FLICKR CIRCLES: MINING SOCIALLY-AWARE AESTHETIC TENDENCY

Luming Zhang and Roger Zimmermann

School of Computing, National University of Singapore
{zhanglm, rogerz}@comp.nus.edu.sg

ABSTRACT

Aesthetic tendency discovery is a useful and interesting application in social media. This paper proposes to categorize large-scale Flickr users into multiple *circles*. Each circle contains users with similar aesthetic interests (*e.g.*, landscapes or abstract paintings). We notice that: 1) an aesthetic model should be flexible as different visual features may be used to describe different image sets, and 2) the numbers of photos from different users varies significantly and some users have very few photos. Therefore, a regularized topic model is proposed to quantify user's aesthetic interest as a distribution in the latent space. Then, a graph is built to describe the similarity of aesthetic interests among users. Obviously, densely connected users are with similar aesthetic interests. Thus an efficient dense subgraph mining algorithm is adopted to group users into different circles. Experiments show that our approach accurately detects circles on an image set crawled from over 60,000 Flickr users.

1. INTRODUCTION

Flickr is a popular photo/video hosting website. There are many *groups* in Flickr where users with same aesthetic interests can share photos and exchange opinions. There are many popular public groups (*e.g.*,"architecture" and "silhouette") with thousands of members and over one million shared photos. Users can freely join in/leave a public group or launch a new group. However, the grouping mechanism is less intelligent since the groups are constructed and maintained manually. In practice, we want a system that automatically categorizes Flickr users into different communities based on users' aesthetic tendency. However, building such a system is challenging due to two reasons:

- In many computational aesthetic models [1, 2], different visual features are employed to represent different image sets. For example, if an image set contains portraits, then the active shape model (ASM) [3] can be used to localize human faces. This requires the designed system to be highly extensible. Thereby different visual features can be integrated flexibly.
- The number of photos of different users varies significantly. Some users have uploaded over 50,000 photos while others have only less than 10 photos. This brings severe overfitting problem in the model training stage.

To solve the above problems, a regularized topic model is proposed to model users' aesthetic tendency. We first extract a set of visual features to describe each image at both low-level and high-level. Then, a topic model is designed to seamlessly integrate these low&high-level features and further represent user's aesthetic interest by a distribution of latent topics. To alleviate the overfitting caused by the photo scarcity of some users, a regularized term is added into the topic model. Using KL-divergence to measure the distribution between users, an affinity graph is constructed to describe the similarity of aesthetic interests among users. Users with similar aesthetic interests are densely distributed on the graph. They are categorized into different Flickr circles by a dense graph mining algorithm.

The main contributions of this paper are two fold. 1) We propose to discover circles from large-scale Flickr users in terms of their aesthetic tendency. 2) A regularized topic model is developed that describes a user's aesthetic interests as a mixture of latent topics.

2. RELATED WORK

2.1. Computational Aesthetic Models

There are many computational aesthetic models in multimedia [26, 27, 30] and computer vision [28, 29, 31]. Datta *et al.* [4] proposed 58 low-level visual features to capture photo aesthetics. Wong *et al.* [5] proposed three global aesthetic features: low-level features such as exposure extracted from the overall image and the salient regions, as well as the difference between low-level features extracted from subject and background regions. Luo *et al.* [1] proposed a hue distribution and a prominent line-based texture distribution to represent the photo global composition. Dhar *et al.* [6] proposed a set of high-level attributes to evaluate photo aesthetics. To capture the process of human viewing images, Zhang *et al.* [27] proposed to learn human gaze shifting pathes for evaluating photo aesthetics. Cheng *et al.* [8] proposed omni-range context, *i.e.*, spatial distributions of arbitrary pairwise image patches, to model photo aesthetics. Zhang *et al.* [9] introduced graphlets and designed a probabilistic model to transfer them from the training photos into the cropped one. Further, Zhang *et al.* [2] proposed to optimally fuze visual features from multiple channels to access photo aesthetics. It is worth emphasizing that these methods can only access the aesthetics of a single image. They cannot measure the aesthetic discrepancy of image sets belonging to different Flickr users.

2.2. Community Detection using Probabilistic Models

Probabilistic topic models such as latent Dirichlet allocation (LDA) [11] and its variants [12, 13] have been applied to detect social communities recently. Based on a probabilistic topic model, a social link graph can be considered as a generative process. The model categorizes users into different communities though a sampling process, given the distribution of communities in the graph, the distribution of users in communities, and the distribution of social links among users. Communities are detected based on the fact that users belonging to one community have similar link patterns in the graph. Some work [14, 15] applies probabilistic models to detect communities where each community is considered as a combination of semantic topics. Zhou *et al.* [14] proposed a model that extracts e-communities from email corpus. The model employs social interactions and topical similarity extracted from documents to search communities. A recent work by Yin *et al.* [15] constructs text-associated graphs. The model combines the generation of links between users and words of users to extract communities, where both the link structures and the users' semantic topics are exploited.

3. THE PROPOSED METHOD

3.1. Low-&High-level Visual Descriptors

Following [10], we use 9-D color moment [19] and 128-D HOG [18] to describe a photo at low-level. To capture the high-level visual cues, we use the weakly supervised learning [10] to describe each image region by a 128-D semantic vector. Afterward, we use max pooling [7] to integrate region-level semantic vectors into an image-level one. In total, each photo is represented by a 9+128+128=265-D vector.

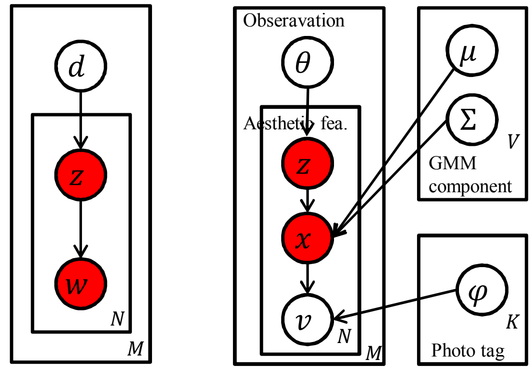
3.2. Regularized Latent Topic Model

The proposed regularized topic model is built upon the Gaussian mixture models (GMMs) and the latent semantic analysis (LSA) [20]. GMMs provide a richer class of density modeling than a single Gaussian distribution over the continuous variables. The latent variable v is the index of a Gaussian component that generates the observation $\mathbf{x}$ (*i.e.*, a Flickr photo). GMMs encode the distribution of observations $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ as:

$$p(\mathbf{X}|\varphi, \mu, \Sigma) = \prod_{j=1}^{N} \sum_{v=1}^{V} p(v|\varphi) p(\mathbf{x}_j|\mu_v, \Sigma_v), \quad (1)$$

where $p(v|\varphi)$ is the mixture coefficient, and $p(\mathbf{x}|\mu_v, \Sigma_v)$ is the multivariate Gaussian distribution.

The LSA [20] provides a probabilistic way to model top-



(a) LSA model **(b) Regularized topic model**

Fig. 1. The LSA model (left) and the proposed regularized topic model (right). The red nodes are observable variables.

ics over a document and a corpus. An illustration of LSA is shown in Fig. 1. Given the hidden variable z (*i.e.*, latent topic), each discrete word w (*i.e.*, aesthetic feature) is assumed to be independent of the document d containing it. The joint distribution of the observable variables is calculated as:

$$p(d_i, w_j) = p(d_i) \sum_{z=1}^{K} p(z|d_i) p(w_j|z), \quad (2)$$

where $p(z|d_i)$ is the weight reflecting the probability that topic z occurs in document d, and $p(w_j|z)$ is the probability that discrete word w occurs in topic z.

One disadvantage of LSA is that it only models discrete variables. Thus we cannot use it to describe the aesthetic distribution of Flickr users. Inspired by GMMs that can naturally describe continuous variables, we propose a Gaussian latent topic model. It replaces discrete words in LSA by continuous features modeled by a mixture of multivariate Gaussian distributions. As shown on the right of Fig. 1, the node $\mathbf{x}$ represents the 265-D aesthetic feature in photo d along with the photo tag z and its Gaussian component assignment v. For the entire data set, the joint distribution can be formulated as:

$$p(\mathbf{X}|\mathbf{\Phi}) = \prod_{d=1}^{M} \prod_{i=1}^{N} \sum_{z=1}^{K} \sum_{v=1}^{V} p(z^{(d,i)|\theta_d}) \\ p(v^{(d,i)}|\phi_z) p(\mathbf{x}^{(d,i)}|\mu_{v,z}, \Sigma_{v,z}), \quad (3)$$

where $p(z|\theta_d)$ and $p(v|\phi_z)$ are the multinomial distributions, and $p(\mathbf{x}|\mu, \Sigma)$ is the multivariate Gaussian distribution with a diagonal covariance matrix.

As we claimed, some Flickr users have very few photos, which results in overfitting during model training. To alleviate this, a regularized term is integrated in the parameter learning process. More specifically, the expectation maximization (EM) algorithm is applied to iteratively learn model parameters. The parameters of the regularized topic model are learned by maximizing the auxiliary function as:

$$
\begin{aligned}
&Q(\Psi|\Psi^t,\Psi_G^t) = p(\mathbf{X}|\Phi) \\
&+\sum_{d=1}^{M}\sum_{i=1}^{N}\sum_{z=1}^{K}\sum_{v=1}^{V} p(z,v|\Psi^t)\log\frac{p(\mathbf{x}^{(d,i),z,v|\Phi})}{p(\mathbf{x}^{(d,i),z,v|\Phi^t})} \\
&+\lambda\sum_{d=1}^{M}\sum_{i=1}^{N}\sum_{z=1}^{K}\sum_{v=1}^{V} p(z,v|\Psi_G^t)\log\frac{p(\mathbf{x}^{(d,i),z,v|\Phi})}{p(\mathbf{x}^{(d,i),z,v|\Phi_G^t})} \\
&\propto E_{p_r(z,v|\mathbf{X},\Psi^t,\Psi_G^t)}[\log p(\mathbf{X},z,v|\Psi)],
\end{aligned} \tag{4}
$$

where $\lambda \in [0,\infty]$ is the regularization factor. $\Psi_G = \{\theta_G, \varphi\}$, $\Phi_G = \{\theta_G, \varphi, \mu, \Sigma\}$. $p_r(z,v|\mathbf{X},\Psi^t,\Psi_G^T)$ is the regularized distribution over the latent variables, *i.e.*,

$$
p_r(z,v|\mathbf{X},\Psi^t,\Psi_G^t) = \frac{p(z,v|\mathbf{X},\Psi^t) + \lambda p(z,v|\mathbf{X},\Psi_G^t)}{1+\lambda}. \tag{5}
$$

In the E-step, given the data and the current parameter values, the posterior distributions over the latent variables are computed as:

$$
l_{z,v}^{(d,i)} = p_r(z^{(d,i)}, v^{(d,i)}|\mathbf{x}^{(d,i),\Psi^t,\Psi_G^t}), \tag{6}
$$

where $l_{z,v}^{(d,i)}$ denotes a set of latent variables.

In the M-step, new optimal parameters are computed based on the re-estimated latent variables. In particular, the parameters are calculated as:

$$
\begin{aligned}
\Psi^{t+1} &= \arg\max_{\Psi} Q(\Psi|\Psi^t,\Psi_G^t) \\
&+\sigma_d\sum_{d=1}^{M}(1-\sum_{z=1}^{K}\theta_{d,z}) + \sigma_z\sum_{z=1}^{K}(1-\sum_{v=1}^{V}\psi_{z,v}),
\end{aligned} \tag{7}
$$

where the second and the third terms are the Lagrange multipliers. By solving (7), we obtain:

$$
\theta_{d,z}^{t+1} = \frac{\sum_{i=1}^{N}\sum_{v=1}^{V} l_{z,v}^{(d,i)}}{\sum_{z=1}^{K}\sum_{i=1}^{N}\sum_{v=1}^{V} l_{z,v}^{(d,i)}}, \tag{8}
$$

$$
\varphi_{z,v}^{t+1} = \frac{\sum_{d=1}^{M}\sum_{i=1}^{N} l_{z,v}^{(d,i)}}{\sum_{v=1}^{V}\sum_{d=1}^{M}\sum_{i=1}^{N} l_{z,v}^{(d,i)}}, \tag{9}
$$

Finally, the regularized parameter θ_G is derived as:

$$
\theta_G^{t+1} = \frac{\exp\left(\frac{1}{MN}\sum_{d=1}^{M}\sum_{i=1}^{N}\log\theta_{d,z}^{t+1}\right)}{\sum_{z'=1}^{K}\exp\left(\frac{1}{MN}\sum_{d=1}^{M}\sum_{i=1}^{N}\log\theta_{d,z'}^{t+1}\right)}, \tag{10}
$$

which is essentially the geometric mean of the observation dependent $\theta_{d,z}$ with the same tag. The log scale is applied to make the computation tractable when $\theta_{d,z} \to 0$.

3.3. Flickr Circles Discovery by Dense Graph Mining

3.3.1. *Affinity Graph Construction*

To construct an affinity graph that describes the aesthetic similarity between Flickr users, a similarity measure is required. Based on the regularized topic model, user's aesthetic interest is represented by a mixture of Gaussian distribution. To measure the similarity between distributions, KL-divergence [17] $D_{KL}(\mathcal{N}||\mathcal{N}')$ is adopted. $\mathcal{N}$ and $\mathcal{N}'$ denote the learned aesthetic distribution of two users respectively.

Due to the non-symmetry of KL-divergence, it is difficult to integrate it into a semi-definite matrix for grouping task (*e.g.*, spectral clustering [21]). Instead, we use the square root of Jensen-Shannon divergence [17], a metric derived form KL-divergence:

$$
D_{JS}^{1/2}(\mathcal{N}||\mathcal{N}') = \sqrt{\frac{1}{2}\left(D_{KL}(\mathcal{N}||\mathcal{N}') + D_{KL}(\mathcal{N}'||\mathcal{N})\right)}. \tag{11}
$$

The above metric reflects the aesthetic similarity between Flickr users. It is integrated into a dense graph mining framework for detecting users with similar aesthetic interests. Firstly, we construct an affinity matrix $\mathbf{W}$ where the ij-th element is calculated as:

$$
\mathbf{W}_{ij} = \exp\left(-\frac{D_{JS}(\mathcal{N}_i||\mathcal{N}_j)}{2\psi^2}\right), \tag{12}
$$

where $\mathcal{N}_i$ and $\mathcal{N}_j$ denote aesthetic distribution of the i-th and the j-th users. On the basis of the affinity matrix, we construct an affinity graph as shown in Fig. 2.

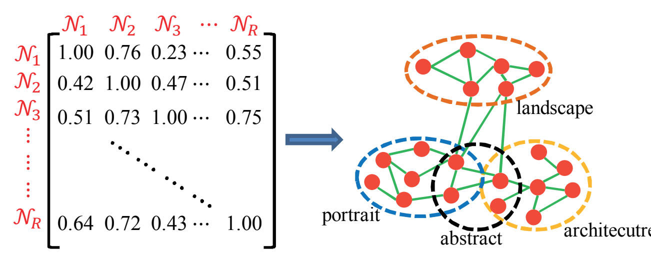


Fig. 2. Generating the affinity graph based on $\mathbf{W}$ (left) and discovering Flickr circles using dense graph mining (right).

3.3.2. *Graph Shift-based Flickr Circles Detection*

Obviously, Flickr users with similar aesthetic interests are densely distributed in the affinity graph. To effectively discover those dense subgraphs, two conditions are required. 1) *Compatibility with graph representation*: many similarity metrics are defined based on binary relation, such as our color+texture+semantics similarity. Only graph-based clustering can utilize the pairwise relation directly. 2)*Robustness to outliers*: a few users are with very particular aesthetic interests (*e.g.* skull photos) and they may not belong to any circles.

Methods insisting on partitioning all input data into coherent circles without outliers may fail to preserve the structure of the multiple circles.

Conventional clustering algorithms are not suitable for discovering circles from Flickr users, as they insist on partitioning all the input data. Comparatively, graph shift [22], which is efficient and robust for graph mode seeking, is suitable for mining the densely distributed Flickr users. It directly works on graph, supports arbitrary number of clusters, and leaves the outlier points ungrouped. Formally, we generate the affinity graph $\mathbf{G} = (\mathbf{U}, \mathbf{W})$. $\mathbf{U} = \{u_1, u_2, \cdots, u_M\}$ is a set of vertices denoting the Flickr users. $\mathbf{W}$ is a symmetric matrix detailed in (12). The modes of a graph $\mathbf{G}$ are defined as local maximizers of graph density function $g(y) = y^T \mathbf{A} y$, $y \in \Delta^M$, where $\Delta^M = \{y \in \mathbb{R}^M : y \geq 0$ and $||y||_{l_1} = 1\}$. More specifically, the similarity between users is expressed as the edge weights of graph $\mathbf{G}$. Those strongly connected subgraphs correspond to large local maxima of $g(y)$ over simplex, which is an approximate measure of the average affinity score of these subgraphs.

The target patterns are the local maximizers of $g(y)$. They represent the users in each Flickr circle, which are calculated by solving a quadratic optimization problem:

$$\max_{y} g(y) = y^T \mathbf{A} y, \;\; s.t. \;\; y \in \Delta^M, \tag{13}$$

Note that obtaining an analytic solution of (13) is difficult. Therefore, we employ replicator dynamics to find the local maxima of (13). Given an initialization $y(0)$, the local solution y^* can be iteratively computed by the discrete-time version of the first-order replicator equation, *i.e.*,

$$y_i(t+1) = y_i(t) \frac{(\mathbf{A}y(t))_i}{y(t)^T \mathbf{A} y(t)}. \tag{14}$$

4. EXPERIMENTS AND ANALYSIS

4.1. Data Set Compilation

Although the proposed method detects Flickr circles in an unsupervised manner, we require the labeled ground-truth data to evaluate its performance. We expended significant time, effort, and resources to crawl photos from 20 well-known public groups from Flickr. Each group contains more than 300,000 photos from 10,000 users. For each group, we crawled 50,000∼70,000 photos from nearly 5,000 Flickr users. The statistics of our data set is given in Fig. 3. For different Flicker groups, the number of photos of each users varies significantly, as shown in Table 1. This observation is the motivation of our regularized topic model. Fig. 4 shows the extent to which our 20 Flickr groups overlap with each other[1]. In summary, nearly 20% of the ground-truth groups are relatively independent to the other groups. About 60% of the groups moderately intersect with the other groups. The rest 20% of the groups are highly correlated with the others.

[1]Overlaps occur frequently because a few users are belonging to two or more groups simultaneously.

Table 1. Max/Min/Ave photo numbers of Flicker user from the 20 groups and the standard variance (SV)

Flickr group	Max No.	Min No.	Ave No.	SV
The light Fan.	2132	12	212	43.213
Film noir Mood	1765	22	196	36.764
Graphic designers	2543	7	267	46.541
Aesthetics failure	3567	43	231	24.356
Green is beautiful	2865	21	186	34.251
Colors	5643	41	324	65.674
Closer	3245	6	243	46.784
Less is more	2132	12	134	32.228
Field guide	2657	24	178	46.783
Night lights	4465	3	249	56.887
Black and white	3214	76	147	31.183
Stick figure	2654	16	227	40.654
Writing mach.	1342	11	103	23.355
Through glass	3421	42	245	46.678
Fog and rain	2885	51	215	38.897
Architecture	3146	36	195	37.769
Window seat	2989	11	245	56.782
Movement	4564	32	277	36.689
Orange and blue	3245	9	214	44.325
Jump Project	3105	14	227	43.236

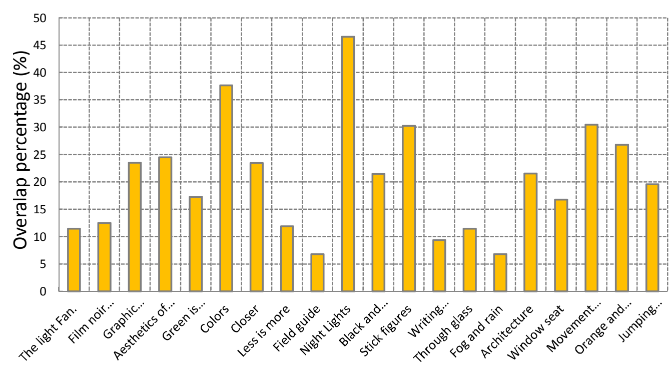


Fig. 4. The overlaps among the 20 Flickr groups.

4.2. A Comparative Study

We can evaluate the detected Flickr circles $\mathcal{C} = \{C_1, C_2, \cdots, C_Z\}$ by comparing them with the ground-truth Flickr circles (*i.e.*, groups) $\bar{\mathcal{C}} = \{\bar{C}_1, \bar{C}_2, \cdots, \bar{C}_{\bar{Z}}\}$. Our observation is that for an optimal communities discovery algorithm, the predicted circles should closely align with the ground-truth circles. To measure the alignment between a predicted circle C and a ground-truth circle $\bar{C}$, we compute the balanced error rate (BER) [23] between two circles.

We compare our approach with seven well-known clustering algorithms, including those considering only the graph/network structure, those exploring only the profile information, and those combining the both. 1)K-means Clustering (KC); 2) Hierarchical Clustering (HC) [16]; 3) Link Clus-

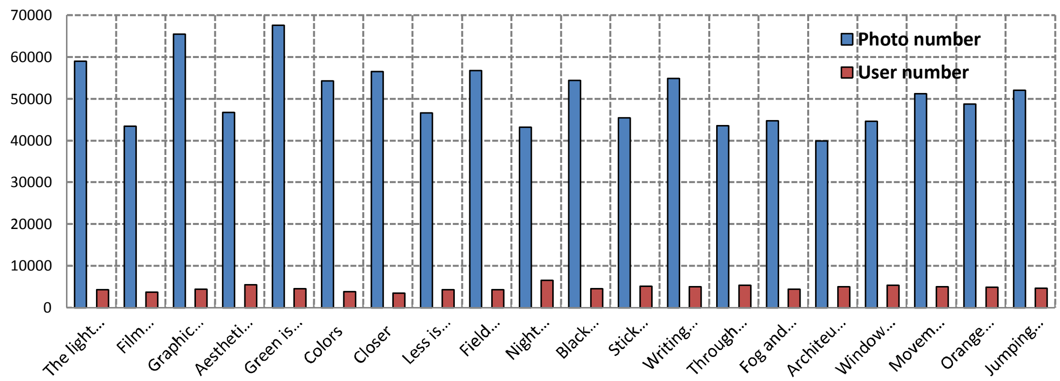


Fig. 3. The groups (the horizontal axis) and the number of Flicker users (the vertical axis) in each of the 20 groups.

Table 2. BER scores of the Seven Clustering Algorithms

Flickr group	KC	HC	LC	CP	LRE	MAC	Ours
The light Fan.	0.6334	0.4119	0.5448	0.5225	0.4448	0.5018	**0.5993**
Film noir Mood	0.6447	0.5079	0.6339	0.5745	0.5335	0.6077	**0.7144**
Graphic designers	0.4669	0.4553	0.4757	0.5043	0.4669	0.4811	**0.5836**
Aesthetics failure	0.6789	0.7004	0.6656	0.6741	0.7339	0.6875	**0.7993**
Green is beautiful	0.4118	0.4664	0.4338	0.4698	0.4934	0.5036	**0.5448**
Colors	0.8223	0.7786	0.7814	0.7643	0.7331	0.7654	**0.8337**
Closer	0.6697	0.6854	0.6659	0.6841	0.6758	0.6916	**0.7741**
Less is more	0.6049	0.6213	0.5887	0.6083	0.5865	0.6059	**0.6653**
Field guide	0.7113	0.7204	0.6779	0.6887	0.7014	0.7116	**0.7559**
Night lights	0.6896	0.7032	0.6884	0.6886	0.7013	0.6654	**0.7867**
Black and white	0.3098	**0.4334**	0.3448	0.3118	0.2985	0.3552	0.3342
Stick figure	0.6685	0.6773	0.6936	0.6853	0.6774	0.6819	**0.7528**
Writing mach.	0.7669	0.7883	0.8215	0.7665	0.8114	0.7748	**0.8665**
Through glass	0.7118	0.7336	0.6887	0.7033	0.7129	0.7559	**0.7783**
Fog and rain	0.3879	0.3821	0.3946	0.4119	0.4228	0.4315	**0.4652**
Architecture	0.7179	0.7226	0.7012	0.6894	0.6884	0.6943	**0.7332**
Window seat	0.7659	0.7932	0.8087	0.7946	0.7945	0.8128	**0.8441**
Movement	0.4937	0.4653	0.5222	0.5134	0.4667	0.4894	**0.5448**
Orange and blue	0.5339	0.5337	0.5032	0.5119	0.4875	0.4334	**0.5943**
Jump Project	0.6415	0.6229	0.5949	0.6112	0.6034	**0.6213**	0.6049
Average	0.6131	0.6311	0.6125	0.5968	0.5462	0.6205	**0.7485**

tering (LC) [25]; 4) Clique Percolation (CP) [24]; 5) Low-Rank Embedding (LRE) [32]; and 6) Multi-Assignment Clustering(MAC) [33]. We compare our approach with the seven baseline clustering algorithms described above. For all the algorithms, we fix the cluster number to 20. The low&high-level visual features of the seven algorithms are the same as ours. As shown in Table 2, the following observations can be made. 1) For 18 out of the 20 Flickr groups, our approach outperforms all its competitors, as the corresponding BER s-cores are the highest. This observation shows the advantage of our regularized topic model, which can optimally capture users' aesthetic interest. 2) For Flickr circles describing specific concepts, *e.g.*, the architecture and the window seat, they can be more accurately detected. Comparatively, for circles describing abstract concepts, *e.g.*, the jump project, all the algorithms are difficult to detect the circles. This is because photos with abstract concepts do not have a regular visual appearance. Therefore, the low&high-level features cannot well describe them. 3) As shown in Table 1, our proposed method performs significantly better on Flickr groups containing users with very few photos (*e.g.*, the graphic designers and the night lights). This again demonstrates the necessity to use a regularized term to model the aesthetic interests of users with few photos.

5. CONCLUSIONS

This paper proposes to learn circles from a large number of Flickr users, where a circle contains users with similar aesthetic interests. A regularized term is incorporated into the topic model to describe the aesthetic distribution of each user. Next, an affinity graph is constructed to describe the aesthetic relationships of users,. Finally, users densely distributed on the affinity graph are categorized into different circles.

6. ACKNOWLEDGMENTS

This research has been supported by the Singapore National Research Foundation under its International Research Centre@ Singapore Funding Initiative and administered by the IDM Programme Office through the Centre of Social Media Innovations for Communities (COSMIC).

7. REFERENCES

[1] W. Luo, X. Wang, and X. Tang, "Content-based photo quality assessment", *in Proc. IEEE Int'l Conf. Comput. Vis.*, Nov 2011, pp. 2206–2213.

[2] L. Zhang, Y. Gao, R. Zimmermann, Q. Tian, and X. Li, "Fusion of multichannel local and global structural cues for photo aesthetics evaluation", *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1419–1429, March 2014.

[3] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models-their training and application", *Comput. Vis. and Image Understanding*, vol. 61, no. 1, pp. 38 – 59, 1995.

[4] R. Datta, D. Joshi, J. Li, and J. Wang, "Studying aesthetics in photographic images using a computational approach", *in Proc. Euro. Conf. Comput. Vis.*, 2006, pp. 288–301.

[5] L.-K. Wong and K.-L. Low, "Saliency-enhanced image aesthetics class prediction," *in Proc. IEEE Int'l Conf. Image Process.*, Nov 2009, pp. 997–1000.

[6] S. Dhar, V. Ordonez, and T. Berg, "High level describable attributes for predicting aesthetics and interestingness", *in Proc. IEEE Int'l Conf. Comput. Vis.*, June 2011, pp. 1657–1664.

[7] Y-L. Boureau, J. Ponce, and Y. LeCun, "A Theoretical Analysis of Feature Pooling in Visual Recognition", *in Proc. IEEE Int'l Conf. Mach. Learn.*, 2010.

[8] B. Cheng, B. Ni, S. Yan, and Q. Tian, "Learning to photograph", *ACM Multimedia*, 2010, pp. 291–300.

[9] L. Zhang, M. Song, Q. Zhao, X. Liu, J. Bu, and C. Chen, "Probabilistic graphlet transfer for photo cropping", *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 802–815, Feb 2013.

[10] L. Zhang, Y. Gao, C. Zhang, H. Zhang, Q. Tian, R. Zimmermann, "Perception-Guided Multimodal Feature Fusion for Photo Aesthetics Assessment", *ACM MM*, 2014.

[11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation", *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[12] G. Costa and R. Ortale, "A bayesian hierarchical approach for exploratory analysis of communities and roles in social networks," *ASONAM*, 2012, pp. 194–201.

[13] E. Yao, G. Zheng, O. Jin, S. Bao, K. Chen, Z. Su, and Y. Yu, "Probabilistic text modeling with orthogonalized topics", *SIGIR*, 2014, pp. 907–910.

[14] D. Zhou, E. Manavoglu, J. Li, C. L. Giles, and H. Zha, "Probabilistic models for discovering e-communities", *in Proc. Int'l Conf. World Wide Web*, 2006, pp. 173–182.

[15] Z. Yin, L. Cao, Q. Gu, and J. Han, "Latent community topic analysis: Integration of community discovery with topic modeling", *ACM TIST*, vol. 3, no. 4, p. 63, 2012.

[16] S. Johnson, "Hierarchical clustering schemes", *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.

[17] W. Press, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Cambridge University Press, 2007.

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2005, pp. 886–893.

[19] M. A. Stricker and M. Orengo, "Similarity of color images," *Storage and Retrieval for Image and Video Databases*, 1995, pp. 381–392.

[20] T. Hofmann, "Probabilistic latent semantic analysis", *in Proc. of Uncert. in artifi. intelli.*, 1999, pp. 289–296.

[21] X. Li, W. Hu, C. Shen, A. Dick, and Z. Zhang, "Context-aware hypergraph construction for robust spectral clustering", *IEEE Trans. Knowledge and Data Eng.*, pp. 1–1, 2013.

[22] H. Liu and S. Yan, "Robust graph mode seeking by graph shift", *in Proc. Int'l Conf. Mach. Learn.*, 2010, pp. 671–678.

[23] Y. wei Chen, "Combining svms with various feature selection strategies", *Taiwan University*, Springer-Verlag, 2005.

[24] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society", *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.

[25] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks", *Nature*, vol. 466, no. 7307, pp. 761–764, 2010.

[26] L. Zhang, Y. Gao, R. Ji, Q. Dai, X. Li, "Actively Learning Human Gaze Shifting Paths for Photo Cropping", *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2235–2245, 2014.

[27] L. Zhang, M.i Song, Y. Yang, Q. Zhao, Z. Chen, N. Sebe, "Weakly Supervised Photo Cropping", *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 94–107, 2014.

[28] L. Zhang, R. Ji, Y. Xia, X. Li, "Learning a Probabilistic Topology Discovering Model for Scene Categorization", *IEEE Trans. Neur. Network and Learn. System*, 2014.

[29] L. Zhang, Y. Gao, R. Hong, Y. Hu, R. Ji, Q. Dai, "Probabilistic Skimlet Fusion for Summarizing Multiple Consumer Landmark Videos", *IEEE Trans. Multimedia*, 71(1): 40–49, 2015.

[30] L. Nie, Y.-L. Zhao, M. Akbari, J. Shen, T.-S. Chua, "Bridging the Vocabulary Gap between Health Seekers and Healthcare Knowledge", *IEEE Trans. Knowl. Data Eng.*, 27(2): 396–409, 2015.

[31] L. Nie, Y.-L. Zhao, X. Wang, J. Shen, T.-S. Chua, "Learning to Recommend Descriptive Tags for Questions in Social Forums", *ACM Trans. Inf. Syst.*, 32(1): 5, 2014.

[32] T. Yoshida, "Toward finding hidden communities based on user profile", *Workshops of IEEE Int'l Conf. on Data Mining*, 2010, pp. 380–387.

[33] M. Frank, A. P. Streich, D. Basin, and J. M. Buhmann, "Multi-assignment clustering for boolean data," *J. Mach. Learn. Res.*, vol. 13, pp. 459–489, Feb. 2012.