Image Re-Ranking Based on Topic Diversity

Xueming Qian, *Member, IEEE*, Dan Lu, Yaxiong Wang, Li Zhu, Yuan Yan Tang, *Fellow, IEEE*, and Meng Wang, *Member, IEEE*

***Abstract*— Social media sharing Websites allow users to annotate images with free tags, which significantly contribute to the development of the web image retrieval. Tag-based image search is an important method to find images shared by users in social networks. However, how to make the top ranked result relevant and with diversity is challenging. In this paper, we propose a topic diverse ranking approach for tag-based image retrieval with the consideration of promoting the topic coverage performance. First, we construct a tag graph based on the similarity between each tag. Then, the community detection method is conducted to mine the topic community of each tag. After that, inter-community and intra-community ranking are introduced to obtain the final retrieved results. In the inter-community ranking process, an adaptive random walk model is employed to rank the community based on the multi-information of each topic community. Besides, we build an inverted index structure for images to accelerate the searching process. Experimental results on Flickr data set and NUS-Wide data sets show the effectiveness of the proposed approach.**

***Index Terms*— Social media, tag-based image retrieval, topic community, image search, re-ranking.**

I. Introduction

With the development of social media based on Web 2.0, amounts of images and videos spring up everywhere on the Internet. This phenomenon has brought great challenges to multimedia storage, indexing and retrieval. Generally speaking, tag-based image search is more commonly used in social media than content based image retrieval and content understanding [2]–[10], [13], [16], [22], [32], [38], [40], [44], [46], [47]–[51], [62], [63]. Thanks to the low relevance and diversity performance of initial retrieval results, the ranking problem in the tag-based image retrieval has gained researchers' wide attention [42], [43], [47]–[50].

Nonetheless, the following challenges block the path for the development of re-ranking technologies in the tag-based image retrieval.

A. Tag Mismatch

Social tagging requires users to label their uploaded images with their own keywords and share with others [25]. Different from ontology based image annotation, there is no predefined ontology or taxonomy in social image tagging. Every user has its own habit to tag images. Even for the same image, tags contributed by different users will be of great difference [25], [43], [47]. Thus, the same image can be interpreted in several ways with several different tags according to the background behind the image. In this case, many seemingly irrelevant tags are introduced.

B. Query Ambiguity

Users cannot precisely describe their request with a single word and tag suggestion systems always recommend words that are highly correlated to the existing tag set. Besides, polysemy and synonyms are the other causes of the query ambiguity.

Thus, a fundamental issue in the ranking of the tag-based social image retrieval is how to solve these problems reliably. As far as the "tag mismatch" problem is concerned, tag refinement [1], [20], [24]–[26], [45], tag relevance ranking [17], [35], [45] and image relevance ranking [3], [7], [15], [21], [27], [33], [34] approaches have been dedicated to overcome it. As for the "query ambiguity" problem, an effective approach is to provide diverse retrieval results that cover multiple topics underlying a query. Currently, image clustering [10], [43] and duplicate removal [5], [6], [9], [28], [29]–[31] are the major approaches in settling the diversity problem. However, most of the literature regards the diversity problem as to promote the visual diversity performance, but the promotion of the semantic coverage is often ignored. To diversify the top ranked search results from the semantic aspect, the topic community belongs to each image should be considered.

In recent years, more and more scholars pay attention to retrieval result's diversity [47], [48] [51], [54]–[61]. In [54], the authors first apply graph clustering to assign the images to clusters, then utilize random walk to obtain the final result. The diversity is achieved by set the transition probability of two images in different clusters higher than that in the same cluster. Tian *et al*. think the topic structure in the initial list is hierarchical [55]. They first organize images to different leaf topic, then define the topic cover score based on topic list, and finally use a greedy algorithm to obtain the highest

Manuscript received June 22, 2016; revised December 30, 2016 and April 9, 2017; accepted April 18, 2017. Date of publication April 28, 2017; date of current version June 7, 2017. This work was supported in part by the NSFC under Grant 61373113, Grant 61332018, and Grant u1531141, in part by the Microsoft Research, in part by the Humanities and Social Sciences Foundation of Ministry of Education under Grant 16XJAZH003, in part by the Special funding for Scientific Research Projects of the Central University under Grant sk2016013, and in part by Guangdong Science and Technology Department under Grant 2016A010101005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dimitrios Tzovaras. *(Corresponding author: Xueming Qian.)*

X. Qian is with the Key Laboratory for Intelligent Networks and Network Security, Ministry of Education, with the Smiles Laboratory, Xi'an Jiaotong University, Xi'an 710049, China, and with the Research Institute of Xi'an Jiao Tong University, Shunde, Guangdong (e-mail: qianxm@mail.xjtu.edu.cn).

D. Lu is with Smile Laboratory, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: ludandan@stu.xjtu.edu.cn).

Y. Wang is with Smile Laboratory, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: 764113779@qq.com).

L. Zhu is with software engineering, Xi'an Jiaotong Univeristy, Xi'an 710049, China (e-mail: zhuli@mail.xjtu.edu.cn).

Y. Y. Tang, is with Macau University, Macau 999078, China (e-mail: yytang@umac.mo).

M. Wang, is with Hefei University of Technology, Hefei 230009, China (e-mail eric.wangmeng@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2699623

1057-7149 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

topic cover score list. Dang-Nguyen et al. [56] first propose a clustering algorithm to obtain a topic tree, and then sort topics according to the number of images in the topic. In each cluster, the image uploaded by the user who has highest visual score is selected as the top ranked image. The second image is the one which has the largest distance to the first image. The third image is chosen as the image with the largest distance to both two previous images, and so on. In our previous work [47], the diversity is achieved based on social user re-ranking. We regard the images uploaded by the same user as a cluster and we pick one image from each cluster to achieve the diversity.

Most papers consider the diversity from visual perspective and achieve it by applying clustering on visual features [47], [48], [54]–[57]. In this paper, we focus on the topic diversity. We first group all the tags in the initial retrieval image list to make the tags with similar semantic be the same cluster, then assign images into different clusters. The images within the same cluster are viewed as the ones with similar semantics. After ranking the clusters and images in each cluster, we select one image from each cluster to achieving our semantic diversity.

In this paper, we propose to construct the tag graph and mine the topic community to diversify the semantic information of the retrieval results. The contributions of this paper are summarized as follows:

1) We propose a topic diverse ranking approach considering the topic coverage of the retrieved images. The inter-community ranking method and intra-community ranking methods are proposed to achieve a good trade-off between the diversity and relevance performance.

2) The tag graph construction based on each tag's word vector and community mining approach are employed in our approach to detect topic community. The mined community can represent each sub-topic under the given query. Besides, in order to represent the relationship of tags better, we train the word vector of each tag based on the English Wikipedia corpus with the model word2vec.

3) We rank each mined community according to their relevance level to the query. In the inter-community ranking process, an adaptive random walk model is employed to accomplish the ranking based on the relevance of each community with respect to the query, pair-wise similarity between each community, and the image number in each community. With the adaptive random walk model, the community that possesses the bigger semantic relevance value with the query and larger confidence value will be ranked higher.

Both the goals of this paper and our previous work [47] are to diversify the top ranked retrieval results. However they have considerable differences, which are summarized as follows:

First, in [47], we aim at diversifying the retrieval results by social user oriented re-ranking. We make the final result list contain images from different users as many as possible to achieve the diversity. While in this paper, our goal is to diversify the topics for the top ranked retrieval results. We apply (topic) community detection to make the final result list contain images with different semantics as many as possible.

Second, [47] computes the similarity between the user-oriented image set and query based on the co-occurrence tag mechanism, while this paper calculates the similarity between the tag community and query based on all of the tags in the community.

Third, the grouping step is not required in [47], because in the dataset every image has a user-id. However, in this paper, grouping images into different topic properly is a major problem.

Fourth, the relevance measurement approach for image and query is different. In [47], the relevance between image and query is represented by the average google distance of co-occurrence tags in tag collection of the image. In this paper, it is measured by the average google distance of all tags of the image.

Fifth, the ranking of image groups is different. In [47], we sort the image collections of different users according to their contributions, i.e. the number of co-occurrence tags of query in users' tag sets. In this paper, we sort the communities based on relevance scores obtained by random walk.

The remainder of this paper is organized as follows. In section II, we review the related work on the re-ranking of the tag-based image retrieval. The system overview is illustrated on section III. Section IV demonstrates the details of each process in our system. Experiments on Flickr dataset are setup and shown in section V. Finally, conclusion and future work are given in section VI.

II. Related Work

Social networks allow users to annotate their shared images with a set of descriptors such as tags. The tag-based image search can be easily accomplished by using the tags as query. However, the weakly relevant tags, noisy tags and duplicated information make the search results unsatisfactory. Most of the literature focuses on tag processing, image relevance ranking and diversity enhancement for the retrieval results. The following parts present the existing works related to the above three aspects respectively.

A. Tag Processing Strategy

It has been long acknowledged that tag ranking and refinement play an important role in the re-ranking of tag-based image retrieval, for they lay a firm foundation on the development of re-ranking in tag based image retrieval (TBIR). For example, Liu *et al*. [1] proposed a tag ranking method to rank the tags of a given image, in which probability density estimation is used to get the initial relevance scores and a random walk is proposed to refine these scores over a tag similarity graph. Similar to [1], and [26] sort the tag list by the tag relevance scores which are learned by counting votes from visually similar neighbors. The applications in tag-based image retrieval also have been conducted. Based on these initial efforts, Lee and Neve [64] proposed to learn the relevance of tag and image by visually weighted neighbor voting, a variant of the popular baseline neighbor voting algorithm. Agrawal and Chaudhary [17] proposed a relevance tag ranking algorithm, which can automatically rank tags according to their relevance with the constraint of image content. A modified probabilistic relevance estimation method is proposed by

taking the size of object into account. Furthermore, random walk based refinement is utilized to improve final retrieval results. Li [24] presented a tag fusion method for tag relevance estimation to solve the limitations of a single measurement on tag relevance. Besides, early and late fusion schemes for a neighbor voting based tag relevance estimator are conducted. Zhu *et al*. [34] proposed an adaptive teleportation random walk model on the voting graph which is constructed based on the images relationship to estimate the tag relevance. Moreover, many research efforts about the tag refinement emerged. Wu *et al*. [19] raised a tag completion algorithm to complete the missing tags and correct the erroneous tags for the given image. Qian et al. [42] proposed a retagging approach to cover a wide range of semantics, in which both the relevance of a tag to image as well as its semantic compensations to the already determined tags are fused to determine the final tag list of the given image. Gu et al. [45] proposed an image tagging approach by latent community classification and multi-kernel learning. Yang et al. [20] proposed a tag refinement module which leverages the abundant user-generated images and the associated tags as the "social assistance" to learn the classifiers to refine noisy tags of the web images directly. Qi et al. proposed a collective intelligence mining method to correct the erroneous tags [50].

B. Relevance Ranking Approach

To directly rank the raw photos without undergoing any intermediate tag processing, Liu et al. [3] utilized an optimization framework to automatically rank images based on their relevance scores to a given tag. Visual consistency between images and semantic information of tags are both considered. Gao et al. [7] proposed a hypergraph learning approach, which aims to estimate the relevance of images. They investigate the bag-of-words and bag-of-visual words of images, which is extracted from both the visual and textual information of image. Chen et al. [21] proposed a support vector machine classifier per query to learn relevance scores of its associated photos. Wu et al. [15] proposed a two-step similarity ranking scheme that aims to preserve both visual and semantic resemblance in the similarity ranking. In order to achieve this, a self-tune manifold ranking solution that focuses on the visual-based similarity ranking and a semantic-oriented similarity re-ranking method are included. Hu et al. [27] proposed an image ranking method which represents image by sets of regions and apply these representations to the multiple-instance learning based on the max margin framework. Yu et al. [35] proposed a learning based ranking model, in which both the click and visual feature are adopted simultaneously in the learning process. Specially, Haruechaiyasak and Damrongrat [33] proposed a content-based image retrieval method to improve the search results returned by tag-based image retrieval. In order to give users a better visual enjoyment, Chen et al. [18] proposed relevance-quality re-ranking approach to boost the quality of the retrieval images.

C. Diversity Enhancement

The relevance based image retrieval approaches can boost the relevance performance, but the diversity performance of searching is also very important. Many researchers dedicated their extensive efforts to make the top ranked results diversified. Leuken et al. studied three visually diverse ranking methods to re-rank the search results [10]. Different from clustering, Song et al. [9] proposed a re-ranking method to meet users' ambiguous needs by analyzing the topic richness. A diverse relevance ranking algorithm to maximize average diverse precision in the optimization framework by mining the semantic similarities of social images based on their visual features and tags is proposed in [5]. Sun et al. [28] proposed a social image ranking scheme to retrieve the images to meet the relevance, typicality and diversity criteria. They explored both semantic and visual information of images on the basis of [5]. Ksibi et al. [31] proposed to assign a dynamic trade-off between the relevance and diversity performance according to the ambiguity level of the given query. Based on [31], they further proposed a query expansion approach [6] to select the most representative concept weight by aggregating the weights of concepts from different views. Wang et al. [29] proposed a duplicate detection algorithm to represent images with hash code, so that large image database with similar hash codes can be grouped quickly. Qian et al. [48] proposed an approach for diversifying the landmark summarization from diverse viewpoints based on the relative view point of each image. The relative viewpoint of each image is represented with a 4-dimensional viewpoint vector. They select the relevant images with large viewpoint variations as top ranked images. Tong et al. achieved the diversity by introducing a diversity term in their model whose function is to punish the visual similarity between images [59], [60].

However, most of the above literatures view the diversity problem as to promote the visual diversity but not the topic coverage. As reported in [14], most people said they preferred the retrieval results with broad and interesting topics. So, many literatures about topic coverage are emerged [9] [23], [30], [49]. For instance, Agrawal et al. [23] classify the taxonomy over queries to represent the different aspects of query. This approach promotes documents that share a high number of classes with the query, while demoting those with classes already well represented in the ranking.

III. System Overview

Our system includes five main parts: 1) Tag graph construction based on the tag information of image dataset. Tag graph is constructed to mine the topic community. 2) Community detection. Affinity propagation clustering method is employed to detect topic communities. 3) Image community mapping process. We assign each image to a single community according to the tag overlap ratio between the topic community and image. 4) Inter-community ranking. We introduce the adaptive random walk model to rank topic communities according to the semantic relevance between the community and query. 5) Intra-community ranking. A regularization framework is proposed to determine the relevance of each image to the query by fusing the visual, semantic and view information into a unified system. We sequentially select the most relevant image in each ranked community as our final re-ranking results.

IV. The Proposed Method

Hereinafter the detail of each part is given. Some of the notations and their definitions are provided in APPENDIX A of the paper.

A. Tag Graph Construction

To realize fast retrieval, we build fast inverted index structure for the collected images as that utilized in our previous works [25], [43]. The inverted index structure is based on tags. Each tag corresponds to the images uploaded by different users. Let o denote the total number of tags in our image dataset and the corresponding tag set is denoted by $\mathbf{\Gamma} = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_o\}$. The term Γ_i denotes the i-th tag that users used to annotate their shared photos. The inverted index structure of the image dataset is described as $\boldsymbol{ID} = \{ID_1, ID_2, \ldots, ID_o\}$. ID_i is the image collection of tag Γ_i. That is to say, all images in ID_i have been tagged with Γ_i. For simplicity, we denote the image set containing query q by X. The corresponding image number in X is denoted by N. The tag set of X is denoted by $V = \{v_1, v_2, \ldots, v_N\}$. Thus, for each query q, we only need to conduct in dataset X.

In order to construct the tag graph, the representation of each tag must be learned. Word2vec [41] is a group of related models that are used to produce word embeddings. It has garnered a lot of interest in the text mining area.

In order to get a better representation for each tag, we employ the Word2vec based on the English Wikipedia dataset [52] to train each tag's word vector. To generate the word vectors well, we employ the Skip-gram model. After training, each word is represented by a vector with 100-dimension. Finally, we construct a high dimension word vector matrix $\boldsymbol{FW} = \{fw_1, fw_2, \ldots, fw_N\}$. Each row in matrix $\boldsymbol{FW}$ represents a training tag sample in $V = \{v_1, v_2, \ldots, v_N\}$, and the columns are the generated word vectors. As a consequence, the word has multiple degrees of similarity. It can be computed via a linear calculation. For example, vector ("Beijing") – vector ("China") + vector ("America") equals vector ("Washington").

After obtaining the word vector of each tag, we construct the undirected graph $\mathrm{G} = \{V_t, E\}$ based on the word vector similarities between each tag. In the graph $\mathrm{G} = \{V_t, E\}$, the elements of vertex set V_t are tags from $V = \{v_1, v_{2,} \ldots, v_N\}$. Two tags v_i and v_j are connected by edge e_{ij}. The weight of the edge e_{ij} is noted by c_{ij} which is determined by the cosine similarity between the word vectors of two tags as follows:

$$c(v_i, v_{\mathrm{j}}) = c_{ij} = \frac{< fw_i, fw_j >}{\|fw_i\| * \|fw_j\|} \tag{1}$$

where $< fw_i, fw_j >$ means the inner product of the two word vectors, $\|fw_i\|$ denotes the magnitude of the vector fw_i.

B. Community Detection

After we have constructed the Graph $G = \{V_t, E\}$, we employ the affinity propagation clustering method to mine the topic community based on this graph.

AP clustering has been successfully used in a series of areas [53], e.g., face recognition and document clustering. The affinity propagation clustering method can be conducted as follows:

Step 1: Initialization. Through the equation (1), we calculate the tag similarity matrix $\boldsymbol{C} = \{c_{11}, c_{12}, \ldots, c_{NN}\}$, and make the value in the diagonal line equal to the medium value of other values in $\boldsymbol{C}$. The "responsibility" $r(i,k)$, sent from data point i to candidate exemplar point k, reflects the accumulated evidence for how well-suited point k is to serve as the exemplar for point i, taking into account other potential exemplars for point i. The "availability" $a(i,k)$, sent from candidate exemplar point k to point i, reflects the accumulated evidence for how appropriate it would be for point i to choose point k as its exemplar, taking into account the support from other points that point k should be an exemplar. We initialize the responsibilities $r(i,k) = 0$ and the availabilities $a(i,k) = 0$.

Step 2: The responsibilities and availabilities are iteratively computed as follows:

$$r(i,k) = c(i,k) - \max_{k' \neq k} \left(a\left(i,k'\right) + c\left(i,k'\right) \right) \tag{2}$$

$$a(i,k) = \begin{cases} \min\left\{0, r(k,k) + \sum\limits_{i' \neq \{i,k\}} \max\left(0, r\left(i',k\right)\right)\right\}, & i \neq k \\ \sum\limits_{i' \neq k} \max\left(0, r\left(i',k\right)\right), & i = k \end{cases} \tag{3}$$

Step 3: Responsibilities and availabilities update as equation (2) and (3) till convergence. Then the exemplar of point i can be obtained by

$$\underset{k}{argmax}\left\{r(i,k) + a(i,k)\right\} \tag{4}$$

Through the above clustering method, we can obtain m detected communities $S = \{s_1, s_{2,} \ldots, s_m\}$.

After we obtain the m communities $S = \{s_1, s_{2,} \ldots, s_m\}$, the tag elements in community s_i can be described as $\{t_{i1}, t_{i2,} \ldots, t_{iZ_i}\}$, Z_i is the tag number in community s_i. Based on the tag set $V = \{v_1, v_2, \ldots, v_N\}$, the tag vector of community s_i can be rewritten as $ST_i = (st_{i1}, st_{i2,} \ldots, st_{iN})$, where each of the component st_{il} can be rewritten as follows:

$$st_{il} = \begin{cases} \log\left(\frac{Y}{R(v_l)}\right), & the\ community\ s_i \text{ contains tag } v_l \\ 0, & otherwise \end{cases} \tag{5}$$

where $R(v_l)$ is the number of images which tagged with v_l in image dataset X. Y is the image number of whole image dataset X. Here we choose log transform to mitigate the frequencies of different tags in image set and make components of ST_i change smoothly. This is benefit for our subsequent processing.

C. Image Mapping to Community

In this part, we aim to map each image $A \in X$ to a single community. The VSM (Vector Space Model) is employed to measure the tag overlap ratio h_i between the image A and the community s_i. The tag vector of image A can also be

rewritten as $IT = (it_1, it_{2,} \ldots, it_N)$ based on the tag set $V = \{v_1, v_{2,} \ldots, v_N\}$. Similar to the Eq.(5), the item it_l can be rewritten as follows:

$$it_l = \begin{cases} \log\left(\frac{Y}{R(v_l)}\right), & the\ image\ A \text{ contains tag } v_l \\ 0, & otherwise \end{cases} \quad (6)$$

The overlap ratio h_i of image A and community s_i can be calculated as follows:

$$h_i = \frac{\langle IT, ST_i \rangle}{\|IT\| * \|ST_i\|} \quad (7)$$

After we obtain the $h_i, i\epsilon(1, 2, \ldots, \mathrm{m})$, we assign the image A to the community which possess the highest overlap ratio. Then, the community s_i not only contains the tag set $\{t_{i1}, t_{i2,} \ldots, t_{iZ_i}\}$, but also contains the image set $X_i = \{x_{i1}, x_{i2,} \ldots, x_{il_i}\}$, l_i is the image number in community s_i.

D. The Inter-Community Ranking

In this part, we rank the m detected communities $S = \{s_1, s_{2,} \ldots, s_m\}$ according to their relevance scores with the query q. An adaptive random walk model [34] is employed in the semantic similarity matrix of communities to obtain the relevance value of each community. In the random walk model, we define a confidence factor to control the propagation direction of random walk model. We give a larger propagation weight to the community which has more images. The confidence factor λ_i of community s_i is defined as the normalized form of l_i, which is the image number of community s_i.

In order to rank these detected communities, we firstly calculate the semantic relevance Sq_i between the tag set $\{t_{i1}, t_{i2,} \ldots, t_{iZ_i}\}$ of each community s_i and the query q, which is defined as the mean cosine similarity between them. Besides, we count the tag histogram of images in each community s_i, and denote it as his_i. Then, the semantic similarity between communities s_i and s_j is SS_{ij}, which is cosine similarity of his_i and his_j. Thus, the semantic relevance Sq_i is as follows:

$$Sq_i = \frac{1}{Z_i} \sum_{l=1}^{Z_i} cos(t_{il}, q) \quad (8)$$

where $cos(t_{il}, q)$ is the cosine similarity between the word vectors of tags t_{il} and q. We denote p_{ij} as the normalized form of SS_{ij}:

$$p_{ij} = \frac{SS_{ij}}{\sum_{k=1}^{m} SS_{ik}} \quad (9)$$

Based on the above information, an adaptive random walk model is employed to obtain the relevance value $\boldsymbol{rs} = (rs_1, rs_{2,} \ldots, rs_m)$ of each community with the query q, where component rs_i represents the relevance score of community s_i and can be described as follows:

$$\mathrm{rs}_j(t) = \alpha \sum_i \lambda_\mathrm{i} p_{ij} \mathrm{rs}_i(t-1) + \alpha Sq_j \sum_i (1-\lambda_\mathrm{i}) \mathrm{rs}_i(t-1) + (1-\alpha) Sq_j \quad (10)$$

where $\alpha \in (0, 1)$ is the propagation factor and t is iteration times. We can obtain $\boldsymbol{rs} = (rs_1, rs_{2,} \ldots, rs_m)$ by solving (10) according to [34] as follows:

$$\boldsymbol{rs}_\pi = (1-\alpha)(\boldsymbol{I}_m - \alpha(\boldsymbol{P}^T \Lambda + \boldsymbol{Sqe}^T(\boldsymbol{I}_m - \Lambda)))^{-1} \boldsymbol{Sq} \quad (11)$$

where $\boldsymbol{rs}_\pi$ is the final answer of $\boldsymbol{rs} = (rs_1, rs_{2,} \ldots, rs_\mathrm{m})$. $\boldsymbol{P}$ is the matrix form of p_{ij}. Λ is the diagonal form of $\lambda_\mathrm{i}, i \in (1, 2, \ldots, m)$. $\boldsymbol{Sq}$ is matrix form of $Sq_i, i \in (1, 2, \ldots, \mathrm{m})$. $\boldsymbol{e} = \{1, 1, \ldots, 1\}^T$ with the dimension m. $\boldsymbol{I}_m$ is the identity matrix with the size $m \times m$. Matrix $(\boldsymbol{I}_m - \alpha(\boldsymbol{P}^T \Lambda + \boldsymbol{Sqe}^T(\boldsymbol{I}_m - \Lambda)))$ in Eq.(11) is always invertible, detailed explanation can be found in [34].

After obtaining $\boldsymbol{rs} = (rs_1, rs_{2,} \ldots, rs_m)$, the communities $S = \{s_1, s_2, \ldots, s_m\}$ can be ranked by their relevance value, then we can get the ranked community set.

E. Intra-Community Ranking

After inter-community ranking, we implement intra-community re-ranking to select the image which has the highest relevant score among each community's image set. We take the image set X of a community $s \in \mathrm{S}$ as an example to demonstrate our intra-community ranking process.

Similar to our previous work [47], Our regularization framework is defined as follows:

$$\mathrm{Q}(\boldsymbol{rm}) = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} \left\{ \frac{rm_i}{\sqrt{D}_{ii}} - \frac{rm_j}{\sqrt{D}_{jj}} \right\}^2 + \beta \sum_{i=1}^{n} (rm_i - Sc_i)^2 + \mu \sum_{i=1}^{n} (rm_i - v t_i)^2 \quad (12)$$

where $\mathrm{Q}(\boldsymbol{rm})$ is the cost function; rm_i is the relevance score of image $x_i \in \mathrm{X}, i = 1, 2, \ldots n$, $D_{ii} = \sum_{j=1}^{n} w_{ij}$. Here, w_{ij} can be directly calculated using Gaussian kernel function with a radius parameter σ as follows:

$$w_{ij} = exp\left(-\frac{||v_i - v_j||^2}{2\sigma^2}\right) \quad (13)$$

where $||\cdot||^2$ stands for the l_2-norm of the vector. Furthermore, σ represents the radius parameter which is set to be the mean value of all pairwise Euclidean distance between images.

Sc_i is the semantic relevance score of image x_i, which is defined based on Google distance [11] as follows:

$$Sc_i = \frac{1}{B} \sum_{k=1}^{B} GD_k \quad (14)$$

where B is the tag number in image x_i, GD_k is the similarity based on google distance between query and the k-th tag of image, defined as :

$$GD_k = \exp\left(-\frac{\max\{\log R(q), \log R(v_k)\} - \log R(q, v_k)}{\log Y - \min\{\log R(q), \log(v_k)\}}\right) \quad (15)$$

where $R(q, v_k)$ represents the number of image tagged by query q and v_k. In this paper, we choose Eq.(14) to calculated the semantic relevance score between image and the query is based on the fact that tags of image are ranged in random order

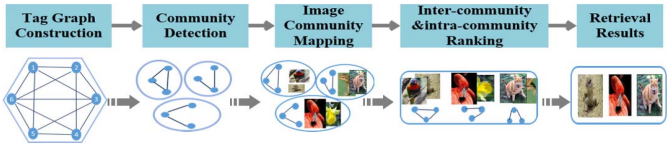


Fig. 1. The framework of our proposed method.

Fig. 2. An exemplary image from Flickr and its associated information.

as shown in Fig. 2. Google distance has only relationship with the tag collection and is irrelevant to the order of tags. It is appropriate for semantic metric.

More explanation for Eq.(12) can be found in [47]. Our optimization problem is to minimize cost function defined by Eq.(12) as follows:

$$\boldsymbol{rm}_{\pi} = argmin\left(Q\left(\boldsymbol{rm}\right)\right) \tag{16}$$

To get $\boldsymbol{rm}_{\pi}$, we can use iterative optimization algorithm in our previous work [47] to solve this problem. $Q(\boldsymbol{rm})$ can be rewritten as the matrix form as follows:

$$\begin{aligned} \mathrm{Q}\left(\boldsymbol{rm}\right) = \boldsymbol{rm}^{T}\left(\boldsymbol{I}_{n} - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{W}\boldsymbol{D}^{-\frac{1}{2}}\right)\boldsymbol{rm} + \beta||\boldsymbol{rm} - \boldsymbol{Sc}||^{2} \\ + \mu||\boldsymbol{rm} - \boldsymbol{VT}||^{2} \end{aligned} \tag{17}$$

where $\boldsymbol{D} = \mathrm{Diag}(D_{11}, D_{22}, \ldots, D_{nn})$, $\boldsymbol{Sc} = (Sc_1, Sc_2, \ldots, Sc_n)$, and $\boldsymbol{VT} = (vt_1, vt_2, \ldots, vt_n)$, $\boldsymbol{I}_n$ is a unit matrix with size $n \times n$. This approach avoids the intensive computation brought by the direct matrix inversion in Eq.(11). We give the detailed iterative steps for solving Eq.(17) in APPENDIX B.

V. Experiments

In order to demonstrate the effectiveness of the proposed topic diverse ranking (denoted by TDR) based image retrieval approach, we conduct experiments on our crawled Flickr images [43], [47] and NUS-wide. We will give the detailed descriptions of our dataset in next subsection. In order to evaluate the performance of different methods, we utilize following 20 tags as query: airplane, beach, bird, blue, buildings, Christmas, forest, reflection, garden, girl, ocean, orange, sea, sky, animal, and etc. We systematically make comparisons for the following five tag-based image retrieval approaches:

1) RR: Relevance-based ranking [3], an optimization framework is applied to automatically re-rank images based on visual and semantic information.

2) DRR: Diverse relevance ranking [5], which optimizes an ADP measure with the consideration of the semantic and visual information of images.

3) DR: Diverse ranking [9]. First, the topic coverage of each image is calculated. Then, PageRank model based on the topic coverage is utilized to re-rank the initial retrieval results.

4) SR: Social ranking [43], [47]. User information is utilized to boost the diversity performance. A regularization framework which fuses the semantic, visual and views information is introduced to improve the relevance.

5) TDR: Topic Diverse Ranking. Tag graph and community detection method are utilized to boost the diversity performance. A regularization framework which fuses the semantic, visual and view information is introduced to improve the relevance performance. In order to train the word vector of each tag, Word2vec model is conducted to train each tag's word vector.

We utilize the 45-D Color moment and 170-D hierarchical wavelet packet descriptor [12] to represent the visual feature of each image. For Color moment, an image is divided into four equal sized blocks and a centralized image with equal-size. For each block, a 9-D color moment is computed, thus the dimension of color moment for each image is 45.

View information is not only important for image retrieval performance [35], [39] but also for documents relevance estimation [36], [37]. In this paper, $view_i$ represents the view times of the image i. Its normalized form vt_i can be described as follows:

$$vt_i = \frac{view_i - view_{min}}{view_{max} - view_{min}} \tag{18}$$

where $view_{max}$ and $view_{min}$ are the maximum and minimum "views" of the images [47].

A. Flickr Dataset

In order to evaluate the performance of our method, we randomly crawled more than 6 million images together with their associated information from Flickr through its public API [43], [47]. This data set contains 7,279 users and 6,593,096 images, but only 7,090 users upload images and 5,318,503 images contain tags and view information. These images cover lots of categories including scenery, object, cartoon, art, person, behavior, buildings, trademark, portrait and so on.

B. Performance Evaluation

The performance evaluation of our method is voted by five volunteers who are invited to assign the scores for the retrieval results under each query. The average score is used to measure the correlation between the query and the retrieval results.

In each specific query, the five volunteers are asked to give the relevance score to each of the top ranked image according to their judgment for the relationship between each image and query. The relevance score is confined with in the following three categories: 2-relevant, 1-hard to tell, 0-irrelevant. The three categories are good for volunteers to distinguish. More categories will result in more deviation. Then, the relevance score of the image i under query q is obtained by averaging the assigned relevance scores. We denote the relevance value of image i by rel_i.

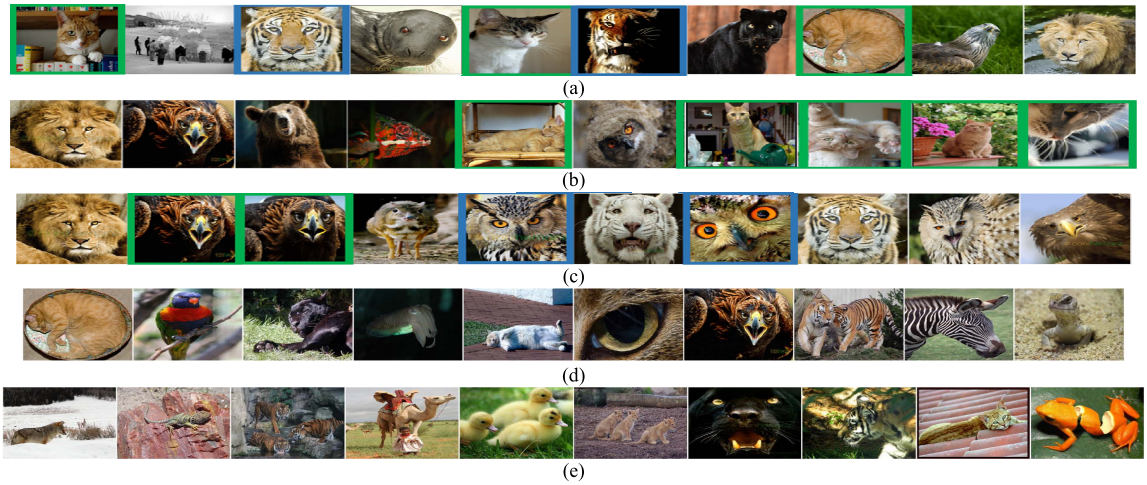


Fig. 3. Top 10 Ranking results of different methods for query *animal*. (a) Searching results using RR. (b) Searching results using DRR. (c) Searching results using DR. (d) Searching results using SR. (e) Searching results using TDR.

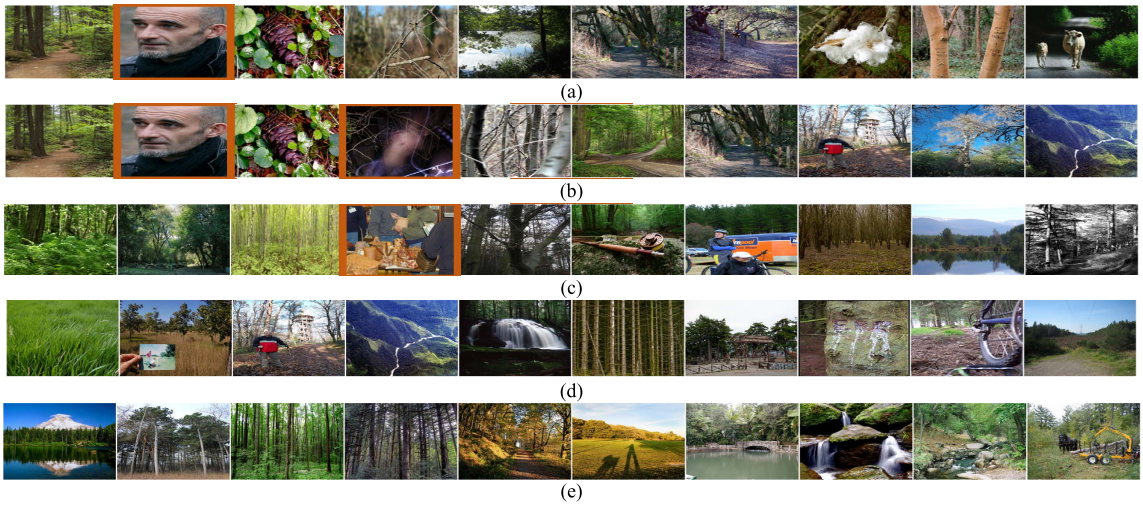


Fig. 4. Top 10 Ranking results of different methods for query *forest*. (a) Searching results using RR. (b) Searching results using DRR. (c) Searching results using DR. (d) Searching results using SR. (e) Searching results using TDR.

1) Criteria of Performance Evaluation: We use the *NDCG* [43] and average precision under depth n (denoted as $AP@n$) to measure the relevance performances of retrieval results, which are expressed as follows:

$$NDCG@n = \frac{1}{W}\sum_{i=1}^{n}\frac{2^{rel_i}-1}{log(1+i)} \tag{19}$$

$$AP@n = \frac{1}{2n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i}\frac{rel_j}{i}\right) \tag{20}$$

where W is a normalization constant. It makes the optimal ranking's NDCG score to be 1 and makes AP be in [0,1].

Suppose the top n retrieval image results are represented by $I = \{i_1, i_2, \ldots i_n\}$. And the image i_k has M_k tags. Then the diversity score is used to measure the topic coverage of the top ranked images in I is denoted as follows:

$$DS@n = \frac{1}{n}\sum_{k=1}^{n} DS(i_k) \tag{21}$$

$$DS_I(i_k) = \frac{1}{M_k}\sum_{j=1}^{M_k}\frac{1}{N_{t_j}^I} \tag{22}$$

where $DS(i_k)$ is the diversity score of image i_k in I, and $N_{t_j}^I$ denotes the image number in the top ranked list which is associated with tag t_j. $DS@n$ represents the average diversity score. It is used to evaluate topic coverage of the top n results.

Moreover, we can get the average diverse precision under depth n (denoted as $ADP@n$) as follows:

$$ADP@n = \frac{1}{2n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i}\frac{rel_j}{i}\right) * DS@i \tag{23}$$

2) Exemplar Search Results: The top 10 results of exemplar queries: animal and forest on Flickr database of the five different ranking algorithms are shown in Fig. 3 and Fig. 4 respectively. The images marked with the red border are irrelevant with the query. Besides, we mark the similar images by the borders with the same color.

We find that the top ranked images determined by RR, DR and DRR all suffer from the lack of diversity. Their retrieval results all contain images from the same topic as shown in Fig. 3. The RR method aims to rank the images based on their relevance scores, ignoring the diversity. For example, the

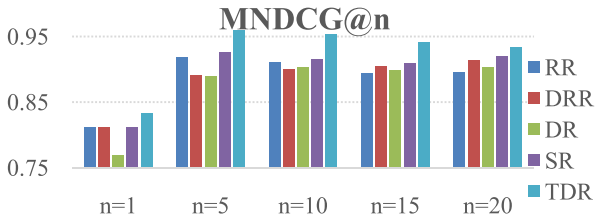


Fig. 5. The NDCG of all 5 ranking methods under different depths.

search results of RR are shown in Fig. 3 (a), from which we find that the second and the third images, the fifth and the seventh images are the same animal species. Besides, RR method introduces the irrelevant images in the top ranked retrieval results of "forest" as shown in Fig. 4 (a). The DRR introduces the semantic similarity restriction to enhance the diversity. DR method introduces re-ranking which improves the topic coverage. The retrieval results of "animal" for DRR and DR have the same topic image which are illustrated in Fig. 3 (b) and (c). The retrieval results of "forest" contain the irrelevant images as shown in Fig. 4 (b) and (c).

From Fig. 3 (a) $\sim$ (c), we find that the top ranked images are with similar topics. From Fig. 4 (a) $\sim$ (c), we find that there are some irrelevant images in the top ranked results. However, SR and TDR which employ the clustering idea to improve the diversity performance. They remove redundant images with the same topic images and irrelevant images successfully from the top ranked results.

From the examples shown in Fig. 3 and Fig. 4, we can acknowledge that SR and TDR solve the deficiencies of existing tag based image retrieval approaches and make a better trade-off between the diversity and relevance.

3) Performance Analysis: To make fair comparisons for the methods RR, DR, DRR, SR and TDR, the parameter β in the RR is set to be 1 (method suggested), the parameter β in the DRR is set to be 0.1 (method suggested), the parameter β and μ in SR are set to $\beta = 10$, $\mu = 1$ (method suggested), the parameters β and μ in TDR are set to $\beta = 5$, $\mu = 1$. The discussions on β and μ are illustrated in the next subsection.

Let MAP@n and MNDCG@n denote the mean values of AP@n and NDCG@n for all of the 20 query tags. Let MDS@n and MADP@n denote the mean values of DS@n and ADP@n for all the 20 query tags. The NDCG@n, MAP@n, MDS@n and MADP@nwith n $= 1, 5, 10, 15,$ and 20 are shown in Fig. 5, Fig. 6, Fig. 7 and Fig. 8 respectively. For example, the MNDCG@20 of RR, DRR, DR, SR and TDR are 0.89, 0.91, 0.90, 0.92 and 0.93 respectively, while their MDS@20 values are 0.11, 0.12, 0.17, 0.28 and 0.47 respectively.

We find that the DRR achieves a little higher under NDCG and MAP, and much higher under MDS than those of RR. Besides, the DR has a little higher NDCG@20 and a little bigger MDS@20 than the RR method. From this, we can acknowledge that using the constructed mathematical formula to diversify the tag information of the retrieval images contributes to the promotion of the diversity, but adds no contribution to improve the relevance.

Using the clustering idea and user information not only add great contribution to the relevance performance but also

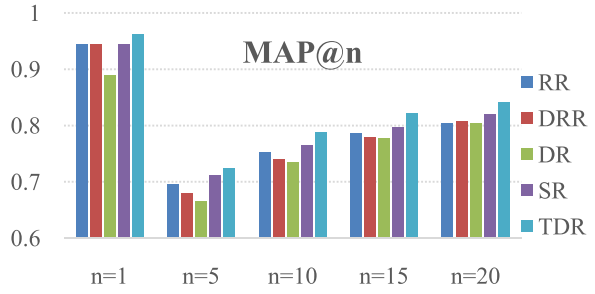


Fig. 6. The MAP of all 5 ranking methods under different depths.

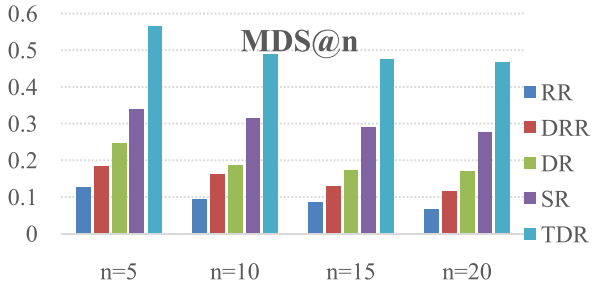


Fig. 7. The MDS of all 5 ranking methods under different depths.

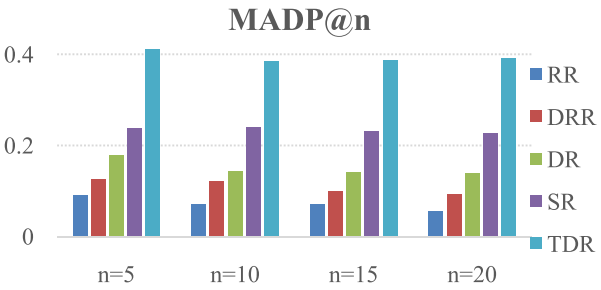


Fig. 8. The MADP of all 5 ranking methods under different depths.

the diversity performance, just as Fig. 5, Fig. 6 and Fig. 7 shown. Besides, TDR employs the constructed graph and the community detection method to make the relevance and diversity performance reaching the local optimal.

From the experimental results, we can find that the SR and TDR both get better diversity performance as shown in Fig. 7. However, SR aims to enhance the visual diversity performance.

TDR aims to improve retrieval performances by enhancing the topic coverage. They both make a better trade-off between the relevance and diversity performance by employing the clustering idea.

4) Experiment on NUS-Wide: In this part, we show our simple experiment results on NUS-Wide experiment. We only make comparisons for four tag-based image retrieval approaches: RR, DR, DRR, and TDR Fig. 9 and Fig. 10 show the results of MAP and MADP respectively. As the ground truth provided by NUS-Wide only have two level:1-relevant, 0-irrelevant, we remove the constant $\frac{1}{2}$ in Eq.(20) and Eq.(23) to make MAP and MADP be in [0,1].

From Fig. 9, we can see that our relevance is outstanding than compared method. From Fig. 10, we find that our diversity also outperforms the compared approaches by a large margin.

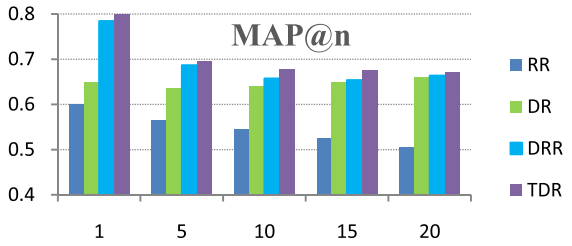


Fig. 9. The MAP of all 4 ranking methods under different depths.

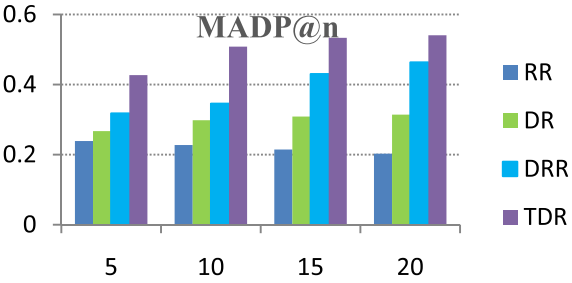


Fig. 10. The MADP of all 4 ranking methods under different depths.

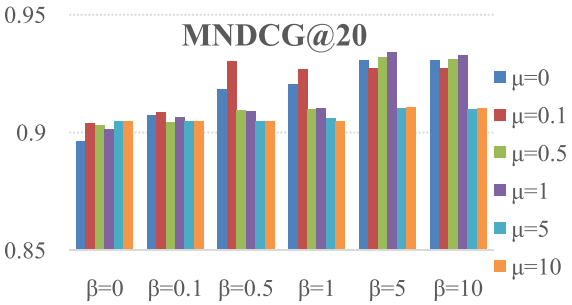


Fig. 11. Impact of parameters to MNDCG@20 of TDR under $\beta = \{0, 0.1, 0.5, 1, 5, 10\}$ with fixed $\mu = \{0, 0.1, 0.5, 1, 5, 10\}$.

VI. Discussions

In this section, we completely discuss the impact of different parameters and the metric methods involved in our topic diversity re-ranking method. For parameter, we will discuss the impacts of β and μ in Eq.(12) and α in Eq.(10) to the image retrieval performance. For metric method, we will discuss about the relevance metric in Eq.(8) and Eq.(14). Besides parameters and metric methods, simper community detection and different diversity methods are also discussed in this section.

A. Discussions on Weight β and μ Selection

In this part, the impacts of the regularization parameter β and μ with fixed $\alpha = 0.2$ on the performance of TDR areshown. Fig. 11-Fig.14 demonstrate the MNDCG@20, MAP@20, MDS@20 and MADP@20 performances of TDR under $\beta = [0, 0.1, 0.5, 1, 5, 10]$ and $\mu = [0, 0.1, 0.5, 1, 5, 10]$.

As can be seen, the MAP@20 and MNDCG@20 of TDR with fixed $\mu = 0$ (under the case that$\beta \neq 0$ and $\mu = 0$) is the biggest when $\beta = 5, 10$; the MAP@20 and MNDCG@20 of TDR under fixed $\beta = 0$ (under the case that $\beta = 0$ and $\mu \neq 0$) change slightly under each various μ. Besides, TDR achieves the local maxima of MDS@20 at $\beta = 5$ and $\mu = 1$. In image retrieval, the relevant performance is the first to be

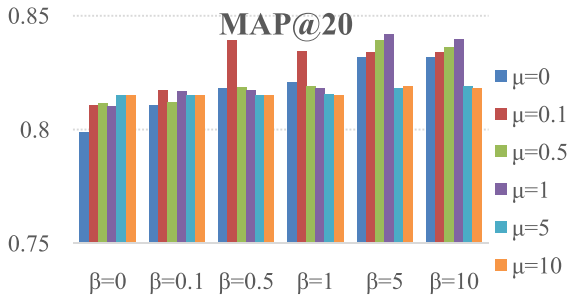


Fig. 12. Impact of parameters to MAP@20 of TDR under $\beta = \{0, 0.1, 0.5, 1, 5, 10\}$ with fixed $\mu = \{0, 0.1, 0.5, 1, 5, 10\}$.

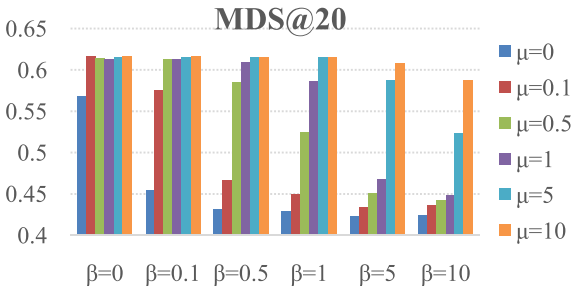


Fig. 13. Impact of parameters to MDS@20 of TDR under $\beta = \{0, 0.1, 0.5, 1, 5, 10\}$ with fixed $\mu = \{0, 0.1, 0.5, 1, 5, 10\}$.

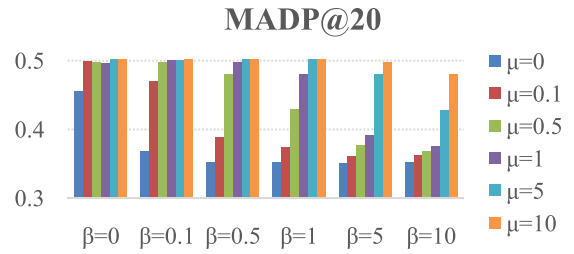


Fig. 14. Impact of parameters to MADP@20 of TDR under $\beta = \{0, 0.1, 0.5, 1, 5, 10\}$ with fixed $\mu = \{0, 0.1, 0.5, 1, 5, 10\}$.

considered, the diversity performance is the second. Hence, the parameters selection in TDR should be $\beta = 5$, $\mu = 1$.

From Fig. 11 and Fig. 12, we find that TDR under the case that $\beta = 0$ and $\mu = 0$ is with lowest NDCG@20 and MAP@20. This indicates that the image ranking performances only with visual information are not satisfactory. When utilizing the semantic information but without the view information (under the case that $\beta \neq 0$ and $\mu = 0$), with the β grows, MAP@20 and NDCG@20 becomes higher and higher, MDS@20 becomes lower and lower. When utilizing the view information (under the case that $\beta = 0$ and $\mu \neq 0$), with the μ grows, MAP@20 and NDCG@20 change slightly, the variation of MDS@20 is stable.

From Fig. 11 and Fig. 12, $\beta = 5$ suggests that the regularization, $\sum_{i=1}^{l_h} (rm_i - vt_i)^2$, seems not so important. However, from Fig. 14 and Fig. 15, we find that the regularization indeed enhance diversity. For example, fix $\beta = 5$, when μ takes value $\{0, 0.1, 0.5, 1.5, 10\}$, the corresponding MDS@20 values are 0.422, 0.434, 0.451, 0.468, 0.588, 0.608 respectively in Fig.14, and the corresponding MADP@20 values are 0.350, 0.361, 0.377, 0.392, 0.481, 0.499 respectively in Fig. 14. When we

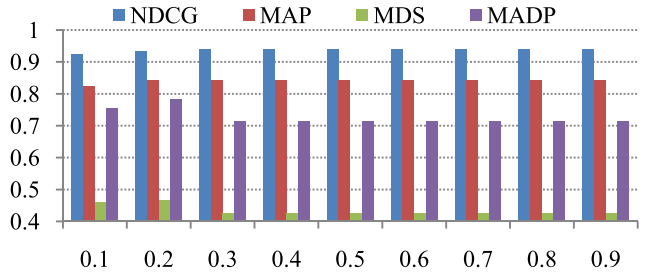


Fig. 15. Impact of parameters to the TDR overall performance under $\alpha = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ in the depth 20.

fix $\beta = 10$, the MDS ranges from 0.437 to 0.587, and MADP ranges from 0.352 to 0.481, when the μ takes value $\{0, 0.1, 0.5, 1.5, 10\}$. As β takes $0, 0.1$, and 0.5, the diversity raises by different levels if we use the view information. So the parameter μ is crucial for diversity.

When utilizing the semantic and view information, under the case that $\beta \neq 0$ and $\mu \neq 0$, the MAP@20 and NDCG@20 achieve the highest performances. This is likely to be caused by the following three aspects: a) The semantic information extracted from the user annotated tags isthe key information in the tag based image retrieval, which contributes to the relevance performance of the image retrieval results. b) The user marked view information can be viewed as high level semantic information which is an assistant factor in image retrieval. When adding the view information, the relevance performance is enhanced. But with the μ grows, the relevance performance ceases to increase. By combining both the view information and semantic information, they reinforce each other in the performance gain. c) The relevance and diversity are two important criteria in image retrieval, with the increase of the relevance performance, the diversity performance may decrease. But the relevance performance is the fundamental criteria of image retrieval. Thus, in the image retrieval, we should consider the relevance performance at the first hand, and then the diversity performance.

B. Discussions About α

In this part, the impact of the regularization parameter α (can be found in Eq. (10)) with fixed $\beta = 5$ and $\mu = 1$ on the performance of TDR is discussed. Fig. 15 shows the MNDCG@20, MAP@20, MDS@20 and MADP@20 of TDR with α in the range[0.1, 0.9].

We find that all of the MAP@20, DS@20,MADP@20 are the biggest when $\alpha = 0.2$. With the α grows, MAP@20, DS@20, MADP@20 reach the biggest at $\alpha = 0.2$, then falls to be stable. With the α grows, MNDCG@20 reaches the biggest at $\alpha = 0.3$ and becomes stable.

C. Discussions About Semantic Metric Between Query and Community

In the Eq. (8), we use average cosine distance to compute the relevance score between query and community. In this part, we make a comparison on three semantic relevance computing methods. In Eq.(8), the semantic relevance between the image and the query is determined by the mean cosine distance between the query and all tags in the community. Our first metric is based on the Google distance:

$$Sq_i = \frac{1}{Z_i}\sum_{l=1}^{Z_i}\exp(-GD(t_{il}, q)) \tag{24}$$

where GD is the google distance defined by Eq. (15). The second metric is based on Gaussian kernel:

$$Sq_i = \frac{1}{Z_i}\sum_{l=1}^{Z_i}\exp(-\frac{||fw_{t_{il}} - fw_q||^2}{2\sigma^2}) \tag{25}$$

where fw is word vector of tag and σ is the Gaussian parameter. Here we set $\sigma = 0.5$, The performance on different semantic relevance computing method is illustrated in Fig. 16.

We find that cosine, Gaussian, and Google distance based similarity measurement approaches are with little variance. Their performances under MDCG, MAP, MADP and MDS are very close.

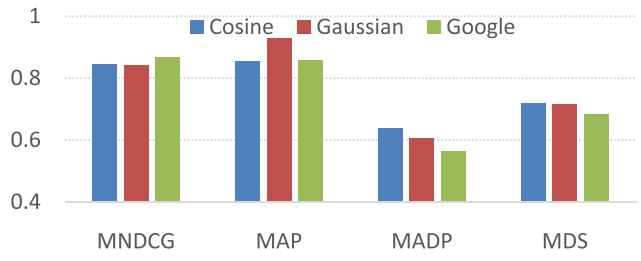


Fig. 16. Impact of different semantic metric of community and query for final results.

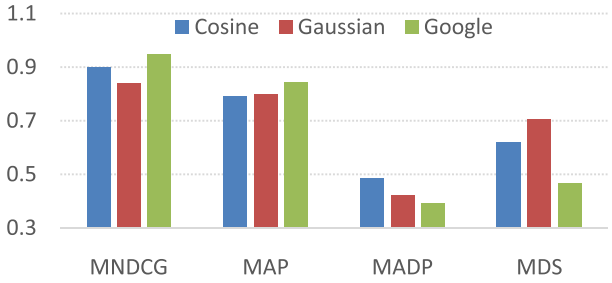


Fig. 17. Impact of different semantic metric between image and query for final result.

D. Discussions About Semantic Metric Between Query and Image

In the Eq.(14) , we introduce the semantic relevance matrix *Sc* to evaluate the relevance score of each image in each community. In this part, we make a comparison on three semantic relevance computing methods. In Eq.(14), the semantic relevance between the image and the query is the mean google distance between all tags of this image and the query. The two metrics are both based on word vector: cosine similarity defined by Eq.(7) and Gaussian kernel defined by Eq.(25). The performance on different semantic relevance computing method is illustrated in Fig. 17.

From the Fig. 17, we find that the google distance is better than the other two metrics based on word vectors in the

TABLE I
RESULT OF KMEANS AND AP ON NUS-WIDE

	MAP	MNDCG	MDS	MADP
AP	0.672	0.674	0.700	0.541
num=20	0.592	0.608	0.719	0.447
num=50	0.529	0.536	0.696	0.443
num=100	0.623	0.625	0.704	0.512
num=150	0.625	0.637	0.717	0.509
num=200	0.631	0.628	0.711	0.509
num=250	0.601	0.592	0.702	0.505
num=500	0.584	0.598	0.704	0.462

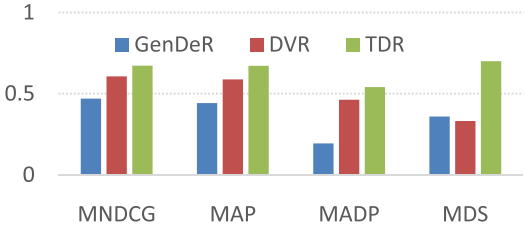


Fig. 18. Performance of three method under depth 20.

Eq. (14) and in the MAP, NDCG performance. The google distance aims to calculate the co-occurred probability between two tags. The co-occurred relationship in Google distance is introduced to judge whether the images are relevant. However, the word vectors are trained on English Wikipedia corpus, which represents semantic distance between two tags. Two metrics Cosine and Gaussian based on word vector are more diversiform. The Google distance is more outstanding under relevant metric, while the Cosine and Gaussian are with better diversity.

E. Discussion of Different Method for Community Detection

In this part, we discuss the impact of Kmeans and AP cluster for our algorithm. We conduct experiment on NUS-Wide using Kmeans and AP respectively.

TABLE I shows the results,where num denotes the Kmeans algorithm with "num" clusters (i.e. num $= 50, 100, 150, 200, 250, 500$). From TABLE I, we can see that the clustering approach AP outperforms Kmeans under MAP, MNDCG and MADP, while for MDS, AP is lightly worse than Kmeans. We can also see that the cluster number of Kmeans is important for the performance. When the cluster number is 50, the performance is much worse than the one with cluster number 100, 150, 200; the performance of 250 and 500 is also unsatisfied. It's very hard to choose a universal number for all query. Since different tags have different number of communities, by AP clustering, we can adaptively determine the cluster number. This also makes the approach extendable.

F. Discussion About the Diversity Method

In our paper, we achieve diversity by clustering. [59] and [60], introduce a diversity term in mathematical model, whose function is to penalize the similarity between images, to diverse results. They select one image at a time by proposed greedy algorithm. We denote "GenDeR" and "DVR" the algorithm proposed by [59] and [60] respectively, in this part, we compare our TDR with these two methods.

Fig. 18 shows the performance of three methods under depth $= 20$. From Fig. 18, we can find that the accuracy and diversity of TDR are both better than GenDeR and DVR. For example, MNDCG of GenDeR is 0.470 and DVR's is 0.607,

TABLE II
NOTATIONS AND DEFINITIONS

The symbol	Meaning
q	The query
o	The total number of tags in our crawled Flickr dataset
$\Gamma = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_o\}$	The tag set in our crawled Flickr dataset
$ID = \{ID_1, ID_2, \ldots, ID_o\}$	The inverted index files in our crawled Flickr dataset Each member ID_{I} Represents the image set which all tag with Γ_{I}
X	The image set which all tag with q
$V = \{v_1, v_2, \ldots, v_N\}$	The tag set in Image dataset X
N	The tag number in X
$FW = \{fw_1, fw_2, \ldots, fw_N\}$	The word vector of each tag in tag set V
G={V, E}	The constructed tag graph based on the tag set V
e_{ij}	The edge which connect v_i And v_j
$\boldsymbol{C} = \{\mathrm{c}_{11}, \mathrm{c}_{12}, \ldots, \mathrm{c}_{\mathrm{NN}}\}$	The tag similarity matrix, c_{ij} is the weight of e_{ij}
$r(i,k)$	The "responsibility" in AP CLUSTERING
$a(i,k)$	The "availability" in AP CLUSTERING
$S = \{s_1, s_2, \ldots, s_m\}$	The detected communities in image dataset X
m	The number of the detected communities in image dataset X
$t_i = \{t_{i1}, t_{i2}, \ldots, t_{iZ_i}\}$	The tag member in community s_i
$ST_i = (st_{i1}, st_{i2}, \ldots, st_{iN})$	The tag vector of community s_i; the mathematical form of t_i
Y	The image number of whole image dataset X
A	The random image in X
h_{i}	The overlap ratio between the image A and t_i
$IT = (it_1, it_2, \ldots, it_N)$	The tag vector of image A in the mathematical form
$\mathrm{X}_i = \{x_{i1}, x_{i2}, \ldots, x_{il_i}\}$	The image set in community s_i
l_i	The image number in community s_i
λ	Confidence factor matrix to control the propagation direction of random walk model
Sq_i	The semantic relevance between the tag set t_i Of community $\mathrm{S_I}$ And the query q
his_i	The tag histogram of image set X_i
SS_{ij}	The cosine similarity of his_i And his_j
p_{ij}	The normalized SS_{ij}
$rs = (\mathrm{rs}_1, \mathrm{rs}_2, \ldots, \mathrm{rs}_{\mathrm{m}})$	$\mathrm{Rs_i}$ is the relevance value of each community $\mathrm{s_i}$ with the query q
α	The propagation factor in the adaptive random walk model
$rm = (rm_1, rm_2, \ldots, rm_{l_h})$	rm_i is the relevance value between each image x_{hi} and the query q
Sc_i	The semantic relevance score of image x_{Hi} With respect to the query q
w_{ij}	The visual distance of image $x_{\mathrm{hi}}$and x_{hj},
vt_i	The normalization views of image x_{hi}
β, μ	Two positive parameters in the intra-community ranking
x_h^*	The representative image of the community S_h
GD_i	The semantic similarity between each tag v_i and the query q

while TDR can reach 0.672. The MDS of GenDeR and DVR is 0.358 and 0.332 respectively, while TDR is 0.700, which is much higher than the other two approaches.

VII. Conclusion and Future Work

In this paper, we propose a topic diverse re-ranking method for tag-based image retrieval. In this topic diverse re-ranking method, inter-community ranking and intra-community ranking are carried out to get satisfactory retrieved results. Tag graph construction and community detection are two effective ways to enhance the diversity. Besides, each tag's word vector is trained by using the Word2vec model based on the English Wikipedia corpus to enhance the relevance performance of the retrieved results.

However, we consider the community similarity in the inter-community ranking process while the topic similarity of representative images is ignored. In addition, much information in social media image set, such as Flickr dataset are still unutilized, such as title, time stamp and so on. For future work, we will investigate the similarity among representative images. Besides, we may fuse these relationships to enhance the diversity performance of image ranking system.

Appendix A

See Table II.

Appendix B
Steps for Solving Eq.(17)

The steps for solving Eq.(17) are as follows:

1) Compute the matrix $\boldsymbol{D} = \mathrm{Diag}(D_{11}, D_{22}, \ldots, D_{nn})$, semantic relevance scores $\mathrm{Sc} = [Sc_1, Sc_2, \ldots, Sc_n]$ and view times $\boldsymbol{VT} = [vt_1, vt_2, \ldots, vt_n]$. Initialize the $rm(0) = [\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}]$.

2) Compute the $rm\,(\mathrm{t}+1) = \frac{1}{1+\beta+\mu} D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \mathrm{r}\,(\mathrm{t}) + \frac{\beta * \mathrm{Sc} + \mu * \mathrm{VT}}{1+\beta+\mu}$ iteratively, until convergence, and the optimization value of $\boldsymbol{rm}$ can be obtained.

From above, we can obtain the optimization relevance score $\boldsymbol{rm}^*$ for every community $s_h, h \in (1, 2, \ldots, \mathrm{m})$. Then we select the image of the highest one among X_h, as the representative image of the community S_h, which denoted by x_h^*. Finally, we re-rank the image set $\{x_1^*, x_2^*, \ldots, x_m^*\}$ by the order of their communities obtained in the inter-community ranking process and get our final ranked image list.

References

[1] D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang, "Tag ranking," in *Proc. WWW*, 2009, pp. 351–360.

[2] X. Qian *et al.*, "Image location inference by multisaliency enhancement," *IEEE Trans. Multimedia*, vol. 19, no. 4, pp. 813–821, Apr. 2017.

[3] D. Liu, X.-S. Hua, M. Wang, and H. Zhang, "Boost search relevance for tag-based social image retrieval," in *Proc. ICME*, 2009, pp. 1636–1639.

[4] X. Lu, X. Zheng, and X. Li, "Latent semantic minimal hashing for image retrieval," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 355–368, Jan. 2017.

[5] M. Wang, K. Yang, X.-S. Hua, and H.-J. Zhang, "Towards a relevant and diverse search of social images," *IEEE Trans. Multimedia*, vol. 12, no. 8, pp. 829–842, Dec. 2010.

[6] A. Ksibi, A. Ben Ammar, and C. Ben Amar, "Adaptive diversification for tag-based social image retrieval," *Int. J. Multimedia Inf. Retr.*, vol. 3, no. 1, pp. 29–39, 2014.

[7] Y. Gao, M. Wang, H. Luan, J. Shen, S. Yan, and D. Tao, "Tag-based social image search with visual-text joint hypergraph learning," in *Proc. 19th ACM Int. Conf. Multimedia*, 2011, pp. 1517–1520.

[8] X. Li, B. Zhao, and X. Lu, "A general framework for edited video and raw video summarization," *IEEE Trans. Image Process.*, to be published, doi: 10.1109/TIP.2017.2695887, Apr. 19, 2017.

[9] K. Song, Y. Tian, T. Huang, and W. Gao, "Diversifying the image retrieval results," in *Proc. ACM Multimedia Conf.*, 2006, pp. 707–710.

[10] R. H. van Leuken, L. Garcia, X. Olivares, and R. van Zwol, "Visual diversification of image search results," in *Proc. WWW*, 2009, pp. 341–350.

[11] R. L. Cilibrasi and P. M. B. Vitanyi, "The Google similarity distance," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 3, pp. 370–383, Mar. 2007.

[12] X. Qian *et al.*, "HWVP: Hierarchical wavelet packet descriptors and their applications in scene categorization and semantic concept retrieval," *Multimedia Tools Appl.*, vol. 69, no. 3, pp. 897–920, Apr. 2014.

[13] X. Lu, Y. Yuan, and X. Zheng, "Joint dictionary learning for multispectral change detection," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 884–897, Apr. 2017.

[14] J. Carbonell and J. Goldstein, "The use of MMR, diversity-based reranking for reordering documents and producing summaries," in *Proc. SIGIR*, 1998, pp. 335–336.

[15] D. Wu, J. Wu, M.-Y. Lu, and C.-L. Wang, "A two-step similarity ranking scheme for image retrieval," in *Proc. 6th Int. Symp. Parallel Archit., Algorithms Programm.*, 2014, pp. 191–196.

[16] G. Ding, Y. Guo, J. Zhou, and Y. Gao, "Large-scale cross-modality search via collective matrix factorization hashing," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5427–5440, Nov. 2016.

[17] G. Agrawal, R. Chaudhary, and P. K. Singh, "Relevancy tag ranking," in *Proc. Int. Conf. Comput. Commun. Technol.*, 2011, pp. 169–173.

[18] L. Chen, S. Zhu, Z. Li, and J. Hu, "Image retrieval via improved relevance ranking," in *Proc. 33rd Chin. Control Conf.*, 2014, pp. 4620–4625.

[19] L. Wu, R. Jin, and A. K. Jain, "Tag completion for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 716–727, Mar. 2013.

[20] Y. Yang, Y. Gao, H. Zhang, J. Shao, and T.-S. Chua, "Image tagging with social assistance," in *Proc. ICMR*, 2014, p. 81.

[21] L. Chen, D. Xua, I. W. Tsang, and J. Luo, "Tag-based image retrieval improved by augmented features and group-based refinement," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1057–1067, Aug. 2012.

[22] Z. Lin, G. Ding, J. Han, and J. Wang, "Cross-view retrieval via probability-based semantics-preserving hashing," *IEEE Trans. Cybern.*, vol. 10, no. 99, pp. 1963–1974, Sep. 2016.

[23] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong, "Diversifying search results," in *Proc. WSDM*, 2009, pp. 5–14.

[24] X. Li, "Tag relevance fusion for social image retrieval," *Multimedia Syst.*, vol. 23, no. 1, pp. 29–40, Feb. 2017.

[25] X. Qian, X. Liu, C. Zheng, and X. Hou, "Tagging photos using users' vocabularies," *Neurocomputing*, vol. 111, pp. 144–153, Jul. 2013.

[26] D. Mishra, U. P. Singh, and V. Richhariya, "Tag relevance for social image retrieval in accordance with neighbor voting algorithm," *Int. J. Comput. Sci. Netw. Secur.*, vol. 14, no. 7, pp. 50–57, 2014.

[27] Y. Hu, M. Li, and N. Yu, "Multiple-instance ranking: Learning to rank images for image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[28] F. Sun, M. Wang, D. Wang, and X. Wang, "Optimizing social image search with multiple criteria: Relevance, diversity, and typicality," *Neurocomputing*, vol. 95, pp. 40–47, Oct. 2012.

[29] B. Wang, Z. Li, M. Li, and W.-Y. Ma, "Large-scale duplicate detection for Web image search," in *Proc. ICME*, 2006, pp. 353–356.

[30] R. L. T. Santos, C. Macdonald, and I. Ounis, "Exploiting query reformulations for Web search result diversification," in *Proc. WWW*, 2010, pp. 881–890.

[31] A. Ksibi, G. Feki, A. Ben Ammar, and C. Ben Amar, "Effective diversification for ambiguous queries in social image retrieval," in *Computer Analysis of Images and Patterns*. Berlin, Germany: Springer, 2013, pp. 571–578.

[32] Y. Guo, G. Ding, L. Liu, J. Han, and L. Shao, "Learning to hash with optimized anchor embedding for scalable retrieval," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1344–1354, Mar. 2017.

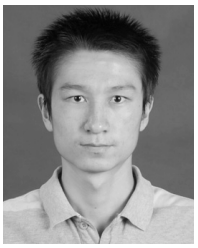
[33] C. Haruechaiyasak and C. Damrongrat, "Improving social tag-based image retrieval with CBIR technique," in *Proc. Role Digit. Libraries Time Global Change, 12th Int. Conf. Asia–Pacific Digit. Libraries* , 2010, pp. 212–215.

[34] X. Zhu, W. Nejdl, and M. Georgescu, "An adaptive teleportation random walk model for learning social tag relevance," in *Proc. ACM SIGIR*, 2014, pp. 223–232.

[35] J. Yu, D. Tao, M. Wang, and Y. Rui, "Learning to rank using user clicks and visual features for image retrieval," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 767–779, Apr. 2015.
[36] S. Ji *et al.*, "Global ranking by exploiting user clicks," in *Proc. ACM SIGIR*, 2009, pp. 35–42.
[37] G. Dupret and C. Liao, "A model to estimate intrinsic document relevance from the clickthrough logs of a Web search engine," in *Proc. ACM Int. Conf. Web Search Data Mining*, 2010, pp. 181–190.
[38] X. Lu, X. Li, and L. Mou, "Semi-supervised multitask learning for scene recognition," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1967–1976, Sep. 2015.
[39] X.-S. Hua, M. Ye, and J. Li, "Mining knowledge from clicks: MSR-bing image retrieval challenge," in *Proc. Multimedia Expo Workshops*, 2014, pp. 1–4.
[40] X. Lu and X. Li, "Multiresolution imaging," *IEEE Trans. Cybern.*, vol. 44, no. 1, pp. 149–160, Jan. 2014.
[41] *Word2Vec Source Code*. [Online]. Available: https://code.google.com/p/word2vec/
[42] X. Qian, X.-S. Hua, Y. Y. Tang, and T. Mei, "Social image tagging with diverse semantics," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2493–2508, Dec. 2014.
[43] X. Qian, D. Lu, and X. Liu, "Image retrieval by user-oriented ranking," in *Proc. ICMR*, 2015, pp. 511–514.
[44] Y. Zhang, X. Qian, X. Tan, J. Han, and Y. Tang, "Sketch-based image retrieval by salient contour reinforcement," *IEEE Trans. Multimedia*, vol. 18, no. 8, pp. 1604–1615, Aug. 2016.
[45] Y. Gu, X. Qian, Q. Li, M. Wang, R. Hong, and Q. Tian, "Image annotation by latent community detection and multikernel learning," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3450–3463, Nov. 2015.
[46] X. Yang, X. Qian, and Y. Xue, "Scalable mobile image retrieval by exploring contextual saliency," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1709–1721, Jun. 2015.
[47] D. Lu, X. Liu, and X. Qian, "Tag-based image search by social re-ranking," *IEEE Trans. Multimedia*, vol. 18, no. 8, pp. 1628–1639, Aug. 2016.
[48] X. Qian, Y. Xue, X. Yang, Y. Y. Tang, X. Hou, and T. Mei, "Landmark summarization with diverse viewpoints," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1857–1869, Nov. 2015.
[49] R. L. T. Santos, C. Macdonald, and I. Ounis, "Selectively diversifying Web search results," in *Proc. ACM CIKM*, 2010, pp. 1179–1188.
[50] G.-J. Qi, C. C. Aggarwal, J. Han, and T. Huang, "Mining collective intelligence in diverse groups," in *Proc. WWW*, 2013, pp. 1041–1052.
[51] X. Qian, X. Tan, Y. Zhang, R. Hong, and M. Wang, "Enhancing sketch-based image retrieval by re-ranking and relevance feedback," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 195–208, Jan. 2016.
[52] *English Wiki Training Corpus*. [Online]. Available: https://dumps.wikimedia.org/enwiki/latest/enwiki-latest-pages-articles.xml.bz2
[53] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.
[54] Y. Yan, G. Liu, S. Wang, J. Zhang, and K. Zheng, "Graph-based clustering and ranking for diversified image search," *Multimedia Syst.*, vol. 23, no. 1, pp. 41–52, 2014.
[55] X. Tian, L. Yang, Y. Lu, Q. Tian, and D. Tao, "Image search reranking with hierarchical topic awareness," *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2177–2189, Oct. 2015.
[56] D.-T. Dang-Nguyen, L. Piras, G. Giacinto, G. Boato, and F. G. B. De Natale, "Retrieval of diversity images by pre-filtering and hierarchical clustering," in *Proc. MediaEval*, 2014, pp. 1–2.
[57] H.-M. Hou, X.-S. Xu, G. Wang, and X.-L. Wang, "Joint-rerank: A novel method for image search reranking," *Multimedia Tools Appl.*, vol. 74, no. 4, pp. 1423–1442, 2015.
[58] S. Liu, P. Cui, H. Luan, W. Zhu, S. Yang, and Q. Tian, "Social visual image reranking for Web image search," in *Proc. MMM*, 2013, pp. 239–249.
[59] J. He, H. Tong, Q. Mei, and B. K. Szymanski, "*GenDeR*: A generic diversified ranking algorithm," in *Proc. Adv. Neural Inf. Process Syst.*, vol. 2. 2012, pp. 1142–1150.
[60] H. Tong, J. He, Z. Wen, R. Konuru, and C.-Y. Lin, "Diversified ranking on large graphs: An optimization viewpoint," in *Proc. SIGKDD*, 2011, pp. 1028–1036.
[61] X. Li, S. Liao, W. Lan, X. Du, and G. Yang, "Zero-shot image tagging by hierarchical semantic embedding," in *Proc. ACM SIGIR*, 2015, pp. 879–882.
[62] D. Zhang, J. Han, C. Li, J. Wang, and X. Li, "Detection of co-salient objects by looking deep and wide," *Int. J. Comput. Vis.*, vol. 120, no. 2, pp. 215–232, 2016.
[63] D. Zhang, J. Han, J. Han, and L. Shao, "Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1163–1176, Jun. 2016.
[64] S. Lee, W. De Neve, and Y. M. Ro, "Visually weighted neighbor voting for image tag relevance learning," *Multimedia Tools Appl.*, vol. 72, no. 2, pp. 1363–1386, 2013.

Xueming Qian (M'10) received the B.S. and M.S. degrees from the Xi'an University of Technology, Xi'an, China, in 1999 and 2004, respectively, and the Ph.D. degree from the School of Electronics and Information Engineering, Xi'an Jiaotong University, in 2008. He was a Visiting Scholar with Microsoft Research Asia from 2010 to 2011. He was an Assistant Professor with Xi'an Jiaotong University, where he was an Associate Professor from 2011 to 2014, and is currently a Full Professor. He is also the Director of the Smiles Laboratory, Xi'an Jiaotong University. His current research interests include social media big data mining and search. His research is supported by the National Natural Science Foundation of China, Microsoft Research, and the Ministry of Science and Technology. He received the Microsoft Fellowship in 2006. He received the Outstanding Doctoral Dissertations of Xi'an Jiaotong University and Shaanxi Province, in 2010 and 2011, respectively.

Dan Lu received the B.S. degree from Chang'an University, Xi'an, China, in 2013, and the M.S. degree from the School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, China, in 2016.



Yaxiong Wang received the B.S. degree from Lanzhou University, Lanzhou, China, in 2015. He is currently pursuing the Ph.D. degree with the School of software, Xi'an Jiaotong University, Xi'an China. He is currently the Post-Graduate Researcher with the SMILES Laboratory, Xi'an Jiaotong University. His current research interests include tag-based image retrieval.

Li Zhu received the B.S. degree from Northwestern Polytechnic University in 1989, and the M.S. and Ph.D. degrees from Xi'an Jiaotong University in 1995 and 2000, respectively. He is currently an Associate Professor with the School of Software, Xi'an Jiaotong University. His main research interests include multimedia processing and communication, parallel computing, and networking.

Yuan Yan Tang (F'04) received the B.E. degree in electrical and computer engineering from Chongqing University, Chongqing, China, the M.Eng. degree in electric engineering from the Beijing Institute of Post and Telecommunications, Beijing, China, and the Ph.D. degree in computer science from Concordia University, Montreal, QC, Canada. He is currently a Chair Professor with the Faculty of Science and Technology, University of Macau, Macau, China, and also a Professor, an Adjunct Professor, and an Honorary Professor with several institutes, including Chongqing University, Consordia University, and Hong Kong, Baptist University, Hong Kong, China. His current research interests include wavelet theory and applications, pattern recognition, image processing, document processing, artificial intelligence, and Chinese computing. He is a fellow of the International Associate of Pattern Recognition.

Meng Wang (M'09) received the B.E. degree and Ph.D. degree in the Special Class for the Gifted Young from the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China, in 2003 and 2008, respectively. He is currently a Professor with the Hefei University of Technology, China. His current research interests include multimedia content analysis, computer vision, and pattern recognition. He has authored over 200 book chapters, journal and conference papers in these areas. He was a recipient of the ACM SIGMM Rising Star Award 2014. He is an Associate Editor of the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.