FRIENDTRANSFER: COLD-START FRIEND RECOMMENDATION WITH CROSS-PLATFORM TRANSFER LEARNING OF SOCIAL KNOWLEDGE

Ming Yan[1,2], *Jitao Sang*[1,2], *Tao Mei*[3], *Changsheng Xu*[1,2]

[1]National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China
[2]China-Singapore Institute of Digital Media, Singapore, 139951, Singapore
[3]Microsoft Research Asia, Beijing, P.R. China
{ming.yan, jtsang, csxu}@nlpr.ia.ac.cn,tmei@microsoft.com

ABSTRACT

The emergence of various and disparate social media platforms has opened opportunities for the research on cross-platform media analysis. This provides huge potentials to solve many challenging problems which cannot be well explored in one single platform. In this paper, we investigate into cross-platform social relation and behavior information to address the cold-start friend recommendation problem. In particular, we conduct an in-depth data analysis to examine what information can better transfer from one platform to another and the result demonstrates a strong correlation for the bidirectional relation and common contact behavior between our test platforms. Inspired by the observations, we design a random walk-based method to employ and integrate these convinced social information to boost friend recommendation performance. To validate the effectiveness of our cross-platform social transfer learning, we have collected a cross-platform dataset including 3,000 users with recognized accounts in both Flickr and Twitter. We demonstrate the effectiveness of the proposed friend transfer methods by promising results.

Index Terms— cross-platform, social network analysis, friend recommendation, random walk

1. INTRODUCTION

Imagine you are an experienced user in Facebook, where a friend network is carefully maintained and social activities are updated everyday. When Google released its Social Network Service (SNS) platform, Google Plus, it will be awkward if you desire to try this new service, since no ready friends are available and the new platform knows nothing about your social interests for intelligent recommendation. These awkwardness will make your SNS experience "non-social" at all. The question is, how can we exploit social relations and behaviors of active users in one platform, to help another platform perform the cold-start friend recommendation task as well as help users to construct their initial friend networks.

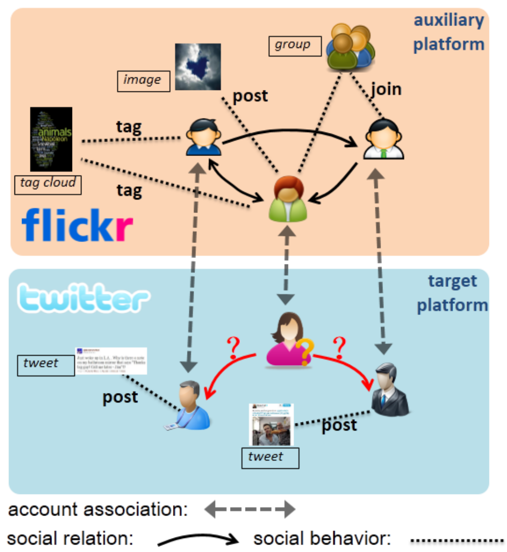

Fig. 1. A brief illustration for our cross-platform user data

As social media continues to lead a surge of interest, most users join in multiple social media platforms simultaneously, providing the possibility to perform user-centric cross-platform data analysis. In this paper, we observe that users' social relation and behavior in one platform offers important knowledge about social interest in another platform. Based on the observations, we integrate social information transfer to address the cross-platform friend recommendation problem.

1.1. Related Work

Extensive works have been conducted on analyzing social relations and recommending friends based on user's rich social context information in the same platform. [1] proposed to learn an optimal linear combination of many different relations by leveraging multiple data types and social contexts. [2] recommended friends based on user's visual similarity and geographic similarity in online photo-sharing community. [3] investigated the relationship between semantic similarity of user profile entries and the social network topology and finds the consistency between the user keyword profile and their social relation. These methods work well when the target user already have a stable network and closely related friends or rich social context information is available. However, in the situation of entering into a totally new social network, no user information is known except for little noisy registration profile, where the above methods tend to fail.

Recently, cross-platform user data has attracted industrial as well as academic attentions. In industry, many account fusion or management tools have been released to facilitate individuals and organizations to aggregate multi-platform news or activities. In academia, researchers have also begun to realize the importance of cross-platform user data in social media analysis and applications. [4] investigated tag profiles for the same user in Flickr, Twitter and Delicious, and discovered consistency and replication characteristics in cross-platform user behavior. [5] utilize the rich text information in news domain to enrich the user semantic representation in twitter and in turn boost the performance of personalized news recommendation. Social topology structure has been ignored in the aforementioned works. So far as we know, no work has systematically analyzed what social relation and behavior can better transfer user preferences between different platforms, and how to combine them for improved applications.

1.2. Overview of Our Approach

In this work, we employ Flickr and Twitter [6] as the test platforms: utilizing users' rich social information in the source platform Flickr, to initialize the construction of new friend network in the target platform Twitter. We assume users in Twitter have no available social relations or behaviors, which forms as a *cross-platform cold-start friend recommendation* problem. First, we crawled a cross-platform dataset with user account association between Flickr and Twitter. Social relations, such as *contact* in Flickr and *follow* in Twitter, and social behaviors, such as *tag* in Flickr and *tweet* in Twitter are also collected in each platform (shown in Fig. 1). Comprehensive global as well as local analysis is performed in the crawled dataset, where we observed what social relation and behavior better contribute to knowledge transfer under the cross-platform scenario. Based on the observations, corresponding information is utilized for cross-platform friend recommendation and a random walk-based method is designed to combine the strong social relation with convinced social behavior. Improved recommendation performance is shown in experiments on the real-world dataset.

Therefore, the main contributions can be summarized as follows:

- We address the cross-platform friend recommendation problem by employing the rich social relation and behavior information in one platform to help the initialization of new user network in a new platform. To the best of our knowledge, this represents one of the first attempts to deal with the notorious cold-start problem from the perspective of cross-platform user data utilization.
- We conduct an in-depth data analysis on our crawled cross-platform user dataset. Several observations are made about what social information better contributes to knowledge transfer, which inspires our friend recommendation implementation. The observations also provide valuable guidance for future researches on cross-platform applications.
- A random walk-based method is proposed to combine the convinced social relation and behavior information, whose effectiveness is validated in experimental results. This simply integrated scheme opens up opportunities to extend the scope of cross-platform data in user-based applications.

2. CROSS-PLATFORM SOCIAL ANALYSIS

In this section, we first describe how we collect our cross-platform user dataset. Then we make some quantitative analysis from two aspects, i.e., social relation and social behavior. Conclusions are made on what information can be better transferred to another platform, which further inspires the following cross-platform friend recommendation methods.

2.1. Data Collection

To obtain a collection of users who have both accounts in Flickr and Twitter, we started from Google+ website where people provide many external links to their other social network homepages and collected about 40K users in total. Then we kept only those who have both Flickr and Twitter accounts and got 3,003 users. We further filtered the user dataset so that every user has at least one friend with the others in the dataset resulting in the final 1,457 user dataset. The users' rich social context information are also downloaded from their Flickr and Twitter accounts respectively. Most of the social analysis work is done in this dataset.

In the dataset, we can represent every cross-platform user as a 2-dimensional map first:

$$R := < U_F, U_T >$$

where U_F represents the user information in Flickr and U_T represents his information in Twitter.

Focusing only on the social relation and behavior information investigated in this paper, we further represent the user account U_F and U_T as follows:

$$U_F := < U, C, T, G >, U_T := < U, F_r, F_l >$$

where U is the user ID in both representations, C and G are the contact list and interested group list respectively in Flickr, T is the tag cloud obtained from the whole image set uploaded by the Flickr user U, F_r and F_l are the friend list and follower list in Twitter.

2.2. Social Relation Analysis

We find that the online photo-sharing site Flickr and micro-blogging site Twitter are both unidirectional social networks. This means that people can add any person they are interested in to their friend list without confirmation by both sides. As a result, people may have three types of relations with others, i.e., no relation, unidirectional relation(A follows B but B doesn't follow back) and bidirectional relation(A follows B and B also follows A back). Under cross-platform situation, we also define an advanced fourth social relation, cross-follow relation, which is defined as follows:

Definition 1 (Cross follow relation)**.** *Given two users A and B both in two social platforms, if user A follows B in both platforms, we call A cross follows B.*

2.2.1. Flickr is more closely connected than Twitter

To investigate the differences of relations between Flickr and Twitter, we first made some comparison on the global relation situation between them. The result is shown in Table 1. We can see that Twitter has a much larger social circle than Flickr, about 10 times larger, and users in Flickr often have a bigger probability to follow back which indicates a stronger relationship in Flickr than Twitter. Moreover, 55 percent of friend relationship in Flickr is cross-follow relationship in both platforms which means that on average among the 111.93 friends for every user in Flickr about 60 of them are still the user's friends in Twitter platform. This gives us a clear sign that social relation plays an important role in our cold-start friend recommendation and the transfer from Flickr to Twitter may be more reliable for the closer relationship in Flickr.

social platform	bidirectional follow ratio (in friends)	cross follow ratio (in friends)	Avg(Contact Num)
Flickr	0.5843	0.55	111.93
Twitter	0.4503	0.315	1273.84

Table 1. Global relation situation between Flickr and Twitter

2.2.2. Bidirectional relation is more reliable

As users have so many different types of relations with others in these social platforms, it's easy to understand the relation type may also be a critical factor influencing the social transfer. So we take one step further to figure out what type of social relations can be better transferred to another platform. Therefore, we split the whole friend pairs into two parts, i.e., unidirectional friend pairs and bidirectional friend pairs, in Flickr and Twitter respectively. Then we counted how many of the friend pairs in one social platform become bidirectional relation, unidirectional relation or no relation in the other

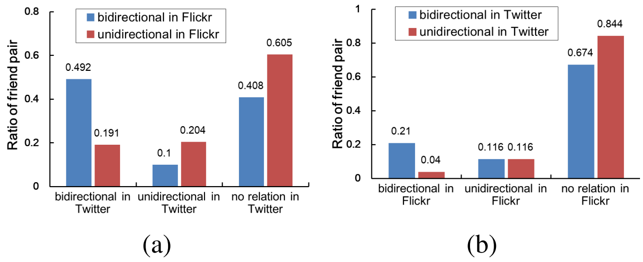

Fig. 2. Comparison of relation type transfer situation.

social platform in each part of our splitted friend pair respectively. The result is shown in Figure 2. We can see that no matter in what direction, bidirectional follow relation has a bigger probability to keep its relation than the unidirectional follow relation when transferring to another social platform which indicates that bidirectional relation can be better transferred to another social platform. Especially in the direction from Flickr to Twitter, about half of the total bidirectional friend pairs in Flickr still keep the bidirectional follow relations in Twitter and about 10% change to the unidirectional follow relations. It means that about 6 out of 10 bidirectional relations in Flickr are cross follow relations in both Flickr and Twitter. Moreover, by comparing the situation in two directions we can see that it's easier to keep the relation alive when transferring from Flickr to Twitter which again verifies our observation in the previous subsection.

2.3. Social Behavior Analysis

Other than the multiple social relations, users will also have rich social behavior information in the social platform. In this subsection, we investigate: (1) whether there exists consistency between the user's rich social behaviors and their social relations; (2) what social behaviors can be better transferred to another platform.

2.3.1. Social behavior and social relation have some sort of consistency

As our work mainly focus on using rich social relation and behavior information in Flickr to help construct the new user network in Twitter, we only made extensive analysis on the user behavior information in his Flickr network. The related behavior information we leverage for user modeling has been described in section 2.1. Given two users i and j in their Flickr accounts, where $U_{Fi} :=< U_i, C_i, T_i, G_i >$, $U_{Fj} :=< U_j, C_j, T_j, G_j >$, we model the user similarity s_{ij} in three ways as follows:

- By Common Contact: $s_{ij} = \frac{|C_i \cap C_j|}{|C_i|}$
- By Common Interested Group: $s_{ij} = \frac{|G_i \cap G_j|}{|G_i|}$
- By Tag Cloud: In this method, we adopt the bag-of-word model to represent the user in text space. With a tag dictionary built from the total tags in our dataset, we map every

user's tag cloud T into the dictionary space and represent them by the traditional *tf-idf* weighting method. In this way, we assume that every user is represented as a vector $t \in R^d$ via his tags and the final user similarity is calculated by the common cosine measure: $s_{ij} = \frac{{t_i}^T t_j}{\sqrt{({t_i}^T t_i)({t_j}^T t_j)}}$

It's easy to understand that users with social relations may conduct some behavior in a more similar way. To explore this point of view, we split all the Flickr user pairs into two parts, i.e., with relations and without relations, respectively and calculate the average of the common contact number(CCN), common interested group number(CGN) and tag-based similarity(TBS) in the respective user pairs. Table 2 shows the result. We can see that strong consistency exists between the user's social behaviors listed in the table and their social relations, users generally have more common contacts and common interested groups with their friends than other people. Besides, friends may also use more similar tags in their uploaded images which may be influenced by each other.

relation or not	Avg.CCN	Avg.CGN	Avg.TBS
with relation	20.2777	4.8649	0.0550
without relation	2.4259	1.9430	0.0211

Table 2. Comparison of the social behaviors between the user pairs with and without relations

2.3.2. Common contact and tag-based profile can promote cross-follow relations

To further facilitate our cross-platform transfer analysis, we again collected all the friend pairs in Fickr and split them into two parts, i.e., the pairs that only be friends in Flickr and the pairs that have cross-follow relationship in both Flickr and Twitter. Then we made the same data analysis as in the previous subsection and the result is shown in Table 3. We can see that the common contact number and tag-based similarity shows a bigger value in the cross-follow relation pairs while the common interested group number is a little smaller instead. Therefore we can assume that the common contact and tag-based profile behavior can promote the establishment of the cross-follow relations at some extent and there is a bigger probability for the friends in one platform to cross follow each other in a new platform when they share more common contacts or their tag-based profile is more similar.

relation type	Avg.CCN	Avg.CGN	Avg.TBS
only follow in Flickr	15.7250	5.3340	0.0403
cross follow both in Flickr and Twitter	23.0743	4.5768	0.0913

Table 3. Comparison of the social behaviors between the friend pairs with and without cross-follow relations

The above user data analysis inspires us the design of the cold-start friend recommendation problem by jointly leveraging the rich social relations especially the bidirectional follow relation and the convinced user behaviors such as common contact and tag-based profile in another social platform. The comparison of the social behaviors for the bidirectional friend pairs in Table 4 more or less gives us a sign of this. In the next, we propose a random walk-based method to better combine the two types of information.

relation type (all bidirectional in Flickr)	Avg.CCN	Avg.CGN	Avg.TBS
only follow in Flickr	17.6505	5.7348	0.0407
cross follow both in Flickr and Twitter	26.1505	4.8380	0.0996

Table 4. Comparison of the social behaviors between the bidirectional friend pairs with and without cross follow

3. CROSS-PLATFORM FRIEND RECOMMENDATION

In our application, the goal is to recommend the top-k friends for experienced users in Flickr who just come to a new social platform Twitter. We first introduce the random walk ranking formulation, and then provide the implementation for social relation and behavior collaboration.

3.1. Random Walk On User Graph

Random walk methods have been widely used in machine learning and information retrieval fields [7, 8, 9]. Here we formulate the top-k friend recommendation solution as a random walk with restart over the user graph, where users are nodes and the edges between them are weighted by their similarity on social behavior.

Assuming we have a user graph with n nodes in the random walk process, we use $r_k(i)$ to denote the relevance score of node i at iteration k and the final recommendation results are obtained by ranking the user nodes according to their relevance score $r_k(i)$ in the stable state. A transition matrix $P = [p_{ij}]_{n \times n} = [p(j|i)]_{n \times n}$ is used to control the transition of a random walk process, p_{ij} is the probability of the transition from node i to node j. Meanwhile, we also want to give an approximate relevance score to every node in the random walk process as their initial score and this can be fulfilled by the initial bias vector v in the random walk with restart method. Thus the random walk process can be formulated as

$$r_k(j) = \alpha \sum_i r_{k-1}(i) p_{ij} + (1 - \alpha) v_j \quad (1)$$

where α is a tradeoff parameter that belongs to (0,1) which controls the effectiveness of the transition matrix P and the initial bias vector v. The iteration of Eq. (1) converges to a stable state r_π. Here we give the final stable state result directly

$$r_\pi = (1 - \alpha)(\mathrm{I} - \alpha P)^{-1} v \quad (2)$$

3.2. Social Relation and Behavior Collaboration

Now we focus on how to jointly utilize the convinced social information in Flickr to help construct the new user network for the same person in Twitter by our random walk method. In our method, the key question turns to how to design the transition Matrix P and initial bias vector v properly.

Our data analysis in section 2.2 shows the effectiveness of the social relation especially the bidirectional relation in one platform for the establishment of the cross-follow relation. It's also easy to understand this because users who have strong relationship in one social platform are more likely to become friends in another social platform. Therefore, given a query user u_i, we design the initial vector v as an indicator for the follow relation between the query user u_i and the candidate user u_j as follows

$$v_j = \begin{cases} 0, if\ u_i\ has\ no\ relation\ with\ u_j \\ 1, if\ u_i\ has\ unidirectional\ relation\ with\ u_j \\ 2, if\ u_i\ has\ bidirectional\ relation\ with\ u_j \end{cases}$$

Then we regularize v_j with its L1-norm so that v_j belongs to (0,1).

We can also see in section 2.3 that common contact number and tag-based user profile can enhance the cross follow relationship. Inspired by this, we model the transition matrix P with these two kinds of social behavior respectively, its element p_{ij} can be computed as

$$p_{ij} = \frac{s_{ij}}{\sum\limits_{k} s_{ik}}$$

where s_{ij} denotes the pairwise user similarity computed by their common contact number or tag-based user profile illustrated in section 2.3.1.

In this way, we not only consider the explicit binary follow relation information but also connect all the users by their specific implicit social behavior. The random walk process will more likely jump to the nodes with higher pairwise similarity. Therefore, we can get a list of users strongly correlated by their social behavior with higher relation scores and these users are more likely to be friends with the query user. The promising results of the random walk methods will be presented in the next section.

4. EXPERIMENTS

4.1. Experimental Settings

We conducted experiments on the collected 1457 user dataset mentioned in section 2.1. To facilitate our cold-start friend recommendation application, we selected a dense user dataset with 328 users in which each user has at least 3 friends with the others in this dataset. On average, users have 8.01 friends in Flickr and 13.34 friends in Twitter in this dataset. The total 315k images with tags uploaded by the 328 users were also downloaded to conduct our tag-based user modeling. In the random walk process, the tradeoff parameter α is set to 0.9.

The top-k average precision, recall and F-score are utilized as the evaluation metrics [10] and user's friendlist in Twitter is used as the ground truth. For every user in our dataset we recommend the top k users to their Twitter account by leveraging their rich social information in Flickr account, the final evaluation value is the average over all the users.

We compared our random walk-based method with the following baseline methods:

- Friend(*Frd*): Directly recommend the friends in Flickr, the bidirectional friends have a priorty to the unidirectional friends;
- Common Contact(*CC*): Recommend the users who have more common contacts in Flickr with the query user;
- Tag: Recommend the users who have bigger tag-based similarity in Flickr with the query user;
- Group(*Grp*): Recommend the users who have more common interested groups in Flickr with the query user;
- Friend + Common Contact (*Frd*+*CC*): Recommend the users who are friends and have more common contacts in Flickr with the query user;
- Friend + tag (*Frd*+*Tag*): Recommend the users who are friends and have bigger tag-based similarity in Flickr with the query user;
- Friend + group (*Frd*+*Grp*): Recommend the users who are friends and have more common interested groups in Flickr with the query user;

4.2. Experiment Results

Figure 3 shows the F-score of all the examined methods. The *Our 1*, *Our 2*, *Our 3* methods represent our random walk methods which combines the common contact behavior, tag-based profile and common interested group behavior with the friend relation respectively. We can draw several conclusions from these results.

First, among all the methods that only utilize one type of social information, the *Frd* and *CC* methods have significantly better results than the *Tag* and *Grp* methods. This indicates the effectiveness of the friend relation and the common contact behavior information in our cross-platform application. Second, all the methods that combine the social behavior information with the social relation information achieve superior performance than the methods which only utilize one type of information. This validates the power of the fusion of social information. Besides, our random walk method can always improve the final recommendation performance using the same types of social information as in our baseline which verifies a strong ability of our random walk method to fuse different kinds of social information. Third, our method that combines the friend relation and common contact behavior gets the best result among all methods that combine two types of information. It again proves the effectiveness of the friend

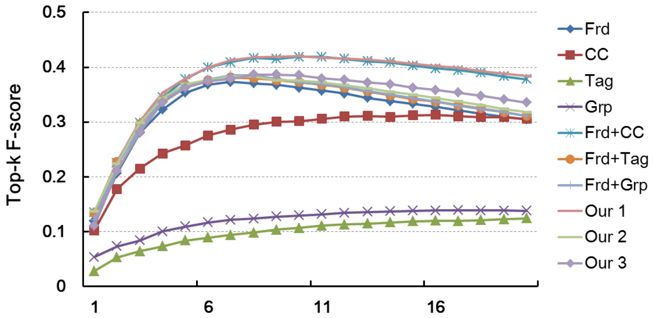

Fig. 3. The top-k F-score for all the examined methods.

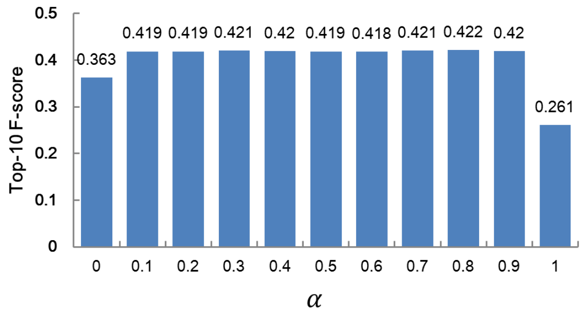

Fig. 4. The top-10 F-score for different α.

relation and common contact behavior in our application.

4.3. Discussions

We first analyzed the sensitivity of the tradeoff parameter α in Eq. (1). Figure 4 shows the result with respect to different values of α utilizing the friend and common contact information. We can see that the performance is smooth when α varies in a range from 0 to 1 and only a little smaller when it is exactly 0 or 1 where only one type of social information is considered. The performance is not sensitive to the change of parameter and the best result is achieved when $\alpha = 0.8$. We set our parameter $\alpha = 0.9$ because the result is better when considering our overall result by utilizing different social information which means our best result can be better improved if we only focus on this result.

In our experiment results, we can see that when $k > 5$ the *Our 3* method can get even better results than the *Our 2* method which means the common interested group behavior is more valuable than the tag-based profile which is against our observation in section 2.3.2. It may be due to the fact that although the tag information can contribute more to the cross-follow relation in the friend pairs, it can't separate the friend pairs from the pairs without any relation as well as the common interested group behavior illustrated in Table 2. In other words, the common interested group behavior may find more friend pairs in one platform in which the cross follow relation have a bigger probability to be established. Therefore, although the cross follow ratio in friend pairs may be smaller, the number of friend pairs found by this behavior is larger and the final result is thus better.

5. CONCLUSION

In this paper, we have introduced the idea of leveraging the rich social information in one platform to help the construction of a new user network in another platform on user level. It is shown that friend relations (especially bidirectional relation) and common contact behavior can be better transferred to another social platform based on our data analysis. The proposed random walk-based method shows a robust fusion for different social information in all situations. Future work includes the fusion of multiple social information in different modalities across multiple social platforms. We are also expecting to find some general cross platform social patterns in nature in our future work.

6. ACKNOWLEDGMENT

This work is supported in part by National Basic Research Program of China (No. 2012CB316304), National Natural Science Foundation of China (No. 61225009), and Beijing Natural Science Foundation (No. 4131004). This work is also supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

7. REFERENCES

[1] D. Cai, Z. Shao, X. He, X. Yan, and J. Han, "Mining hidden community in heterogeneous social networks," in *LinkKDD*, 2005, pp. 58–65.

[2] Ting Yao, Chong-Wah Ngo, and Tao Mei, "Context-based friend suggestion in online photo-sharing community," in *ACM Multimedia*, 2011, pp. 945–948.

[3] Prantik Bhattacharyya, Ankush Garg, and Shyhtsun Felix Wu, "Analysis of user keyword similarity in online social networks," *Social Netw. Analys. Mining*, pp. 143–158, 2011.

[4] F. Abel, E. Herder, G.J. Houben, N. Henze, and D. Krause, "Cross-system user modeling and personalization on the social web," *UMUAI*, vol. 22, no. 3, pp. 1–42, 2011.

[5] F. Abel, Q. Gao, G.J. Houben, and K. Tao, "Analyzing user modeling on twitter for personalized news recommendations," *User Modeling, Adaption and Personalization*, pp. 1–12, 2011.

[6] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon, "What is Twitter, a social network or a news media?," in *WWW '10*. 2010, pp. 591–600, ACM.

[7] Dong Liu, Xian-Sheng Hua, Linjun Yang, Meng Wang, and Hong-Jiang Zhang, "Tag ranking," in *WWW'09*, pp. 351–360.

[8] Winston H. Hsu, Lyndon S. Kennedy, and Shih-Fu Chang, "Video search reranking through random walk over document-level context graph," in *ACM Multimedia*, 2007, pp. 971–980.

[9] Ting Yao *, Yuan Liu, Chong-Wah Ngo, and Tao Mei, "Unified entity search in social media community," in *Proc. of International World-Wide Web Conference (WWW)*, 2013.

[10] Chunjie Zhang, Jing Liu, Qi Tian, Changsheng Xu, Hanqing Lu, and Songde Ma, "Image classification by non-negative sparse coding, low-rank and sparse decomposition," in *Computer Vision and Pattern Recognition*, 2011, pp. 1673–1680.