

Social-oriented visual image search



Shaowei Liu^{a,*}, Peng Cui^{a,1,2}, Huanbo Luan^b, Wenwu Zhu^{a,1,2}, Shiqiang Yang^{a,1,2}, Qi Tian^c

^a Department of Computer Science, Tsinghua University, Beijing, China

^b School of Computing, National University of Singapore, Singapore

^c Department of Computer Science, University of Texas at San Antonio, United States

ARTICLE INFO

Article history:

Received 15 September 2012

Accepted 27 June 2013

Available online 30 August 2013

Keywords:

Social image search

Image reranking

Social relevance

ABSTRACT

Many research have been focusing on how to match the textual query with visual images and their surrounding texts or tags for Web image search. The returned results are often unsatisfactory due to their deviation from user intentions, particularly for queries with heterogeneous concepts (such as “apple”, “jaguar”) or general (non-specific) concepts (such as “landscape”, “hotel”). In this paper, we exploit social data from social media platforms to assist image search engines, aiming to improve the relevance between returned images and user intentions (i.e., social relevance). Facing the challenges of social data sparseness, the tradeoff between social relevance and visual relevance, and the complex social and visual factors, we propose a community-specific Social-Visual Ranking (SVR) algorithm to rerank the Web images returned by current image search engines. The SVR algorithm is implemented by PageRank over a hybrid image link graph, which is the combination of an image social-link graph and an image visual-link graph. By conducting extensive experiments, we demonstrated the importance of both visual factors and social factors, and the advantages of social-visual ranking algorithm for Web image search.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Image search engines play the role of a bridge between user intentions and visual images. By simply representing user intentions with textual query, many existing research works have been focusing on how to match the textual query with visual images and their surrounding texts or tags. However, the returned results are often unsatisfactory due to their deviation from user intentions. Let's take the image search case “jaguar” as an example scenario, as shown in Fig. 1. Different users have different intentions when inputting the query “jaguar”. Some are expecting leopard images, while others are expecting automobile images. This scenario is quite common, particularly for queries with heterogeneous concepts (such as “apple”, “jaguar”) or general (non-specific) concepts (such as “landscape”, “hotel”). This raises a fundamental but rarely-researched problem in Web image search: how to understand user intentions when users conducting image search?

In the past years, this problem is very difficult to resolve due to the lack of social (i.e., inter-personal and personal) or personal data to reveal user intentions. On one hand, the user search logs, which contain rich user information, are maintained by search engine

companies and kept confidential; on the other hand, the lack of ID (user identifier) information in the user search logs makes them hard to be exploited for intention representation and discovery. However, with the development of social media platforms, such as Flickr and Facebook, the way people can get social (including personal) data has been changed: users' profiles, interests and their favorite images are exposed online and open to public, which are crucial information sources to implicitly understand user intentions.

Thus, let's imagine a novel and interesting image search scenario: what if we know users' Flickr ID when they conducting image search with textual queries? Can we exploit users' social information to understand their intentions, and further improve the image search performances? In this paper, we exploit social data from social media platforms to assist image search engines, aiming to improve the relevance between returned images and user intentions (i.e., user interests), which is termed as *Social Relevance*.

However, the combination of social media platforms and image search engine is not easy in that:

- (1) *Social data sparseness*. With respect to image search, the most important social data is the favored images of users. However, the large volume of users and images intrinsically decide the sparseness of user-image interactions. Therefore most users only possess a small number of favored images, from which it is difficult to discover user intentions. This problem can be

* Corresponding author.

E-mail addresses: liu-sw11@mails.tsinghua.edu.cn (S. Liu), cui@tsinghua.edu.cn (P. Cui), luanhuanbo@gmail.com (H. Luan), wwzhu@tsinghua.edu.cn (W. Zhu), yangshq@tsinghua.edu.cn (S. Yang), qitian@cs.utsa.edu (Q. Tian).

¹ Tsinghua National Laboratory for Information Science and Technology, China.

² Beijing Key Laboratory of Networked Multimedia, Tsinghua University, China.

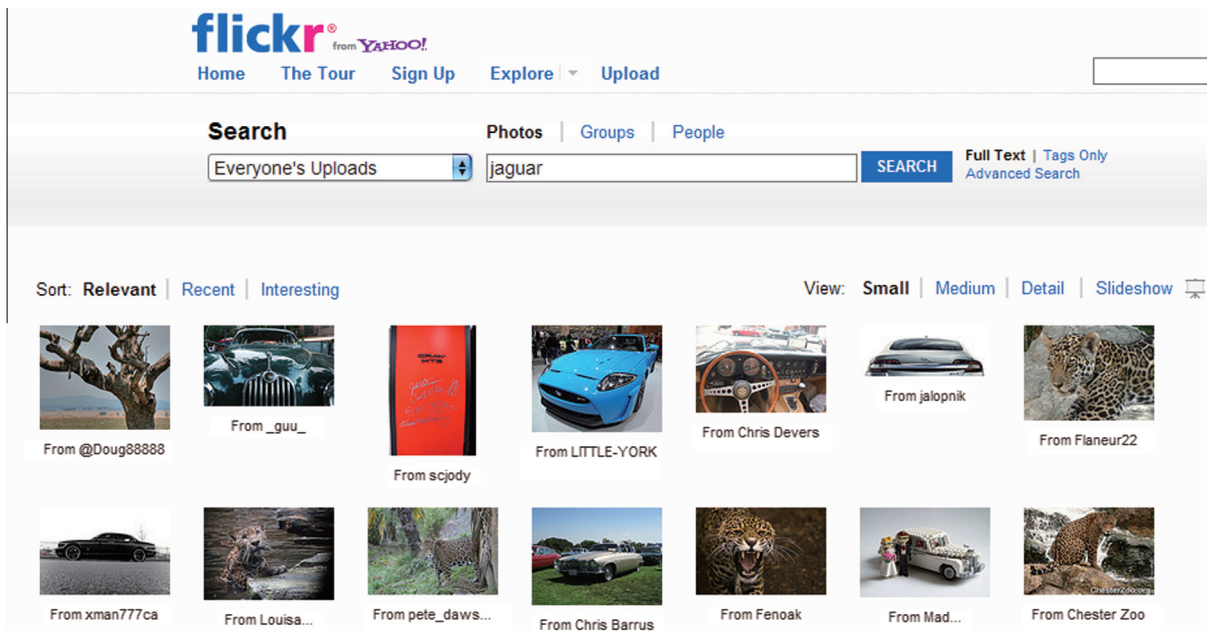


Fig. 1. The results returned by Flickr for the query "jaguar", recorded on April, 10th, 2012.

alleviated by grouping users into communities, with the hypothesis that users in the same community share similar interests. Thus, a community-specific method is more practical and effective than a user-specific method.

- (2) *The tradeoff between social relevance and visual relevance.* Although this paper aims to improve the social relevance of returned image search results, there still exists another important aspect: the *Visual Relevance* between the query and returned images. The visual relevance may guarantee the quality and representativeness of returned images for the query, while the social relevance may guarantee the interest of returned images for the user, both of which are necessary for good search results. Thus, both social relevance and visual relevance are needed to be addressed and subtly balanced.
- (3) *Complex factors.* To generate the final image ranking, we need to consider the user query, returned images from current search engines, and many complex social factors (e.g. interest groups, group-user relations, group-image relations, etc.) derived from social media platforms. How to integrate these heterogeneous factors in an effective and efficient way is quite challenging. In order to deal with the above issues, in this paper, we propose a community-specific Social-Visual Ranking (SVR) algorithm to rerank the Web images returned by current image search engines. More specifically, given the preliminary image search results (returned by current image search engines, such as Flickr search and Google Image) and the user's Flickr ID, we will use group information in social platform and visual contents of the images to rerank the Web images for a group that the user belongs to, which is termed as *the user's membership group*. The SVR algorithm is implemented by PageRank over a hybrid image link graph, which is the combination of an image social-link graph and an image visual-link graph. In the image social-link graph, the weights of the edges are derived from social strength of the groups. In the image visual-link graph, the weights of the edges are based on visual similarities. Through SVR, the Web images are reranked according to their interests to the users while maintaining high visual quality and representativeness for the query.

It is worthwhile to highlight our contributions as follows:

- (1) We propose a novel image search scenario by combining the information in social media platforms and image search engines to address the user intention understanding problem in Web image search, which is of ample significance to improve image search performances.
- (2) We propose a community-specific social-visual ranking algorithm to rerank Web images according to their social relevances and visual relevances. In this algorithm, complex social and visual factors are effectively and efficiently incorporated by hybrid image link graph, and more factors can be naturally enriched.
- (3) We have conducted intensive experiments, indicated the importance of both visual factors and social factors, and demonstrated the advantages of social-visual ranking algorithms for Web image search. Except image search, our algorithm can also be straightforwardly applied in other related areas, such as product recommendation and personalized advertisement.

The rest of the paper is organized as follows. We introduce some related works in Section 2. Image link graph generation and image ranking is presented in Section 3. Section 4 presents the details and analysis of our experiments. Finally, Section 5 concludes the paper.

2. Related work

Aiming at improving the visual relevance, a series of methods are proposed based on incorporating visual factors into image ranking. The approaches can be classified into three categories: classification [1–3], clustering [4] and link graph analysis [5–7]. An essential problem in these methods is to measure the visual similarity [8], assuming that similar images should have similar ranks. Besides, many kinds of features can be selected to estimate the similarity, including global features such as color, texture, and shape [9,10], and local features such as Scale Invariant Feature Transform (SIFT) feature [11]. Although there are different

performance measurements of a reranking algorithm [12], the relevance to the query is still the most recognized measurement [13].

As an effective approach, VisualRank [5] determines the visual similarity by the number of shared SIFT features [11], which is replaced by visual words in the later works [14–17]. The similarity between two images is evaluated by the co-occurrence of shared visual words. After a similarity based image link graph was generated, an iterative computation similar to PageRank [18] is utilized to rerank the images. A latent assumption in VisualRank is, similar SIFT features represent shared user interests. Intuitively, if an image captures user's intention, the similar images also should be of user's interest. By this hypothesis, VisualRank obtains a better performance than text-based image search in the measurement of relevance for queries with homogeneous visual concepts. However, for queries with heterogeneous visual concepts, VisualRank does not work well. For the returned results often include multiple categories, it is difficult for us to estimate which one best captures user's intention. Though VisualRank can better rank the images from same category, it is still hard to determine which category is expected by the user. Thus, VisualRank is applied mainly for product search, where the queries are usually with homogeneous visual concepts.

With the development of social media platform, the concept of social image retrieval was proposed, which brings more information and challenges to us. Most of works in social image search focus on tags of social image, such as evaluating the tag relevance [19,20], and measuring the quality of the tag [21,22]. It is obvious that search result quality is very low for queries with heterogeneous visual concepts. Although we are aware of this, lots of images on the Internet are still be tagged by these low-quality queries. It's desired to recommend users some better queries to choose. However, the quality of recommendation is based on the technique of tag annotation [23], which is not mature enough. Overall, understanding user intention is significant but challengeable in social media platform.

Many social media sites such as Flickr offer millions of groups for users to share images with others. There are tons of works based on improving the user experience in this type of information exchange, such as group recommendation [24] and social strength [25]. The basic idea of these works is the groups and the images are all based on users' interests. Therefore, a series of social information can be used to help us understand user intention better [26]. Based on the hybrid graph of social information, some work are proposed for recommendation and retrieval [27]. However, because of the complexity of the social factors, it is difficult to evaluate the importance of the social information.

3. Social-visual reranking

Fig. 2 illustrates the framework of our social-visual ranking algorithm. In this framework there are four major intermediate results: global group link graph for group ranking, local group link graph, image social-link graph and image visual-link graph. Images are reranked by PageRank based on the linear combination of the image social-link graph and the image visual-link graph. We will first analyze the factors in random walk, then we will show the details of the definition of each graph. Finally, the overall iterative PageRank process is introduced. To clarify the symbols in our formulations, we list all the symbols in Table 1.

3.1. Random walk and image ranking

An image link graph is a graph with images as vertices and a set of weighted edges to describe the relationship between pairwise images. When an image link graph is given, the method of Eigen-

vector Centrality can be utilized to combine the vertex and its neighbors together for the consideration of ranking. The basic idea is, the vertex being neighbor to the high-rank vertex will also have a high rank order. As one of the most successful applications of Eigenvector Centrality, PageRank [18] computes a rank vector by a link graph based on hyperlinks of the webpages. Each of the elements in rank vector indicate the importance of each webpage. For image search there are not direct relationships as hyperlinks.

In our approach, a random walk model based on PageRank is utilized for image ranking. Although for image search there are not direct relationships as hyperlinks, we can intuitively imagine that there are some invisible links driving the user to jump from one image to another. Therefore, the essential problem is to define the weights of the links. In our model, the weight $p(I_i, I_j)$ of the link from image I_i to image I_j represents the probability that a user will jump to I_j after viewing I_i . This procedure can be considered in both social factor and visual factor. From the social point of view, when the user is interested in I_i , he will also be interested in I_i 's group G_p . Then he found a group G_q which is very similar to G_p . Finally he decides to visit the images in G_p including I_j . This type of jump is driven by social relevance. From the visual point of view, a user may be attracted by some contents of I_i and then decide to view I_j which also contains these contents. This jump is driven by visual similarity. As a result, these two factors will both have significant effects in image ranking. For incorporation, we define our image link graph as the linear combination of the visual-link graph and the social-link graph. I.e.,

$$P_G = \alpha \cdot P_G^S + (1 - \alpha) \cdot P^V \quad (1)$$

where P_G is the adjacency matrix of the hybrid image link graph. P_G^S is the matrix for image social-link and P^V is the matrix of image visual-link graph. α is a parameter to balance these factors. The estimation of α will be discussed in Section 4. In this equation, P_G and P_G^S are relevant to the user's membership group G . Therefore they have a subscript as 'G'. The symbols with the subscript 'G' in our algorithm have the same meaning.

In the next sections, we will show how to generate image link graph for final ranking step by step.

3.2. Image social-link graph

For image social-link graph is based on the social strength, we will first generate a global group link graph based on group similarity. Then, PageRank method is utilized for this graph to evaluate the group importance. Next, we generate a local group link graph specific for group G . The edge weights of the group link graph are determined by social strength. Finally, an image social-link graph can be constructed based on the local group link graph.

3.2.1. Global group ranking

First, a global group link graph is generated in preprocessing phase of our algorithm based on group similarity. The similarity between two groups is determined by two factors: image factor and user factor. It can be assumed that an image belongs to a category of interest, a user has some interests and a group is a community where users share interests with others. Based on this hypothesis, if two groups have many users in common, it is of high probability that these two groups are of same interests. Similarly, if two groups have many images in common, they are also very probable to share the common interests. Therefore, group similarity in user interests can be measured by the overlap of user sets and data sets, which is defined as:

$$S(G_u, G_v) = \lambda \cdot \text{overlap}(\mathcal{M}_u, \mathcal{M}_v) + (1 - \lambda) \text{overlap}(\mathcal{I}_u, \mathcal{I}_v) \quad (2)$$

where \mathcal{M}_i is the member set of group G_u and \mathcal{I}_u is the image set of group G_u . λ is a parameter to balance the user factor and the image

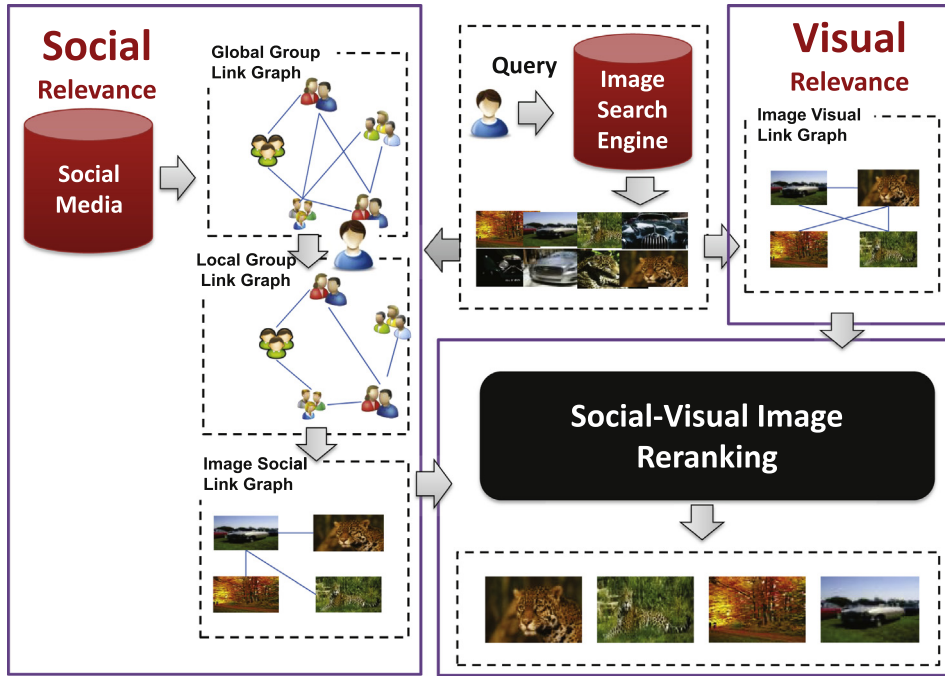


Fig. 2. The framework of our approach. Among the four generated graphs, global group link graph evaluates the global importance of each group. Local group link graph reflects the relationship of the groups w.r.t. the interests of the current user. Image social link graph evaluates the social similarity of the images. Image visual link graph calculates the visual similarity of the images. Our ranking method is based on the last two graphs.

Table 1
Symbols used in social-visual ranking.

Symbol	Meaning
G	The user membership group for image ranking
G_u	The u th group in our group set
I_i	The i th image in our image set
\mathcal{M}_u	The member set of G_u
\mathcal{I}_u	The image set of G_u
$p_G^S(I_i, I_j)$	The image social similarity between I_i and I_j for G
P_G^S	The image social-link graph matrix constructed by $p_G^S(I_i, I_j)$
$p^V(I_i, I_j)$	The visual similarity between I_i and I_j
P^V	The image visual-link graph matrix constructed by $p^V(I_i, I_j)$
P_G	The hybrid image link graph matrix for G
$S(G_u, G_v)$	The similarity of G_u and G_v
$T_G(G_u, G_v)$	The social strength of G_u and G_v
$C(I_i, I_j)$	The number of co-occurrence visual words of I_i and I_j
$A(I_i, G_u)$	The indicator function of I_i belongs to G_u

factor. It will be studied by our experiments. The overlap of \mathcal{M}_i and \mathcal{M}_j can be described as Jaccard similarity:

$$\text{overlap}(\mathcal{M}_u, \mathcal{M}_v) = \frac{|\mathcal{M}_u \cap \mathcal{M}_v|}{|\mathcal{M}_u \cup \mathcal{M}_v|} \quad (3)$$

so is the overlap of \mathcal{I}_u and \mathcal{I}_v . We utilize overlap function rather than the number of common elements for the consideration of normalization.

After the pair-wised group similarities are computed, we define the importance of a group as its centrality in the global group link graph. Intuitively, a group similar to many other groups should be important. The iterative computation based on PageRank can be utilized to evaluate the centrality of the groups:

$$gr = d \cdot S \cdot gr + (1 - d)e_0, \quad e_0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{N_G \times 1} \quad (4)$$

where S is a column-normalized matrix constructed by $S(G_u, G_v)$. N_G is the number of groups. d is called damping factor, which denotes

the probability user stay in visiting the images along the graph links. In practice, $d = 0.8$ has a good performance with small variance around this value.

As illustrated in Fig. 3, PageRank over global group link graph models the random walk behavior of a user on these groups. First, he randomly accesses a group. Then, for the interests in some members or some images of this group, he visits another group which also contains these members or images. After a period of time, he is tired of visiting the images along the link of the group, then he randomly access another group again.

Global group link graph and group rank describe the global properties (similarity and importance) of groups. It is not specific for the query and the user's membership group G . Therefore, $S(G_u, G_v)$ and gr can be computed off line and updated at regular time.

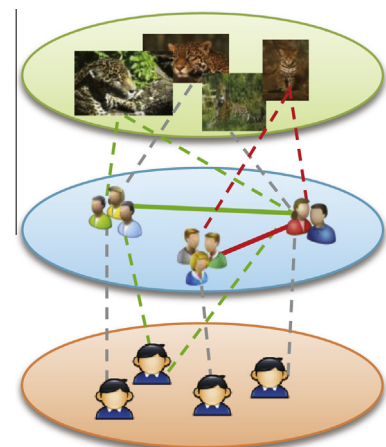


Fig. 3. An illustration of group similarity (middle layer), which is determined by images' overlap (top layer) and users' overlap (bottom layer).

3.2.2. Local group link graph

Since the global group link graph and group ranking are obtained, the local group link graph can be generated based on social strength of pairwise groups. The social strength of group G_u and group G_v , which is represented as $T_G(G_u, G_v)$ describes the correlation G_u and G_v in G 's intentions. In other words, $T_G(G_u, G_v)$ denotes the probability that an user in G will jump to the images of G_v after viewing the images of G_u .

For instance, when G is a group about animals, one group about zoos and another about birds should be of tight social strength. Meanwhile, if a user in G has viewed the images of a group about IT, they may be more probably interested in the images about cartoon production than computer hardware because he is more interested in animals. Although the group about IT and another group about hardware are similar globally, the global social strength of them for G should be very weak. For quantitative analysis, There are some basic considerations about the social strength as following:

- The group similarity $S(G_u, G_v)$ denotes the degree that G_u recommend G_v to G .
- Users in G will not accept all the recommendations for it has a specific intention. If they are interested in G_v indeed, they may decide to visit it.
- In another case, When users in G are interested in G_u and G_u recommend G_v to G , G may decide to accept the recommendation to visit G_v .
- If other conditions are the same, the users in G will trust the recommendation of the most important group.

Under these considerations, we can formulate the social strength $T_G(G_u, G_v)$:

$$T_G(G_u, G_v) = (S(G, G_u) + S(G, G_v)) \cdot S(G_u, G_v) \cdot f(gr(G_u)) \cdot f(gr(G_v)) \quad (5)$$

where $f(gr(G_u))$ is a function of the group rank value of G_u in the rank vector calculated in Eq. (4). It denotes the weight of group importance.

It has been prove that there is a power law between the scale (importance) of the groups and the quality of the images [28]. In this paper, we consider the function $f(x)$ in the form of power function, which is proved to be valid in previous work [15], i.e.:

$$f(x) = x^r \quad (6)$$

where r is a parameter which will be estimated by experimental study.

3.2.3. Image social-link graph

Image social-link graph can be generated based on the social strength of group. For images and groups, we first construct a basic image-group graph. The edge from an image to a group denotes the image belonging to the group, which can be formulated as:

$$A(I_i, G_u) = \begin{cases} 1 & I_i \text{ belongs to } G_u \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Fig. 4 is an illustration of image social similarity. Based on local group link graph and image-group graph, we can define the weight of the edge in image social-link graph as:

$$p_G^S(I_i, I_j) = \frac{Z_1}{\left(\sum_{u=1}^{N_G} A(I_i, G_u)\right) \left(\sum_{u=1}^{N_G} A(I_j, G_u)\right)} \cdot \left(\sum_{u=1}^{N_G} \sum_{v=1}^{N_G} A(I_i, G_u) \cdot A(I_j, G_v) \cdot T(G_u, G_v)\right) \quad (8)$$

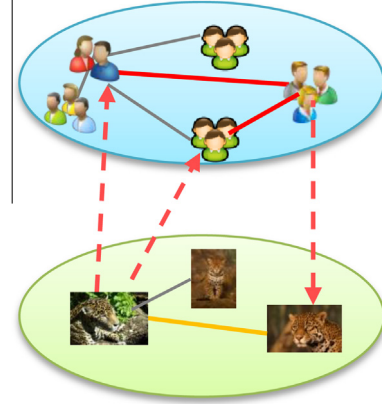


Fig. 4. An illustration of image social-link graph (bottom layer) generation based on local group link graph (top layer).

where Z_1 is a column-normalization factor to normalize $\sum_j p_G^S(I_i, I_j)$ to 1. $p_G^S(I_i, I_j)$ denotes the probability that group G will visit I_j after viewing I_i .

In traditional random-walk model such as topic-sensitive Page-Rank [29], there is a basic assumption that the probabilities of the user's jump from one vertex to another is only determined by the global correlation of the vertices. However, users usually surf on the Internet with some specific intention. By assuming this, users' random walks are usually not on a global graph but a local one for a specific intention. Therefore, it is of great significance for us to generate the image social-link graph by a local group link graph.

However, there will be some problems if we only consider social-link graph. When the value of $p_G^S(I_i, I_j)$ is high, the user will be very probable to jump from I_i to I_j . However, when the value is low, it just indicates that we cannot estimate the probability based on our social knowledge. In other words, low social similarity cannot represent low transition probability. An extreme case is, when the current group G has no correlation to the query at all, social image ranking will not work. For instance, if a user in an IT group wants to search some foods one day, all the values in P_G^S may be close to zero. This is an important reason why we incorporate the visual factor to the social factor as well as to improve visual relevance.

3.3. Image visual-link graph

In VisualRank [5], the visual image link is weighted as the number of common SIFT descriptors. In our approach, we improve this method by a BoW (bag of words) representation. Since the SIFT descriptors of each image are extracted, a hierarchical visual vocabulary tree [30] based on hierarchical k-means clustering is built by all the descriptors. The leaf nodes of the hierarchical vocabulary tree are defined as visual words. After visual vocabulary is generated, an image can be regarded as a documents including some words. We can efficiently count the co-occurrence of visual words in two images. Therefore, the weight of the edge in visual image link graph can be defined as:

$$p^V(I_i, I_j) = \frac{C(I_i, I_j)}{\sum_i C(I_i, I_j)} \quad (9)$$

where $C(I_i, I_j)$ is the count of co-occurrence of visual words in image I_i and I_j . $p^V(I_i, I_j)$ denotes the possibility that a user transits from I_i to I_j in visual factor. Since visual words can represent the local contents of images, $C(I_i, I_j)$ can effectively measure the visual similarity. When a user is visiting some images, after he visits I_i , he will be interested in one or more objects in I_i which contains some local

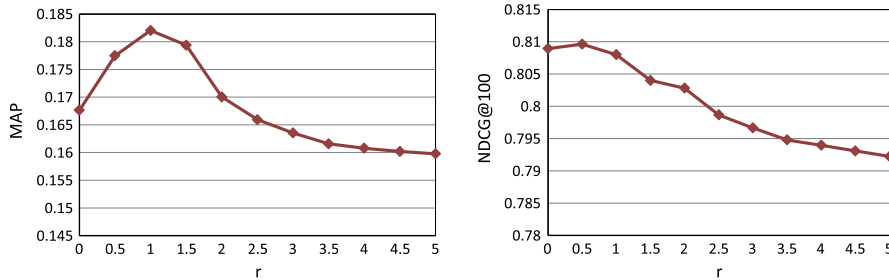


Fig. 5. Performance for different values of r with $\alpha = 0.3$, $\lambda = 0.4$ and $e = e_G$. Best performance is obtained with $r = 0.5$.

features and jump to an image I_j which includes similar local features. If the image's local areas are shared by many other images, it will obtain a high ranking order by VisualRank. A lot of works show the effectiveness of VisualRank in visual relevance. In this paper, we explore the effect of VisualRank in social relevance.

3.4. Image social-visual ranking

After two image link graphs are generated, hybrid image link graph can be constructed by Eq. (1). Then, the iteration procedure based on PageRank can be formulated as:

$$r_G = d \cdot P_G \cdot r_G + (1 - d)e \quad (10)$$

where $d = 0.8$ as in Eq. (4). e is a parameter to describe the probability a user jumps to another image without links when they are tired of surfing by links. In the basic implementation of PageRank, each element of e is taken an equal value, as the jump without link is expected to be random. However, for a specific user, e should be relevant to the user's interests. Therefore, in personalized PageRank [31], e is defined as the probability that user randomly access a document based on his interest. In this method, a good measurement of the user's interests of image I_i is the average similarity of the user's group G and the groups including I_i . Thus we have two choices of e :

$$e_1(i) = \frac{1}{N_i} \quad (11)$$

where N_i is the number of images, and

$$e_G(i) = Z_2 \frac{\sum_{u=1}^{N_G} A(I_i, G_u) \cdot S(G, G_u)}{\sum_{u=1}^{N_G} A(I_i, G_u)} \quad (12)$$

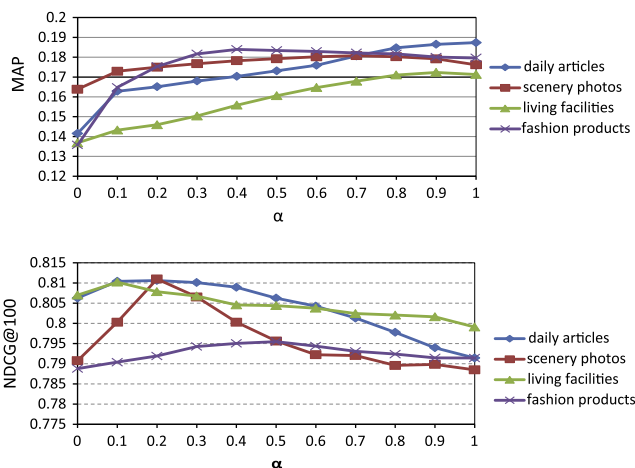


Fig. 6. Performance for different values of α with $\lambda = 0.4$, $r = 0.5$ and $e = e_G$. Best performance is obtained with $\alpha = 0.3$.

where Z_2 is the factor to normalize the sum of $\sum e_G(i)$ to 1. In parameter analysis of PageRank [18], e is an important parameter for personalized search. Next session we will discuss the effect of e for e_1 and e_G in our algorithm.

4. Experiments

4.1. Dataset and settings

To implement our algorithm, we conduct experiments with data including images, groups, users, group-user relations and group-image relations from Flickr.com. Thirty queries are collected and 1000 images are downloaded for each query. These selected queries cover a series of categories tightly related to our daily life, including:

1. Daily articles with no less than two different meanings, such as "apple", "jaguar" and "golf".
2. Natural scenery photos with multiple visual categories, such as "landscape", "scenery" and "hotel".
3. Living facilities with indoor and outdoor views, such as "restaurant" and "hotel".
4. Fashion products with different product types, such as "smart phone" and "dress".

Each image downloaded must belong to at least one group. Social data are collected through Flickr API. If the size of a set (both user set and image set) is larger than 1000, we just keep the first 1000 elements.

The SIFT feature is extracted by a standard implementation [11]. The hierarchical visual vocabulary tree is of 4 layers and 10 branches for each layer. Each image is of normal size in Flickr, i.e., the length of the longest edge is no more than 400 pixels. According to our statistics, each image has about 400 SIFT features in average. Therefore, there are about 400 thousand SIFT descriptors in the image set of 1000 images for the query. Besides, the number of clustered visual words is no more than 10 thousand.

In our experiment, we compare our algorithm SVR with other three rank methods: VisualRank (VR), SocialRank (SR) and Flickr search engine by relevance (FR) as baseline. Among them, VR is the special case for SVR when $\alpha = 0$, and SR is the special case for $\alpha = 1$. We evaluate our algorithm in social relevance and visual relevance respectively.

4.2. Metrics

We estimate the performance of our approach by two measurements. One is the relevance of current group's intentions, i.e., social relevance. The other is image quality, which is represented by visual relevance. The target of our algorithm is to improve the social relevance as much as possible with maintaining the visual relevance.

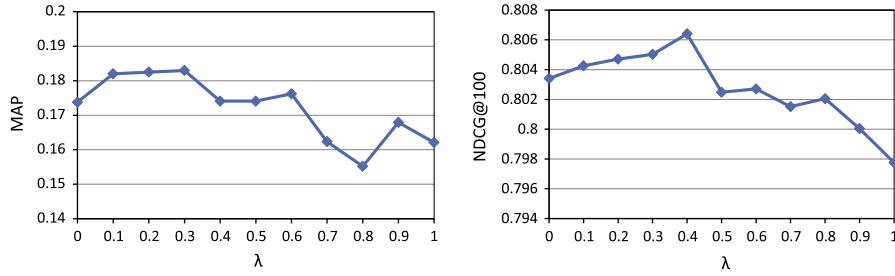


Fig. 7. Performance for different values of λ with $\alpha = 0.3$, $r = 0.5$ and $e = e_G$. Best performance is obtained with $\lambda = 0.4$.

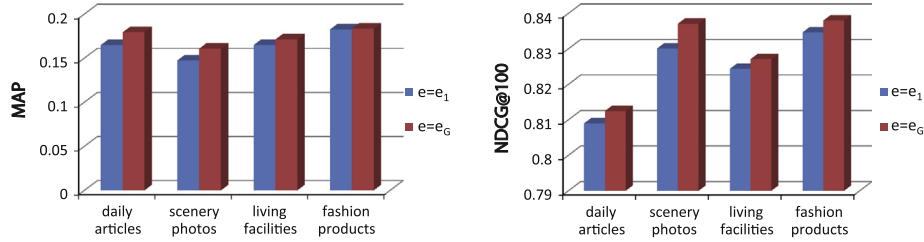


Fig. 8. The Performance for $e = e_1$ and $e = e_G$ for different categories of queries, with $\alpha = 0.3$, $\lambda = 0.4$ and $r = 0.5$. and. Best performance is obtained with $e = e_G$.

4.2.1. Social relevance

Defined as the relevance of user intentions, social relevance is an important measurement in our experiments. To estimate whether the results of our approach is of group G 's intention, we use *MAP* (Mean Average Precision) as the metric. Sharing behaviors are used as the ground truth. For a query, we randomly select n testing groups from the dataset. For each group, if an image contains the query and belongs to the group, we can believe this image is relevant to the intention of the group. After the definition of the ground truth, *MAP* can be calculated as:

$$MAP = \frac{1}{|Q|} \sum_{q_i \in Q} \frac{1}{|\mathcal{G}_{q_i}|} \sum_{G_j \in \mathcal{G}_{q_i}} AP(q_i, G_j) \quad (13)$$

where Q is the set of queries; \mathcal{G}_{q_i} is the set of the testing groups for the query q_i ; and $AP(q_i, G_j)$ is the average precision of the ranking result when group G_j searches for query q_i . In our experiments, there are 30 queries and we select 20 testing groups for each query. Therefore, $|Q| = 30$ and $|\mathcal{G}_{q_i}| = 20$. *MAP* reflects the social relevance, i.e., whether an image is relevant the interests of the group. However, we need to consider about the over-fitting problem: it is possible for a mac user to find an image about fruit apple. Therefore, we can sacrifice the performance of AR a little as a trade off to visual relevance.

4.2.2. Visual relevance

For all images in our dataset are labeled according to their relevance, Normalized Discounted Cumulative Gain (*NDCG*) is adopted to measure the visual relevance [13,32]. Giving a ranking list, the score *NDCG@n* is defined as

$$NDCG@n = Z_n \sum_{i=1}^n \frac{2^{r(i)} - 1}{\log(1 + i)} \quad (14)$$

$r(i)$ is the score of the image in the i th rank. Z_n is the normalization factor to normalize the perfect rank to 1. In our experiment, we set $n = 100$, which means the user usually find the target image in the first 100 results returned by the search engine.

To evaluate the visual relevance, all the images are scored with ground truth according to the relevance to the corresponding query. The scores are of four levels, 0: irrelevant, 1: so-so, 2: good,

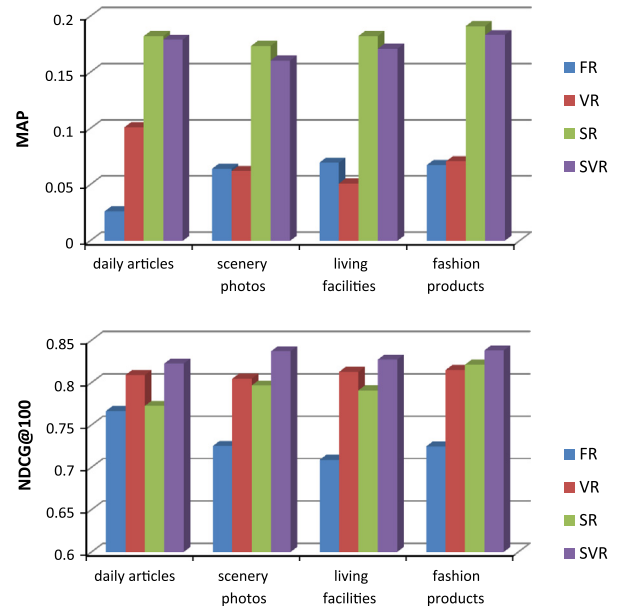


Fig. 9. The performance of our approach compared to other two methods FlickrRank and Visual Rank by the measurements *MAP* and *NDCG@100*. Significance is tested by paired *t*-test, where $p < 0.05$.

3: excellent. The score of an image is just determined by the quality of the image. For the query with heterogeneous visual concepts, all categories will be equal. For instance, both a leopard and a Jaguar brand car will be scored 3 if they are complete and clear.

In the rest of the this session, we'll show the effectiveness of the proposed graphs in our approach and evaluate the performance based on these two measurements.

4.3. Parameter settings

In our approach, there are four parameters: λ in Eq. (2), r in Eq. (6), α in Eq. (1) and e in Eq. (10). In this subsection, we will investigate the effect of different parameter settings. First, we randomly

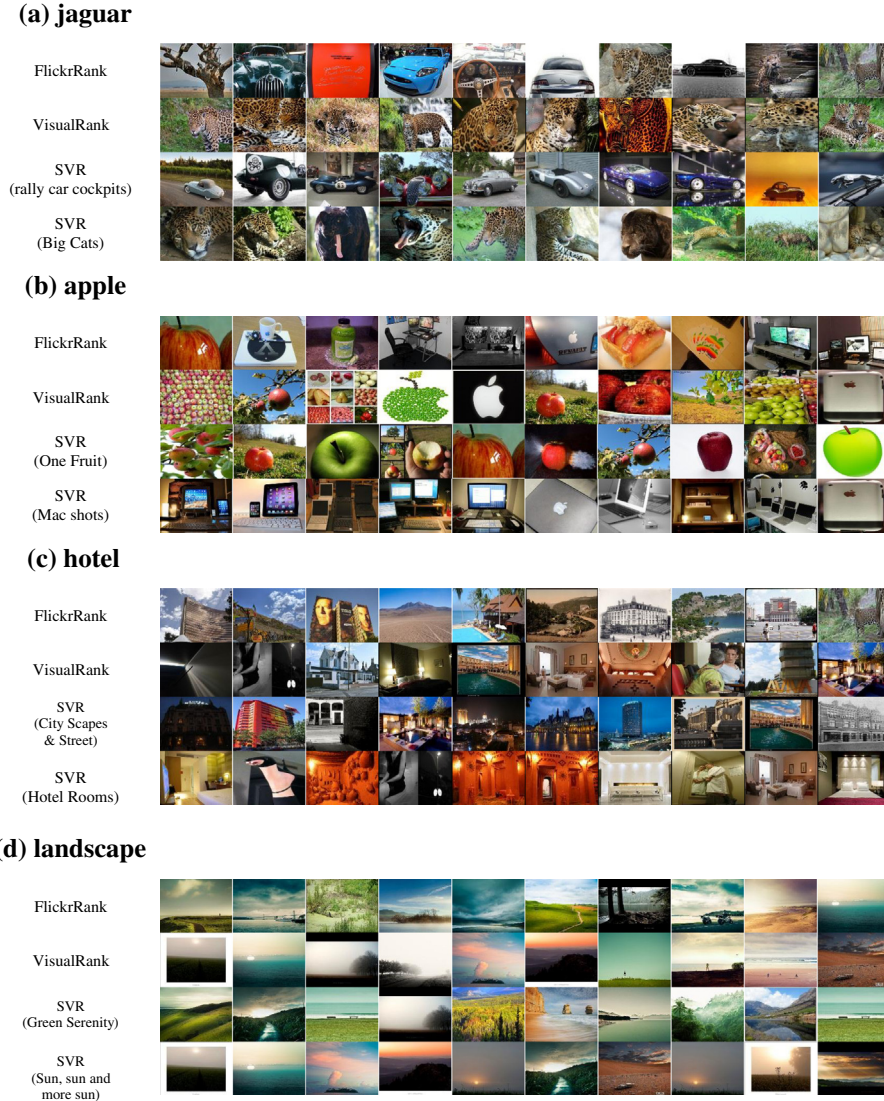


Fig. 10. Top-10 reranking results of our approach for two different groups compared to FlickrRank and VisualRank.

sample a region of parameters and select the best setting. Starting from this setting, we study the effectiveness of each parameter. Iteratively, we fix three other parameters as constants and adjust the other one until there is no change for all parameters. After all parameters are convergent, we draw the curves of *MAP* and *NDCG@100* for each parameter. Testing groups are selected over the whole dataset, as a total of 600. It is observed from Figs. 5–7 that *MAP* and *NDCG@100* usually obtain the best performance for different parameter values. Actually, our hope is to improve the *MAP* while maintaining higher *NDCG*.

4.3.1. Effectiveness of global group link graph

In the above four parameters, r denotes the importance of global group link graph. When $r = 0$, we do not consider this graph in our method. When r is large, the more important groups will effect the user intention more. Fig. 5 shows the performance when r obtains different values. From the figure, it can be observed that the best performance measured by *NDCG@100* is obtained when $r = 0.5$ and $r = 1$ for *MAP*. Besides, Fig. 5 indicates that the performance varies less when r is around 0.5. To guarantee the quality of search results, i.e., visual relevance, we utilize $r = 0.5$ as the near

optimal setting of r , which means the social strength $S(u, v)$ in Eq. (5) finally should be defined as:

$$T_G(G_u, G_v) = (S(G, G_u) + S(G, G_v)) \cdot S(G_u, G_v) \cdot \sqrt{gr(G_u) \cdot gr(G_v)} \quad (15)$$

This result shows that in our algorithm, the importance of a group does help to improve the performance of our approach. However, it is not the decisive condition especially for *NDCG*. In other words, the importance of groups can help us better judge the image quality of the images.

4.3.2. Trade-off between social and visual link graph

α is the parameter that denotes the importance of weigh of social factor in our ranking method. When $\alpha = 0$, the ranking methods is purely a visual method. When $\alpha = 1$, the method is only determined by social factors. It can be imagined that α should have correlation to the query. Therefore, we estimate the setting of α for each of the four categories. Fig. 6 shows the performance of our approach when α obtains different values. From the results, we can observe that:

- For any category, the result of our approach obtains the best performance when α is not zero, i.e., α is of great help for all the queries tested in our experiment. The hypothesis that social factor is effective in image search is proved.
- Measured by *MAP*, α close to 1 produces the best performance. Therefore, if we just consider to give a higher ranking order to the images fit for user intentions, image ranking can be mainly based on social factor.
- The curve of *NDCG* indicates that, as the weight of social factor growing after a critical point, more and more images with low visual relevance are ranked to the front. In most cases, the performance of social image ranking in relevance is much worse than visual image ranking.

Based on these observation, α is determined to be 0.3 in our approach, which can guarantee the effect of *NDCG* with a fairly high *MAP*.

4.3.3. Other parameters

In Eq. (2), we calculate the group similarity in two dimensions: user dimension and image dimension. λ is a trade-off parameter of these two dimensions. When $\lambda = 0$, we evaluate the group similarity by the overlap of their images. When $\lambda = 1$, the similarity depends on the overlap of the images. Fig. 7 shows the performance of our approach for different λ . From the figure, it can be observed that all these two measurements can be of best performance when $\lambda = 0.4$. As a parameter representing the trade off between the users' overlap and the images' overlap, λ shows a user more likely to be interested in a group due to its images than users. With above three parameters being set, we investigate the effect of e in Eq. (10). We compare the performance of $e = e_G$ to $e = e_1$ for four categories of queries. Fig. 8 shows the results by two measurements. It can be observed that the approach with $e = e_G$ is better than $e = e_1$ for all categories. Thus, e_G can indeed improve the performance of our approach.

4.4. Overall performance and search results

To prove the results of SVR can really reflect the user intentions, we randomly select 4 queries in different areas, which have obvious different visual meanings: "apple" (including fruits and mac products), "jaguar" (including leopards and automobiles), "landscape" (including different categories of photos about natural sceneries) and "hotel" (including photos about location and decoration). For each query, we select 2 groups that we can obviously estimate the interests by their names. Then, we show the top 10 images ranked by our approach for the selected two groups with FR and VR as baselines. Fig. 10 shows the results. For each query, each row from top to bottom corresponds to the top 10 results of FR, VR, SVR for the first group and SVR for the second group. From these instances it can be observed that our approach really knows what the group wants and the results are mostly of high quality. For the query "apple" and "jaguar", which has obvious different visual concepts, SVR can find the images fit for the group names fairly well. In contrast, the top-10 results of VisualRank for "jaguar" are all about leopards.

For the quantitative evaluation of the performance, we compare our approach with other three ranking methods FR, VR and SR by the measurements *MAP* and *NDCG* for each category of queries defined in Section 4.1. The parameters are set from previous section. The *NDCG* and *MAP* are calculated by our approach compared with other three methods. Fig. 9 shows the comparison results. It can be observed that our approach achieves the best performance in *NDCG* and has great improvement in *MAP* compared to VisualRank. Although SocialRank performs best on *MAP*, the *NDCG* of SocialRank is much worse than VisualRank. Under the comprehensive consideration, SVR performs best in these four ranking methods.

5. Conclusions and future work

In this paper, we propose a novel framework of community-specific social-visual image ranking for Web image search. We explore to combine the social factor and visual factor together based on image link graph to improve the performance of social relevance under the premise of visual relevance. Comprehensive experiments show effectiveness of our approach. Our proposed method is significantly better than VisualRank and Flickr search engine in social relevance as well as visual relevance. Besides, the importance of both social factor and visual factor is discussed in details. For the query with heterogeneous visual concepts and the group with clear intention, our framework can effectively conduct community-specific image search.

Future work will be carried out on taking more features in social network into our consideration and making the social weight adaptive.

Acknowledgments

This work is supported by National Natural Science Foundation of China, No. 61370022, No. 61003097, No. 60933013, and No. 61210008; International Science and Technology Cooperation Program of China, No. 2013DFG12870; National Program on Key Basic Research Project, No. 2011CB302206. This work is also supported in part to Dr. Qi Tian by ARO grant W911NF-12-1-0057, NSF IIS 1052851, Faculty Research Awards by Google, FXPAL, and NEC Laboratories of America, and 2012 UTSA START-R Research Award respectively. This work is supported in part by NSF 61128007. Thanks for the support of NEXT Research Center funded by MDA, Singapore, under the research grant, WBS:R-252-300-001-490.

References

- [1] R. Yan, A. Hauptmann, R. Jin, Multimedia search with pseudo-relevance feedback, in: E. Bakker, M. Lew, T. Huang, N. Sebe, X. Zhou (Eds.), *Image and Video Retrieval, Lecture Notes in Computer Science*, vol. 2728, Springer, Berlin/Heidelberg, 2003, pp. 649–654. ISBN 978-3-540-40634-1.
- [2] Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, Y. Pan, A multimedia retrieval framework based on semi-supervised ranking and relevance feedback, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (4) (2012) 723–742.
- [3] Y. Yang, Y.-T. Zhuang, F. Wu, Y.-H. Pan, Harmonizing hierarchical manifolds for multimedia document semantics understanding and cross-media retrieval, *IEEE Trans. Multimedia* 10 (3) (2008) 437–446.
- [4] W.H. Hsu, L.S. Kennedy, S.-F. Chang, Video search reranking via information bottleneck principle, in: *Proceedings of the 14th Annual ACM International Conference on Multimedia*, MULTIMEDIA '06, ACM, New York, NY, USA, 2006, pp. 35–44. <http://dx.doi.org/10.1145/1180639.1180654>. ISBN 1-59593-447-2.
- [5] Y. Jing, S. Baluja, VisualRank: applying pagerank to large-scale image search, *IEEE Trans. Pattern Anal. Mach. Intell.* 30 (11) (2008) 1877–1890. <http://dx.doi.org/10.1109/TPAMI.2008.121>. ISSN 0162-8828.
- [6] J. Liu, W. Lai, X.-S. Hua, Y. Huang, S. Li, Video search re-ranking via multi-graph propagation, in: *Proceedings of the 15th International Conference on Multimedia*, MULTIMEDIA '07, ACM, New York, NY, USA, 2007, pp. 208–217. <http://dx.doi.org/10.1145/1291233.1291279>. ISBN 978-1-59593-702-5.
- [7] W.H. Hsu, L.S. Kennedy, S.-F. Chang, Video search reranking through random walk over document-level context graph, in: *Proceedings of the 15th international conference on Multimedia*, MULTIMEDIA '07, ACM, New York, NY, USA, 2007, pp. 971–980. <http://dx.doi.org/10.1145/1291233.1291446>. ISBN 978-1-59593-702-5, 971–980.
- [8] R.I. Kondor, J. Lafferty, Diffusion kernels on graphs and other discrete structures, in: *Proceedings of the ICML*, 2002, pp. 315–322.
- [9] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, X.-S. Hua, Bayesian video search reranking, in: *Proceedings of the 16th ACM International Conference on Multimedia*, MM '08, ACM, New York, NY, USA, 2008, pp. 131–140. <http://dx.doi.org/10.1145/1459359.1459378>. ISBN 978-1-60558-303-7.
- [10] H. Zitouni, S. Sevil, D. Ozkan, P. Duygulu, Re-ranking of web image search results using a graph algorithm, in: *ICPR 2008. 19th International Conference on Pattern Recognition*, 2008, pp. 1–4. doi: 10.1109/ICPR.2008.4761472 (ISSN 1051-4651).
- [11] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener, Graph structure in the Web, *Comput. Netw.* 33 (1C6) (2000) 309–320. [http://dx.doi.org/10.1016/S1389-1286\(00\)00083-9](http://dx.doi.org/10.1016/S1389-1286(00)00083-9) (ISSN 1389-1286).

- [12] B. Geng, L. Yang, C. Xu, X.-S. Hua, S. Li, The role of attractiveness in web image search, in: Proceedings of the 19th ACM International Conference on Multimedia, MM '11, ACM, New York, NY, USA, 2011, pp. 63–72, <http://dx.doi.org/10.1145/2072298.2072308>. ISBN 978-1-4503-0616-4.
- [13] K. Järvelin, J. Kekäläinen, IR evaluation methods for retrieving highly relevant documents, in: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '00, ACM, New York, NY, USA, 2000, pp. 41–48, <http://dx.doi.org/10.1145/345508.345545>. ISBN 1-58113-226-3.
- [14] Shiliang, Q. Zhang, G. Tian, Q. Hua, S. Huang, Li, Descriptive visual words and visual phrases for image applications, in: Proceedings of the 17th ACM International Conference on Multimedia, MM '09, ACM, New York, NY, USA, 2009, pp. 75–84, <http://dx.doi.org/10.1145/1631272.1631285>. ISBN 978-1-60558-608-3.
- [15] W. Zhou, Q. Tian, Y. Lu, L. Yang, H. Li, Latent visual context learning for web image applications, *Pattern Recognit.* 44 (10C11) (2011) 2263–2273. ISSN 0031-3203 [semi-Supervised Learning for Visual Content Analysis and Understanding](http://dx.doi.org/10.1145/1459359.1459387).
- [16] X. Zhou, X. Zhuang, S. Yan, S.-F. Chang, M. Hasegawa-Johnson, T.S. Huang, SIFT-Bag kernel for video event analysis, in: Proceedings of the 16th ACM International Conference on Multimedia, MM '08, ACM, New York, NY, USA, 2008, pp. 229–238, <http://dx.doi.org/10.1145/1459359.1459391>. ISBN 978-1-60558-303-7.
- [17] Y.H. Yang, P.T. Wu, C.W. Lee, K.H. Lin, W.H. Hsu, H.H. Chen, ContextSeer: context search and recommendation at query time for shared consumer photos, in: Proceedings of the 16th ACM International Conference on Multimedia, MM '08, ACM, New York, NY, USA, 2008, pp. 199–208, <http://dx.doi.org/10.1145/1459359.1459387>. ISBN 978-1-60558-303-7.
- [18] L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web, Technical Report 1999-66, Stanford InfoLab, Previous, 1999.
- [19] X. Li, C.G. Snoek, M. Worring, Learning tag relevance by neighbor voting for social image retrieval, in: Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval, MIR '08, ACM, New York, NY, USA, 2008, pp. 180–187, <http://dx.doi.org/10.1145/1460096.1460126>. ISBN 978-1-60558-312-9.
- [20] M. Larson, C. Kofler, A. Hanjalic, Reading between the tags to predict real-world size-class for visually depicted objects in images, in: Proceedings of the 19th ACM International Conference on Multimedia, MM '11, ACM, New York, NY, USA, 2011, pp. 273–282, <http://dx.doi.org/10.1145/2072298.2072335>. ISBN 978-1-4503-0616-4.
- [21] A. Sun, S.S. Bhowmick, Quantifying tag representativeness of visual content of social images, in: Proceedings of the International Conference on Multimedia, MM '10, ACM, New York, NY, USA, 2010, pp. 471–480, <http://dx.doi.org/10.1145/1873951.1874029>. ISBN 978-1-60558-933-6.
- [22] M. Ames, M. Naaman, Why we tag: motivations for annotation in mobile and online media, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07, ACM, New York, NY, USA, 2007, pp. 971–980, <http://dx.doi.org/10.1145/1240624.1240772>. ISBN 978-1-59593-593-9.
- [23] J. Sang, J. Liu, C. Xu, Exploiting user information for image tag refinement, in: Proceedings of the 19th ACM International Conference on Multimedia, MM '11, ACM, New York, NY, USA, 2011, pp. 1129–1132, <http://dx.doi.org/10.1145/2072298.2071956>. ISBN 978-1-4503-0616-4.
- [24] L.A.F. Park, K. Ramamohanarao, Mining web multi-resolution community-based popularity for information retrieval, in: Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, CIKM '07, ACM, New York, NY, USA, 2007, pp. 545–554, <http://dx.doi.org/10.1145/1321440.1321517>. ISBN 978-1-59593-803-9.
- [25] J. Zhuang, T. Mei, S.C. Hoi, X.-S. Hua, S. Li, Modeling social strength in social media community via kernel-based learning, in: Proceedings of the 19th ACM International Conference on Multimedia, MM '11, ACM, New York, NY, USA, 2011, pp. 113–122, <http://dx.doi.org/10.1145/2072298.2072315>. ISBN 978-1-4503-0616-4.
- [26] J. Bu, S. Tan, C. Chen, C. Wang, H. Wu, L. Zhang, X. He, Music recommendation by unified hypergraph: combining social media information and music content, in: Proceedings of the International Conference on Multimedia, ACM, 2010, pp. 391–400.
- [27] D. Liu, G. Ye, C.-T. Chen, S. Yan, S.-F. Chang, Hybrid Social Media Network.
- [28] R.A. Negoescu, D. Gatica-Perez, Analyzing flickr groups, in: Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval, ACM, 2008, pp. 417–426.
- [29] T.H. Haveliwala, Topic-sensitive pagerank: a context-sensitive ranking algorithm for web search, *IEEE Trans. Knowl. Data Eng.* 15 (4) (2003) 784–796.
- [30] D. Nister, H. Stewenius, Scalable recognition with a vocabulary tree, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2006, pp. 2161–2168, doi: 10.1109/CVPR.2006.264 (ISSN 1063-6919).
- [31] S. Chakrabarti, Dynamic personalized pagerank in entity-relation graphs, in: Proceedings of the 16th International Conference on World Wide Web, ACM, 2007, pp. 571–580.
- [32] L. Wang, L. Yang, X. Tian, Query aware visual similarity propagation for image search reranking, in: Proceedings of the 17th ACM International Conference on Multimedia, MM '09, ACM, New York, NY, USA, 2009, pp. 725–728, <http://dx.doi.org/10.1145/1631272.1631398>. ISBN 978-1-60558-608-3.