1-2015

Community discovery from social media by low-rank matrix recovery

Jinfeng ZHUANG
Singapore Management University

Mei TAO
Microsoft Research

Steven C. H. HOI
Singapore Management University, CHHOI@smu.edu.sg

Xian-Sheng HUA
Microsoft Research

Yongdong ZHANG
Chinese Academy of Sciences

DOI: https://doi.org/10.1145/2668110

Published in ACM Transactions on Intelligent Systems and Technology, January 2015, Volume 5 Issue 4, Article No. 67
http://dx.doi.org/10.1145/2668110

Community Discovery from Social Media by Low-Rank Matrix Recovery

Jinfeng Zhuang, Nanyang Technological University
Tao Mei, Microsoft Research
Steven C. H. Hoi, Nanyang Technological University
Xian-Sheng Hua, Microsoft Research
Yongdong Zhang, Institute of Computing Technology, Chinese Academy of Sciences

The pervasive usage and reach of social media have attracted a surge of attention in the multimedia research community. Community discovery from social media has therefore become an important yet challenging issue. However, due to the subjective generating process, the explicitly observed communities (e.g., group-user and user-user relationship) are often noisy and incomplete in nature. This paper presents a novel approach to discovering communities from social media, including the group membership and user friend structure, by exploring a low-rank matrix recovery technique. In particular, we take Flickr as one exemplary social media platform. We first model the observed indicator matrix of the Flickr community as a summation of a low-rank *true* matrix and a sparse *error* matrix. We then formulate an optimization problem by regularizing the *true* matrix to coincide with the available rich context and content (i.e., photos and their associated tags). An iterative algorithm is developed to recover the *true* community indicator matrix. The proposed approach leads to a variety of social applications, including community visualization, interest group refinement, friend suggestion, and influential user identification. The evaluations on a large-scale testbed, consisting of 4,919 Flickr users, 1,467 interest groups, and over five million photos, show that our approach opens a new yet effective perspective to solve social network problems with sparse learning technique. Despite being focused on Flickr, our technique can be applied in any other social media community.

Categories and Subject Descriptors: H.4.m [**Information Systems Applications**] Miscellaneous

General Terms: Algorithms, Experimentation, Performance.

Additional Key Words and Phrases: Social networks, community discovery, low-rank matrix, social media, context information.

1. INTRODUCTION

Recent years have witnessed an explosive growth of digital photos shared in a variety of emerging social media sites, such as *Flickr*, *Picasa*, *Facebook*, and so on. Images shared on such web sites are *social* in nature as they often contain rich context information besides visual images [Mei et al. 2014]. In Flickr, each photo has an owner who uploaded the photo. Every shared photo can be annotated with meaningful textual tags by the owner/other users. Besides the photos themselves, Flickr users

J. Zhuang and S. Hoi are with the Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798 (email: perfectzjf@gmail.com, chhoi@ntu.edu.sg). T. Mei is with Microsoft Research, Beijing, China. X.-S. Hua is with Microsoft Research, USA (email: {tmei, xshua}@microsoft.com). Y. Zhang is with the Institute of Computing Technology, Chinese Academy of Sciences, China (email: zhyd@ict.ac.cn).

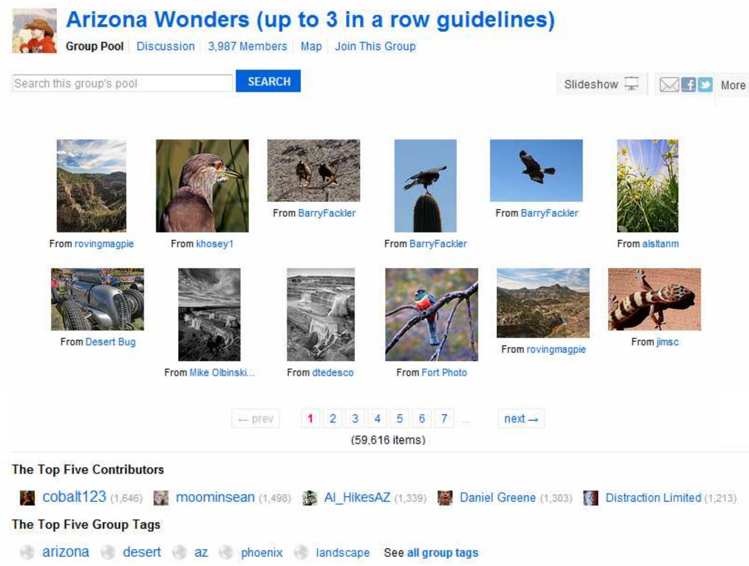


Fig. 1. A Snapshot of Flickr interest group "Arizona Wonders" which contains 3,987 members and 59,616 photos. Photos within this group tend to be related to Arizona in one way or another.

also interact closely. In particular, each Flickr user is allowed to create or join an interest group that is formed by a collection of users together with their shared photos. Each user is also allowed to add others into his/her friend list to facilitate social interactions. Thus, Flickr essentially forms a special structure of social networks. Figure 1 gives an example to illustrate a Flickr interest group.

According to a recent study, there are over five billion photos shared in Flickr and about 3,000 photos uploaded per minute [1]. It is imperative to study how to organize such a huge volume of photos to facilitate browsing and search. To this end, Flickr provides the functionality of *interest group* which contains photos of similar themes. People can create or join groups freely according to their interest. Moreover, a user can add others as friends if they share similar interest. This has improved social experience significantly.

However, due to the subjective generating process of groups, the group membership is noisy and incomplete in nature. For example, a user may happen to join a group which does not match his/her interest well because he/she did not browse the pool or the tags of the group carefully. Another scenario is a user cannot browse all the related groups due to the large amount of groups. Similar problems exist in the generating process of user friend list. In this paper, we study the problem of community discovery from social media. By "community," we refer to the group of interacting people sharing some common interest. Specifically, we analyze *user-oriented* relationships, i.e., the *group-user membership* and *user-user friendship* (not necessarily true friends in real life), taking Flickr as an exemplary social media site.

Researchers have proposed techniques to analyze interest groups [Kennedy et al. 2007; Negoescu et al. 2009; Negoescu and Gatica-Perez 2008; Yu et al. 2009; Yao et al. 2013]. However, to the best of our knowledge, little attention has focused on refining the known user-oriented structures. Most existing works usually treat the user-group and user-user relationship as ground truth to conduct further

[1] http://blog.flickr.net/en/2010/09/19/5000000000/

or deeper analysis. Unlike existing studies, we present a framework to refine the user-provided community information. We recommend groups and latent friends to users by analyzing the rich context information available in social media communities. In Figure 1, the context information includes textual tags provided by users, while the content information refers to the uploaded photos. We conjecture such data provides clues to social networks mining.

In particular, we develop an effective solution by employing a state-of-the-art low-rank matrix recovery technique [Wright et al. 2009], which has been successfully applied to some real applications in computer vision and multimedia [Wang et al. 2009; Zhu et al. 2010]. Specifically, we assume the observed user-group membership matrix is a summation of a *low-rank* true matrix and a *sparse* error matrix. By enforcing the true matrix coincide with the content and context Flickr data, we solve it with proximal gradient descent in an iterative manner. Our contributions can be summarized as follows:

—We propose to discover communities from user-contributed photos, which relies in the interplay between recommendation system and social network mining. To our knowledge, this is the first study to address this particular problem.

—We propose a general community discovery approach to robustly recover group-user and user-user relationships by exploring a powerful low-rank matrix recovery technique to make use of the rich context information available in social media sites. Specifically, we investigate how visual and textual information can be used for community discovery.

—We suggest several important applications based on our techniques, including interest group recommendation, friend suggestion, visualization, and influential user identification.

—We conducted extensive experiments on a large-scale real dataset from Flickr, and presented insights about Flickr community.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 gives formal definitions of social community. Section 4 presents our community discovery approach. Section 5 presents applications. Section 6 shows evaluations, followed by conclusions in Section 7.

2. RELATED WORK

In this section, we briefly review some related work in several domains, including social community discovery, recommendation systems, and Flickr data mining.

2.1 Social Community Discovery

Our target is to find out the Flickr interest group and latent friendship between users. Naturally, it falls into the category of traditional works on social community discovery or social relationship mining [Adamic and Adar 2001; Backstrom et al. 2006; Choudhury et al. 2010; Crandall et al. 2008; Roy et al. 2013; Singla and Richardson 2008; Yang et al. 2009; Zhuang et al. 2011]. In [Backstrom et al. 2006], the basic problems of social group formation, including membership, growth, and evolution are discussed. When it comes to our case, we have the initial Flickr group membership. Instead of discovering unknown groups, we predict the most appropriate groups for each Flickr user. The group formation and evolution problem is not our focus here. On the other hand, due to the special characteristic of Flickr data, the technical perspective could be completely different.

The other task of this work is to recommend latent friends for each user. This connects to social strength modeling or link predictions [Choudhury et al. 2010; Gilbert and Karahalios 2009; Leskovec et al. 2010; Taskar et al. 2003]. However, few existing works make use of the rich content+context data of Flickr images to infer the social relationship between users. Moreover, our motivation here is to help users find out who exhibits similar interest in photo sharing. Thus our technique should be content driven, instead of mutual communication-based approaches by traditional works.

2.2 Recommendation Systems

Our goal is to recommend interest groups and friends to each user. Apparently it is a kind of recommendation system [Guy et al. 2010; Konstas et al. 2009; Mei et al. 2011]. In general, a recommendation system can be categorized into three classes: content-based, collaborative, and hybrid approaches. The approach proposed in this paper falls into the third category, i.e., we consider not only the content of a candidate group or user, but also the mutual affinity between users and groups. With the fast growth of social medias, social recommendation systems have come to the proscenium of social network research. Some typical problems are discussed in [King et al. 2010]. Regarding our work here, the terms to be recommended are interest groups and registered users, which are more complex and difficult to deal with than conventional terms. No off-the-shelf recommendation systems can address the task in this paper.

2.3 Data Mining with Flickr Data

There exists rich research on mining Flickr data. In [Negoescu and Gatica-Perez 2008], the author conducted probabilistic latent semantic analysis (pLSA) on the Flickr interest groups, where each group is abstracted to be a collection of tags annotated to the images belonging to this group. After learning the latent topic, the tag based search and group content analysis can be aided. This work is group-oriented and it is unclear how strong is the relationship between groups and users. Due to the subjective and noisy process in group generating and expanding, Negoescu *et al.* proposed methods to hierarchically organize groups, and model users and groups equally [Negoescu et al. 2009]. These works are also *group-oriented*.

There are some works that focused on how to make use of Flickr metadata to facilitate other applications. For example, geo-tag analysis [Kennedy et al. 2007] [Serdyukov et al. 2009], automatic tag annotation [Wu et al. 2009], concept/tag modeling [Wu et al. 2008]. These techniques are related to our work as they analyze both Flickr images and their annotated tags. However, all these studies are *image-orientated*, which cannot directly solve our problem.

The idea of using Flickr data to infer social relationships has been recently proposed [Singla et al. 2008; Wu and Tretter 2009; Yu et al. 2009]. Wu *et al.* [Wu and Tretter 2009] tried to reveal the closeness of people by face detection techniques. However, this work is severely limited by the range of useable images. Singla *et al.* identified the social relationships between individuals in consumer photos with the principled Markov logic networks [Singla et al. 2008]. Due to the diversity of the images in Flickr, this rule based method is not applicable to our task here. Yu *et al.* intended to recommend a user's images to a known interest group based on supervised classification where the initial group membership is deemed as label information [Yu et al. 2009]. As we have mentioned, the initial group membership is subjective, incomplete, and noisy, which may not be reliable enough to serve as supervised information.

3. PROBLEM DEFINITION

This section presents the problem definition. Table 1 lists the key notations used in the paper. Taking Flickr as our exemplary social media platform in this paper (due to the comprehensive public APIs provided by Flickr), the first basic entity in Flickr community is *Flickr user*.

Definition 3.1 (*Flickr user*). A *Flickr user* U is a tuple $\mathbf{u} := [\mathcal{Y}, \mathcal{M}, \mathcal{N}]$, where $\mathcal{M} = \{\mathbf{g}_i : \mathbf{g}_i \in \mathcal{G}\}$ is a collection of groups that user $\mathbf{u}$ belongs to , $\mathcal{N} = \{\mathbf{v}_i : \mathbf{v}_i \in \mathcal{U}\}$ is a collection of users appearing in the friend list of user $\mathbf{u}$, and $\mathcal{Y} = \{[\mathbf{x}_i, \mathbf{t}_i] : i \in \mathbb{N}_z\}$ is a collection of Flickr photos uploaded by the user $\mathbf{u}$. Here we assume each single Flickr photo $y_i := [\mathbf{x}_i, \mathbf{t}_i]$ has a two-view representation, i.e., the visual photo $\mathbf{x}_i$ and its associated textual tags $\mathbf{t}_i$.

A set of Flickr users can form *an interest group*, which is defined below.

Table I. List of key notations.

Symbol	Description
$\mathbb{N}_u$	$\{1, \ldots, u\}$, a set of integers up to u
N_u	$\lvert\mathbb{N}_u\rvert$, the cardinality of $\mathbb{N}_u$
$\mathbf{u}$	a Flickr user
$\mathcal{U}$	$\{\mathbf{u}_i : i \in \mathbb{N}_u\}$, a collection of N_u users
$\mathbf{g}$	a Flickr interest group
$\mathcal{G}$	$\{g_j : j \in \mathbb{N}_g\}$, a collection of N_g groups
$\mathbf{M}$	$\in \mathbb{R}^{N_u \times N_g}$, observed group indicator matrix
$\mathbf{T}$	$\in \mathbb{R}^{N_u \times N_g}$, true group membership matrix
$\mathbf{E}$	$\in \mathbb{R}^{N_u \times N_g}$, sparse random error matrix

Definition 3.2 (*Flickr interest group*). A *Flickr interest group* $G = \{\mathbf{u}_i\}$ consists of a set of Flickr users who share common interests of photo sharing.

We can represent the group-user membership by a matrix $\mathbf{M} \in \{0,1\}^{N_u \times N_g}$, where the (i, j)-th entry indicates whether user $\mathbf{u}_i$ belongs to group $\mathbf{g}_j$. Similarly, we can also represent the user-user friend relationship by a matrix $\mathbf{N} \in \{0,1\}^{N_u \times N_u}$. Based on the defined notions of Flickr data, given a collection of Flickr users $\mathcal{U}$ and groups $\mathcal{G}$, our goals of Flickr community discovery are two-fold: (1) Flickr group discovery, and (2) latent friend discovery, which are formally defined below, respectively.

Definition 3.3 (*Flickr Group Discovery*). Given a collection of Flickr users $\mathcal{U}$ and predefined Flickr interest groups $\mathcal{G}$, the task of *Flickr group discovery* is to construct a matrix $\mathbf{T}^g \in \mathbb{R}^{N_u \times N_g}$ such that the (i, j)-th entry T^g_{ij} measures the confidence that the user $\mathbf{u}_i$ belongs to group $\mathbf{g}_j$.

Definition 3.4 (*Latent Friend Discovery*). Given a collection of Flickr users $\mathcal{U}$, the task of *latent friend discovery* is to learn a matrix $\mathbf{T}^u \in \mathbb{R}^{N_u \times N_u}$ such that the (i, j)-th entry T^u_{ij} measures the potential of $\mathbf{u}_i$ adding $\mathbf{u}_j$ into his/her friend list.

4. COMMUNITY DISCOVERY

In this section, we present the approach to the community discovery. In particular, we begin with the interest group recommendation. However, the proposed approach can be easily adapted to recommend similar users (i.e., users with common interests), which will be discussed in Section 5.

4.1 Problem Formulation

The goal of the community discovery task is to recover the true group membership matrix $\mathbf{T}$ from the observed group membership matrix M. We notice there are three key observations that motivate the solution of our scheme.

—We notice that the observed group membership M is **noisy** and **incomplete**. This is because Flickr interest groups are usually created by users in a fairly unrestricted manner. A user may randomly browse the list of interest groups and join some of them. On one hand, some of the joined groups may not be really relevant to the user's interests; and on the other hand, many other potentially relevant groups are likely unexplored.

—The true group membership matrix $\mathbf{T}$ is **low-rank**. This is because many popular groups in Flickr are usually semantically related in terms of their shared contents. For example, when you search "Beijing" in Flickr, the results come from a list of groups, including "Beijing Photo Community," "Beijing," "Walking in Beijing," etc. Such interest groups are highly overlapped for their common image contents. There is no essential difference between the themes of these groups.

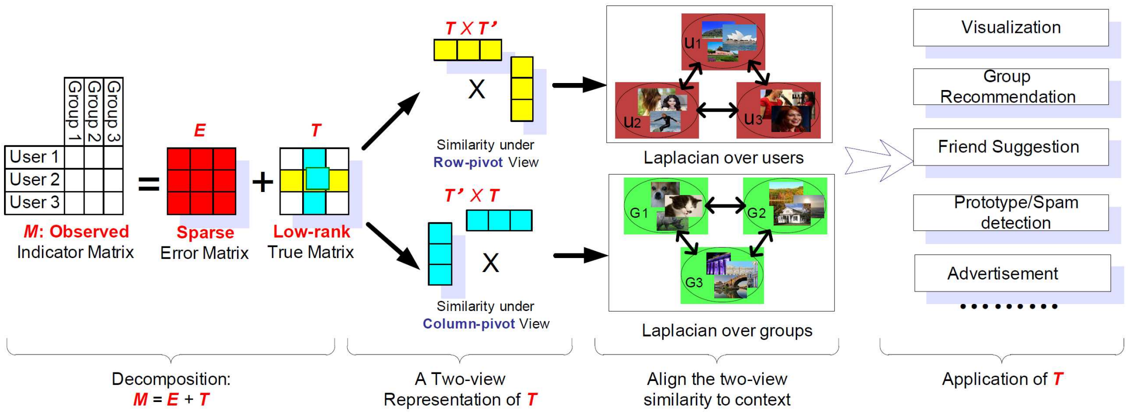


Fig. 2. The proposed approach for community discovery via low-rank matrix recovery and its applications.

—The error matrix $\mathbf{E}$ is **sparse**. We assume only a small fraction of $\mathcal{T}$ is corrupted due to the randomness in group formation and growth.

Motivated by the above observations, we propose to model the relationship between the observed membership matrix M and the true group membership matrix $\mathbf{T}$ as follows:

$$\mathbf{M} = \mathbf{T} + \mathbf{E}, \tag{1}$$

where $\mathbf{E}$ is an unknown sparse matrix of random noise, and matrix $\mathbf{T}$ is assumed to be low-rank. As a result, we suggest an interest group refinement scheme by formulating it as an optimization problem of low-rank matrix recovery from noisy and incomplete observed data M:

$$\min_{\mathbf{T},\mathbf{E}} \quad \text{rank}(\mathbf{T}) + \gamma\|\mathbf{E}\|_0, \;\; \text{s.t.} \;\; \mathbf{M} = \mathbf{T} + \mathbf{E}, \tag{2}$$

where $\|\cdot\|_0$ is the zero-norm, which counts the number of non-zero entries of a matrix, and γ is a tradeoff parameter. In this objective function, we aim to optimize $\mathbf{T}$ by minimizing both the rank of matrix $\mathbf{T}$ and the sparsity of matrix $\mathbf{E}$, which reflect the previous two important facts, respectively.

In general, the above optimization problem is non-continuous, highly non-convex, and hard to solve. Fortunately, under certain conditions (please refer to [Wright et al. 2009]), one can show the above optimization problem can be well-approximated as:

$$\min_{\mathbf{T},\mathbf{E}} \quad \|\mathbf{T}\|_* + \gamma\|\mathbf{E}\|_1, \;\; \text{s.t.} \;\; \mathbf{M} = \mathbf{T} + \mathbf{E}, \tag{3}$$

where the nuclear norm $\|\mathbf{T}\|_*$ (the sum of its singular values) is adopted to approximate the rank, and l_1-norm $\|\mathbf{E}\|_1$ is used to approximate the zero-norm $\|\mathbf{E}\|_0$. We further restrict $\mathbf{T}$ consistent with the content and context information. The proposed approach is shown in Figure 2, which consists of three key components, i.e., decomposition of the observed matrix, a two-view representation of low-rank matrix $\mathbf{T}$, and the alignment of two-view similarity with context. The proposed approach leads to a wide variety of social applications, including community visualization, group recommendation, friend suggestion, and so on.

4.2 Content Consistency for Group Refinement

Besides the low-rank consideration, another concern is that the true group membership matrix $\mathbf{T}$ should reflect the content consistency with respect to the rich information available in the Flickr groups and users. For example, any two users who have uploaded similar photos tend to share the similar groups. In the following, we discuss some formulations to reflect such consistency consideration.

On one hand, we notice that the i-th row of $\mathbf{T}$ essentially provides a high-level abstract of the Flickr user $\mathbf{u}_i$, i.e., the group assignment of $\mathbf{u}_i$ reflects the interest of $\mathbf{u}_i$. Therefore, the similarity among Flickr users induced by $\mathbf{T}$ should be consistent with the one computed from their uploaded photos. Thus, we propose to minimize the following distortion:

$$\frac{1}{2}\sum_{i,j\in\mathbb{N}_u} S_{ij}\left\|\frac{\mathbf{T}_i}{\sqrt{d_i}}-\frac{\mathbf{T}_j}{\sqrt{d_j}}\right\|_2^2 = \mathrm{tr}\ (\mathbf{T}^\top\mathbf{L}\mathbf{T}), \tag{4}$$

where S_{ij} denotes the similarity of two users computed from their uploaded contents, $d_i = \sum_{j\in\mathbb{N}_u} S_{ij}$ is used for normalization, $\mathbf{T}_i$ is the i-th row of $\mathbf{T}$, $\mathbf{L}$ is the normalized graph Laplacian defined as:

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2}, \tag{5}$$

where $\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots)$ is a diagonal matrix. $\mathbf{D}$ is of size $N_u \times N_u$, where N_u is the number of users. The graph Laplacian encodes the local geometry of the data, which has achieved success in semi-supervised learning [Zhu 2005].

On the other hand, the j-th column of $\mathbf{T}$ essentially represents the list of members of group $\mathbf{g}_j$, which should also express the theme of group $\mathbf{g}_j$ to some extent. Therefore, we can also construct the Laplacian matrix based on some similarity measure over the interest groups. The similarity induced from the columns of $\mathbf{T}$ should also be aligned to the Laplacian over groups as much as possible.

By integrating with the above consideration of content consistency, we have the following improved formulation:

$$\begin{aligned} &\min_{\mathbf{T},\mathbf{E}} \|\mathbf{T}\|_* + \gamma_1\|\mathbf{E}\|_1 + \gamma_2\Big(\mathrm{tr}\ \big(\mathbf{T}^\top\mathbf{L}^u\mathbf{T}\big) + \mathrm{tr}\ \big(\mathbf{T}\mathbf{L}^g\mathbf{T}^\top\big)\Big) \\ &\text{s.t.} \quad \mathbf{M} = \mathbf{T} + \mathbf{E}, \mathbf{T} \in \mathbb{R}^{N_u\times N_g}, \mathbf{E} \in \mathbb{R}^{N_u\times N_g}, \end{aligned} \tag{6}$$

where $\mathbf{L}^u$ and $\mathbf{L}^g$ are the graph Laplacian defined on $\mathcal{U}$ and $\mathcal{G}$, respectively. γ_1 and γ_2 are hyper-parameters controlling the trade-off.

4.3 Building the Graph Laplacian

We have two graph Laplacian matrices $\mathbf{L}^g$ and $\mathbf{L}^u$ defined on groups and users, respectively. They are important in rectifying the true group membership. In this section, we describe techniques for similarity measure towards the graph Laplacian by assuming a unified view of groups and users.

4.3.1 *A Unified View of Groups and Users.* In our approach, to simplify the computation, we assume a unified view of both Flickr interest groups and Flickr users, i.e., each of them consists of two components: a collection of photos in visual space $\mathcal{X}$ and an associated tag document in textual space $\mathcal{T}$. For visual feature representation, we adopt the bag-of-word (BoW) model. Specifically, we first extract the SIFT descriptors [Lowe 2004]. All these descriptors are split into d_x groups by a k-means clustering process. Given a photo, we quantize it by assigning each of its SIFT descriptors to a nearest cluster center. Then each photo is converted into a fixed length vector $\mathbf{x} \in \mathbb{R}^{d_x}$, where d_x is the visual vocabulary size.assigned to the i-th cluster.

Regarding the textual representation, we also adopt the BoW model. In particular, we collect all the tags and build a tag dictionary of size d_t. All the tags of an entity is converted into a fixed length vector in $\mathbb{R}^{d_t}$ by the traditional *tf-idf* weighing method. Here the inverse document frequency is the number of entities containing that tag. Next we discuss the approach to computing similarity matrix $\mathbf{S}$ for groups and users, which is the same in either visual space or textual space under the unified view. Once the similarity matrix is obtained, the graph Laplacian can be built by Eqn. (5).

Similarity in Visual Space. We measure image similarity on $\mathcal{X}$ by a Gaussian kernel $s(\mathbf{x}_i, \mathbf{x}_j) = e^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\sigma^2}$, where σ is a kernel parameter. For a specific user, we employ the centroid of its images to represent its visual contents, i.e., $\bar{\mathbf{u}} = \sum_{\mathbf{x}_i \in \mathbf{u}} \frac{\mathbf{x}_i}{|\mathbf{u}|}$, where $|\mathbf{u}|$ is the number of photos belonging to user $\mathbf{u}$. Thus, the similarity between two users $\mathbf{u}$ and $\mathbf{u}'$ can be given by $k^x(\mathbf{u}, \mathbf{u}') := e^{-\|\bar{\mathbf{u}} - \bar{\mathbf{u}}'\|^2/\sigma^2}$.

Similarity in Textual Space. Each photo is usually associated with a set of tags provided by Flickr users. Thus, we have a collection of tags corresponding to the images. For a Flickr user, we collect all of its tags to construct a single document as in [Negoescu and Gatica-Perez 2008]. Then, we compute the *tf-idf* weight for each tag. The textual representation is expressed as a fixed length vector $\mathbf{t} \in \mathbb{R}^{d_t}$, where d_t is the size of the tag vocabulary. We adopt the normalized linear kernel which is widely used for text classification:

$$k^t(\mathbf{u}, \mathbf{u}') = \frac{\sqrt{\sum_{i=1}^{d_t} t_i t_i'}}{\sqrt{\sum_{i=1}^{d_t} t_i^2}\sqrt{\sum_{i=1}^{d_t} t_i'}}. \tag{7}$$

With the above kernel functions in both visual space and textual space, we can compute the similarity between two entities by a simple linear combination:

$$k(\mathbf{u}, \mathbf{u}') = \alpha k^x(\mathbf{u}, \mathbf{u}') + (1 - \alpha) k^t(\mathbf{u}, \mathbf{u}'). \tag{8}$$

We can compute the similarity matrix $\mathbf{S}$ by employing the above kernel function. Furthermore, we find the n-nearest neighbor for each user to truncate $\mathbf{S}$: $S_{ij} = k(\mathbf{u}_i, \mathbf{u}_j)$ if $\mathbf{u}_i$ is among the n nearest neighbors of $\mathbf{u}_j$, or vice versa. Thus, we obtain the graph Laplacian by equation (5).

It is worth noting that there are other possible approaches to calculating the similarity in both spaces. Depending on how to model the visual/textual space, indeed there are a lot of candidate metrics to measure the similarity. For example, one can use global features like color histogram to represent image, then use inner product to compute similarity. One may use n-gram model (e.g., $n = 1, 2, 3 \ldots$) to represent tags, and use string kernels to compute textual similarity.

4.4 Problem Solution

In this section, we present an efficient algorithm to solve the optimization task in equation (6). Specifically, we convert the optimization problem (6) into a non-constrained optimization task, and solve it using the accelerated proximal gradient technique [Lin et al. 2009].

First of all, we consider an unconstrained convex problem:

$$\min_{x \in \mathcal{H}} J(x) := \mu g(x) + f(x), \tag{9}$$

where $\mathcal{H}$ is a real Hilbert space endowed with an inner product $\langle \cdot, \cdot \rangle$ and a corresponding norm $\| \cdot \|$. When $f(x)$ is Lipschitz continuous $\|\nabla f(x_1) - \nabla f(x_2)\| \leqslant L_f \|x_1 - x_2\|$ with the Lipschitz constant L_f, a proximal gradient algorithm minimizes a series of approximations of J, chosen in form of

$$\tilde{J}(x, y) := f(y) + \langle \nabla f(y), x - y \rangle + \frac{L_f}{2}\|x - y\|^2 + \mu g(x). \tag{10}$$

Instead of solving (9) directly, we iteratively solve (10) by

$$x_{t+1} = \arg\min_x \tilde{J}(x, y_t) \ ,$$

Algorithm 1 CALRMR for Flickr interest group refinement

Input: Initial interest group membership $\mathbf{M} \in \mathbb{R}^{N_u \times N_g}$, and the graph Laplacian $\mathbf{L}^g$ and $\mathbf{L}^u$;
Output: Refined membership matrix $\mathbf{T} \in \mathbb{R}^{N_u \times N_g}$ and error matrix $\mathbf{E} \in \mathbb{R}^{N_u \times N_g}$.

```
t = 0, T_0, T_{-1} = 0; E_0, E_{-1} = 0; b_0, b_{-1} = 1; μ̄ = δμ_0;
repeat
    Y_t^T = T_t + ((b_{t-1} - 1)/b_t)(T_t - T_{t-1});
    Y_t^E = E_t + ((b_{t-1} - 1)/b_t)(E_t - E_{t-1});
    P_t^T = Y_t^T - (1/L_f)[μγ_2(Y_t^T L^g + L^u Y_t^T) + Y_t^T + Y_t^E - M];
    T_{t+1} = U[Σ]_{μ_t/L_f} V^⊤, where (U, Σ, V) = SVD(P_t^T);
    E_{t+1} = [Y_t^E - (1/L_f)(Y_t^T + Y_t^E - M)]_{μ_t γ_1 / L_f};
    b_{t+1} = (1 + sqrt(4 b_t^2 + 1))/2, μ_{t+1} = max(ημ_t, μ̄);
    t = t + 1;
until convergence
```

$$t = 0, \mathbf{T}_0, \mathbf{T}_{-1} = \mathbf{0};\ \mathbf{E}_0, \mathbf{E}_{-1} = \mathbf{0};\ b_0, b_{-1} = 1;\ \bar{\mu} = \delta\mu_0;$$

repeat

$$\mathbf{Y}_t^T = \mathbf{T}_t + \tfrac{b_{t-1}-1}{b_t}(\mathbf{T}_t - \mathbf{T}_{t-1});$$
$$\mathbf{Y}_t^E = \mathbf{E}_t + \tfrac{b_{t-1}-1}{b_t}(\mathbf{E}_t - \mathbf{E}_{t-1});$$
$$\mathbf{P}_t^T = \mathbf{Y}_t^T - \tfrac{1}{L_f}\Big[\mu\gamma_2(\mathbf{Y}_t^T\mathbf{L}^g + \mathbf{L}^u\mathbf{Y}_t^T) + \mathbf{Y}_t^T + \mathbf{Y}_t^E - \mathbf{M}\Big];$$
$$\mathbf{T}_{t+1} = \mathbf{U}\big[\boldsymbol{\Sigma}\big]_{\frac{\mu_t}{L_f}}\mathbf{V}^\top, \text{ where } (\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}) = \mathrm{SVD}(\mathbf{P}_t^T);$$
$$\mathbf{E}_{t+1} = \big[\mathbf{Y}_t^E - \tfrac{1}{L_f}(\mathbf{Y}_t^T + \mathbf{Y}_t^E - \mathbf{M})\big]_{\frac{\mu_t\gamma_1}{L_f}};$$
$$b_{t+1} = \tfrac{1+\sqrt{4b_t^2+1}}{2},\ \mu_{t+1} = \max(\eta\mu_t, \bar{\mu});$$
$$t = t + 1;$$

until convergence

where $y_t = x_t + \frac{b_{t-1}-1}{b_t}(x_t - x_{t-1})$ for sequence $\{b_t\}$ satisfying $b_{t+1}^2 - b_{t+1} \leqslant b_t^2$. This iterative algorithm is able to achieve the convergence rate $O(t^{-2})$ [Lin et al. 2009]. Applying this technique we derive the Content-Aware Low-Rank Matrix Recovery (CALRMR) algorithm for social community discovery. We move the details to Appendix A.

5. APPLICATIONS

Our technique benefits a wide variety of potential applications, including interest group recommend, friend suggestion, and visualization, and so on.

5.1 Interest Group Recommendation

As aforementioned, the raw group membership of Flickr users is noisy and incomplete.To help users find potentially interested groups, we study the interest group recommendation problem.We notice that the taste/interest of a user in photos can be implicitly mined from the photos uploaded by the user. Thus, we develop an interest group recommendation scheme by applying the proposed interest group refinement solution. In particular, we consider noisy factors as the sparse error matrix $\mathbf{E}$ and enforce the global content consistency by aligning the true matrix with the graph Laplacian $\mathbf{L}$ when rectifying the group assignment.

Once the true group membership result $\mathbf{T}$ is found by applying the previous algorithm, for the i-th user $\mathbf{u}_i$, we can rank the groups according to the i-row of $\mathbf{T}$, and suggest them to the user. With such a list of top ranked groups, the user could find his/her interest groups quickly so as to facilitate the photo browsing and search tasks.

5.2 Friend Suggestion

Each Flickr user $\mathbf{u}_i$ has a friend list of other users, which can be “real” friends of $\mathbf{u}_i$, i.e., they know each other well in real life. It is also common that a user $\mathbf{u}_i$ adds an unknown user $\mathbf{u}_j$ into his/her friend list simply because he/she is interested in the photos shared by user $\mathbf{u}_j$. Akin to the case of group membership assignment, such kind of friendships are noisy and incomplete.

Table II. The statistics about our crawled Flick dataset *before* (left) and *after* (right) pre-processing.

#user	#group	#image	#tag	#friend	#user	#group	#image	#tag	#friend
4,919	109,205	5,001,601	116,372	81,447	4,919	1,467	5,001,601	13,360	81,447

With the similar motivations, the proposed group refinement framework (6) can be immediately adapted for the friend suggestion task. As a result, we can mine the *latent* friends by refining a user's friend list. Specifically, we aim to solve the optimization task

$$\begin{aligned} \min_{\mathbf{T},\mathbf{E}} \;\; & \|\mathbf{T}\|_* + \gamma_1\|\mathbf{E}\|_1 + \gamma_2 \mathbf{tr}\,(\mathbf{T}^\top \mathbf{L}^u \mathbf{T}) \\ \text{s.t.} \quad & \mathbf{N} = \mathbf{T} + \mathbf{E}, \quad \mathbf{T} \in \mathbb{R}^{N_u \times N_u}, \mathbf{E} \in \mathbb{R}^{N_u \times N_u}, \end{aligned} \tag{11}$$

where $\mathbf{N} \in \{0,1\}^{N_u \times N_u}$. The (i,j)-th entry of $\mathbf{N}$ indicates whether $\mathbf{u}_j$ appears in the friend list of $\mathbf{u}_i$. The potential friend matrix $\mathbf{T}$ becomes a square matrix, of which the entries reflect the confidence if two users are friends. This optimization can be solved by setting $\mathbf{L}^g = \mathbf{0}$ in Algorithm 1. With the learned $\mathbf{T}$, for a particular user $\mathbf{u}_i$, we rank the other users according the value of the i-th row of $\mathbf{T}$ and suggest the top ranked users to $\mathbf{u}_i$. This thus can help user $\mathbf{u}_i$ to find other potential users s/he may be interested in.

5.3 Community Visualization

A good visualization solution is essential to help discover and understand the community distribution between users and groups. Let us denote by $\mathbf{T}^g$ and $\mathbf{T}^u$ the solutions of $\mathbf{T}$ of (6) and (11), respectively by applying the above proposed technique. Accordingly we could 1) re-arrange the columns and rows of $\mathbf{T}^g$ to form block structures. So the clustering structure of interest groups can be visually inspected; 2) plot the link graph between users according to $\mathbf{T}^u$ such that we can visually find the important (the nodes with many edges) and inactive users (the nodes having few edges).

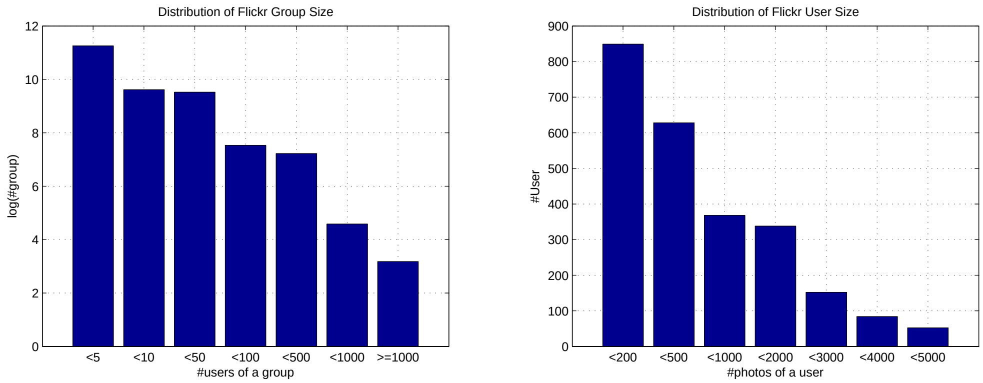
6. EXPERIMENTS

6.1 Experimental Settings

We built a large-scale testbed by crawling real data from Flickr. In our testbed, each user has at least 100 uploaded photos. We do not consider users of very few photos mainly because they are not easy to be evaluated. Moreover, users who have many uploaded photos are often more active. It would be more meaningful to focus on such active users.

Table 2 summarizes the statistics of our testbed, with 4,919 Flickr users and over 5 million images. In particular, the first five cells in the left of Table 2 show the statistics of the raw data without preprocess. We can observe that $\#group$ can be much larger than $\#user$ in the raw data. But the user size of each group varies much. Figure 3(a) shows the distribution of group's $\#user$ in the original testbed. It is clear to see that a large fraction of the groups have fewer than 10 users. Such small groups are usually inactive and not representative. We thus preprocessed the data by filtering out the groups with less than 100 users. The statistics of the data after preprocess are shown on the right of Table 2. Finally, we also show the distribution of user's $\#photo$ in Figure 3(b). On average, each user has about 1,000 photos.

Regarding the feature extraction, we form the collected tag dictionary by dropping the tags that appear in fewer than three users, and use the *tf-idf* [Sebastiani 2001] weighting scheme to represent the tag documents for both users and groups. For the photos, we first extract the SIFT descriptors [Lowe 2004]. After that we sample a subset of the SIFT descriptors and cluster them into 1,000 visual words. With the set of visual words, every image can be represented as a 1,000-dimensional bag-of-visual-word feature vector. There is a variety of similarity measures in visual space depending on the



(a) (b)

Fig. 3. The statistics about (a) the number of users in Flickr groups and (b) the number of photos from users in our dataset.

underlying representation. We evaluated 1) linear kernels, 2) Gaussian kernels, 3) binary k-nearest neighbors based on global features described in [Zhuang and Hoi 2011]. We have found that Gaussian-based on SIFT descriptor is slightly better. Due to capacity limitation, we do list the details here.

There are several hyper-parameters in our algorithm. In our approach, we set them as follows:

—σ: the bandwidth parameter of Gaussian kernel for computing image similarity, which is set to one divided by the average Euclidean distance over a set of randomly sampled 100K photos.

—α: the trade-off parameter in Eqn (8) to balance between visual and textual contents. A larger α emphasizes more on the visual images when computing the similarity between users. We set it to 0.6.

—n: the number of nearest neighbors when constructing graph Laplacian, we set it to 10 for both $\mathbf{L}^g$ and $\mathbf{L}^u$.

—γ_1 and γ_2: the parameters balancing the trade-off between sparse error and content consistency in Eqn (6). We set $\gamma_1 = 1$ and $\gamma_2 = 4$.

6.2 Evaluation of Group Refinement

For a user $\mathbf{u}_i$, assume that we have the ground truth of an optimal group assignment M_i^*. After we obtain the true group indicator matrix $\mathbf{T}$, we can sort $\mathbf{T}_i$ in decreasing order. Then, we extract the top-m ranked groups $\{\pi_1, \ldots, \pi_m\}$ for user $\mathbf{u}_i$, where π_i refers to the index of the i-th recommended group in M^*. We can compute the *precision* of this top-m ranked groups to measure the quality of $\mathbf{T}_i$:

$$precision(\{\pi_1, \ldots, \pi_m\}; \mathbf{M}_i^*) = \frac{\# M_{i\pi_j}^* = 1}{|\mathbf{M}_i^*|}, \quad j \in \mathbb{N}_m, \tag{12}$$

where $|\mathbf{M}_i^*|$ is the number of non-zeros in $\mathbf{M}_i^*$. Finally, we can compute the mean precision for a collection of test users. Please note that π_i is the index of recommended item. $\# M_{i\pi_j}^* = 1$ means we made a correct recommendation for item i. So the numerator is counting how many recommended item is correct according to the ground truth M^*. The denominator is how many items are associated with i. So this precision is always smaller than 1.

6.2.1 *Objective Evaluation.* We cover the observed group membership of a set of 100 randomly picked users, i.e., setting the corresponding 100 rows in the initial indicating matrix M to be zeros.

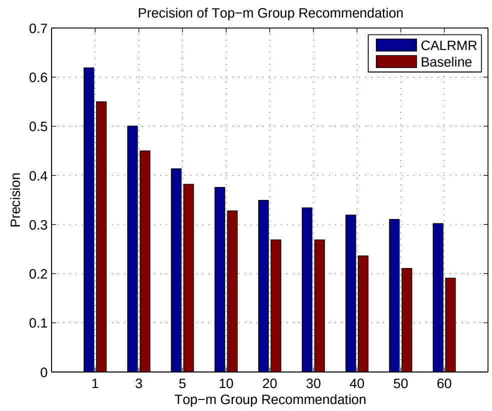
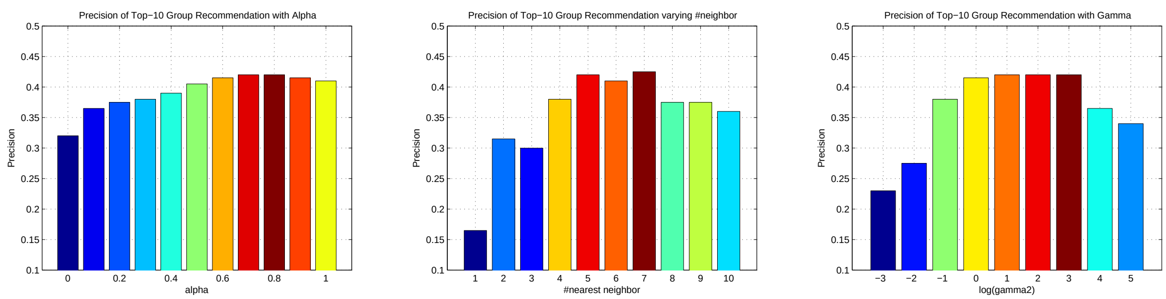


Fig. 4. The *objective* revaluation of the precision of the top-m group recommendation.



(a) α (b) #NN (c) γ_2

Fig. 5. The evaluation of hyper-parameters for the objective group discovery. (a) The group prediction with different α when computing the combined kernel function in (8). (b) The result with different #nearest neighbors (NN) when computing the graph Laplacian in Section 4.3. (c) The result with different weight γ_2 on the content consistency (6).

Then we run Algorithm 1 to recover the true indicating matrix $\mathbf{T}$. We recommend the users the top-m ranked groups according to the recovered $\mathbf{T}$. We have collected the friend lists of all testing users (through Flickr public APIs), and used the friend list for each user as the ground truth for friend suggestion. Since we cannot know the most proper size m, we evaluate the recommendation results of the following two methods by a series of varied m values:

—Baseline: Under the unified view of groups and users introduced in Section 4.3.1, we can compute the similarity between users and groups directly. Thus the (i, j)-th entry of matrix $\mathbf{T}$ is $T_{ij} = k(\mathbf{u}_i, \mathbf{g}_j)$.

—CALRMR: The proposed Content-Aware Low-Rank Matrix Recovery technique in Section 4.

Figure 4 shows the mean precision results for the selected test users. For our solution, we observe that the top-1 recommended group would be joined by the original Flickr user with a probability larger than 60%. The top-3 recommended groups have a mean precision about 50%. As m increases, the mean precision drops, though the trend is not sharp when m is greater than 10. This observation shows that the smaller of the prediction confidence in $\mathbf{T}$, the lower the precision would be obtained. This observation verifies that the learned ranking score is meaningful.

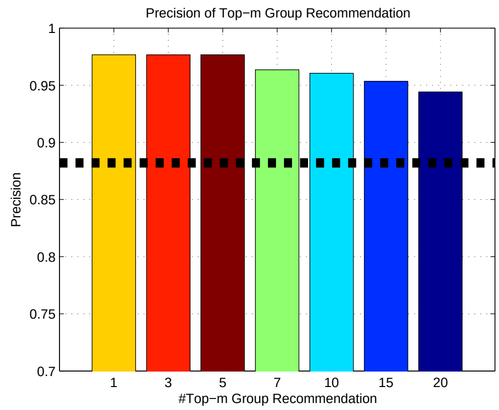


Fig. 6. The *subjective* evaluation of the precision of the top-m group recommendation. The dashed line refers to the original group membership.

Compared with the k nearest neighbor method, the top 10 recommendation accuracy is slightly better. However, as m increases, our algorithm beats baseline significantly. In general, if the content of a user is akin to a group, the user may tend to join the group. This is why the k-nn recommendation works. However, as m increases, the uncertainty of of group membership also increases (refer to the motivation in Section 4.1). The naive nearest neighbor method cannot reflect such noise. Our error matrix $\mathbf{E}$ encodes the error information such that the confidence matrix $\mathbf{T}$ can better capture the group membership information.

We further explore the contribution of each factor that can affect the group refinement in our framework (6), as shown in Figure 5. First, we deem each user has a two-view representation, i.e., photos and textual tags. We use the linear combination of the similarity in these two spaces as in Eqn. (8). The α parameter controls the importance of each view. We examine its influence in Figure 5(a). When $\alpha = 0$ or 1, it corresponds to the case of using photos or tags alone. We observe the best result at $\alpha = 0.8$ which implies visual information could be more important than tags. We only use the similarity to k-nearest neighbors for each user or group when computing the graph Laplacian for the content consistency. Figure 5(b) plots the precision with different k values. The best result is observed at $k = 7$. When $k = 1$ the result is poor. This is because very little content information is taken into account. When k is large, the performance starts to drop, which is probably because the noisy information is involved. The γ parameter in Eqn (6) controls the trade-off of low-rank decomposition and content-awareness. We fix $\gamma_1 = 0.5$ and examine the influence of γ_2. A small γ_2 value implies the refinement does not respect the context much, while a large γ_2 emphasizes the importance of content consistency. The evaluation of γ_2 is shown in Figure 5(c). We can observe that it is not very sensitive for parameter γ_2. Typically, setting γ_2 in the range of $[1, 8]$ is able to achieve fairly good performance.

6.2.2 *Subjective Evaluation.* The objective evaluation may not be fair for the proposed approach as it treats the observed group membership as ground truth. In this section, we randomly pick 100 users who have fewer than 20 groups and run the proposed algorithm to refine their group membership. We invited five subjects to manually judge whether he/she would join the recommended groups. The recommendation is regarded "correct" if three or more subjects answering "yes."

Figure 6 plots the precision of top recommend groups according to the learned matrix $\mathbf{T}$. The dashed line marks the average precision of the original groups joined by the users, which is slightly lower

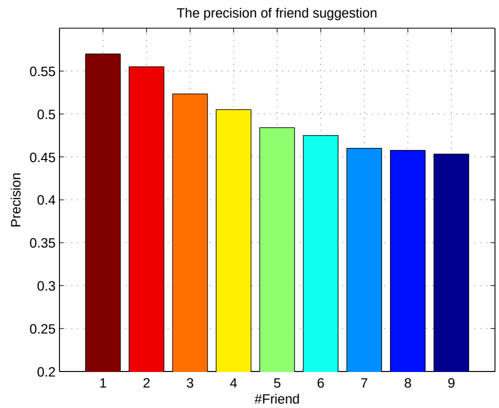


Fig. 7. Precision of friend suggestion.

than 90%. This supports our motivation that the observed groups may not be accurate from the view of content consistency, i.e., the theme of the joined groups is not closely related to the user-generated data. Finally, we observed that the users would very likely join the top 20 ranked groups with a probability higher than 90%. In particular, for the top-1 recommended group, it is almost always related to the user's interest. This promising result again validates the efficacy of the proposed technique.

6.3 Evaluation of Friend Suggestion

The proposed framework can also be used for (implicit) friend suggestion as described in Section 5.2. Since manual labeling is very heavy, we do not evaluate the quality of all recommended friends. Instead, for top m recommended friends $\mathcal{N}_r$, we first compare it with the original friend list $\mathcal{N}$. Then, we label the candidate friends $\mathcal{D} = \mathcal{N}_r - \mathcal{N}$. We randomly selected 100 users and subjectively tell whether it is reasonable to add each candidate into his/her friend list.

Figure 7 plots the precision of friend suggestion. The top 3 recommended users have a chance larger than 50% to be confirmed as a friend. As the number of friends increases, the performance drops gradually. Comparing with the group recommendation result in Figure 4, the recommendation accuracy is worse significantly. We conjecture the possible reason is that the prior probability of adding users as friends is much lower than joining a group. This fact is actually observed from our data statistics in Table 2, where the user-user indicator matrix is more sparse than group-user indicator matrix (before pre-processing).

6.4 Visualization and Prototype Detection

After we have discovered the friendship structure of Flickr users, we can visualize it by 2D embedding. Note here we treat the friendship symmetric, i.e., an edge exists between $\mathbf{u}_i$ and $\mathbf{u}_j$ i.f.f. $\mathbf{u}_i$ appears in the friends of $\mathbf{u}_j$ or vice versa. Figure 8 plots the link structure of 100 randomly sampled Flickr users. Each node in the graph represents a specific registered Flickr user. The size and colors of the nodes are proportional to the degree of the nodes, which in turn reflects the popularity or influence of Flickr users indirectly. The larger the size of a node, the more links starts from it. The distance between nodes measures how similar the contents uploaded by users. Based on such visualization results, one can locate the prototype users intuitively. For example, the red node at the center of the Figure has 15 friends. In general, it plays a more important role in group formation and content propagation than

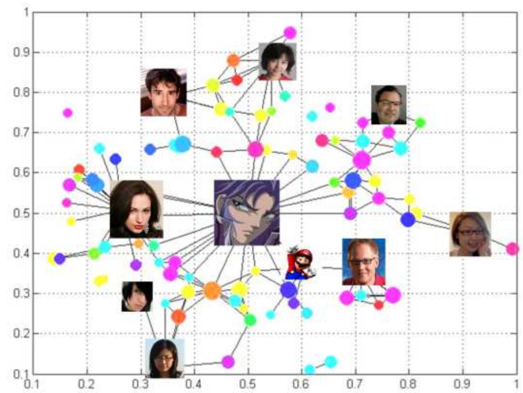


Fig. 8. The 2D embedding results of 100 randomly sampled Flickr users from our dataset.

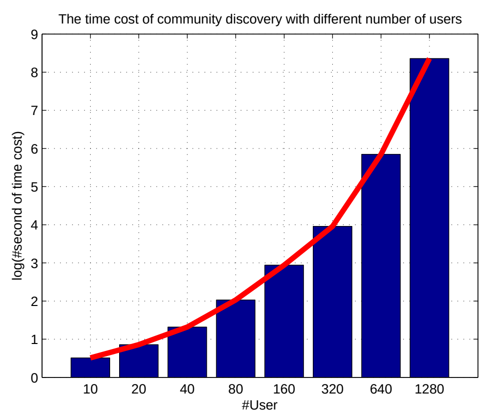


Fig. 9. The computational time costs of the group refinement task with #group=1,467. Note here #user increases exponentially.

other users. On the other hand, the small red node at the upper right corner has no friends. Probably it is an inactive user.

6.5 Complexity Analysis

In this section, we examine the time cost of the presented proximal gradient based Algorithm 1 for community discovery. We only evaluate the group refinement framework. We trace the CPU time with different number of users with 1,467 interest groups. The iteration is stopped when the change of objective values and the violation of $\|\mathbf{M} - \mathbf{T} - \mathbf{E}\|_F$ are both smaller than 10. We test the computation cost on a workstation with Quad CPU of 2.67GHz and 8G memory. We plot the CPU time in Figure 9. The algorithms relies matrix decomposition. It would have complexity of $t \times n^3$, where t is the iteration and n is matrix size. This actually limits applying the proposed algorithm to large-scale applications.

It is shown that the accelerated proximal gradient algorithm has the convergence rate of order $O(t^{-2})$ [Lin et al. 2009]. At each iteration, the computation of Algorithm 1 is dominated by the singular value decomposition with the time complexity of $O(N_u N_g^2)$. Note the #user in Figure 9 increases exponentially. The log-linear trend with N_u fits the theoretical complexity very well. For our overall 4,919 Flickr users, the cost is about 260 minutes. Therefore, it is indeed practical to apply the presented community discovery framework for mining large-scale real Flickr data.

7. CONCLUSIONS AND FUTURE WORK

This paper investigated a novel framework for community discovery by proposing a Content-Aware Low-Rank Matrix Recovery technique. The proposed community discovery framework is able to effectively exploit both textual and visual information for social community analysis. We formulated the problem as a task of decomposing the observed community indicator matrix into a low-rank true matrix and a sparse error matrix, which is then formulated as an optimization problem of low-rank matrix recovery. Further, we proposed to enforce the similarity induced by this representation to coincide with the local geometry of the rich content data, making the true matrix content-aware. To solve the optimization task efficiently, we developed an accelerated proximal gradient descent algorithm, and applied the proposed technique to several applications, including visualization, group recommendation, and friend suggestion on a large Flickr testbed.

In spite of the promising results achieved, our work still has a few limitations. First, the proposed algorithm is transductive and thus is not easy to handle previously unseen data efficiently. As the number of registered Flickr users keeps growing, our future work will extend the current framework by studying online learning techniques. Second, our current framework mainly exploits the textual and visual information, but in Flickr, there are still much other heterogeneous metadata available other than the images and tags. For example, the temporal-spatial information of the photos, the favorite pictures of users, the online forum of the groups, and so on. Such rich metadata could also be beneficial to community discovery by incorporating into our framework with some proper modification. Third, we used the average bag-of-visual-word histogram as the visual representation of a user (over all his/her uploaded photos), which is not a satisfying representation. We will consider to use a few representative or “favorite” photos as the visual presentation.

REFERENCES

ADAMIC, L. A. AND ADAR, E. 2001. Friends and neighbors on the web. *SOCIAL NETWORKS 25*, 211–230.

BACKSTROM, L., HUTTENLOCHER, D. P., KLEINBERG, J. M., AND LAN, X. 2006. Group formation in large social networks: membership, growth, and evolution. In *KDD*. 44–54.

CAI, J.-F., CANDÈS, E. J., AND SHEN, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization 20,* 4, 1956–1982.

CHOUDHURY, M. D., MASON, W. A., HOFMAN, J. M., AND WATTS, D. J. 2010. Inferring relevant social networks from interpersonal communication. In *WWW*. 301–310.

CRANDALL, D. J., COSLEY, D., HUTTENLOCHER, D. P., KLEINBERG, J. M., AND SURI, S. 2008. Feedback effects between similarity and social influence in online communities. In *KDD*. 160–168.

GILBERT, E. AND KARAHALIOS, K. 2009. Predicting tie strength with social media. In *CHI*. 211–220.

GUY, I., ZWERDLING, N., RONEN, I., CARMEL, D., AND UZIEL, E. 2010. Social media recommendation based on people and tags. In *SIGIR*. 194–201.

KENNEDY, L. S., NAAMAN, M., AHERN, S., NAIR, R., AND RATTENBURY, T. 2007. How flickr helps us make sense of the world: context and content in community-contributed media collections. In *ACM Multimedia*. 631–640.

KING, I., LYU, M. R., AND MA, H. 2010. Introduction to social recommendation. In *WWW*. 1355–1356.

KONSTAS, I., STATHOPOULOS, V., AND JOSE, J. M. 2009. On social networks and collaborative recommendation. In *SIGIR*. 195–202.

LESKOVEC, J., HUTTENLOCHER, D. P., AND KLEINBERG, J. M. 2010. Predicting positive and negative links in online social networks. In *WWW*. 641–650.

LIN, Z., GANESH, A., WRIGHT, J., WU, L., CHEN, M., AND MA, Y. 2009. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. Tech. Rep. UILU-ENG-09-2214, UIUC.

LOWE, D. G. 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision 60,* 2, 91–110.

MEI, T., RUI, Y., LI, S., AND TIAN, Q. 2014. Multimedia search reranking: A literature survey. *ACM Computing Surveys 46,* 3.

MEI, T., YANG, B., HUA, X.-S., AND LI, S. 2011. Contextual video recommendation by multimodal relevance and user feedback. *ACM Trans. on Information Systems 29,* 2.

NEGOESCU, R. A., ADAMS, B., PHUNG, D. Q., VENKATESH, S., AND GATICA-PEREZ, D. 2009. Flickr hypergroups. In *ACM Multimedia*. 813–816.

NEGOESCU, R. A. AND GATICA-PEREZ, D. 2008. Analyzing flickr groups. In *CIVR*. 417–426.

ROY, S. D., MEI, T., ZENG, W., AND LI, S. 2013. Towards cross-domain learning for social video popularity prediction. *IEEE Trans. on Multimedia 15,* 6, 1255–1267.

SEBASTIANI, F. 2001. Machine learning in automated text categorization. *CoRR cs.IR/0110053*.

SERDYUKOV, P., MURDOCK, V., AND VAN ZWOL, R. 2009. Placing flickr photos on a map. In *SIGIR*. 484–491.

SINGLA, P., KAUTZ, H., LUO, J., AND GALLAGHER, A. 2008. Discovery of social relationships in consumer photo collections using markov logic. In *CVPR workshop*.

SINGLA, P. AND RICHARDSON, M. 2008. Yes, there is a correlation: - from social networks to personal behavior on the web. In *WWW*. 655–664.

TASKAR, B., WONG, M. F., ABBEEL, P., AND KOLLER, D. 2003. Link prediction in relational data. In *NIPS*.

WANG, C., YAN, S., ZHANG, L., AND ZHANG, H.-J. 2009. Multi-label sparse coding for automatic image annotation. In *CVPR*. 1643–1650.

WRIGHT, J., GANESH, A., RAO, S., AND MA, Y. 2009. Robust principal component analysis: Exact recovery of corrupted low-rank matrices. *CoRR abs/0905.0233*.

WU, L., HOI, S. C. H., JIN, R., ZHU, J., AND YU, N. 2009. Distance metric learning from uncertain side information with application to automated photo tagging. In *ACM Multimedia*. 135–144.

WU, L., HUA, X.-S., YU, N., MA, W.-Y., AND LI, S. 2008. Flickr distance. In *ACM Multimedia*. 31–40.

WU, P. AND TRETTER, D. 2009. Close & closer: social cluster and closeness from photo collections. In *ACM Multimedia*. 709–712.

YANG, T., JIN, R., CHI, Y., AND ZHU, S. 2009. Combining link and content for community detection: A discriminative approach. In *KDD*. 927–935.

YAO, T., LIU, Y., NGO, C.-W., AND MEI, T. 2013. Unified entity search in social media community. In *Proceedings of WWW*. 1457–1466.

YU, J., JOSHI, D., AND LUO, J. 2009. Connecting people in photo-sharing sites by photo content and user annotations. In *ICME*. 1464–1467.

ZHU, G., YAN, S., AND MA, Y. 2010. Image tag refinement towards low-rank, content-tag prior and error sparsity. In *ACM Multimedia*.

ZHU, X. 2005. Semi-supervised learning literature survey. Tech. Rep. 1530, Computer Sciences, University of Wisconsin-Madison.

ZHUANG, J. AND HOI, S. C. H. 2011. A two-view learning approach for image tag ranking. In *Proceedings of ACM WSDM*. 625–634.

ZHUANG, J., MEI, T., HOI, S. C. H., HUA, X.-S., AND LI, S. 2011. Modeling social strength in social media community via kernel-based learning. In *Proceedings of ACM Multimedia*. 113–122.

A. CONTENT-AWARE LOW-RANK MATRIX RECOVERY (CALRMR)

Applying the technique of proximal gradient descent introduced in Section 4.4 to solve (6), we have:

$$g(x) = \mu\|\mathbf{T}\|_* + \mu\gamma_1\|\mathbf{E}\|_1,$$

$$f(x) = \frac{\mu\gamma_2}{2}\Big\{\mathbf{tr}\,(\mathbf{T}^\top\mathbf{L}^u\mathbf{T}) + \mathbf{tr}\,(\mathbf{T}\mathbf{L}^g\mathbf{T}^\top)\Big\} + \frac{1}{2}\|\mathbf{M}-\mathbf{T}-\mathbf{E}\|_F^2,$$

where the term $\frac{1}{2}\|\mathbf{M}-\mathbf{T}-\mathbf{E}\|_F^2$ is added to formulate an unconstrained optimization problem by removing the constraint in (6), $x = \begin{pmatrix}\mathbf{T}\\ \mathbf{E}\end{pmatrix}$, $\mathbf{Y} = \begin{pmatrix}\mathbf{Y}^T\\ \mathbf{Y}^E\end{pmatrix}$ accordingly. It is not difficult to show that $f(x)$

satisfies the Lipschitz continuity [Zhu et al. 2010] with the Lipschitz constant L_f computed as follows:

$$L_f = \sqrt{4\sigma^2_{\max}(\mu\gamma_2\mathbf{L}^g) + 4\sigma^2_{\max}(\mu\gamma_2\mathbf{L}^u) + 6}\ .$$

Thus, we can tackle it by iteratively solving $\mathbf{T}$ and $\mathbf{E}$.

We develop a so-called content-aware low-rank matrix recovery (CALRMR) algorithm to iteratively solve $\mathbf{T}$ and $\mathbf{E}$, i.e., we alternate between optimizing $\mathbf{T}$ and $\mathbf{E}$. By fixing $\mathbf{E}$ to $\mathbf{E}_t$, we solve $\mathbf{T}_{t+1}$ as follows:

$$\begin{aligned}\mathbf{T}_{t+1} = \arg\min_{\mathbf{T}} \quad & \Big(\frac{L_f}{2}\|\mathbf{T} - \mathbf{Y}^T_t + \frac{1}{L_f}\mathbf{P}^T_t\|^2_F + \mu\|\mathbf{T}\|_* \\ & +\mu\gamma_1\|\mathbf{E}_t\|_1 + f(\mathbf{Y}_t) - \frac{1}{2L_f}\|\mathbf{P}^T_t\|^2_F\Big),\end{aligned}$$

where $\mathbf{P}^T_t = \mu\gamma_2(\mathbf{Y}^T_t\mathbf{L}^u + \mathbf{L}^g\mathbf{Y}^T_t) + \mathbf{Y}^T_t + \mathbf{Y}^E_t - \mathbf{M}$. This problem is equivalent to the following:

$$\mathbf{T}_{t+1} = \arg\min_{\mathbf{T}} \frac{\mu}{L_f}\|\mathbf{T}\|_* + \frac{1}{2}\|\mathbf{T} - \mathbf{Y}^T_t + \frac{1}{L_f}\mathbf{P}^T_t\|^2_F. \tag{13}$$

According to [Cai et al. 2010], this problem can be solved by a singular value thresholding algorithm as follows:

$$\mathbf{T}_{t+1} = \mathbf{U}[\mathbf{\Sigma}]_{\frac{\mu}{L_f}}\mathbf{V}^\top, \tag{14}$$

where $\mathbf{U\Sigma V}^\top$ is the SVD of $(\mathbf{Y}^T_t - \frac{1}{L_f}\mathbf{P}^T_t)$, $[x]_\varepsilon$ is the soft-thresholding operatrion defined as

$$[x]_\varepsilon = \begin{cases} x-\varepsilon & \text{if } x > \varepsilon, \\ x+\varepsilon & \text{if } x < -\varepsilon, \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Similarly, by fixing $\mathbf{T}$, we can apply the above approach to computing the sparse error matrix as follows:

$$\mathbf{E}_{t+1} = \Big[\mathbf{Y}^E_t - \frac{1}{L_f}\big(\mathbf{Y}^T_t + \mathbf{Y}^E_t - \mathbf{M}\big)\Big]_{\frac{\mu\gamma_1}{L_f}}. \tag{16}$$

We summarize the details of the proposed CALRMR algorithm for Flickr interest group refinement in Algorithm 1.