

Ranking in Heterogeneous Social Media

Min-Hsuan Tsai
Google
1600 Amphitheatre Pkwy
Mountain View, CA
mhtsai@google.com

Charu Aggarwal
IBM Research
1101 Kitchawan Road
Yorktown Heights, NY
charu@us.ibm.com

Thomas Huang
UIUC
2039 Beckman Institute
Urbana, Illinois
t-huang1@illinois.edu

ABSTRACT

The problem of image search has been studied extensively in recent years because of the large and increasing repositories of images on the web, social media, and other linked networks. Most of the available techniques for keyword-based image search on the web use the text in the surrounding or linked text in order to retrieve related images.

Many image repositories on the web are built upon social media platforms such as *Flickr*. Such platforms provide a rich level of information in terms of the user linkage information to images, tags or other comments which are contributed by the users. It is reasonable to assume that the content of the images, users and other social cues such as tags and comments are often related to one another. Therefore, such cues can be useful for improving the effectiveness of search and ranking algorithms. In this paper, we propose *SocialRank*, which is a technique for using social hints in order to improve the image search and ranking process. Furthermore, we propose a holistic framework to combine social tags, social network text, linkage between actors and images, as well as the actual image features in order to create a ranking technique for image search. We design a *PageRank*-like method which can combine these different methods in order to provide an effective method for image search and ranking in social networks.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

Keywords

heterogeneous network; random walk; image retrieval

1. INTRODUCTION

The problem of image search is an important one in the context of web and social network applications [1, 2]. The growing repository of images on the web, social networks

such as *Facebook* and social media such as *Flickr* has created the need for effective methods for keyword-based search of such images. While the problem of ranking and searching large *text* collections has been studied extensively, the available solutions for keyword-based *image* search have remained relatively primitive. Typically, these methods make use of the limited knowledge of the surrounding text in order to provide query responses for a given set of keywords. These limitations are quite evident in a search engine such as *Google* in which the query results are sometimes unrelated to the keywords.

An important observation is that social networks have a rich level of information in terms of user linkages, user-to-image relationships, tags, comments, and descriptive text, which can be used in order to enhance the search process. In the context of a social network, the linkage relationships between social network actors and images may take on many different forms:

- The contribution of an image to a social network such as *Flickr* constitutes a linkage between that actor and that image.
- A comment, tag or “like” label by a user in any social network also constitutes a link between the user and the image.
- A friendship link between two users in a conventional social network such as *Facebook* may also be considered a link between the embedded images of the pages of the different users.
- The images in social networks may contain *content-based semantic links* (or *visual links*) between each other. This helps encode image similarity information into the search process.

Thus, a social network may also be considered a network of images, text, and users with heterogeneous links of different kinds between different nodes. In fact, such a network can be represented as a graph structure with nodes of different types and links representing the relationships between them. Since, the linkage information may represent user interest in images, and tags may represent image descriptions, the social network structure and content may contain rich information for improving the quality of the search process. In the traditional work on text-search in the IR-domain, such linkage information has often been leveraged in order to improve the quality of the underlying search in the context of the web. This is because similar documents are often

clustered together in the same network locality of most real linked entities such as the web. This class of problems is referred to as *keyword search in graphs* [10]. In the context of image data, the problem is much more complex, because of the greater heterogeneity of the problem in terms of linkage and content.

In this paper, we will use a more general approach to keyword search in graphs and allow the use of *image content information in addition to* social linkage information and the surrounding text, tags and comments for the search process. The goal of this work is to holistically use the social linkage structure with the *image and text content* of the social network.

The use of image content in the search process is still preliminary, but has seen some advances in recent years. For example, the work by [21] proposes a number of methods for using content information directly in the image retrieval process. The work in [15] uses ontological ranking, whereas the work by [4, 11] creates content-based linkages in order to provide visual cues during the search process. The content-based linkages are used in conjunction with the random surfer model of page-rank style algorithms in order to provide responses to queries. We note that since our model is general and uses a social network with heterogeneous linkages of different kinds, it is general enough to model the visual links within this framework.

Thus, the contributions of this work are as follows:

- We design an algorithm which uses the social network structure in addition to the keyword and text surrounding the image for ranking images with the use of a page-rank style algorithm. The algorithm is designed to work with and learn the importance of social links of different kinds for the ranking process. Furthermore, the surrounding or linked text of images is combined with the link structure during the ranking process using a personalized page-rank style method, so as to direct the ranking towards images which are related to the specified query-keywords.
- Since our approach is effective for ranking in networks which contain links of different kinds, it is easy to directly incorporate image content into the ranking process. Specifically, we use visual similarity between images as in [4, 11] in order to create links between images, and incorporate them into the ranking process. As in the previous case, the relative importance of different kinds of social links, keyword links and image content links is learned for the ranking process.
- One additional advantage of this approach is that it can also be used to improve the quality of retrieval in non-social scenarios by combining the effects of links of different kinds. In particular, this approach can also be used to improve the ranking quality to combining both image and keyword content effectively with the underlying network structure.

The remainder of the paper is organized as follows. In section 2, we provide a review of related work. Section 3 details a heterogeneous network model for the problem. The algorithmic design of the ranking algorithm is proposed in Section 4. The conclusions and summary are presented in section 6.

2. RELATED WORK

Content-based image retrieval has been a popular topic in recent years. Many of the earliest systems were developed for querying images with an image object [12, 13, 14] as the target. With the popularity of the web and related networks, keyword-based search has become increasingly relevant because of its natural semantic appeal, and the availability of surrounding text for web images. The focus of this paper is on keyword-based search, rather than image-based querying, though our ranking technique can be easily generalized to both scenarios because of its use of a heterogeneous network abstraction. Most keyword-based search methods rely purely on the use of surrounding text of the images in order to provide search results. On the other hand, for searching web-based *text* corpora, page-rank style methods are used quite often. In particular, even though *topic-sensitive* page rank methods [5, 6, 7, 9] have been used extensively in text search, their applicability has been rather limited [11] in the image domain. A natural technique for improving image ranking in the context of *network-based data* is to incorporate page-rank style random walk methods [3], which effectively combine the text, linkage structure, social cues, and image content in the search process.

Some recent work [11, 4, 15] designs methods for incorporating image features directly in the search process. The work in [15] uses content-based ontologies to improve image ranking, whereas that in [11, 4] uses *visual similarity links* between images in order to improve the ranking process. However, these techniques do not holistically use the text and image content along with network linkages in order to use all available information for ranking. The use of network linkage structure for improving the quality of retrieval has been well documented in text-based applications [10]. Nevertheless, this ability has continued to remain primitive in the context of network-based image retrieval applications. Furthermore, since many large network-based image collections are found in social media such as *Flickr*, it is also desirable to leverage this rich information in order to design an effective retrieval process. Sahbi and Audibert [23, 24] proposed to learn a network-dependent kernel or social network kernel (SNK) from the intrinsic image attributes and network linkage structure primarily for a single (image) domain, where both the target objects and the queried database correspond to images. In this paper we design a holistic framework to combine image and text content, and social and network linkage structure for effective image retrieval. We further note that our approach is not restricted to the social network scenario, but can be used for ranking in any heterogeneous network scenario containing a combination of different kinds of content and linkage information.

3. HETEROGENEOUS NETWORK MODEL

In this section, we define a heterogeneous network model for social media networks, which expressively utilizes multiple types of objects, linkages and their importance. We further note that the general principles of our approach can easily be extended to any kind of heterogeneous network, though the focus of this paper is on keyword-based image ranking in social networks.

Since the social network ranking process will use nodes with different kinds of data (text content, image content and users), we will first define a heterogeneous network model

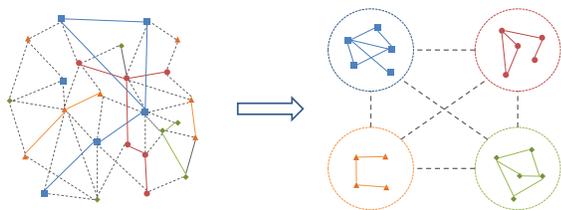


Figure 1: Heterogeneous network can be decomposed into homogeneous sub-networks connected by heterogeneous links (dashed black lines).

with different node *types*. We denote the set of types of nodes by \mathcal{D} with the corresponding cardinality $|\mathcal{D}|$ denoted by N ($N \geq 1$). We define a graph $G = \langle V, E, C \rangle$ with N types of nodes in V , the edges (or relations) between the nodes denoted by E , and the content sets attached to the different nodes by C . We note that C contains a data record for each node in V , which may be either text, image or a user identifier, depending upon the node type. We will see that the content plays a key role in defining the similarities and *implicit* links between nodes, which are helpful in the random walk-based ranking process. We refer to such a heterogeneous network of nodes, edges and weights as an information network. Note that a heterogeneous network has more than one type of node ($N \geq 2$), whereas a homogeneous network has only one type of node ($N = 1$).

In the case of heterogeneous networks, the graph G can be decomposed into the N homogeneous subgraphs $\{G^{T_i}\}_{i=1}^N$ whose nodes are of the same type and the inter-subgraph heterogeneous links E^H , as illustrated in Figure 1. Specifically, we can decompose the sets of nodes, links and the content of G into the constituent homogeneous components and intra-component edges as follows:

$$\begin{aligned} V(G) &= V(G^{T_1}) \cup V(G^{T_2}) \cup \dots \cup V(G^{T_N}) \\ E(G) &= E(G^{T_1}) \cup E(G^{T_2}) \cup \dots \cup E(G^{T_N}) \cup E^H \\ C(G) &= C(G^{T_1}) \cup C(G^{T_2}) \cup \dots \cup C(G^{T_N}) \end{aligned}$$

Here E^H is the set of heterogeneous links between nodes from different subgraphs.

We can use the above model to decompose the heterogeneous network problem into several homogeneous network problems and solve them, as long as we are careful to propagate information between the different subgraphs. We will show that it is possible to achieve this with the use of linkage information between different domains, and an iterative ranking-based information which can effectively leverage the content and linkage information across different domains.

Next, we discuss the instantiation of this general heterogeneous information network model to the problem of image ranking in social networks. Such social media networks can be modeled with three types of nodes, which correspond to *actor nodes*, *text nodes* and *image nodes*. These different kinds of nodes are connected by links of different kinds. We describe the different kinds of nodes and links below.

- The most basic kind of node in a social media application is an *actor* node. This could correspond to an user who may either contribute, comment on, or tag the images in the network, or to a user group whose mem-

bers share the same interests toward some kinds of images. Such an actor may be linked to either text nodes corresponding to their comments, tags, wall posts, or to image nodes depending upon their image contributions, or sharing behavior. The different kinds of links may refer to the relationship between the actor and the underlying text. The actors may also be directly connected to one another with the use of friendship or membership links.

- The image nodes represent the set of images in the social network which we would like to search. Such images are linked to the nodes corresponding to their surrounding text, the tags or comments by users, and wall posts depending upon the association between the underlying text and the image. The image nodes are also linked to the actors based on contribution, or sharing behavior of likeability flags. Furthermore, in order to incorporate the effect of *content-based* similarity, images may also be linked to one another based on a measure of visual similarity.
- The text nodes represent the copious amount of text available in any social or web network. Traditionally, the text content of web networks are the key factor which are used for image search. For example, search engines such as Google use *only* the surrounding text of an image for the ranking process. In this model, the text nodes may be connected to either actor nodes or image nodes. For example, a comment by a user on an image could be considered a text node which links to both that user and the corresponding image. The traditional linkage of an image with its surrounding text or linked text is also captured by this model.

We note that different social networks can be modeled in different ways depending upon the nature of the interconnectedness between the actors, the text content and the images. Nevertheless, most social networks can be modeled with these three kinds of nodes and links of different kinds. This model is quite general and expressive in capturing social linkages, image content linkages as well as the text-to-image relationships. The key is to be able to combine them in a *directed* page-rank style algorithm which can learn the semantic importance of different kinds of linkages and leverage them for the retrieval process.

4. THE SOCIALRANK ALGORITHM

In this section we design an effective algorithm named *SocialRank* to incorporate social, text and image content in the ranking process. In order to achieve this goal, we design an approach, which is based on random walks on our heterogeneous network model of images, text content and users. We use both the explicit links between different domains, as well as the similarities between nodes in order to design an effective ranking method. For convenience, we will denote the set of node types by $\mathcal{D} = \{T, I, A\}$, corresponding to the text nodes (v^T), image nodes (v^I), and actor nodes (v^A) respectively. While these three types suffice to expressively represent most social media networks, we note that our general heterogeneous network model can handle a wide variety of more general scenarios as well. In order to effectively rank the nodes, we introduce a measure r_v for node v , which indicates its *relevance importance*. This is

essentially the long term probability of our biased random walk to visit node v , and therefore provides the actual measure which is used in the ranking process. Furthermore, we use S_{ij}^h to denote the similarity between node v_i^h and v_j^h in domain h , i.e., $S_{ij}^h = \text{sim}(c_i^h, c_j^h)$, with c_i^h and c_j^h being the content of node v_i^h and v_j^h respectively. This similarity is very important in regulating the random walk between nodes, because our random walk process directly encodes these similarity values. In our application, the value of h is an alphanumeric superscript drawn from $\mathcal{D} = \{\text{T}, \text{I}, \text{A}\}$, depending upon whether we are working with text, image, or actor nodes. The similarity measure $\text{sim} : \langle C, C \rangle \rightarrow [0, 1]$ may be chosen differently depending upon the domain at hand, as long as it is nonnegative and normalized in the range $[0, 1]$ such that the pair of nodes with a higher value is more similar either in a visual or semantic sense. For the particular case of our application, we defined the similarities for the different domains as follows:

- **Image Domain:** We use visual similarity as the similarity measure between image nodes. Specifically, we first quantize SIFT descriptors into visual words with the code-books built by hierarchical clustering of randomly selected descriptors from the whole image collection. The obtained image representation is typically a sparse vector and allows the use of the inverted index to accelerate the computation of the similarity scores. In particular, we use cosine similarity with three kinds of weighting for each visual word in this paper, namely (1) number of co-occurrence term (COT); (2) term frequency (TF); and (3) term frequency - inverse document frequency (TF-IDF). For two images v_i^I and v_j^I with \mathbf{w}_i and \mathbf{w}_j as the visual word frequency column vector, the similarity score between them is defined as follows:

1. number of co-occurrence term (COT):

$$S_{ij}^I = \frac{\langle \delta(\mathbf{w}_i), \delta(\mathbf{w}_j) \rangle}{\|\delta(\mathbf{w}_i)\| \|\delta(\mathbf{w}_j)\|} \quad (1)$$

where $\delta(\cdot)$ is an element-wise Heaviside step function and $\langle \cdot, \cdot \rangle$ being the standard Hermitian inner product.

2. term frequency (TF):

$$S_{ij}^I = \frac{\langle \mathbf{w}_i, \mathbf{w}_j \rangle}{\|\mathbf{w}_i\| \|\mathbf{w}_j\|} \quad (2)$$

3. term frequency - inverse document frequency (TF-IDF):

$$S_{ij}^I = \frac{\langle \mathbf{w}_i, \mathbf{w}_j \rangle_K}{\|K^{1/2} \mathbf{w}_i\| \|K^{1/2} \mathbf{w}_j\|} \quad (3)$$

where K is a diagonal matrix whose elements are the inverse document frequency weight for each visual word, given by $K_{ii} = (\log \frac{\# \text{ documents}}{\# \text{ doc containing word } i})^2$. And $\langle \cdot, \cdot \rangle_K$ is the Hermitian inner product characterized by the Hermitian matrix K .

We will use all three of the above measures in our experimental results in order to show the effect of using different similarity measures.

- **Text Domain:** We use the normalized proportion of co-occurring words as the similarity measure in the

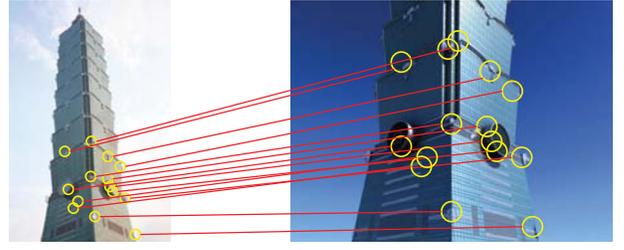


Figure 2: Similarity by matching visual words.

text domain. In particular, for two text node v_i^T and v_j^T with the word frequency represented content c_i and c_j , we have

$$S_{ij}^T = \frac{\langle \delta(c_i), \delta(c_j) \rangle}{\|\delta(c_i)\| \|\delta(c_j)\|} \quad (4)$$

where the numerator is the number of matched words between the two texts.

- **Actor (Social) Domain:** In this case, we made the simplifying assumption that the similarity between different user groups are 0 and the self similarity is 1. That is, $S^A = I$. This is because the actor nodes do not use any attribute-specific image content.

4.1 Hints from heterogeneous domains

The motivation of the proposed approach is to use the hints from other heterogeneous domains to improve the inference process during query processing. For a given query, the nodes in each of the domains can be ranked independently with a personalized-page rank style algorithm. For example, *VisualRank* [11] uses only visual content to rank the images. However, external knowledge has been shown to be very useful in recent research for a variety of mining tasks. For instance, transferring text domain knowledge from community annotated tags has demonstrated good improvement in [22], even when the tags are sometimes noisy. Besides the text domain, social linkages (which may represent user actions) are also very good knowledge sources for bridging the semantic gap by transferring the high level perception of the users into the ranking process. One example of such a phenomenon is that images favored by the same photo sharing group usually share similar visual content and perception of the underlying users.

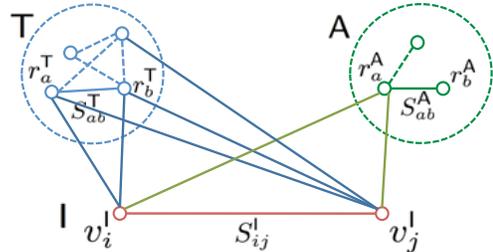


Figure 3: Hints from heterogeneous nodes.

While it is possible to directly use the content-based similarity value S_{ij}^h in order to influence the random walk, this does not incorporate a lot of information which is available in the linkages of the images to the text and actor nodes, as well as the linkages between those nodes themselves. In order to further incorporate this information into the random walk framework, we introduce an augmented similarity value \tilde{S}_{ij}^h between the different objects, which take such linkages into account. The augmented similarity is based on the idea from *SimRank* [17] that two objects are similar if they are related to similar objects. However, different from *SimRank*, the augmented similarity also considers the relevance importance of each object to the query in order to further affect the similarity function (and therefore influence the random walk). This augmented similarity value is a function of the similarity values and node relevance importance values in the other two domains. For example, Figure 3 provides an illustration in which v_i^l and v_j^l are nodes in the image domain, and their visual similarity is S_{ij}^l . In such a case, the augmented similarity value \tilde{S}_{ij}^l is expressed as a function $f(S_{ij}^l, S^A, r^A, S^T, r^T)$ of the original content-based similarity and the relevant parameters in the other two domains. Specifically, it is defined as follows:

$$\begin{aligned} \tilde{S}_{ij}^l &= f(S_{ij}^l, S^A, r^A, S^T, r^T) \\ &= S_{ij}^l + \sum_{h \in \{T, A\}} \beta^h \left(\sum_{a \in V_i^h, b \in V_j^h} S_{ab}^h r_a^h r_b^h \right) \end{aligned} \quad (5)$$

V_i^h and V_j^h are the sets of nodes in domain h that have a link to nodes v_i^l and v_j^l respectively. In particular, $V_i^h = \{v | v \in V^h, (i, v) \in E^{lh}\}$. β^h are weighting parameters for each heterogeneous domain. One can generalize Eq. (5) to other (text and actor) domains as well and write the equation generally in matrix form as follows:

$$\tilde{S}^d = S^d + \sum_{h \in \mathcal{D} - \{d\}} \beta^h L^{dh} R^h S^h R^h L^{dh'} \quad (6)$$

Here L^{dh} is the link map between domain d and h whose entry L_{ij}^{dh} is 1 if node i is connected to node j , otherwise it is 0. R^h is a diagonal matrix where $R_{ii}^h = r_i^h$.

This approach provides different levels of importance to different nodes (and the incident links) in the similarity function, because they are not of equal relevance to the query. Thus, the above augmented similarity function considers not only the link structures between domains but also takes into account the relevance of the nodes such that two images, which are both linked to a more important node would have higher similarity than a pair which are linked with a less important node. Since the random walk equations will be set up with the use of this similarity matrix, this also has a direct bearing on the query results.

After the augmented similarity matrix for each domain is constructed, we use a separate random walk process for each domain in order to determine the relevance scores of the nodes. As we will see later, the separate random walk processes for the different domains will be connected to one another by re-computation of the similarity matrices with the use of inter-domain linkages in an iterative way. The re-start probability of the random walk process is denoted

Algorithm 1 SocialRank with heterogeneous random walk

input similarity matrices S^d for each domain $d \in \mathcal{D}$, heterogeneous domain link E^H , parameters α, β

1. Initialize bias vector p by Eq. (8) or (9) and relevance score vector $r^{(0)} = [\frac{1}{N}]_{N \times 1}$
2. Set $t = 1$
- repeat**
- for** each homogeneous sub-network h **do**
3. Construct augmented similarity matrix \tilde{S}^h by Eq. (6) with $\tilde{S}^{d, (t-1)}$ and $r^{d, (t-1)}$, $d \in \mathcal{D} - \{h\}$
4. Apply random walk (7) to obtain relevance score vector $r^{h, (t)}$
- end for**
5. $t = t + 1$
- until** Convergence on r^d or maximum iteration number achieves.

output relevance score vector r^d for each domain

by α . The probability vector p is the restart vector. As we will discuss later, the value of p is biased by the query. The relevance score at iteration is t is denoted by $r^{(t)}$, and is determined with the use of page rank equations that are biased by the underlying similarity matrix \tilde{S}^d :

$$r^{(t)} = \alpha \tilde{S}^h D^{-1} r^{(t-1)} + (1 - \alpha)p \quad (7)$$

Here, D is a diagonal matrix whose i -th diagonal entry equals to the sum of the i -th row of \tilde{S}^h . As we will see later, the bias vectors play an important role in making the random walk process query-sensitive. We note that the relevance scores are a function of the similarity in the page rank equations (Equation (7)), whereas the similarity values were determined as a function of the relevance scores (and the hints from other domains) in Equation (5). This suggests an iterative approach for computing both from each other, as discussed below.

The pseudo-code of the *SocialRank* algorithm is provided in Algorithm 1. A biased re-start vector is used in the text domain order to encode the query into the ranking process. Specifically, the random walks in the text domain use a bias vector based on the query keywords. The algorithm performs random walks separately in each homogeneous sub-network. However, the random walks in each subnetwork are regulated by the similarity matrix as indicated in Equation (7). Since the similarity values are computed on the basis of the social cues from heterogeneous domains (based on Equation (5)), this ensures that social cues are passed from one domain to the other during the ranking process. At the same time, we re-compute the similarities between nodes in each iteration with the use of the relevance scores generated in the last iteration as indicated in Equation (5). Thus, the similarity values between nodes and the relevance importance values are inter-dependent upon one another in an iterative way. These steps are repeated until the different variables (node similarity scores and node relevance importance values) have converged. The node relevance importance values provide the ranking output for the algorithm.

4.2 Picking the Query-Bias Vector

The bias vector [7] is used to make the proposed *SocialRank* algorithm query-sensitive. For given query keywords

Q , we defined the query-bias vector p in the text domain T with n nodes ($v_1^\top, v_2^\top, \dots, v_n^\top$) and their corresponding content ($c_1^\top, c_2^\top, \dots, c_n^\top$) as follows:

$$p_{v_i}^\top = g \left(\sum_{q_j \in Q} \text{sim}(q_j, c_i^\top) \right) \quad (8)$$

The notation g represents a normalization function, which is used to ensure that the elements of the probability vector p sum to 1. Here $\text{sim}(q_j, c_i^\top)$ is the semantic similarity between the j -th query keyword and the text node v_i^\top which can be WordNet similarity [19], normalized Google distance [20] or simply the 0/1 match indicator. We adopt the binary match indicator in the experiments of this paper for its simplicity.

With the bias vector, we note that the proposed *SocialRank* algorithm will not only work with the keyword type of query, but can also be used with image queries. In order to enable query by images, one needs to start the *SocialRank* algorithm with random walks in the image domain and initialize the bias vector with the visual similarities between the query images and the database images. More specifically, we set the bias vector for query by images as follows:

$$p_i^1 = g \left(\text{sim}(q, c_i^1) \right) \quad (9)$$

This general principle applies to any kind of object query in a generic heterogeneous network model. The main difference is that *one needs to define a bias vector for the appropriate node type*, just as we defined one for the case of text nodes for keyword-based search. This makes the *SocialRank* algorithm much more general in its ability to handle queries and networks of different types. For example, one can use the approach to do group recommendation or image annotation for a given query image. Thus, this approach can be used in a wide variety of heterogeneous web-based scenarios in which it is possible to query and process objects of different types.

4.3 Convergence of SocialRank Algorithm

Next, we will show the convergence of the *SocialRank* algorithm, which uses iterative re-computation of the parameters across different homogeneous subnetworks. For simplicity, we use the case where the heterogeneous network G has only two homogeneous sub-networks, G_1 and G_2 . The steps of the proof can be easily generalized to the case where there are more than two sub-networks.

Before we start the proof, it is worth noting that the augmented similarity described in section 4.1 can be employed using a biased restart vector in the random walk, rather than adjusting transition probability mentioned in section 4.1. Indeed, these two approaches to exploit the hints from heterogeneous domains are equivalent and interchangeable since by letting $\hat{p} = \frac{\alpha}{1-\alpha} \hat{S} D^{-1} r^{(t-1)}$ for given $r^{(t-1)}$ we have

$$\begin{aligned} r^{(t)} &= \alpha(S + \hat{S})D^{-1}r^{(t-1)} + (1-\alpha)p \\ &= \alpha S D^{-1}r^{(t-1)} + \alpha \hat{S} D^{-1}r^{(t-1)} + (1-\alpha)p \\ &= \alpha S D^{-1}r^{(t-1)} + (1-\alpha)(p + \hat{p}) \end{aligned}$$

Now consider the random walks on G_1 with a stochastic matrix \mathbf{M}_{11} . When we choose the damping vector to be $\mathbf{p}_1 + \frac{\alpha}{1-\alpha} \mathbf{B}_{12} \bar{\mathbf{r}}_2$, the update formula for \mathbf{r}_1 is as follows:

$$\mathbf{r}_1 = \alpha \mathbf{M}_{11} \mathbf{r}_1 + ((1-\alpha)\mathbf{p}_1 + \alpha \mathbf{B}_{12} \bar{\mathbf{r}}_2) \quad (10)$$

The unique solution to (10) can be obtained in the form of the following expression:

$$\hat{\mathbf{r}}_1 = (I - \alpha \mathbf{M}_{11})^{-1} ((1-\alpha)\mathbf{p}_1 + \alpha \mathbf{B}_{12} \bar{\mathbf{r}}_2) \quad (11)$$

Similarly, the unique equilibrium state of the random walks on G_2 for a given vector $\mathbf{p}_2 + \frac{\alpha}{1-\alpha} \mathbf{B}_{21} \bar{\mathbf{r}}_1$ is as follows:

$$\hat{\mathbf{r}}_2 = (I - \alpha \mathbf{M}_{22})^{-1} ((1-\alpha)\mathbf{p}_2 + \alpha \mathbf{B}_{21} \bar{\mathbf{r}}_1) \quad (12)$$

As in the *SocialRank* algorithm, we iteratively solve for the equilibrium vectors of each homogeneous subgraph. The equations (11) and (12) can be written with a superscript t to indicate the iteration. Also, we denote $\bar{\mathbf{r}}$ to be the equilibrium vectors from previous iteration $\hat{\mathbf{r}}^{(t-1)}$. Therefore Equations (11) and (12) can be re-written as follows:

$$\hat{\mathbf{r}}_1^{(t)} = (I - \alpha \mathbf{M}_{11})^{-1} \left(\alpha \mathbf{B}_{12} \hat{\mathbf{r}}_2^{(t-1)} + (1-\alpha)\bar{\mathbf{p}}_1 \right) \quad (13)$$

$$\hat{\mathbf{r}}_2^{(t)} = (I - \alpha \mathbf{M}_{22})^{-1} \left(\alpha \mathbf{B}_{21} \hat{\mathbf{r}}_1^{(t-1)} + (1-\alpha)\bar{\mathbf{p}}_2 \right) \quad (14)$$

From (14) we can obtain $\hat{\mathbf{r}}_2^{(t-1)}$ and substitute it into (13) to get the following:

$$\begin{aligned} \hat{\mathbf{r}}_1^{(t)} &= \alpha^2 (I - \alpha \mathbf{M}_{11})^{-1} \mathbf{B}_{12} (I - \alpha \mathbf{M}_{22})^{-1} \mathbf{B}_{21} \hat{\mathbf{r}}_1^{(t-2)} + \\ & (1-\alpha) (I - \alpha \mathbf{M}_{11})^{-1} (\mathbf{p}_1 + \alpha \mathbf{B}_{12} (I - \alpha \mathbf{M}_{22})^{-1} \mathbf{p}_2) \end{aligned} \quad (15)$$

From the above power iteration (15) we can see that the *SocialRank* algorithm will converge to a unique solution for G_1 due to the irreducibility of \mathbf{M}_{11} . This follows the assumption that G_1 is strongly connected.

To see this, we choose $\mathbf{p}_1 = \frac{1}{N} \mathbf{1} = [\frac{1}{N}]_{N \times 1}$ (for ease of illustration only, \mathbf{p}_1 can be any positive vector, i.e., all the elements in \mathbf{p}_1 are all positive) and reorganize (15) as

$$\begin{aligned} \hat{\mathbf{r}}_1^{(t)} &= (I - \alpha \mathbf{M}_{11})^{-1} \left(\alpha^2 \mathbf{A} \hat{\mathbf{r}}_1^{(t-2)} + (1-\alpha)\mathbf{p}_1 \right) + \alpha(1-\alpha)\mathbf{b} \\ &= (I - \alpha \mathbf{M}_{11})^{-1} \left(\alpha^2 \mathbf{A} + \frac{1-\alpha}{N} \mathbf{1} \mathbf{1}^\top \right) \hat{\mathbf{r}}_1^{(t-2)} + \alpha(1-\alpha)\mathbf{b} \end{aligned} \quad (16)$$

where

$$\begin{aligned} \mathbf{A} &= \mathbf{B}_{12} (I - \alpha \mathbf{M}_{22})^{-1} \mathbf{B}_{21} \\ \mathbf{b} &= (I - \alpha \mathbf{M}_{11})^{-1} \mathbf{B}_{12} (I - \alpha \mathbf{M}_{22})^{-1} \mathbf{p}_2 \end{aligned}$$

Note that the matrix $(I - \alpha \mathbf{M}_{11})^{-1} \{ \alpha^2 \mathbf{A} + \frac{1-\alpha}{N} \mathbf{1} \mathbf{1}^\top \}$ is irreducible as \mathbf{M}_{11} and $\mathbf{1} \mathbf{1}^\top$ are irreducible and \mathbf{A} is non-negative. Hence the convergence of *SocialRank* to a unique solution is guaranteed.

To obtain the equilibrium vector of (15), we rearrange it as follows:

$$\begin{aligned} (I - \alpha \mathbf{M}_{11}) \hat{\mathbf{r}}_1^{(t)} &= \alpha^2 \mathbf{B}_{12} (I - \alpha \mathbf{M}_{22})^{-1} \mathbf{B}_{21} \hat{\mathbf{r}}_1^{(t-2)} + \\ & (1-\alpha) (\mathbf{p}_1 + \alpha \mathbf{B}_{12} (I - \alpha \mathbf{M}_{22})^{-1} \mathbf{p}_2) \end{aligned} \quad (17)$$

When (17) reaches the equilibrium state, meaning $\hat{\mathbf{r}}_1^{(t)} = \hat{\mathbf{r}}_1^{(t-2)} = \hat{\mathbf{r}}_1^*$, we will have the following:

$$\begin{aligned} (I - \alpha \mathbf{M}_{11} - \alpha^2 \mathbf{B}_{12} (I - \alpha \mathbf{M}_{22})^{-1} \mathbf{B}_{21}) \hat{\mathbf{r}}_1^* &= \\ (1-\alpha) (\mathbf{p}_1 + \alpha \mathbf{B}_{12} (I - \alpha \mathbf{M}_{22})^{-1} \mathbf{p}_2) \end{aligned} \quad (18)$$

This provides the following solution:

$$\hat{\mathbf{r}}_1^* = (1-\alpha) (I - \alpha \bar{\mathbf{M}}_{11})^{-1} \bar{\mathbf{p}}_1 \quad (19)$$

The variables on the RHS in the above equation are defined as follows:

$$\begin{aligned}\bar{\mathbf{M}}_{11} &= \mathbf{M}_{11} + \alpha \mathbf{B}_{12} (\mathbf{I} - \alpha \mathbf{M}_{22})^{-1} \mathbf{B}_{21} \\ \bar{\mathbf{p}}_1 &= \mathbf{p}_1 + \alpha \mathbf{B}_{12} (\mathbf{I} - \alpha \mathbf{M}_{22})^{-1} \mathbf{p}_2\end{aligned}\quad (20)$$

5. EXPERIMENTAL RESULTS

We evaluated the *SocialRank* algorithm, on a number of tasks which are either directly or indirectly related to the quality of the ranking process. We first describe the data sets.

5.1 Dataset

We evaluated the *SocialRank* algorithm with the use of images from the well known social media site known as *Flickr*. We crawled the *Flickr* site, and constructed the *Flickrgroup* dataset which includes images from 140 selected user groups in *Flickr*. The groups in *Flickr* are communities with people who have the same interests toward a target subject. The group members typically favor photos which are closely related to the target subject. A user group may favor hundreds to thousands of images favored. On the other hand, there are some groups having similar interests. Therefore, the same image may be favored by multiple user groups. The *Flickrgroup* dataset contained around 118,000 images in total, along with more than 150,000 unique tags associated with these images. We employed stop-word removal and stemming, and then selected the top 5,000 most frequent ones as the codebook tags.

For the image representation, we adopt a hierarchical vocabulary tree as the codebook to encode each image into a bag-of-words histogram. Specifically, all the images are first resized to maximum 500×500 . Then we use David Lowe’s SIFT binary¹ to extract 128-D SIFT keypoint descriptors for each image, where each descriptor is then quantized into a million of visual words with a 10^6 hierarchical vocabulary tree (6 layers and branch factor 10, trained by 20 million randomly selected SIFT descriptors). The final image representation is a sparse 10^6 -D vector. Due to its sparsity, the storage requirement for the representation as well as the time complexity for computing similarities between images is quite modest. Moreover, the tree structured codebook made it very suitable for building inverted index files. These files were need to compute the underlying similarities between the objects.

We evaluated the proposed algorithm on three social media related tasks, corresponding to image ranking, group ranking and group recommendation.

Throughout the experiments, we set the parameters as follows:

1. the damping factor in Equation (7):

$$\alpha = 0.85 \quad (21)$$

2. the weighting parameters for each domin in Equation (6):

$$\beta^h = \gamma \cdot \max(S^d) \quad (22)$$

where $\max(S^d)$ is the maximum entry in S^d and γ is a scalar between 0 and 1, which is set to 0.5 if not mentioned.

¹<http://www.cs.ubc.ca/~lowe/keypoints/>

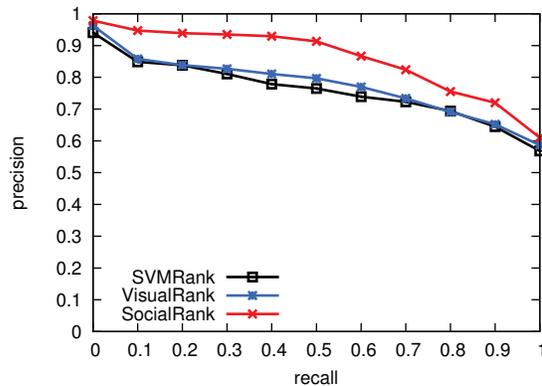


Figure 4: Interpolated precision-recall curve on the Flickrgroup dataset. The TF-IDF weighting scheme is used for similarity metric.

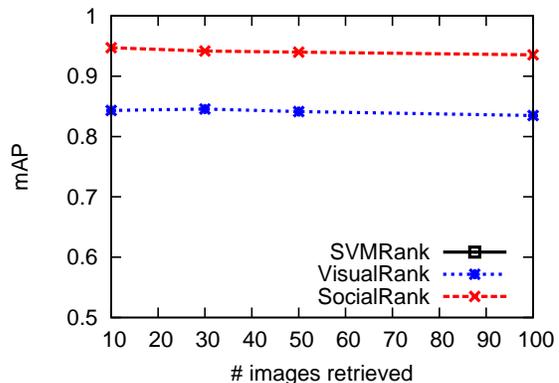


Figure 5: Image ranking performance (mean average precision) with various number of images retrieved. TF-IDF weighting scheme is used.

5.2 Image Ranking

We compare the proposed *SocialRank* algorithm with *SVM-Rank* [16] and *VisualRank* [11] with the standard performance metrics for information retrieval, i.e., average precision (AP). The comparison is built upon ten queries on the *Flickrgroup* dataset. We use the default linear kernel function and set the trade-off parameter for *SVMRank* as $C = 50$. The damping factor for *VisualRank* is set as $\alpha = 0.85$.

Table 1 shows the average precision of top 100 returned images (AP@100) for each query, where the proposed method demonstrates an impressive improvement over the other methods with the aid of social hints. It is interesting to note that *SVMRank* is quite sensitive to the similarity measures whereas the graph-based algorithms *VisualRank* and the proposed *SocialRank* show consistent performance with different similarity metrics. The main reason behind this would be that *SVMRank* considers only the similarities between training and testing samples while graph-based *VisualRank* and *SocialRank* additionally considers the similarities between testing samples, which provides a more robust linkage structure of samples. This reflects the advantage of semi-

Table 1: Image ranking performance (AP@100) on the *Flickr*group dataset. The one with best performance for each query is indicated by an oval box.

Query	<i>SVMRank</i>			<i>VisualRank</i>			<i>SocialRank</i>		
	COT	TF	TF-IDF	COT	TF	TF-IDF	COT	TF	TF-IDF
balloon	0.4896	0.7674	0.8676	0.8694	0.8606	0.8691	0.9593	0.9607	0.9605
bird	0.4621	0.6896	0.9031	0.9032	0.8744	0.8908	0.9847	0.9922	0.9900
box	0.7314	0.8905	0.9313	0.9276	0.9072	0.9062	0.9945	0.9944	0.9954
car	0.4679	0.6113	0.8382	0.8671	0.8609	0.8673	0.9872	0.9206	0.9155
cat	0.4390	0.5157	0.6226	0.6237	0.6517	0.6437	0.9854	0.9805	0.9818
child	0.7037	0.6723	0.8766	0.8739	0.8849	0.8754	0.9404	0.9405	0.9384
flower	0.6123	0.6997	0.7637	0.8046	0.7301	0.7546	0.9633	0.8119	0.9909
guitar	0.6048	0.7380	0.8021	0.8119	0.8047	0.8197	0.8328	0.8409	0.8333
pumpkin	0.7269	0.7416	0.7701	0.8027	0.7816	0.7816	0.7748	0.7748	0.7748
waterfall	0.8000	0.8321	0.8405	0.9389	0.9384	0.9401	0.9720	0.9707	0.9719
average	0.6037	0.7158	0.8215	0.8423	0.8294	0.8348	0.9394	0.9187	0.9352

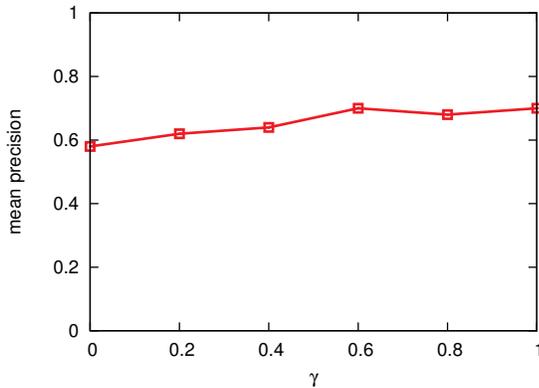


Figure 6: Sensitivity of parameter γ on the group ranking task.

supervised learning over supervised learning when one has no good estimation on the distribution of samples.

Figure 4 illustrates the interpolated precision-recall curves with 500 images retrieved using the 11 cutoff recall values (i.e., the precision at a given recall level is obtained for each query, then an averaged precision over all queries is reported). This figure indicates that our proposed algorithm achieves best performance in all the recall ranges. We also varied the number of images retrieved by plotting the mean average precision (mAP) for the top 10, 30, 50, 100, 200 and 500 images in Figure 5. The mAP is calculated as an average of AP over the 10 queries obtained for each N . It is evident that the proposed *SocialRank* algorithm outperforms the other two methods at all level of images retrieved, with around 12% improvement.

5.3 Group Ranking

One of the biggest differences between traditional multi-media applications and social media applications is that the former is image-centric (or data-centric), whereas the latter is more human-centric (or actor-centric). In other words, from a social perspective, it is also relevant to search for users or groups who may be interested in particular key-

words or images. One interesting aspect of the proposed *SocialRank* algorithm is that its seamless integration of different domains in the ranking process enables the search of objects of any type, including actors in social media. Indeed, the most relevant actors and images are simultaneously found by the *SocialRank* algorithm, with the use of inter-connected random walks in different domains.

Table 2 provides the top-5 results of group ranking from keyword-based queries performed for image ranking in section 5.2. The numbers in the parentheses are the relevance scores of the actor nodes in the ranking process. The groups in black color are considered the relevant groups (ground truth), based on the favored images for each group. The groups in red color are considered the irrelevant groups. The value of the precision with variation of the parameter γ is illustrated in Figure 6. It is evident that the results are not too sensitive to the value of the parameter γ . The best overall precision over 10 queries is around 0.70 for the top 5 results, when the value of γ was set to 0.6.

5.4 Group Recommendation

Recommendation is an important application in the context of a social media platform. On specific application in the context of an image sharing network such as *Flickr* is to recommend groups that may be interested in a newly shared image. Currently, such assignments are performed manually or are based on noisy tags [18]. However, none of these methods use the existing image collection of a group into account, which provide very useful hints in the commendation process. Therefore, we apply the *SocialRank* algorithm, by using a network model in which the actor nodes represent the different groups in the data. For each group, we compute a personalization vector which is proportional to the similarity between the query image and the representative images from each group. This is then used in conjunction with our ranking approach to determine the relevance scores of the group nodes. The titles of these group nodes are reported as the recommendations.

Table 3 demonstrates the group recommendation results. In each case, we have illustrated the query image, together with the titles of the group nodes which are recommended. The group names in blue are considered to be proper assignments while those in black are improper assignments.

Table 2: Top-5 user groups ranking corresponding to queries: balloon, bird, car and flower.

	Balloon	Bird	Flower	Car
COT	Red Balloon (0.803) Disney Balloons (0.032) Balloons n Fireworks (0.023) Painted Furniture (0.009) Party decorating ideas (0.008)	Birds birds birds birds (0.502) Birding In The Wild (0.046) we love wildlife (0.041) Nature Anything Goes (0.022) Painted Furniture (0.020)	Flowers all kinds (0.412) Painted Furniture (0.063) FLOWERS (0.050) Flower Flower Flower (0.048) Flowers Group (0.035)	Autos (CARS) (0.668) Auto Ads (0.063) Auto Glamma (0.054) Auto Shows (0.028) Red Space (0.007)
TF	Red Balloon (0.749) Disney Balloons (0.051) Balloons n Fireworks (0.049) Black 'n' Blue (0.014) UK night photography (0.010)	Birds birds birds birds (0.646) Birding In The Wild (0.025) we love wildlife (0.022) Nature Anything Goes (0.016) Painted Furniture (0.015)	Painted Furniture (0.135) Flowers all kinds (0.127) FLOWERS (0.113) Flower Flower Flower (0.109) Folk Music (0.033)	Auto Ads (0.382) Autos (CARS) (0.334) Auto Glamma (0.109) Auto Shows (0.017) Red Space (0.005)
TF-IDF	Red Balloon (0.806) Disney Balloons (0.027) Balloons n Fireworks (0.026) Black 'n' Blue (0.008) Painted Furniture (0.008)	Birds birds birds birds (0.635) Birding In The Wild (0.026) we love wildlife (0.023) Nature Anything Goes (0.016) Painted Furniture (0.014)	Flowers all kinds (0.319) FLOWERS (0.222) Flower Flower Flower (0.156) Painted Furniture (0.034) Flowers Group (0.017)	Auto Ads (0.313) Autos (CARS) (0.289) Auto Glamma (0.192) Auto Shows (0.026) Red Space (0.007)

Table 3: Top group recommendations corresponding to image queries. The group names in blue are considered to be proper assignments while those in black are improper assignments.

			
Snow Leopards (0.011) 1st Nature n Wildlife... (0.010) Natures waterfall (0.009) Flower Flower Flower (0.009) we love wildlife (0.009) Pet Portraits (0.009)	Tennessee Parks (0.010) Walls Have Paint (0.010) Singin' The Blues (0.010) Balloons And Fireworks (0.010) Birding In The Wild (0.009) Birds birds birds birds (0.009)	Street Lamps (0.026) Sky Chasers (0.011) The best self portraits (0.010) Black 'n' Blue (0.010) (Achromatic) Nature (0.009) City Sunsets (0.009)	Tennessee Parks (0.012) the skies above (0.010) Flower Flower Flower (0.009) FLOWERS (0.009) Sky Chasers (0.009) Flowers all kinds (0.009)
			
San Francisco at Night (0.013) Melbourne at Night (0.012) Corel Paint Shop Pro (0.012) UK night photography (0.011) Flickr's 1st Xmas group (0.011) City Sunsets (0.010)	Electronic Music (0.011) Concert Photographer (0.011) San Francisco at Night (0.011) Concert Photo Score Me (0.010) Music Directory (0.010) Concerts1 (0.010)	the skies above (0.011) Black 'n' Blue (0.010) Flowers all kinds (0.010) FLOWERS (0.010) Flower Flower Flower (0.010) Family Portraiture (0.010)	Black 'n' Blue (0.010) Sky Chasers (0.010) Snow Riders (0.010) natures composition (0.009) the skies above (0.009) City Sunsets (0.009)

It is evident that the *SocialRank* algorithm makes appropriate recommendations in most cases. However, it can sometimes recommend groups, in which the specific interests are not reflected in the visual content of the query image. For instance, a query image with ground truth group as “*San Francisco at Night*” would be visually very similar to a lot of photos in the group “*Melbourne at Night*”. This is of course a challenge of the problem formulation itself (and likely to be a problem for almost any method), because it is inherently hard to make such fine grained distinctions from the subjective information in a visual object. One inherent advantage of the social rank algorithm is that it allows the simultaneous use of *query-bias vectors* from different domains. For example, we can use an *additional* query-bias vector from the text domain in order to incorporate more detailed group

semantics in the ranking process. Thus, both the group semantics and the image visual content can be taken into account, which allows for more fine grained distinctions. This will be a direction of our future efforts.

6. CONCLUSIONS AND SUMMARY

In this paper, we proposed the *SocialRank* algorithm, which uses a heterogeneous network model to perform image ranking effectively with the use of linkages, content and social cues in the network. Our heterogeneous model is quite general, and can be used for network models which are much more complex than the social network scenario. Furthermore, while we have presented the results for the case of keyword-based queries, the method is general enough to be used for content-based queries of different types. An itera-

tive algorithm is proposed to learn the most effective ranking based on the relationships between the different network objects. We presented experimental results which show the effectiveness of our approach over competing methods.

Acknowledgements

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

7. REFERENCES

- [1] C. Aggarwal. Social network data analytics, *Springer*, 2011.
- [2] C. Aggarwal and H. Wang. Managing and mining graph data, *Springer*, 2010.
- [3] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *WWW Conference*, 1998.
- [4] L. Cao, A. D. Pozo, X. Jin, J. Luo, J. Han, and T. S. Huang. RankComplete: Simultaneous Clustering and Ranking of Web Photos, *WWW Conference*, 2010.
- [5] S. Chakrabarti. Dynamic Personalized Pagerank in Entity-relation Graphs, *WWW Conference*, 2007.
- [6] D. Fogaras, B. Racz, K. Csalogany, and T. Sarlos. Towards scaling fully personalized PageRank: Algorithms, lower bounds, and experiments. *Internet Mathematics*, 2(3):333-358, 2005.
- [7] T. H. Haveliwala. Topic-sensitive PageRank. *WWW Conference*, pp. 517-526, 2002.
- [8] X. He, W.-Y. Ma, and H. Zhang. ImageRank: Spectral Techniques for Structural Analysis of Image Database, *ICME*, vol 1, pp. 25-28, 2002.
- [9] G. Jeh and J. Widom. Scaling personalized web search. *WWW Conference*, pp. 271-279, 2003.
- [10] H. He, H. Wang, J. Yang, and P. Yu. BLINKS: Ranked Keyword Searches on Graphs, *SIGMOD Conference*, 2007.
- [11] Y. Jing, and S. Baluja. VisualRank: Applying PageRank to Large-Scale Image Search, *IEEE Transactions on Pattern Mining and Intelligence*, Vol. 30, No. 11, 2008.
- [12] R. P. A. Pentland, S. Sclaroff. Content-based Manipulation of Image Databases, *International Journal on Computer Vision*, pp. 233-254, 1996.
- [13] W.-Y. Ma, and B. S. Manjunath. A Toolbox for Navigating Large Image Databases, *Multimedia Systems* 3(7), pp. 184-198, 1999.
- [14] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based Image Retrieval in the Early Years, *IEEE Trans. on Pattern Mining and Intelligence*, 22(12), pp. 1349-1380, 2000.
- [15] S.-F. Tsai, M.-H. Tsai, and T. Huang. Content-based Image Retrieval with Ontological Ranking, *SPIE*, 2010.
- [16] T. Joachims. Training Linear SVMs in Linear Time, *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2006.
- [17] G. Jeh and J. Widom. SimRank: a measure of structural-context similarity, *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 538-543, 2002.
- [18] D. Liu, X.-S. Hua, L. Yang, M. Wang, H.-J. Zhang. Tag Ranking, *WWW Conference*, pp. 351-360, 2009.
- [19] T. Pedersen, S. Patwardhan, J. Michelizzi. WordNet::Similarity: measuring the relatedness of concepts, *HLT-NAACL*, pp. 38-41, 2004.
- [20] R. Cilibrasi and P. Vitanyi. The Google Similarity Distance, *IEEE Trans. on Knowl. and Data Eng.*, pp. 370-383, 2007.
- [21] Y. Rui, T. S. Huang and S.-F. Chang. Image Retrieval: Current Techniques, Promising Directions, and Open Issues, *Journal of Visual Communication and Image Representation.*, pp. 39-62, 1999.
- [22] G.-J. Qi, C. Aggarwal and T. Huang. Towards semantic knowledge propagation from text corpus to web images, *WWW Conference*, pp. 297-306, 2011.
- [23] H. Sahbi and J-Y. Audibert. Network-dependent kernels for image ranking, *ICIP Conference*, pp. 2357-2360, 2010.
- [24] H. Sahbi and J-Y. Audibert. Social network kernels for image ranking and retrieval, *Research Report, TELECOM ParisTech*, 2009.