

No.	**RMSE**	**MAE**	**Rank**
1	0.295 ± 0.008	0.210 ± 0.005	1.864 ± 0.124
2	0.371 ± 0.005	0.276 ± 0.003	2.667 ± 0.221
3	0.280 ± 0.007	0.195 ± 0.0040	1.386 ± 0.090
4	0.331 ± 0.014	0.230 ± 0.010	2.0890 ± 0.170
5	0.384 ± 0.006	0.283 ± 0.004	2.619 ± 0.153
6	0.331 ± 0.010	0.244 ± 0.009	1.748 ± 0.145
All	0.327 ± 0.002	0.234 ± 0.002	1.969 ± 0.037

Table 3: Category prediction performances.

Evaluation Criteria. In the following experiments, we will use RMSE (Root Mean Square Error) and MAE (Mean Average Error) to calculate the reconstruction loss of $\mathbf{Y}$ and $\mathbf{M}$ on the testing entries, which evaluates the prediction performance in value aspect. We will also evaluate the ranking performance by using Average Rank. Finally, we calculate the Precision@K to evaluate the prediction accuracy on the top recommended options, which is important in applications such as group recommendation and search.

Parameter Setting. In the proposed REXER method, we have 3 parameters in total, including λ_1, λ_2 and σ. For the parameter setting, we use grid search to get the optimal parameters $\lambda_1 = 1.3, \lambda_2 = 0.2, \sigma = 0.13$.

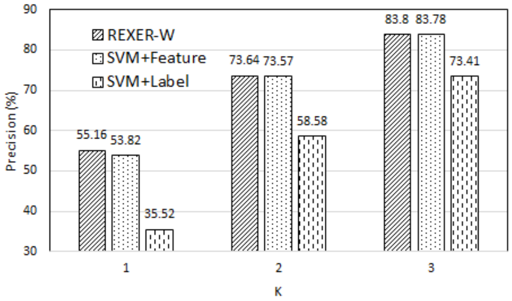

Figure 1: Precision@K for category prediction from REXER-W, SVM+Label, and SVM+Feature.

Predicting Group Category

We first evaluate the performance of REXER in category prediction. The experimental results are shown in Table 3. It can be seen that REXER can produce good results in both value approximation and ranking aspect. A perfect prediction will result in the average rank to be 1, our result 1.97 is satisfactory, especially considering that the predictions are purely based on social and behavioral information without any assistance from the labels and conversational contents of groups. Among the 6 categories, better predictions can be achieved for categories on Friends, Game and Industry. Comparatively, categories on Housing and Education are more challenging to predict.

No.	**RMSE**	**MAE**	**Rank**
All	0.198 ± 0.002	0.099 ± 0.003	12.8 ± 0.4

Table 4: Label set prediction performances.

In order to demonstrate the predictive power of collective social and behavioral information as well as the advantage of REXER model, we conduct comparative study by using SVM, an standard and effective multi-class classification tool, as the baseline method. Here we use SVM in two ways. First, we represent a group with its labels' distribution on the label sets which is derived from LDA, and denote it as SVM+Label. Also, we represent a group with its collective social and behavioral features, and denote it as SVM+Feature. As the group similarity information implied in matrix $\mathbf{W}$ cannot be straightforwardly incorporated into SVM model, in order to make the comparison fair, we remove the $\mathbf{W}$ from REXER to form REXER-W that is purely based on social and behavioral features. Then we calculate the Precision@K for REXER-W, SVM+Feature and SVM+Label methods, and the experimental results are shown in Figure 1. It is obvious that SVM+Feature can achieve much higher precision than SVM+Label. We attribute the weak predicting power of group labels to their deficiencies of noisy and free-style, especially in the social network environment. Comparatively, collective social and behavioral features show robust and good performance on category prediction because of the commonly existed and distinguishable collective social and behavioral synchronies among groups of different themes. Also, REXER-W achieves

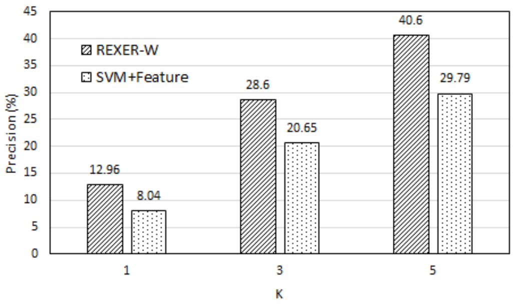

Figure 2: Precision@K for Label set prediction from REXER-W and SVM+Feature.

the best performances among the three models, especially in Precision@1.

Predicting Group Labels

We then evaluate the performance of REXER on label set prediction. The experimental results are shown in Table 4. Note that the overall results are averaged from 30 label sets. It can be seen that REXER can produce good results on RMSE and MAE, but the ranking results are not very satisfactory. According to our observation, the potential cause of this is due to the severe unbalance between label sets. For example, the label set No. 13 is a mixed label set and does not show very distinguishable social and behavioral information, but it includes 17% of all groups, thus the prediction results of groups are prone to be this label set, which would cause a better average rank in this label set while sacrificing other label sets.

The Precision@K for label set prediction is also shown in Figure 2. Here we compare REXER-W with SVM+Feature, and the results demonstrate that the proposed REXER-W model can achieve much higher precision than SVM+Feature in all cases, which demonstrates the advantages of REXER in predicting fine-granularity group themes. We attribute this significant improvement of REXER to the consideration of the duality between category and label sets. The group categories impose constraints on label sets by $\mathbf{R}$, and the label sets impose constraints on social and behavioral features by $\mathbf{A}$ so that the generation process of the observed data can be better approximated.

In order to look into the label set prediction, we plot the prediction confusion matrix in Figure 3. Here the confusion matrix is generated in the following way. Suppose we have a confusion matrix $\widehat{\mathbf{C}}$, where $\widehat{\mathbf{C}}_{i,j}$ means the number of groups that are with groudtruth label set i and are correctly (if $i = j$) or erroneously (if $i \neq j$) classified into label set j. For a predicted group, if the groundtruth labelset k is among the top 3 predicted label sets l, m, n, then $\widehat{\mathbf{C}}_{k,k} + 1$. Else, $\widehat{\mathbf{C}}_{k,l} + 1$, $\widehat{\mathbf{C}}_{k,m} + 1$, and $\widehat{\mathbf{C}}_{k,n} + 1$. Finally, we normalize $\widehat{\mathbf{C}}$ by columns into $\mathbf{C}$ so that the sum of each column in $\mathbf{C}$ is equal to 1. Thus, $\mathbf{C}_{i,j}, (i \neq j)$ means the probability of a group with label set i being misclassified into label set j. From diagonal elements in Figure 3, we can see that the group label sets are well predicted in most cases. Of course there are some obvious confusion pairs. For example, $\mathbf{C}_{3,6}$ indicates that groups on Online Game are easy to be misclassified into Stock Market, as they share several common social and behavioral synchronies, such as low friendship density, and high degree of conversation dominance. The groups on Middle School are easy to be misclassified into University, Organizations and School Alumni (indicated by $\mathbf{C}_{4,5}$, $\mathbf{C}_{4,7}$, and $\mathbf{C}_{4,11}$).

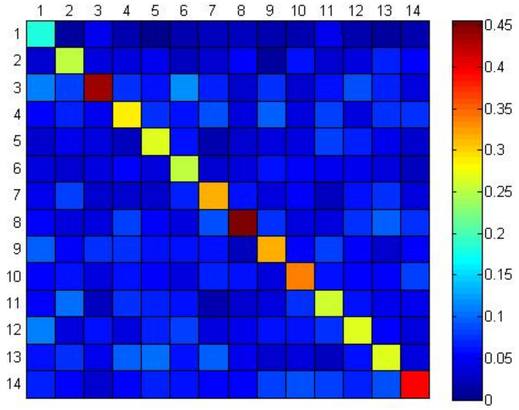

Figure 3: The confusion matrix of label set prediction.

Conclusion

In this paper, we aim to understand the interplay between semantics and collective social and behavioral information, and explore the possibility of perceiving semantics from the content-agnostic social and behavioral information. By studying a large scale real online group dataset, we discover the commonly existed social and behavioral information among groups in each theme and further propose the REXER method to predict hierarchical group themes, including both group categories and label sets, from collective social and behavioral information. We extensively evaluate the proposed REXER and achieve satisfactory prediction performances on both category prediction and label set prediction.

Acknowledgement

This work was supported by National Program on Key Basic Research Project, No. 2015CB352300, No. 2011CB302206; National Natural Science Foundation of China, No. 61370022, No. 61472444 and No. 61210008; International Science and Technology Cooperation Program of China, No. 2013DFG12870. Thanks for the support of NExT Research Center funded by MDA, Singapore, under the research grant, WBS:R-252-300-001-490 and the research fund of Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology.

References

Backstrom, L.; Huttenlocher, D.; Kleinberg, J.; and Lan, X. 2006. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM*

SIGKDD international conference on Knowledge discovery and data mining, 44–54. ACM.

Barsade, S. G. 2002. The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly* 47(4):644–675.

Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the 22nd international conference on World Wide Web*, 307–318. International World Wide Web Conferences Steering Committee.

Kairam, S. R.; Wang, D. J.; and Leskovec, J. 2012. The life and death of online groups: Predicting group growth and longevity. In *Proceedings of the fifth ACM international conference on Web search and data mining*, 673–682. ACM.

Qiu, M.; Zhu, F.; and Jiang, J. 2013. It is not just what we say, but how we say them: Lda-based behavior-topic model. In *2013 SIAM International Conference on Data Mining (SDM'13)*. SIAM.

Sachan, M.; Contractor, D.; Faruquie, T. A.; and Subramaniam, L. V. 2012. Using content and interactions for discovering communities in social networks. In *Proceedings of the 21st international conference on World Wide Web*, 331–340. ACM.

Shaw, M. E. 1971. Group dynamics: The psychology of small group behavior.