Non-user Generated Annotation on User Shared Images for Connection Discovery

Ming Cheung*, James She†, Xiaopeng Li‡

*†‡HKUST-NIE Social Media Lab, Hong Kong University of Science & Technology, Hong Kong
*cpming@ust.hk, †eejames@ust.hk, ‡xlibo@connect.ust.hk

***Abstract*—Social graphs, representing the online friendships among users, are one of the most fundamental types of data for many social media applications, such as recommendation, virality prediction and marketing. However, this data may be unavailable due to the privacy concerns of users, or kept privately by social network operators, which makes such applications difficult. One of the possible solutions to discover user connections is to use shared content, especially images on online social networks, such as Flickr and Instagram. This paper investigates how non-user generated labels annotated on shared images can be used for connection discovery with different color-based and feature-based methods. The label distribution is computed to represent users, and followee/follower relationships are recommended based on the distribution similarity. These methods are evaluated with over 200k images from Flickr and it is proven that with non-user generated labels, user connections can be discovered, regardless of the method used. Feature-based methods are also proven to be 95% better than color-based methods, and 65% better than tag-based methods.**

***Index Terms*—recommendation; online social network; annotation; big data; connection discovery.**

I. Introduction

Social graphs, representing the online friendships among users, are fundamental data for many social media applications, such as recommendation, virality prediction and marketing. However, this data may be unavailable due to the privacy concerns of users, or kept privately by social network operators and these applications become challenging with an incomplete set of data. Providing a solution to this problem, user connections are also reflected in the abundance of social content, especially images, shared on social networks. Two users with a connection will share characteristics, be follower/followee, or be member of the same community. Through discovering user connections from their shared images the connections can be used to identify user gender, or recommend followers/followees to users [1]. One of the possible ways to do this is using the user annotation tags on their shared images. Fig. 1(a) to (d) is a set of images and their tags from different users on Instagram. The tags can include the name of the object, location, time and even the feelings of the user. Such annotations are the major types of tags [2].

One problem is that tags can be inaccurate, unsuitable for analysis or even unavailable [3][4], which may affect the accuracy of the connection discovery. For example, some users are not interested in annotating their shared images because of the considerably longer time it takes, as shown for the user in Fig. 1 (b). For annotated images, the tags may not be a good description of the content. Users may type the tag wrongly, such as the tag "Sar" in Fig. 1 (c). Sometimes irrelevant tags are added intentionally to increase popularity of a shared image, such as the tag "Car" in the picture of flowers in Fig. 1 (d). As a result, the connections discovered from the tags shown in Fig. 1 (e) have wrongly connected user D with user A, while leaving users B and C without any connections. These are some common examples of how users annotate their shared images on social networks, and how this affects the accuracy of connection discovery.

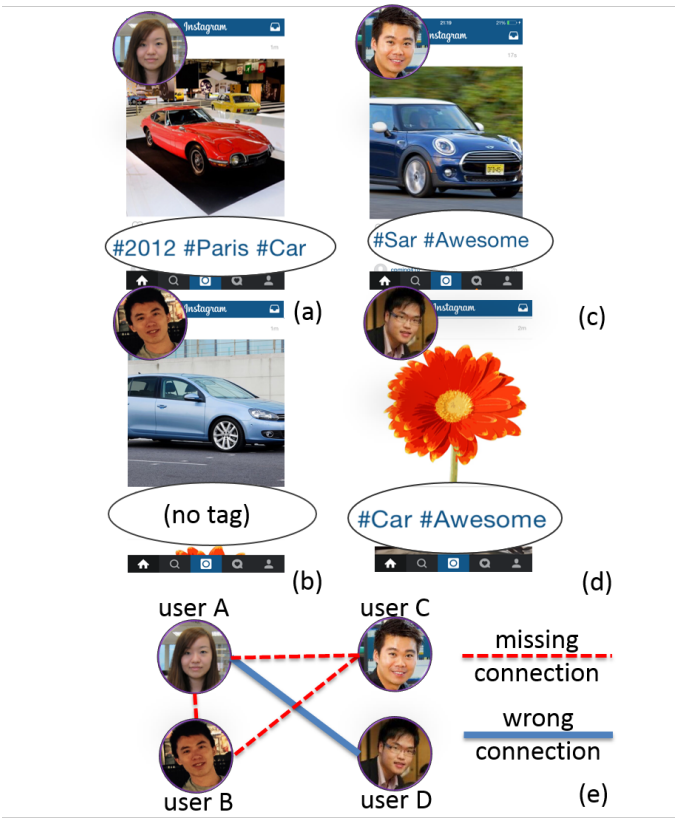

Fig. 1: Examples of bad tags: (a) diverse, (b) missing, (c) mistyped, (d) wrong, (e) and the discovered connections.

Another solution is to annotate user shared images with unbiased non-user generated labels from non-user generated

annotation, and represent each user with their label distribution by Bag-of-features Tagging (BoFT) [4]. As user A, B and C in Fig. 1 share images of cars, the visual similarity of those images is high, and connection can be discovered accordingly. Two users having very similar distributions not only indicates that they are likely to have a follower/followee relationship, but also suggests that they share some characteristics, such as gender [1].

This paper evaluates non-user generated annotation to discover user connections for follower/followee recommendation. Instead of using scale-invariant feature transform (SIFT) as in [4], we examine the use of non-user generated labels with different color-based and feature-based methods. The approach is evaluated using a dataset of 542 users and 201006 images, as well as the actual relationship among users. The results prove the effectiveness of non-user generated annotation.

The main contributions of this paper are the following: 1) we evaluate a novel approach to non-user generated annotation, with the actual relationships of the scraped data with over 200k images; 2) we prove that non-user generated annotation can discover connections for recommendation, regardless of the visual method used to represent images; and 3) we confirm that the feature-based approach is 95% and 65% better than the color-based and tag-based methods, repectively. To the best of our knowledge, this is the first paper to prove that non-user generated annotation is not limited by the method used and that feature-based approaches are better for connection discovery.

This paper are organized as follows. Section II discusses previous works. Section III introduces non-user generated annotation, followed by how the connection discovery works in Section IV. Section V shows the experimental results and Section VI concludes the paper.

II. Previous Works

As discussed above, one of the possible ways to understand images for connection discovery is through user annotated tags, a common feature in social networks that are a set of text on social content. However, those tags can be unreliable, and the performance of the discovered connections can be affected [4]. Another possible way of discovering connections is a content-based approach, in which the visual appearance of an image is considered to generate a label for the image [5][6]. However, determining the relationship between the appearance and the label is not a trivial task because the same object can be visually different among images. To overcome these drawbacks a Bag-of-Features-based approach, BoF Tagging (BoFT) is proposed in [4]. This approach makes use of computer vision techniques in object recognition tasks to recommend follower/followee relationship, and is proven to be better than approaches using user annotated tags. Non-user generated labels are generated by clustering the vectors generated by the BoF, and the similarity of two users can be calculated by the distribution of the labels. The connections can be obtained from the similarity. Two users with a higher similarity are more likely to be follower/followee [1]. Identifying gender and recommending followers/followees are possible with the discovered connections.

Unlike [1][4], this paper evaluates other feature-based and color-based methods to generate vectors for connection discovery, and the efficiency is proven by follower/followee relationship recommendation. Color is the first and one of the most widely used visual feature. Proposed in [7], color histogram has been effectively used for image indexing and retrieval. By quantizing the color space, e.g. RGB and HSV, the color histogram descriptor has a low dimension and low computational complexity. However, the color histograms have limitations when they are used to classify images with a similar color distribution but different content [8]. Different from above, the GIST descriptor, as proposed in [9], is meant to build a holistic and rough representation of the scene, . It is used to summarize gradient information for different parts of the image. The GIST descriptor has been shown very effective in scene categorization [9] and image search [10]. In contrast, SIFT based BoF approach focuses on local features of an image and is robust to occlusion and clutter [11].

III. Generating Non-user Generated Labels

This section first introduces how non-user generated annotation is conducted, followed by different computer vision methods that can be used to represent images as vectors for generating non-user generated labels.

A. Non-user Generated Annotation

In non-user generated annotation, each user shared image is represented by a non-user generated label to indicate the visual appearance of the image. An example can be found in Fig. 2. Given a set of images, the images are first encoded into a vector with R dimensions, as shown in step 1 of Fig. 2. This can be conducted by using different computer vision methods. The details can be found in the next subsection. Then the vectors are clustered by clustering techniques, such as k-means, as shown in step 2 of Fig. 2. Each cluster corresponds to a non-user generated label, as shown in step 3 of Fig. 2.

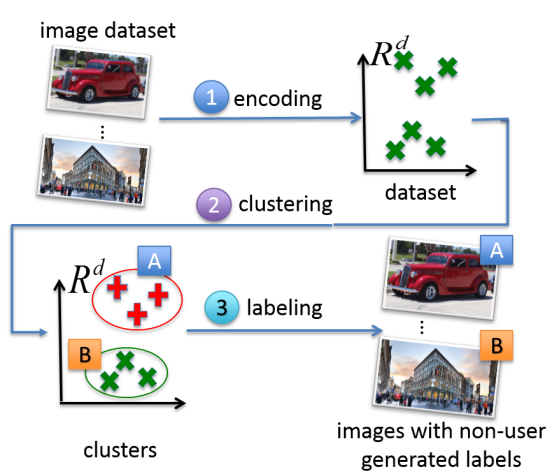

Fig. 2: Flow chart for non-user generated annotation.

B. SIFT

Lowe [11] proposed a scale invariant feature transform (SIFT), which combines a scale invariant region detector and a descriptor based on the gradient distribution in the detected regions. The descriptor is represented by a 3D histogram of gradient locations and orientations within a local region of the image. The contribution to the location and orientation bins is weighted by the gradient magnitude. SIFT descriptor has been shown to be invariant to image rotation and scale and robust across a substantial range of affine distortion, addition of noise, and change in illumination.

C. GIST

The GIST descriptor was initially proposed in [9] and it has shown good results for scene categorization [9] and image search [10]. The idea is to build a holistic and low dimensional representation of a scene, which does not require explicit segmentation of image regions and objects. The method filters each pixel with a sequence of (Gabor) filters, producing feature maps for the image. Each feature map is divided into blocks and the GIST descriptor is obtained by concatenating the averaged value of each block in all feature maps. The GIST descriptor summarizes the gradient information (scales and orientations) for different parts of an image, which provides a rough description of the scene.

D. Color Transformation

Color is one of the most widely used visual elements in image retrieval and indexing. Given a discrete color space defined by some color axes (e.g., red, green, blue), the color histogram is obtained by discretizing the image colors and counting the number of times each discrete color occurs in the image array [7]. The color histogram gives color statistic information of the image. It is scale-invariant to light intensity changes over several color spaces, e.g. HSV space [13].

Based on the vectors obtained in each method, non-user generated labels can be obtained by clustering those vectors. The details are introduced in the next section.

IV. Connection Discovery using Non-user Generated Labels

This section introduces how to find similarities among people from non-user generated labels. The first part introduces how to obtain the label distribution from those labels, while the second part discusses how to discover connections and make recommendations based on the distribution.

A. Non-user Generated Labels and Label Distribution

The label distribution, which reflects the visual appearance of their images, is the key. Two users with a similar distribution indicate that they are follower/followee [1]. Once the user generated labels are obtained with one of the methods, as in step 1 of Fig. 3, the distribution is calculated form the non-user generated labels on the images, as shown in step 2 of Fig. 3. In this work, the occurrences of each non-user generated label are used as the distribution.

B. Label Distribution and Connection Discovery

When the label distribution is obtained, the next step is to discover user connections from the similarity of their distributions. Applications can be based on those discovered connections, even without the access to social graphs. The similarity of two distributions can be calculated by the following formula:

$$S_{i,j} = S(L_i, L_j) = \frac{L_i \cdot L_j}{||L_i|| \cdot ||L_j||} \tag{1}$$

where L_i is the set of non-user generated labels in the images uploaded by user i and L_j is the set of non-user generated labels in the images uploaded by user j. It is proven that two users with higher similarity are more likely to be follower/followee, or share other characteristics, such as gender. The tie is assumed to be undirected, which means that it is the same from user A to B and user B to A. Different types of applications are possible with the discovered connections. In particular, the focus of this paper is evaluating the precision and recall with different methods for follower/followee recommendation.

V. Experimental Results

In this section, the dataset, the experiments and the results are discussed. Our work focuses on discovering the connections among users by the tendency of people to make friends with others who share similar interests. The results show that it is possible to infer connections by using BoF tagging.

A. Dataset and Experimental Setup

The setting of the experiment is shown in Fig. 4. A set of 201006 images uploaded by 542 users is scraped from Flickr,

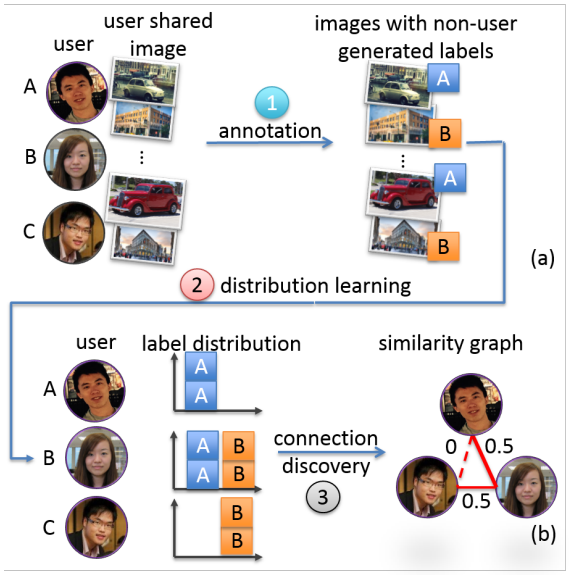

Fig. 3: Flow chart of connection discovery: (a) generating non-user generated labels, (b) connection discovery.

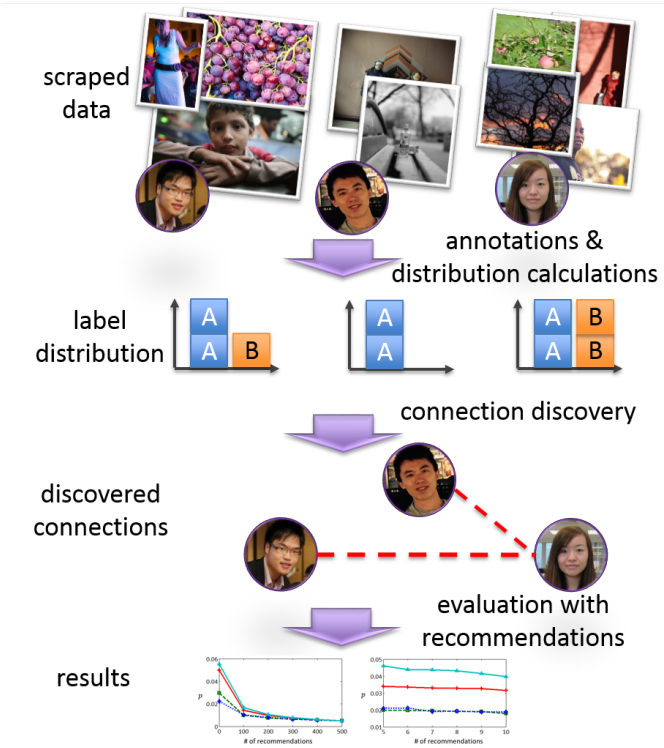

Fig. 4: Experimental settings

an online social network for image sharing with millions of images uploaded, and different methods are used to annotated those images. The 542 users are selected randomly from images under the same tag query page to provide diversity. The label distributions for different methods are obtained, and the similarities are calculated accordingly Then connections among users are inferred from the similarities. Tables I and II show the attributes scraped for the users and the images.

The average number of friends of a user is 170, for which

Fig. 5: Examples of images in clusters of methods: (a) RGB; (b) HSV; (c) SIFT; (d) GIST

TABLE I: MAJOR ATTRIBUTES FOR IMAGES

Attribute	Description
ImageID	the unique ID for the image
Tag	the set of user annotated tags

TABLE II: MAJOR ATTRIBUTES FOR USERS

Attribute	Description
UserID	the unique ID for the user
ImageUploaded	the set of images uploaded by the user
FriendList	the user ID of the user friends

there are 902 connections among the 542 users. The goal of the experiment is to infer those connections using the set of images uploaded by the users, even without access to the social graph.

The images are encoded into vectors with different approaches for comparison. The first two approaches are color-based methods, Color Histogram Descriptors, where both the HSV color space and RGB color space are used. The HSV color space is uniformly discretized as 16x4x4 and the RGB color space is discretized as 16x16x16. Examples of clusters can be found in Fig. 5 (a) and (b). The colors of the images are similar within the same cluster. The other two methods are feature-based, one based on SIFT and the other based on GIST. SIFT is based on BoF and in this method the SIFT descriptor is obtained for each image and the vocabulary is learned by clustering similar features. The image is then represented by the histogram of the image over the vocabulary, as in [4]. Two sample clusters are shown in Fig. 5 (c), in which similar objects are presented in the images of the same cluster. The last method is the GIST descriptor. It is configured with a common setting, namely the number of blocks is 4x4 and the number of orientations for all four scales is 8 [12]. An example can be found in Fig. 5 (d), in which images of the same cluster are similar sences, for example ocean. After the vectors for all the images are obtained, k-means clustering is used to group images with respect to each method. Each cluster obtained is assigned with the same non-user generated label to reflect that the images in the cluster are visually similar and belong to the same group. The label distribution of each method is then calculated by counting the occurrences of each label.

The results are evaluated by two popular rates, the top N recall rate and top N precision rate, as follows:

$$Precision = \frac{T_p}{(T_p + F_p)} \tag{2}$$

$$Recall = \frac{T_p}{(T_p + F_n)} \tag{3}$$

where T_p is the true positive (the recommended follower/followee relationship is an actual follower/followee relationship) and F_p and F_n are the false positive and false negative, respectively. F_p represents the case that the recom-

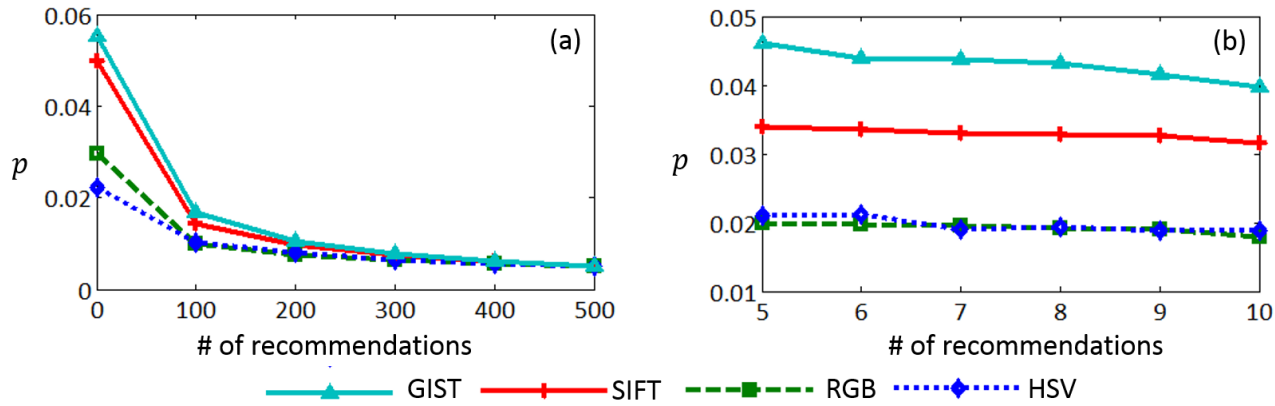

Fig. 6: Precision for: (a) $N \subset$[1,500], (b) $N \subset$[5,10].

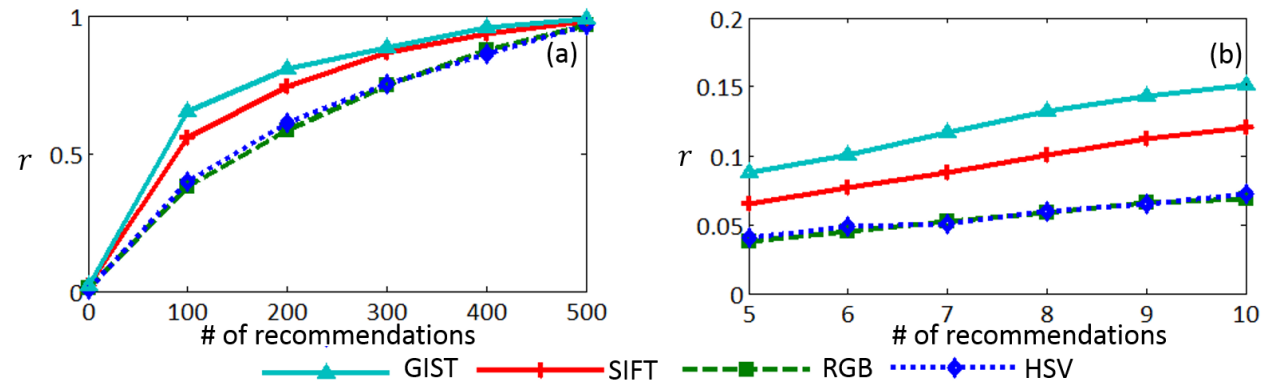

Fig. 7: Reall for: (a) $N \subset$[1,500], (b) $N \subset$[5,10].

mended follower/followee relationship is not an actual follower/followee relationship, while F_n is the follower/followee relationships that are not recommended. The physical meaning of the precision rate is the percentage of recommended items is actual follower/followee relationship. The recall is the percentage of actual follower/followee relationships that are recommended. The higher the values are, the better the method is. When more items are recommended, the recall rate is increased, but the precision rate decreases. A list of recommendations is generated for each user with the different methods. The methods are evaluated with the top N per user recall rate and the top N precision rate, in which the top N users with the highest similarity, are recommended.

B. Results

Fig. 6 and Fig. 7 show the top-N precision and recall rates of different algorithms. As shown in Fig. 6 (a) and Fig. 7 (a), GIST has a higher precision rate and recall rate than the other methods for $N \subset$[1,500]. Note that as there are only 542 users in the dataset, the recommended users of different methods are similar when N is close 542, and a similar value of p and r is obtained. In practical situations, a recommendation system can only recommend 5 to 10 items, and it is interesting to investigate the performance with 5 to 10 recommendations per user, as shown in Fig. 6 (b) and Fig. 7 (b). In terms of precision rate, GIST is 20% better than SIFT. And both GIST and SIFT are much better than RGB and HSV, with a higher rate of more than 100% and 70% respectively. It is also interesting to investigate how much feature-based methods are better than color-based methods and approaches using user annotated tags, UserT [4]. The area under curve (AUC) of each methods for $N \subset$[5,10] is obtained, and the averages of the feature-based and color-based methods are calculated, respectively. The AUC of UserT for $N \subset$[5,10] is also obtained. The results are shown in Fig. 8. It is demonstrated that in terms of precision rate and recall rate, the feature-based methods are 95% better than the color-based methods, and 65% better than UserT. Furthermore, color-based methods achieve only 85% of UserT. The results will be discussed in the next section.

C. Discussion

In the experiment, it is observed the color-based methods have worse performance than UserT by 15%. Although the performance of the color-based methods is worse than UserT, non-user generated labels from the color-based methods are more accessible because they do not have any language barriers and are easy to compute. The color-based methods can serve as a baseline when computational resources are limited, or user

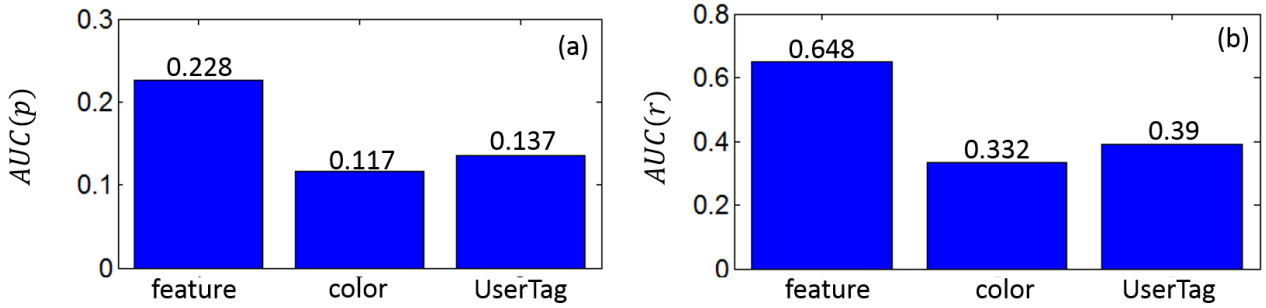

Fig. 8: AUC of feature-based and colour-based methods for $N \subset$[5,10]: (a) precision, $AUC(p)$, (b) recall, $AUC(r)$.

annotated tags are not available. The feature-based methods are also 95% better than the color-based methods. One of the reasons behind this is that the color-based methods do not include any spatial information but different objects with the same color could have the same histogram, such as a blue car and the ocean. In contrast, feature-based methods can capture spatial information. GIST gives a rough glimpse of the whole image by capturing the holistic information, while SIFT-based BoF focuses on local features. Therefore the GIST descriptor should be more suitable for scene categorization, especially in image-based social networks, such as Flickr. By the nature of Flickr, it is expected to contain more images of scenes, which makes the GIST descriptor more appropriate as it exploits more information from the global scene than from local objects. In other social networks where users share images of their daily lives, such as WeChat, SIFT could be more suitable to capture the nature in their shared images. Further investigation is needed on the selection of a good descriptor for a given a social network and combining color information into the feature-based methods. And a more extended experiment can be conducted to a larger social network data to prove the effectiveness of the proposed method.

VI. Conclusion

This paper investigates how non-user generated labels annotated on shared images can be used for connection discovery with different color-based and feature-based methods. As user social graphs are not always available, this paper proves that discovered connections from non-user generated labels is an alternative to social graphs. The label distributions represent the visual features in the shared images, and hence allow recommendation. The proposed approach is evaluated by followere/followee relationship recommendation with a scraped dataset from Flickr with over 500 users and 200k images. It is proven that followee/follower relationships can be recommended based on the label distribution, regardless of the methods used to generate the labels. By using feature-based methods to generate non-user generated labels, the recall and the precision rate are 95% better than the color-based methods.

Acknowledgment

This work is supported by HKUST-NIE Social Media Lab., HKUST

References

[1] M. Cheung, J. She., and Z. Jie., "Connection discovery using Big Data of user-shared images in social media," in IEEE Transactions on Multimedia, vol.17, no.9, pp.1417-1428, Sept. 2015

[2] T. Hammond, T. Hannay, B. Lund, and J. Scott, "Social bookmarking tools (I) a general review," D-Lib Magazine, vol. 2, 2005.

[3] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury, "How flickr helps us make sense of the world: Context and content in community-contributed media collections," in Proceedings of the 15th International Conference on Multimedia, September 2007, pp. 631-640.

[4] M. Cheung and J. She. "Bag-of-features tagging approach for a better recommendation with social Big Data." in Proceedings of the 5th Intel. Conference on Advances in Information Mining and Management, 2014, pp. 83-88.

[5] E. Moxley, J. Kleban, J. Xu, and B. Manjunath, "Not all tags are created equal: Learning flickr tag semantics for global annotation," in Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on, June 2009, pp. 1452-1455.

[6] X. Zhang et al., "Social image tagging using graph-based reinforcement on multi-type interrelated objects," Signal Processing, vol. 93 no. 8, pp. 2178-2189, August 2013.

[7] M. J. Swain and D. H. Ballard, "Color indexing," International Journal of Computer Vision, vol. 7, pp. 11-32, 1991.

[8] J. Poveda Figueroa and V. Robles Bykbaev, "Combination of HSV and RGB histograms and MPEG-7 Descriptor: Preliminary results and future work," 2012.

[9] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," International Journal of Computer Vision, vol. 42, pp. 145-175, 2001.

[10] M. Douze, H. Jgou, H. Sandhawalia, L. Amsaleg and C. Schmid, "Evaluation of gist descriptors for web-scale image search," in Proceedings of the ACM International Conference on Image and Video Retrieval, 2009, pp. 19.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, pp. 91-110, November 2004.

[12] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," Prog. Brain Res., vol. 155, pp. 23-36, 2006.

[13] V. D. Sande, E. A. Koen, T. Gevers and C. G. Snoek, "Evaluating color descriptors for object and scene recognition," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, pp. 1582-1596, 2010.