12-2011

Modeling Social Strength in Social Media Community via Kernel-based Learning

Jinfeng ZHUANG
Nanyang Technological University

Tao MEI
Microsoft Research Asia, Beijing, China

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

Xian-Sheng HUA
Microsoft Redmond

Shipeng LI
Microsoft Research Asia, Beijing, China

Modeling Social Strength in Social Media Community via Kernel-based Learning*

Jinfeng Zhuang †, Tao Mei ‡, Steven C. H. Hoi †, Xian-Sheng Hua ‡, Shipeng Li ‡
† Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798
‡ Microsoft Research Asia, Beijing 100080, China
{zhua0016, chhoi}@ntu.edu.sg, {tmei, xshua, spli}@microsoft.com

ABSTRACT

Modeling continuous social strength rather than conventional binary social ties in the social network can lead to a more precise and informative description of social relationship among people. In this paper, we study the problem of social strength modeling (SSM) for the users in a social media community, who are typically associated with diverse form of data. In particular, we take Flickr—the most popular online photo sharing community—as an example, in which users are sharing their experiences through substantial amounts of multimodal contents (e.g., photos, tags, geolocations, friend lists) and social behaviors (e.g., commenting and joining interest groups). Such heterogeneous data in Flickr bring opportunities yet challenges to the research community for SSM. One of the key issues in SSM is how to effectively explore the heterogeneous data and how to optimally combine them to measure the social strength. In this paper, we present a kernel-based learning to rank framework for inferring the social strength of Flickr users, which involves two learning stages. The first stage employs a kernel target alignment algorithm to integrate the heterogeneous data into a holistic similarity space. With the learned kernel, the second stage rectifies the pair-wise learning to rank approach to estimating the social strength. By learning the social strength graph, we are able to conduct collaborative recommendation and collective classification. The promising results show that the learning-based approach is effective for SSM. Despite being focused on Flickr, our technique can be applied to model social strength of users in any other social media community.

Categories and Subject Descriptors

H.4.m [**Information Systems**]: Information Systems Applications—*Miscellaneous*

General Terms

Algorithms, Experimentation, Performance

Keywords

Social networks, kernel-based learning, learning to rank

1. INTRODUCTION

Social network mining, from communication networks, to friendship networks, to professional and organizational networks, has attracted a surge of interests in both industrial and academic communities. Most traditional studies focus on detecting *binary* relation ties between people (e.g., friends or not). Such a coarse indicator is not precise enough to give insight about the strength of social relationship among people. Some recent work, e.g., the study in [29], has attempted to address the problem of modeling the strength of social connections instead of simple binary linkage. Inferring precise social strength can facilitate a variety of applications, including friendship linkage prediction, item recommendation, social search, and so on.

In this paper, we investigate the challenging problem of Social Strength Modeling (SSM) of users in the social media communities. In particular, we take Flickr, one of the most popular online photo-sharing sites, as the social media platform in our study. Flickr contains rich user-generated contents, including shared photos, user-annotated tags, comments, and so on. Analogous to other social networking sites, e.g., Facebook and LinkedIn, each Flickr user can add other users into his contact list to indicate their friendship. Users can also create and join interest groups where users share photos of common interests and comments with each other. Besides the explicit mutual linkage between users, the uploaded photos and their associated metadata (e.g., tags, comments, etc.) can also be leveraged to infer the implicit relationship between users. The multimodal information available on Flickr poses opportunities yet challenges for the research on SSM.

One key challenge of SSM is to effectively explore and combine heterogeneous data from multiple modalities to measure the social strength in a principled way. Previous work on Flickr data mining has predominately focused on image-only or tag-only analysis. Other rich metadata has not been well exploited. To overcome this limitation, we present a novel framework for social strength modeling by unifying multi-modal heterogeneous data through a kernel-based machine learning approach. In particular, we suggest to use *kernel machine* [6] to model user similarity in each modal-

* This work was performed at Microsoft Research Asia and supported by Microsoft Research grant.
Area chair: Cees Snoek.

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.

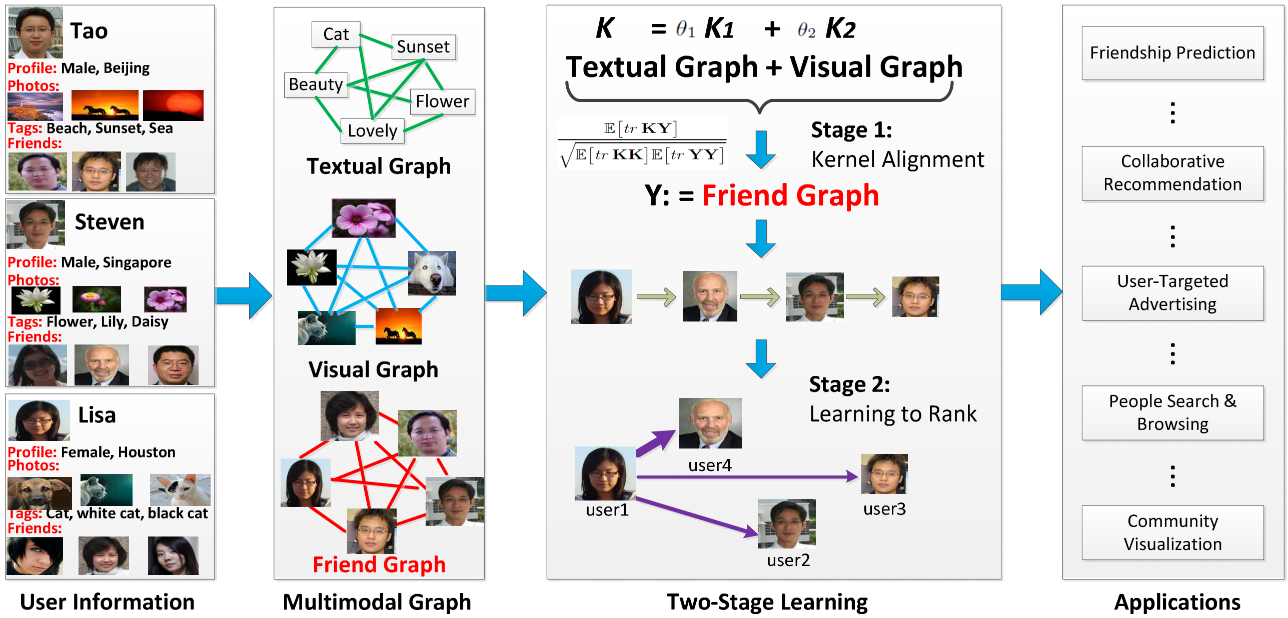


Figure 1: The proposed kernel-based learning to rank framework. The first panel shows the data associated with users, based on which we build three graphs in the second panel (only three kinds of graphs presented for illustration). In the third panel, we first learn the weight θ by maximally aligning the combination of textual and visual graphs to the friend graph, and then we adopt learning to rank framework with logistic loss to estimate the social strength. The learned social strength can lead to a wide variety of applications in the fourth panel.

ity, and then propose a two-stage learning scheme to measure the social strength by kernel-based techniques.

Specifically, we assume the final proximity graph of users is a linear combination of multiple proximity graphs, which are derived from the multiple modalities in Flickr. Each proximity graph is built by defining a kernel function on one modality. Figure 1 shows the proposed framework. At the first learning stage, we propose to combine the multiple proximity graphs by learning the optimal combination weights by following the kernel target alignment (KTA) principle [5, 8]. At the second stage, we propose a kernel-based learning to rank approach to model the social strength with the optimal kernel learned from the first stage. The proposed two-stage learning approach is able to model the social strength of users by exploring multimodal heterogeneous data in a systematic and comprehensive way. It is worth noticing that although we take Flickr as an example of social media community (as Flickr opens a public way to access its rich metadata), the proposed learning-based approach can be applied to any kind of community, such as Facebook and LinkedIn, in which users are also associated with rich multimodal metadata. The learned continuous social strength between users can facilitate a number of social media applications. In particular, we apply our technique to the social tasks of item recommendation and collective classification.

In summary, this paper makes the following contributions:

- We study the problem of modeling continuous social strength among users in a social media community. To the best of our knowledge, there are few works towards this target in the multimedia research community.

- We propose a novel two-stage kernel-based learning framework for social strength modeling, which effectively integrates heterogeneous data by optimally combining multiple kernels, and learns to rank social strength by a kernel-based learning to rank approach.

- We conduct experiments to evaluate the performance on a real-world dataset, and propose a wide variety of applications based on the learned social strength.

The rest of the paper is organized as follows. Section 2 formulates the problem of social strength modeling on Flickr data. Section 3 introduces the framework. Section 4 shows the experiments. Section 5 reviews related work, followed by the conclusion in Section 6.

2. PROBLEM FORMULATION

This section gives the problem definition of social strength modeling. Table 1 lists the key notations. First, we formally define the problem of social strength modeling with Flickr data as follows.

Definition 1 (Social Strength Modeling). *Given a collection of Flickr users $\mathcal{U}$, the goal of a* social strength modeling *problem is to learn a function $f : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}_+$ such that $f(u_i, u_j)$ measures the strength of social relationship between user u_i and user u_j.*

In the above definition, a basic element is a Flickr user $u_i \in \mathcal{U}$, which can be defined below.

Definition 2 (Flickr user). *Each* Flickr user *$u_i \in \mathcal{U}$ is modeled as a three-dimensional tuple $u_i := [\mathcal{P}, \mathcal{N}, \mathcal{G}]$, where $\mathcal{P} = \{p_i : i \in \mathbb{N}_z\}$ is a collection of Flickr photos uploaded by user u_i, $\mathcal{N} \subset \mathcal{U}$ is the collection of users who appear in the contact list of user u_i, $\mathcal{G}$ is the collection of interest groups joined by user u_i.*

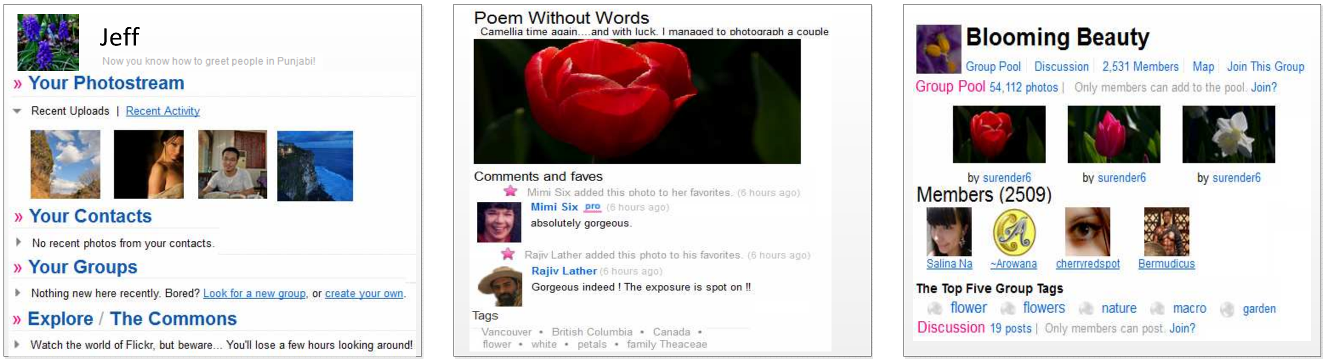


Figure 2: Example of the multimodal information associated with a typical Flickr user (left), a Flickr image (middle), and a Flickr interest group (right). The definition in Section 2 is a formalization of these data.

Table 1: List of key notations.

Notation	Description
$\mathbb{N}_u$	$\{1, \ldots, u\}$, a set of integers up to u †
N_u	$\|\mathbb{N}_u\|$, the cardinality of $\mathbb{N}_u$
u_i	the i-th Flickr user
$\mathcal{U}$	$\{u_i : i \in \mathbb{N}_u\}$, a collection of N_u users
$X, T, D, L, \mathcal{C}$	attributes of a Flickr image: content, tags, date, location, and comments
$\mathcal{P}, \mathcal{G}, \mathcal{N}$	a collection of Flickr photos, groups, and friends, respectively
$\#$	the number of some object
$K, \mathbf{K}$	a similarity (kernel) function and matrix
$\mathrm{tr}\, \mathbf{K}$	the trace of $\mathbf{K}$
$\\|\mathbf{K}\\|_F$	$:= \sqrt{\mathrm{tr}\, \mathbf{KK}}$, the Frobenius norm of $\mathbf{K}$
$\mathbf{K}^\top$	the transpose of $\mathbf{K}$

†: We generalize the notation $\mathbb{N}_u$ and N_u to other objects, i.e., the sub-script u can be k, g, and c for indexing the *kernel*, *group*, and *comment*, respectively.

In the above definition, $\mathcal{N}$ and $\mathcal{G}$ indicate the explicit social relationship and organizations of Flickr users, and $\mathcal{P}$ may be beneficial to discovering some implicit relationship between users. Figure 2 shows the example of a typical Flickr user. "Your Photostream," "Your Contacts," and "Your Groups" correspond to $\mathcal{P}$, $\mathcal{N}$, and $\mathcal{G}$, respectively.

The contact list $\mathcal{N}$ indicates the binary social ties between users. Note that the friend relationship given by $\mathcal{N}$ is always asymmetric, which can be naturally handled by our proposed learning to rank framework in the next section. We employ $\mathcal{N}$ as the training data of known social strength for building our model.

The set $\mathcal{G}$ is the interest groups that user u_i has joined. For each $g \in \mathcal{G}$, it is created and self-organized by registered Flickr users. Users belonging to the same group tend to share photos of common interests. The right picture in Figure 2 shows a popular interest group.

The images $\mathcal{P}$ are useful to find implicit social relationship since they contain rich context information that expresses interests and social behaviors between users. We formally define a Flickr image as follows.

Definition 3 (Flickr image). *A* Flickr image *is defined as a 5-dimensional tuple $P := [X, T, D, L, \mathcal{C}]$, where*

- *$X \in \mathcal{X} \subset \mathbb{R}^{d_x}$ is the visual content with some fixed-length feature representation, d_x is the dimension of the visual descriptors;*
- *$T \in \mathcal{T} \subset \mathbb{R}^{d_t}$ is a vector representing the tags associated with X, d_t is the size of tag vocabulary;*
- *D is the date that X is created;*
- *$L := [latitude, longitude] \in \mathbb{R}_+ \times \mathbb{R}_+$ is the location where P is created;*
- *$\mathcal{C} := \{[U, C]_i \in \mathcal{U} \times \mathbb{R}^{d_t} : i \in \mathbb{N}_c\}$ is a collection of comments of P, where the first component U is the user who posts the comment and the second component C is the content of the comment.*

The rich context information $(T, D, L, \mathcal{C})$ besides the uploaded photo X encodes the social behaviors among different users. The middle panel of Figure 2 shows the typical behaviors of a Flickr user. In this paper, we incorporate this context information into a unified discriminative framework to model the social strength.

3. SOCIAL STRENGTH MODELING

In this section, we propose a kernel-based learning framework for social strength modeling.

3.1 Motivations

As shown in Figure 2, the first challenge is how to effectively combine the heterogeneous data associated with a user. Note that we can compute the similarity K under different modality, which is akin to a kernel function in the kernel machines, such as support vector machine and kernelized logistic regression. This inspires us to adopt the multiple kernel learning (MKL) scheme to integrate the multiple modalities, which has been actively studied in recent years [13]. Although not always yielding better results than single kernel chosen by cross-validation [4], it is regarded as one of the principle way to combine heterogeneous data sources. Therefore, we adopt the state-of-the-art MKL algorithm to weight each modality [5].

Second, we notice the social strength essentially ranks the degree of affinity between a pair of people. Therefore, we adopt the pair-wise learning to rank framework to further rectify the social strength based on the learned K in the first stage. Compared with generative models, such a discriminative model often enjoys better generalization ability [24]. Moreover, it avoids both the latent variable assumption and the parametric-form assumption of generative model functions, which make the learning more compact and precise.

We first discuss how to measure the similarity of Flickr users by defining a variety of kernel functions $\{K\}$ on different modalities of Flickr data. We then present a kernel learning technique to determine the optimal combination weights of multiple kernels, by following the kernel target alignment principle [5, 8]. Finally, based on the learned kernel, we formulate the social strength modeling problem in a pair-wise kernel-based learning to rank task, which can be efficiently solved in an iterative manner as the import vector machine (IVM) [32].

3.2 Kernels for Measuring User Similarity

For machine learning algorithms, the kernel trick is an important technique for mapping observations from a general set into a much higher- and possibly infinite dimensional inner product space without explicitly computing the mapping. Typically, each kernel $k : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ on a set $\mathcal{S}$ essentially defines the way of measuring similarity between data instances in $\mathcal{S}$. In this section, we present a series of candidate kernel functions in different modalities for measuring the similarity of Flickr users, which will be further explored with other kernel methods for inferring social strength.

3.2.1 User Similarity in Visual Space

For visual feature representation, we adopt the bag-of-(visual) word (BoW) model [19]. Specifically, we first extract the local descriptors of Scale-Invariant Feature Transform (SIFT) for each image [15]. All these descriptors are quantized into d_x groups by a K-means clustering process. Given an image, we assign each of its SIFT descriptors to a nearest cluster. Then, each image is converted into a fixed length of feature vector $\mathbf{x} \in \mathbb{R}^{d_x}$, where d_x is the size of visual vocabulary. The i-th component of this vector counts the frequency of SIFT descriptor assigned to cluster i. We measure the visual similarity between image $\mathbf{x}_i$ and $\mathbf{x}_j$ by a Gaussian kernel:

$$s(\mathbf{x}_i, \mathbf{x}_j) = \exp\Big\{ - \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2} \Big\},$$

where σ is a kernel parameter. For a specific user u, we use the centroid of the images belonging to u to represent this user in the visual space, i.e.,

$$\bar{\mathbf{u}} = \sum_i \frac{\mathbf{x}_i}{|u|},$$

where $|u|$ is the number of images belonging to user u, $\mathbf{x}_i$ is one image uploaded by u. Thus, the similarity $K^1(u_i, u_j)$ between user u_i and u_j in the visual space can be given by

$$K^1(u_i, u_j) := \exp\Big\{ - \frac{\|\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j\|^2}{\sigma^2} \Big\}.$$

As it is difficult to obtain an optimal bandwidth parameter σ, we set it as the average Euclidean distance empirically.

3.2.2 User Similarity in Text Space

Note that each uploaded photo can be associated with a set of tags provided by the user. We adopt the bag-of-word model to represent the textual information of a user [2]. Specifically, we collect all the tags and build a tag dictionary with size d_t. The tags of a user are converted into a feature vector in $\mathbb{R}^{d_t}$ by the traditional *tf-idf* weighing method. Here, the inverse document frequency is the number of users containing that tag. In this way, a user u_i can be represented by a vector $\mathbf{z}_i \in \mathbb{R}^{d_t}$. We employ the normalized linear kernel for measuring the user similarity in the text space which is widely used for text classification:

$$K^2(u_i, u_j) := \frac{\mathbf{z}_i^\top \mathbf{z}_j}{\sqrt{\mathbf{z}_i^\top \mathbf{z}_i}\sqrt{\mathbf{z}_j^\top \mathbf{z}_j}}.$$

Compared with the visual kernel K^1, the tag-based kernel K^2 carries more semantic information.

3.2.3 User Similarity by Mutual Comments

The interaction between users reflects the social connection between each other. For example, if two users post comments on each other's photos frequently, probably strong social ties exist between them. In general, people who are friends in real life would communicate more frequently. We can collect the mutual comment information between users to construct a symmetric link graph, in which each vertex represents a user and the edge weight is the number of comments between two users. Thus, we have

$$K^3(u_i, u_j) := \#\mathit{comment} \text{ between } u_i \text{ and } u_j.$$

When $i = j$, the kernel value is the frequency that u_i has ever posted to himself.

3.2.4 User Similarity by Common Interest Groups

The Flickr users can create or join in *Flickr interest groups*, which consists of a collection of users who share common interest or taste on photos of particular styles. Such interest groups can help users to find people or photos of their interest. Researchers have proposed techniques (e.g., [18, 16]) to analyze the structure and themes of the groups due to their importance in organizing Flickr users. Intuitively, people among which strong social connections exist are more likely to join the same interest groups, as they may affect each other and share similar interests. Therefore, we can use the number of common interest groups to measure the similarity between people:

$$K^4(u_i, u_j) := \#\mathit{group} \text{ both } u_i \text{ and } u_j \text{ joined.}$$

When $i = j$, the kernel value is the number of groups that u_i has joined.

3.2.5 User Similarity by Mutual Friends

Each Flickr user has a contact list, which can be deemed as "friends." When two users share many common friends, it is reasonable to infer that they have strong social connection to each other. Naturally, we can count this number as the kernel value to quantify this modality:

$$K^5(u_i, u_j) := \#friend \text{ belong to both } u_i \text{ and } u_j.$$

3.2.6 User Similarity via Geo-tags

The images in a social media site are usually associated with geo-tags, which indicates the latitude and longitude parameters where the photos are taken. If two users have traveled to the same place frequently, it is reasonable to conclude they are similar to each other (since they would like to travel to the same places). Similar to the representation in the text space, we use the bag-of-geotag to compute the similarity between users:

$$K^6(u_i, u_j) \;\; := \;\; \#\textit{location} \text{ where both } u_i \text{ and } u_j \text{ took photos.}$$

At the diagonal position of K^6, the value is the number of places the user has been to. The raw geo-tags are in form of longitude and latitude pairs. We discretize it by dividing the whole globe into tiles. Each location is represented by its corresponding tile.

3.2.7 User Similarity via Favorite Photos

One important function provided by Flickr is that users can mark their favorite photos (fave). Assuming the number of faves correlates to the social strength, we define

$$K^7(u_i, u_j) := \#\textit{photo} \text{ both } u_i \text{ and } u_j \text{ favor.}$$

The rationale underlying this assumption is that users who like the same photos tend to be friends. This may not be true in real case. However, through the KTA algorithm we learn the weights of these kernels, which provides us insight about which modalities are indeed informative for detecting social strength in the multimedia research community.

The similarity measures $K^{1\sim7}$ are not necessarily positive semi-definite (p.s.d.). We can force these measures to be p.s.d. to construct kernels by adding a properly scaled identity matrix to the corresponding similarity matrix.

3.3 Optimal Combination of Multiple Kernels

In the above, we define various kernel functions for similarity measures in different modalities. The next challenge is to find the best way of combining these modalities, which plays a key role in social strength modeling. Specially, our goal is to determine a linear combination of multiple kernels to fuse all modalities for measuring the similarity, parameterized by a weight vector $\boldsymbol{\theta} \in \mathbb{R}^{N_k}$:

$$K(u_i, u_j; \boldsymbol{\theta}) = \sum_{t=1}^{N_k} \theta_t K^t(u_i, u_j), \qquad (1)$$

where K^t is the kernel defined under the t-th view of the users, and N_k is the number of modalities (or views).

One naive way is to manually set the weights of different modalities, which however highly relies on domain knowledge and cannot find the optimal combination. In this section, we present a kernel-based learning technique to find the optimal combination of multiple kernels by following the principle of KTA [5, 8].

Specially, given the target matrix $\mathbf{Y}$ ($\mathbf{Y} \in \mathbb{R}^{N_u \times N_u}$), which encodes the existing known relationship among users (through explicit friend lists), we adopt the *kernel alignment* [8] to measure the quality of kernel $\mathbf{K}$ with respect to the target matrix $\mathbf{Y}$ as follows.

Definition 4 (Kernel Alignment). *Let $\mathbf{K}, \mathbf{Y} \in \mathbb{R}^{N_u \times N_u}$ be two kernel matrices such that $\|\mathbf{K}\|_F \neq 0$ and $\|\mathbf{Y}\|_F \neq 0$. Then, the* alignment *between $\mathbf{K}$ and $\mathbf{Y}$ is defined by*

$$\rho(\mathbf{K}, \mathbf{Y}) = \frac{\mathbb{E}\big[tr\, \mathbf{KY}\big]}{\sqrt{\mathbb{E}\big[tr\, \mathbf{KK}\big]\mathbb{E}\big[tr\, \mathbf{YY}\big]}}.$$

Note that the kernel matrices usually have to be centered ($\mathbb{E}[K_{ij}] = 0$) [5]. This centering step can be computed by

$$[\mathbf{K}]_{ij} := K_{ij} - \frac{1}{N_u}\sum_{i=1}^{N_u} K_{ij} - \frac{1}{N_u}\sum_{j=1}^{N_u} K_{ij} + \frac{1}{N_u^2}\sum_{i,j=1}^{N_u} K_{ij}.$$

Given the target graph represented by matrix $\mathbf{Y}$, we maximize the alignment ρ over $\mathbf{K}$ to solve the kernel. The matrix $\mathbf{Y}$ is observed from the Flickr platform. For example, in a friend prediction task, $\mathbf{Y}$ is the mutual contact graph constructed from the profile of each Flickr users. We assume the target kernel matrix is in form of $\mathbf{K} = \sum_{t=1}^{N_k} \theta_t \mathbf{K}^t$ consistent with equation (1), where $0 \leqslant \theta_t \leqslant 1, \sum_t \|\theta_t\|_2 = 1$. Thus the target variable is reduced from $\mathbb{R}^{N_u \times N_u}$ to $\mathcal{M} := \big\{\|\boldsymbol{\theta}\|_2 = 1 \text{ and } \boldsymbol{\theta} \geqslant 0\big\}$. The following theorem guarantees the optimal solution can be computed efficiently [5].

Theorem 1. *The solution $\boldsymbol{\theta}^*$ of the optimization problem*

$$\max_{\boldsymbol{\theta} \in \mathcal{M}} \rho(\mathbf{K}, \mathbf{Y}) \;:\; \mathbf{K} = \sum_{t=1}^{N_k} \theta_t \mathbf{K}^t$$

is given by $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*/\|\boldsymbol{\theta}^*\|$, where $\boldsymbol{\theta}^*$ is the solution of the following quadratic program:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta} \geqslant \mathbf{0}} \boldsymbol{\theta}^\top \mathbf{M} \boldsymbol{\theta} - 2\boldsymbol{\theta}^\top \mathbf{a},$$

where $\mathbf{a}$ is the vector $\big[tr\, \mathbf{K}^1\mathbf{Y}, \ldots, tr\, \mathbf{K}^{N_k}\mathbf{Y}\big]^\top$ and $\mathbf{M}$ is the matrix $[\mathbf{M}]_{kl} := tr\, \mathbf{K}^k\mathbf{K}^l$, for $k, l \in \mathbb{N}_k$.

3.4 Kernel-based Learning to Rank

After obtaining a similarity measure $K_u : \mathcal{U} \times \mathcal{U} \mapsto \mathbb{R}$ in forms of (1), we consider a discriminative model to estimate the social strength between two users. Let y be a $\{1, -1\}$-valued latent variable to indicate whether there is a connection between two users. The social connection strength inference is about to estimate the probability $P(y_{ij} = 1|u_i, u_j)$. To this purpose, we build a logistic normal model based on training pairs. The connection between Flickr users come from the mutual contact $\mathcal{N}$ in Flickr user profile, i.e., $y_{ij} = 1$ if and only if u_j is in the contact list of u_i.

We first introduce a linear model $f(u_i, u_j) = \mathbf{w}^\top \Phi(u_i, u_j)$ parameterized by $\mathbf{w} \in \mathbb{R}^d$ to predict the social strength between u_i and u_j, where the function $\Phi(\cdot, \cdot) : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}^d$ maps a user pair to a feature representation under some specific view or modality. Thus we solve $\mathbf{w}$ by minimizing the regularized loss:

$$\min_{\mathbf{w}} \frac{\lambda}{2}\|\mathbf{w}\|_2^2 + \sum_{u_i, u_j \in \mathcal{U}} l\Big(y_{ij}, \mathbf{w}^\top \Phi(u_i, u_j)\Big), \qquad (2)$$

where λ is a hyper-parameter controlling the trade-off between regularization and loss of prediction. The function $l(\cdot, \cdot) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_+$ measures the loss of prediction. We adopt the logistic loss $l(y_1, y_2) = \ln(1 + e^{-y_1 y_2})$ instead of hinge loss as in traditional SVM-based model, since it allows a natural estimation of the prediction probability. Thus, we can estimate the social strength of two users by

$$P(y_{ij} = 1|u_i, u_j) = \frac{e^{f(u_i, u_j)}}{1 + e^{f(u_i, u_j)}}.$$

To ease the discussion, we denote by $v \in \mathcal{U} \times \mathcal{U}$ a pair of users. According to the representor theorem [21], the solution of the above problem can be expressed in form of

$f(v) = \sum_j \alpha_j K(v, v_j)$, where $\boldsymbol{\alpha}$ is the weight of training pairs, $v_j := [u_{j1}, u_{j2}]$ is an ordered "support" pair of users. Thus the dual objective can be written as

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^{N_p}} \frac{1}{N_p} \mathbf{1}^\top \ln\left(\mathbf{1} + e^{-\mathbf{y} \circ (\widehat{\mathbf{K}} \boldsymbol{\alpha})}\right) + \frac{\lambda}{2} \boldsymbol{\alpha}^\top \widehat{\mathbf{K}} \boldsymbol{\alpha}, \quad (3)$$

where N_p is the number of training pairs, $\widehat{\mathbf{K}} \in \mathbb{S}^{N_u^2 \times N_u^2}$ is the kernel matrix evaluated from a kernel function $\widehat{K} : (\mathcal{U} \times \mathcal{U}) \times (\mathcal{U} \times \mathcal{U}) \to \mathbb{R}$, i.e., defined on the user pairs. The kernel $\widehat{K}$ of two pairs v_i and v_j is computed as

$$\begin{aligned} \widehat{K}(v_i, v_j) &= \Phi(u_{i1}, u_{i2})^\top \Phi(u_{j1}, u_{j2}) \\ &= K(u_{i1}, u_{j1}) * K(u_{i2}, u_{j2}) \\ &\quad + K(u_{i1}, u_{j2}) * K(u_{i2}, u_{j1}), \end{aligned}$$

where we adopt the pairwise kernel [3] on user pairs, and $K(\cdot) : \mathcal{U} \times \mathcal{U} \mapsto \mathbb{R}$ is in the form of (1). The dual objective (3) can be solved efficiently as the import vector machine (IVM) [32].

3.4.1 Training Pair Selection

Constructing the training pair is not trivial for both efficacy and efficiency purpose. When a user u_j appears in the contact list of u_i, it is likely that u_j is a friend of u_i in real life or that the content uploaded by u_j is of the interest of u_i. Therefore, we can treat (u_i, u_j) as a positive pair directly when training the model. However, it is not so trivial for the case of negative pairs. It is a common case that u_i has not noticed his friend u_j has registered. Thus u_i does not add u_j into his contact list. If we simply treat (u_i, u_j) as negative training pairs, the learned model would fail to predict such latent friends effectively. Actually one of our important applications is to reveal such potential friends. It calls for elaborate strategies for sampling negative pairs.

Moreover, there are computational necessity for training pair sampling. Suppose each user has N_p friends on average, we would have $O(N_p \times (N_u - N_p) \times N_u)$ training pairs in total. Such a large scale makes it prohibitive to directly apply pairwise learning to rank algorithm. On the other hand, some training pairs are not helpful for learning the models. The study in active learning shows a small fraction of discriminative training data can often yield satisfying performance. Thus, we propose a two-stage scheme for sampling pairs.

First, we choose the pairs of users that are least likely to be friends to construct negative training pairs. To this end, for each user u_i, we sort the values $K^t(u_i, u_j), j \in \mathbb{N}_u$ with $t \in \mathbb{N}_k$ in ascending order. The users with small similarity to u_i are excluded from the potential friend list of u_i. Since we have multiple N_k kernels defined on users, it is unclear to adopt which one for the sampling purpose. Here we employ a very conservative scheme, that is, we choose the top N_e users that appear in the top N ($N > N_e$) users of u_i under all the similarity measures K^t. For constructing positive pairs, we use K^3 after filtering out the friends according to the contact list, i.e., the kernel counting the mutual comments, to determine the ones most frequently communicating with u_i as positive pairs.

Second, we adopt the active sample selection scheme in IVM [32] during the training phase. Due to the logistic loss function of (2), the solution is non-sparse. However, many of the samples are not necessary to yield a useful solution. We can actively select a pair to expand the set of support vectors at each iteration. In this way, the learning can be significantly sped up while the efficacy of the learned model is preserved [32].

To summarize, the proposed framework relies on a series of similarity measures from various modalities without any assumption about how these measures are obtained. Therefore, any extension of this work only needs to design new features or add new factors to instantiate these measures.

Table 2: The statistics of the collected Flick data.

#user	#group	#image	#tag	#contact
5,000	109,205	5,001,601	116,372	81,447

3.5 Applications

The proposed social strength modeling can lead to a variety of applications, including but not limited to:

- **Friendship Prediction.** In Flickr, each user can add other users into his/her contact list. However, when users are known to each other in real life, or they share quite similar interest in photo styles, the friendship link may not exist explicitly due to the limited searching and browsing functionality. Our framework can predict the implicit links between Flickr users by leveraging the various content and context information.

- **Collaborative Recommendation.** It is very useful to improve user experience by recommending proper objects to users, e.g., interest groups and favorite photos. Such item recommendation tasks can be benefited from the modeled strength as the popularity of these terms are correlated with the social strength among people. Affinity propagation algorithms can be designed based on the learned social strength graph.

- **User-Targeted Advertising.** Similar to recommendation, we can provide user-targeted advertising to a connected component consisting of similar users, so that the advertisements are relevant to user interest through a propagation among similar users.

- **User Search & Browse.** We can rank the user search results according to their social strengths with the one who conducted the query. The user is more likely to find out the targets of his interest. This technique can compensate the lack of informative words matching with the query. Therefore, we expect the results based on traditional simple keyword matching can be greatly improved.

- **Community Visualization.** The applications of visualizing people's social network could be improved by scaling/shading links according to the estimated relationship strengths.

4. EXPERIMENT

We evaluated the proposed social strength modeling techniques on a real-world data set collected from Flickr. We start from a random user as seed and expand the crawling according to its friend list in a breadth-first search manner. We stopped at 5,000 users as we consider such a scale is enough to work with to conduct reliable conclusions. All

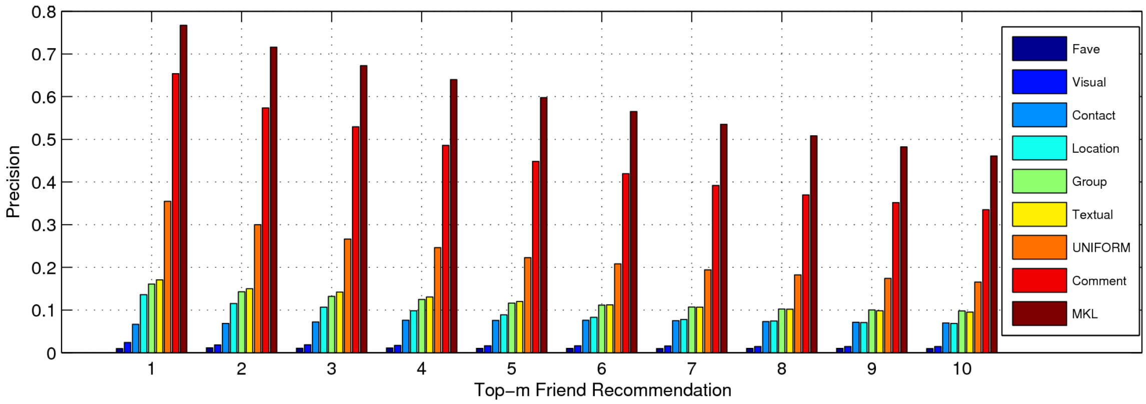


Figure 3: The average accuracy of top-10 friend recommendation by single and combined multiple kernels.

Table 3: The base kernels and the weights by KTA in the friend and group recommendation task.

Kernel	1	2	3	4	5	6	7
Name	Visual	Textual	Comment	Group	Contact	Location	Fave
Friend	.0042	.1007	.6488	.0637	.1042	.0499	.0285
Group	.0183	.1687	.5065	N/A	.2030	.0397	.0299

the associated metadata (as defined in Section 2) are downloaded in the XML format. We evaluate the effectiveness of social strength modeling on recommendation tasks.

4.1 Friend Recommendation

The essence of SSM is to infer the acquaintance between two people. Therefore, the most direct criterion of SSL is to measure the friend recommendation accuracy. It is also an important application for social network sites (including Flickr, though it is not initially designed for social purpose).

We randomly choose 4,000 users for training purpose. The target kernel matrix $\mathbf{Y}$ in Section 3.3 is the friend indicating matrix. The i-th row of $\mathbf{Y}$ is a binary vector indicating whether a user is in the contact list of user u_i. We adopt the learning to rank to infer the social strength based on kernels maximally aligned to $\mathbf{Y}$. For each kernel matrix $\mathbf{K}$, we normalize it by dividing each row with the maximal value at that row. Then we use $\mathbf{K} = (\mathbf{K} + \mathbf{K}^\top)/2$ to make it symmetric. For the ones not satisfying the positive semi-definite (p.s.d.) property (this can be examined by eigen-decomposition), we add $\delta\mathbf{I}$ to make them p.s.d. The valuated kernels include:

- *Single*: A single kernel defined in Section 3.2. By examining the kernel one-by-one, one can observe the effect of each modality clearly;
- *Uniform*: Each kernel is assigned the same kernel weight. This is the baseline showing the result of simple combination, which serves as baseline method;
- *MKL*: The the proposed multiple kernel learning method introduced in Section 3.3 for combining the modalities.

Given a test user, we sort the values in a descending order and extract the top users as recommended friends. The learned weights of the kernels are presented in Table 3. The self-explanatory name of kernel implies its definition in Section 3.2. The top-10 friend recommendation results are plotted in Figure 3, from which we can have the following observations.

First, MKL yields the best performance among all the evaluated kernels. Its top-1 accuracy approaches 80%, and top-10 accuracy approaches 50%. Such relatively high accuracy confirms that inferring social strength on multimedia site is possible. It could open a new perspective for social network mining. For m from 1 to 10, MKL significantly outperforms other kernels. Specially, the naive unform combination only reports half of the accuracy of MKL. This fact verifies the efficacy of the proposed method. Moreover, it implies that the underlying single kernels are complementary. It coincides with the previous conclusion that MKL is effective at concatenating heterogeneous data.

Second, the *Comment* kernel (K^3) reports the second best result. Recall the definition of K^3, it's the number of mutual comments among Flickr users. Its good results show that the *mutual communication* is still the most informative modality when inferring social strength. This fact is consistent with the large volume of works on social network analysis, most of which are mutual communication based. Moreover, note that K^3 is also better than uniform combination. It means some of the modality is actually noisy. Therefore, learning the modality weight in a principle manner is crucial.

Third, we can analyze the contribution of each modality from the results of every single kernel. Specifically:

- The *Textual* kernel (K^2) reports the second best accuracy among all the single kernels, though it is worse than K^3. We can conclude that the photos of friends exhibit stronger semantically similarity.
- The *Visual* kernel (K^1) reports very poor performance. This could be possibly explained by the semantic gap. Besides, we use the mean of photos to measure visual similarity due to the difficulty of dealing with a large scale of photos. This also could lose much information.
- The *Fave* kernel (K^7) produces the worst performance among the 7 base kernels. With m varying from 1 to 10, its accuracy is most zero consistently. This fact means it is not useful when inferring social strength.

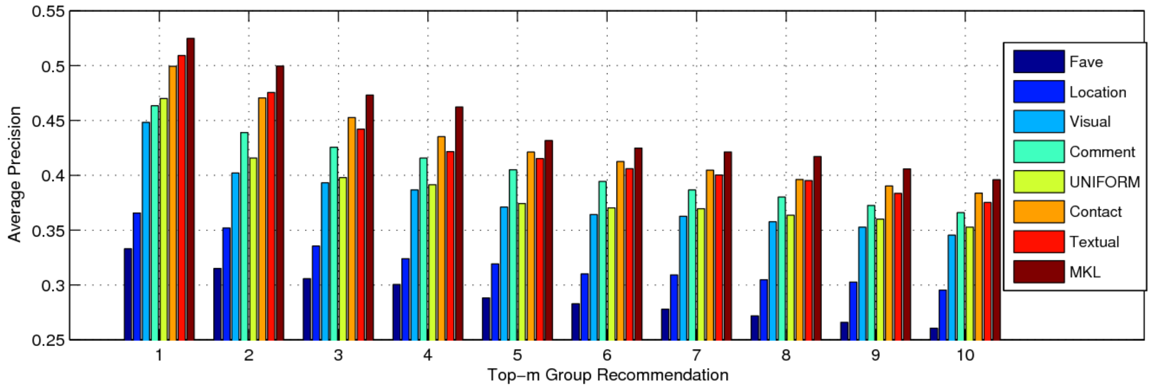


Figure 4: The average accuracy of top-10 group recommendation by single and combined multiple kernels.

- The *Group* kernel (K^4) is quite similar to *Textual* kernel. Their performance can be explained in the same manner as the group membership indicates the semantics of users' photos.

- The *Location* kernel (K^6) is helpful as friends have higher probability residing or taking pictures at the same place.

4.2 Group Recommendation

In this section, we evaluate the interest group recommendation task based on social strength modeling. For the user u_i and group g_k, the recommendation score p is

$$p(u_i, g_k) = \sum_j f(u_i, u_j)\delta(u_j, g_k),$$

where $\delta(u_j, g_k) \in \{0, 1\}$ indicates whether u_j belongs to g_k, f is the social strength predicted by MKL method. We use the *Group* kernel as the target kernel matrix $\mathbf{Y}$ for alignment. We filter out the 1,000 most popular groups for prediction (actually the number of groups is larger than the number of users). We use the average recommendation accuracy to measure performance, as shown in Figure 4. The weights of the six base kernels are listed in Table 3.

We have the following observations from the results. First, MKL outperforms the other candidate kernels as in the case of friend recommendation. The top-10 accuracy is about 40%. It is practical to recommend interest groups to users. This fact confirms the efficacy of MKL again. The *uniform* combination only ranks the 4-th position among the eight methods. It is worse than two base single kernels. Therefor, careless kernel combination can hurt the performance as some kernels are noisy.

For the single kernels, the *Textual* kernel performs the best among the six kernels. We conclude that semantic information is the most important for predicting group membership. This is in contrast to friend recommendation, where mutual friendship is more informative. Each interest group has focused themes. The user annotated tags express the semantics of the photos and thus are more useful.

Difference from friend recommendation, the *Contact* kernel is almost as good as the Textual kernel. We conjecture that user-user relationship is correlated with user-group membership. Both content and social behavior contribute to social strength modeling. The accuracy of *Fave* and *Location* kernels is far from satisfactory. It is a bit surprising that the *Fave* kernel is even worse than the *Location* kernel. Intuitively, the location info is not for group recommendation as it is too coarse, while the *Fave* kernel carries the taste of photos of users, which is related to group membership. The kernel weight learned in Table 4 is quite different from the results in Table 3. Our framework omits the latent variable modeling process such that it can be task dependent.

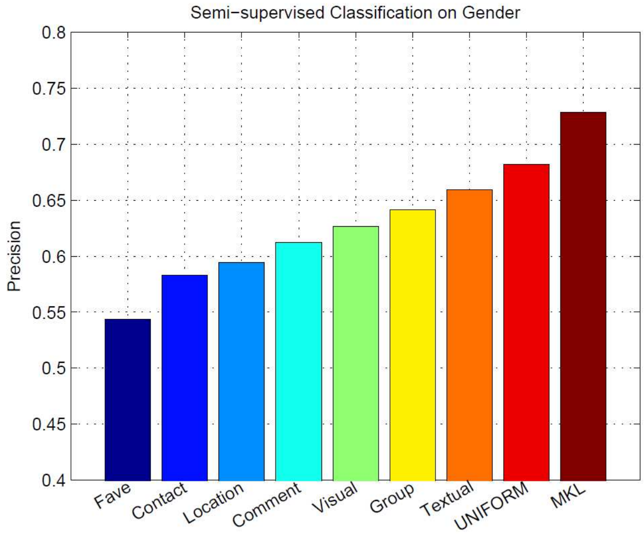
4.3 Semi-supervised Classification

A prior similarity matrix defined over the input samples is crucial for the semi-supervised learning (SSL) (e.g., [31, 33]) as it characterizes the clustering or the manifold structure of the data. Such unsupervised information is helpful for classification purposes as similar data usually share similar labels. The social strength graph learned in this paper can play an important role for the SSL. We can plug it into [33], which is one of the most influential SSL algorithms, to predict the *gender* and the *location* of a Flickr user. For the first task, we predict the *gender* of a user (i.e., male or female). For the second, we first detect the most frequent location from all the locations associated with a user, and then predict whether s/he has been there. The target matrix $\mathbf{Y}$ is set as the contact graph for both tasks. We set the ratio of the supervised samples to the overall samples as 0.7. The parameters in the SSL algorithm are tuned by the cross-validation.

Figure 5 shows the results. We see that MKL performs the best in both tasks. This verifies that the modalities are complementary for the classification tasks. Regrading the single kernel, the *Textual* kernel performs the best, even comparable with MKL for the location prediction task. Recall that this also holds for the friend and group recommendation tasks. We see that the tag-based kernel is the most effective for SSM. The reason is that the semantic information carried by tags is the most meaningful for characterizing a user. On the other hand, among the four tasks, the *Fave* kernel performs the worst. Recall that it is defined by the count of photos users like. However, it is usually dominated by the quality of photos, instead of social relationship.

5. DISCUSSION AND RELATED WORK

Our work in this paper is closely related to the following research topics: 1) Flickr data mining, 2) social strength modeling, and 3) learning to rank.



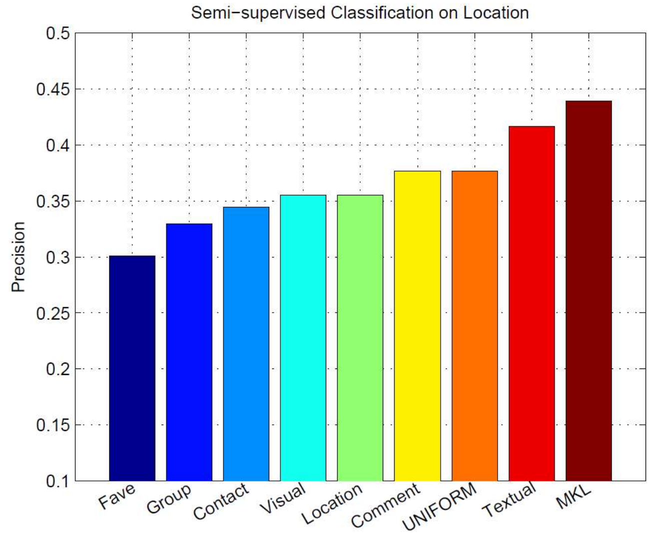


Figure 5: The evaluation of inferred social strength graph for semi-supervised classification. The left figure is the accuracy of *gender* prediction, the right figure is the accuracy of *location*.

5.1 Data Mining with Flickr Data

There exists rich research on mining Flickr data. In [17], the author conducted probabilistic latent semantic analysis on the Flickr interest groups, where each group is abstracted to be a collection of tags annotated to the images belonging to this group. This work is group-oriented and it is unclear how strong the relationship between users is. 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 [16] [18]. These works are also *group-oriented*.

There are works focused on how to make use of the metadata of Flickr images to facilitate other applications. For example, geo-tag analysis [11] [22], automatic tag annotation [26], tag modeling [27]. 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 solve our *user-orientated* problem which involves heterogenous forms of data. Our proposed framework by predicting the social strength incorporates the rich content and context information on Flickr effectively, whereas previous works cannot make use of multiple modalities effectively.

5.2 Social Strength Modeling

Our task in this work is to infer the social strength among Flickr users. This connects to social strength modeling or link predictions, where quite a few representative works exists [1, 9, 10, 23, 25, 29]. However, few existing works make use of the rich content and context data of Flickr images to infer the social relationship between Flickr 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 on traditional works.

The unique characteristics of Flickr data provides more flexibility in social network mining. The idea of using Flickr data to infer social relationships has been recently proposed [20, 28, 30]. Wu *et al.* tried to reveal the closeness of people by face detection techniques [28]. 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 [20]. Due to the diversity of the images in Flickr, this rule-based method is not applicable for our task. 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 [30]. As aforementioned, the initial group membership is subjective, incomplete, and noisy, which may not be reliable enough to serve as supervised information.

Recently, visual similarity, especially face similarity, has been leveraged to discover the relationship between roles in a movie [9] [25]. Our work is different from those in that we are investigating the heterogeneous forms of data in a more complex context, i.e., social communities. Xiang *et al.* proposed a hybrid generative-discriminative model which assumes a latent variable measuring the social strength computed from the user profile similarities [29]. The interactive activities are results of such latent variables. However, the profile data of Flickr is not so complete since it is not initially designed to be a social network site. It is not sufficient to infer the social strength solely based on the profile similarities. In addition, separating the interaction graph from profile similarity graph increases modeling complexity.

5.3 Multiple Kernel Learning to Rank

Our main technique is built on two topics, *kernel learning* and *learning to rank*. The most crucial element of a kernel method is *kernel*, which is in general a function that defines an inner product between any two examples in some induced Hilbert space [7]. Recent years have witnessed the active research of learning effective kernels automatically from data [13]. The most popular example technique for kernel learning is *Multiple Kernel Learning* [13], which aims at learning a linear (or convex) combination of a set of predefined kernels in order to identify a good target kernel for the applications. Besides the work of improving the efficiency of MKL, a number of extended MKL techniques have been proposed to improve the regular linear MKL method (e.g., [5, 12]). Here we adopt the two-stage method proposed by Cortes *et al.* [5] as it is computationally simple and works well both theoretically and practically. We learn the weight of each modality by maximizing the inner product between the combined kernel and the target kernel, which can be task dependent to make our solution more flexible.

Learning to rank is an active research topic in machine learning and information retrieval community [14]. It provides a principled and effective paradigm for IR applications. Here we are aware that the social strength essentially ranks the degree of affinity among people. If we can rank such relationship properly, then we can solve the SSM problem. Thus, we borrow the key idea of learning to rank to use and rectify the kernels learned in the first stage.

6. CONCLUSION

Instead of detecting binary social ties in traditional social networking, this paper studies how to infer the continuous social strength in the social media communities. The learned graph is more delicate and informative and can be applied in a lot of important applications, including friend prediction, item recommendation, visualization, and so on. Our key ideas are threefold: 1) leveraging the multiple data sources to compute multiple similarity functions, which can be enforced to satisfy positive semi-definite property to construct valid kernels, 2) employing the kernel target alignment algorithm to learn the weight of each modality and using the weighted summation of base kernels as the ideal kernel, and 3) devising the learning to rank framework based on the learned kernel to infer the social strength. Our techniques are discriminative and do not require any assumption of the underlying parametric statistical models that governs the social strength. Empirical results verified its effectiveness. We hope this work could call for more attention to the social strength modeling in this community.

7. REFERENCES

[1] L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social Networks*, 25: 211–230, 2001.

[2] R. A. Baeza-Yates and B. A. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.

[3] A. Ben-Hur and W. S. Noble. Kernel methods for predicting proteinÍCprotein interactions. In *Proc. of ISMB (Supplement of Bioinformatics)*, page 38–46, 2005.

[4] C. Cortes. Invited talk: Can learning kernels help performance? In *Proc. of ICML*, page 161, 2009.

[5] C. Cortes, M. Mohri, and A. Rostamizadeh. Two-stage learning kernel algorithms. In *ICML*, pages 239–246, 2010.

[6] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[7] N. Cristianini and J. Shawe-Taylor. *An introduction to support Vector Machines: and other kernel-based learning methods*. Cambridge University Press, New York, NY, USA, 2000.

[8] N. Cristianini, J. Shawe-Taylor, A. Elisseeff, and J. S. Kandola. On kernel-target alignment. In *Proc. of Advances in Neural Information Processing Systems*, pages 367–373, 2001.

[9] L. Ding and A. Yilmaz. Learning relations among movie characters: A social network perspective. In *Proc. of ECCV*, pages 410–423, 2010.

[10] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proc. of CHI*, pages 211–220, 2009.

[11] L. S. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury. How flickr helps us make sense of the world: context and content in community-contributed media collections. In *ACM Multimedia*, pages 631–640, 2007.

[12] M. Kloft, U. Brefeld, S. Sonnenburg, P. Laskov, K.-R. Muller, and A. Zien. Efficient and accurate l_p-norm multiple kernel learning. In *Proc. of Advances in Neural Information Processing Systems*, 2009.

[13] G. R. G. Lanckriet, N. Cristianini, P. L. Bartlett, L. E. Ghaoui, and M. I. Jordan. Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, 5:27–72, 2004.

[14] T.-Y. Liu. Learning to rank for information retrieval. In *Proc. of SIGIR*, page 904, 2010.

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[16] R. A. Negoescu, B. Adams, D. Q. Phung, S. Venkatesh, and D. Gatica-Perez. Flickr hypergroups. In *Proc. of ACM Multimedia*, pages 813–816, 2009.

[17] R. A. Negoescu and D. Gatica-Perez. Analyzing flickr groups. In *Proc. of CIVR*, pages 417–426, 2008.

[18] R. A. Negoescu and D. Gatica-Perez. Topickr: flickr groups and users reloaded. In *Proc. of ACM Multimedia*, pages 857–860, 2008.

[19] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2161–2168, June 2006.

[20] J. L. P Singla, H Kautz and A. Gallagher. Discovery of social relationships in consumer photo collections using markov logic. In *Proc. of CVPR workshop*, 2008.

[21] B. Schölkopf, R. Herbrich, and A. J. Smola. A generalized representer theorem. In *Proc. of COLT/EuroCOLT*, pages 416–426, 2001.

[22] P. Serdyukov, V. Murdock, and R. van Zwol. Placing flickr photos on a map. In *Proc. of SIGIR*, pages 484–491, 2009.

[23] P. Singla and M. Richardson. Yes, there is a correlation: - from social networks to personal behavior on the web. In *Proc. of WWW*, pages 655–664, 2008.

[24] V. N. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

[25] Y. Wang, T. Mei, and X.-S. Hua. Community discovery from movie and its application to poster generation. In *Proc. of International MultiMedia Modeling Conference*, pages 107–117, 2011.

[26] L. Wu, S. C. H. Hoi, R. Jin, J. Zhu, and N. Yu. Distance metric learning from uncertain side information with application to automated photo tagging. In *Proc. of ACM Multimedia*, pages 135–144, 2009.

[27] L. Wu, X.-S. Hua, N. Yu, W.-Y. Ma, and S. Li. Flickr distance. In *Proc. of ACM Multimedia*, pages 31–40, 2008.

[28] P. Wu and D. Tretter. Close & closer: social cluster and closeness from photo collections. In *Proc. of ACM Multimedia*, pages 709–712, 2009.

[29] R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *Proc. of WWW*, pages 981–990, 2010.

[30] J. Yu, D. Joshi, and J. Luo. Connecting people in photo-sharing sites by photo content and user annotations. In *Proc. of ICME*, pages 1464–1467, 2009.

[31] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schlkopf. Learning with local and global consistency. In *Advances in Neural Information Processing Systems*, 2005.

[32] J. Zhu and T. Hastie. Kernel logistic regression and the import vector machine. In *Proc. of Advances in Neural Information Processing Systems*, pages 1081–1088, 2001.

[33] X. Zhu, Z. Ghahramani, and J. D. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proc. of International Conference on Machine Learning*, pages 912–919, 2003.