Connection Discovery using Shared Images by Gaussian Relational Topic Model

Xiaopeng Li[*], Ming Cheung[†], James She[‡]

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

Abstract—Social graphs, representing online friendships among users, are one of the fundamental types of data for many applications, such as recommendation, virality prediction and marketing in social media. However, this data may be unavailable due to the privacy concerns of users, or kept private by social network operators, which makes such applications difficult. Inferring users' interests and discovering users' connections through their shared multimedia content has attracted more and more attention in recent years. This paper proposes a Gaussian relational topic model for connection discovery using user shared images in social media. The proposed model not only models users' interests as latent variables through their shared images, but also considers the connections between users as a result of their shared images. It explicitly relates user shared images to user connections in a hierarchical, systematic and supervisory way and provides an end-to-end solution for the problem. This paper also derives efficient variational inference and learning algorithms for the posterior of the latent variables and model parameters. It is demonstrated through experiments with over 200k images from Flickr that the proposed method significantly outperforms the methods in previous works.

Keywords-Bayesian, topic model, variational inference, user shared images, connection, discovery, recommendation, social network analysis.

I. INTRODUCTION

Social graphs (SGs), representing online friendships among users, are fundamental data for many applications, such as recommendation, virality prediction and marketing in social media. However, this data may be unavailable due to the privacy concerns of users, or kept private by social network operators, and these applications become challenging with an incomplete set of data. Providing a potential solution to this problem, user connections are also reflected in the abundant social content, especially images, shared on social networks. Inferring users' interests and discovering users' connections through their shared multimedia content has attracted more and more attention in recent years. A common but unreliable approach is using user annotated tags (or user tagging) associated with each shared image to discover user connections when the SG is not accessible. However, user annotated tags may be unavailable or images may be incorrectly labeled, as shown in Fig. 1. Instead of indirectly taking the image content information through tags, a more direct method is to consider the image's visual content. Users with connections of follower/followee relationships
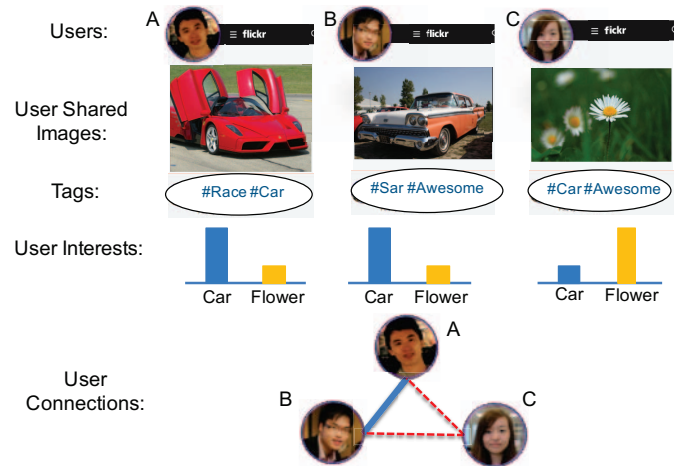

Figure 1: Examples of user shared images, image tags, user interest reflected and user connections.

are found to have relatively higher visual content similarities among their shared images. An simplistic example of user generated images on Flickr is shown in Fig. 1: Both users A and B share images of cars and user C shares an image of a flower. The follower/followee relationship between users A and B can possibly be detected from the higher similarity of visual features in their shared images. When more shared images from each of users A, B and C are accessible for evaluation, the actual follower/followee relationships should become reliably and accurately detectable, though such a task is becoming challenging with the number of shared images and user connections in social networks growing larger every day.

The effectiveness of connection discovery using user shared images is mainly determined by two elements: effective extraction of information from images and an effective method to connect the image content to user connections. On the one hand, with the recent development of convolutional neural network (CNN), the analysis and understanding of image content has become much more effective [1] through hierarchical representation, closing the semantic gap between pixels and content. It is therefore possible to extract rich image content information through CNN. On the other hand, how to effectively discover user connections using the image content remains a challenge. Summarizing the

methods or frameworks proposed in previous works such as [2], [3], [4] and [5], the process is generally divided into two stages: first, construction of user profile by counting the occurrences of image labels or summing the image feature vectors, and second, prediction of connections between users based on the constructed user profile. The limitation of these methods is two-fold. First, aggregating all items in a simple way for each user might not be solid enough to capture users' interests. An analogy can be found in text analytics, where the bag-of-words model is compared with topic models, especially latent Dirichlet allocation (LDA)[6]. Second, user profiling through images is completely separated from connection discovery. Separating the two potentially results in untight relatedness of image content and social links, meaning that the observation of partial social links lends no help toward connection discovery using image content.

In this paper, a Gaussian relational topic model (GRTM) is proposed for connection discovery using user shared images in order to overcome the abovementioned limitations. GRTM is an end-to-end hierarchical model specifically designed to not only model users' interests through image content but also to supervise the modeling in such a way that the content of shared images is statistically connected to the links between users, inspired by hierarchical relational model in text document domain [7]. GRTM effectively extracts rich image content information through CNN and models each semantic topic as a Gaussian topic. GRTM also models each user's interests as a latent factor and assumes that the action of the user sharing an image is probabilistically motivated by his or her interests. Furthermore, the links between users are modeled based on the images each user shares. Combining these in a coherent probabilistic generative process, the proposed GRTM provides a systematic way to close the gap between the actions of users sharing images and users connecting to each other. The main contributions of this paper are the following:

- proposes an end-to-end Gaussian relational topic model for connection discovery using user shared images, closely relating user shared image content to user connections.
- derives efficient variational inference for the proposed model to approximate the posterior of latent variables and learn model parameters.
- evaluates the performance of the proposed model with real data and proves the significantly better performance of the proposed method.

The rest of the paper is organized as follows. Section II discusses previous works. Section III introduces the GRTM for connection discovery using shared images. Section IV describes in detail the inference of latent variables and model parameters, as well as the prediction method. Section V presents experimental results and discussion, and Section VII concludes the paper.

II. RELATED WORKS

One content-based approach to discover connections through images is to generate a label for an image based on its visual elements [8][9]. However, determining the relationship between the visual elements and the label is not a trivial task because the same object can be visually different among images, and denoting each image by a single label loses a lot of information. [2] and [10] propose to first generate labels for user shared images by clustering and describe users' characteristics by counting the occurrences of image clusters appearing in their collections. The prediction is then made through analyzing the similarities among the users' histograms. Both methods assume each image contains one and only one topic, neglecting potentially rich information. On the other hand, [3] and [4] represent users' interests by summing the feature vectors of their images, achieving a similar effect to the other methods. However, all these methods do not actually connect shared images to the links between users, or only connect shared images to links, and not the other way around.

Relational topic model was first proposed in [7] for document networks in the text analytics domain. It is an extension of the latent Dirichlet model (LDA) [6] and supervised LDA [11]. This family of approaches model each document as topic proportions, and each word in the document is drawn from a set of topics, which are distributions over a fixed vocabulary. Furthermore, the relational topic model models the links between documents as a binary random variable that is conditioned on their contents. More recent papers, [12] and [13], share a similar methodology with this paper, also proposing a hierarchical topic model to process user images and infer users' interest as latent factors from Bayesian inference. However, [12] starts with low-level pixels and describes image regions with visual words, while [13] models the visual descriptor as drawn from a vocabulary, just like in the text domain. Also, [12] considers no social information and hence, as with other works, does not relate image content to user relationships for connection discovery, and [13] considers the social influence of existing links to the users rather than predicting new ones. Our work extracts rich information of image content through CNN and models users' interests and users' connections simultaneously. As an end-to-end model, ours closely relates users' connections to user shared images, with the goal being connection discovery.

III. THE PROPOSED MODEL FOR CONNECTION DISCOVERY USING SHARED IMAGES

This section introduces the problem of connection discovery using shared images and the proposed Gaussian relational topic model. Given a social network, there are N users, $\{u_1, u_2, \ldots, u_N\}$. Each of the users, u, shares a collection of images, $I_u = \{x_{u,1}, x_{u,2}, \ldots, x_{u,N_u}\}$, where N_u is the number of images the user u shares and $x_{u,n}$ is the
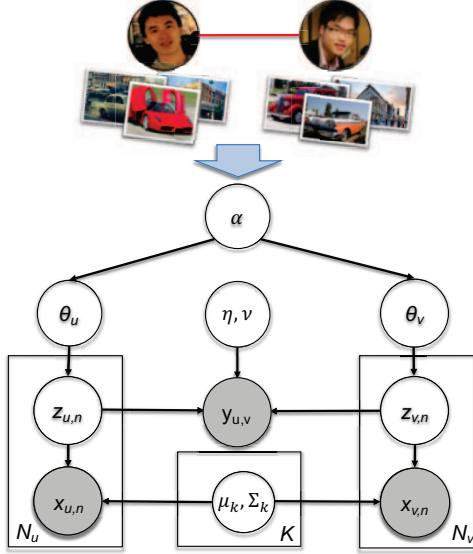
Figure 2: The proposed Gaussian Relational Topic Model (GRTM) (shaded nodes are observed.)

nth image shared by the user u. The connections between users are not available due to privacy concerns or proprietary information protection. However, the connections can still be discovered through the images shared by the users. The objective is to predict the possible connections between users based on the user shared images.

The proposed GRTM model, is based on Gaussian mixture model and relational topic model, as shown in Fig. 2. It is a generative probabilistic model in which a user is described by a topic distribution that is reflected through his or her collection of shared images. It models that the action of a user sharing an image follows a generative process: a user's preferences or the topic proportion in his or her collections is generated through a Dirichlet-distributed vector, and he shares an image by first drawing a topic assignment from his or her preferences then drawing an image from the corresponding topic distribution. Unlike LDA and relational topic model, where text documents are the context and words are generated from a vocabulary, this paper extracts rich information from images through a pre-trained CNN and Gaussian topic is adopted in order to consider and preserve the rich information of the images, as shown in Fig. 3. The links between users are then modeled as binary variables and are determined by the users' preferences. In this way, the user shared images and the links between users are statistically connected.

Formally, the generative process of the proposed GRTM is as follows:

1) For each user u:
   a) Draw topic proportions $\theta_u | \alpha \sim Dir(\alpha)$.
   b) For each image $x_{u,n}$:
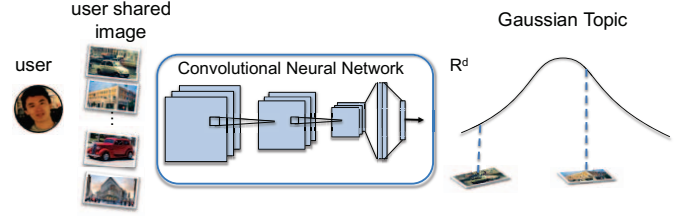   i) Draw topic assignment $z_{u,n} | \theta_u \sim Mult(\theta_u)$.

Figure 3: CNN feature extraction for images and Gaussian topic

   ii) Draw the image $x_{u,n} | z_{u,n}, \{\mu, \Sigma\}_{1:K} \sim \mathcal{N}(\mu_{z_{u,n}}, \Sigma_{z_{u,n}})$.

2) For each pair of users u, v:
   a) Draw binary link indicator between users $y | z_u, z_v \sim \psi(\cdot | z_u, z_v)$.

Fig. 2 illustrates the graphical model for the generative process. Assuming there are a total of K topics, generating an image is done by first picking the topic it belongs to. The apparent choice of distribution for picking the topic assignment $z_{u,n}$ from K possible values is multinomial distribution, parameterized by the k-dimensional user preference θ_u. The conjugate prior for multinomial distribution is Dirichlet distribution. Therefore, the user preference θ_u is assumed to be drawn from Dirichlet distribution, parameterized by α.

Under each topic, an image is assumed to be drawn from a multivariate Gaussian distribution, parameterized by mean μ and covariance Σ:

$$p(x_{u,n}|z_{u,n}, \{\mu, \Sigma\}_{1:K}) = \prod_{k=1}^{K} \mathcal{N}(x_{u,n}|\mu_k, \Sigma_k)^{\mathbb{1}(z_{u,n}=k)}. \tag{1}$$

where $\mathbb{1}(\cdot)$ is an indicator function. Therefore, given the parameters α and topic distribution $\{\mu, \Sigma\}_{1:K}$, the generative probability of a user shared image is given by

$$\begin{aligned} p(x_{u,n}, &z_{u,n}, \theta_u | \alpha, \{\mu, \Sigma\}_{1:K}) \\ &= p(\theta_u|\alpha)p(z_{u,n}|\theta_u)p(x_{u,n}|z_{u,n}, \{\mu, \Sigma\}_{1:K}). \end{aligned} \tag{2}$$

Given two users and the topic assignments of all their images, the link between the two users is determined by the link probability function ψ. Specifically, the exponential function is adopted due to the linear form of the log likelihood [7]:

$$\psi(y = 1|z_u, z_v) = \exp(\eta^T(\bar{z}_u \circ \bar{z}_v) + \nu), \tag{3}$$

where $\bar{z}_u = \frac{1}{N_u} \sum_n z_{u,n}$, η and ν are parameters, and the notation $\circ$ denotes the element-wise product.

The overall joint likelihood of the observation, i.e., the user shared images and links between users, and the latent variables, i.e., the user preference and topic assignment, is

determined by

$$p(U, y, \theta, z | \alpha, \{\mu, \Sigma\}_{1:K}) = \prod_u p(\theta_u | \alpha)$$

$$\prod_n p(z_{u,n} | \theta_u) p(x_{u,n} | z_{u,n}, \{\mu, \Sigma\}_{1:K}) \prod_{(u,v)} p(y | z_u, z_v).$$

$$(4)$$

IV. INFERENCE, ESTIMATION AND PREDICTION

In this section, the Bayesian variational inference for the proposed GRTM is presented.

A. Inference

The posterior distribution of the latent variables and parameters given the observations and topic parameters is inferred by:

$$p(\theta, z, \eta, \nu | U, y, \alpha, \{\mu, \Sigma\}_{1:K})$$
$$= \frac{p(U, y, \theta, z | \alpha, \{\mu, \Sigma\}_{1:K})}{\int \sum_{z_{u,n}} p(U, y, \theta, z | \alpha, \{\mu, \Sigma\}_{1:K}) d\theta}. \quad (5)$$

The exact posterior, however, is intractable since the denomenator involves a complex summation and integral. Instead, like previous works, variational inference with free parameters is used to approximate the exact posterior. Specifically, the mean field approximation with the form

$$q(\theta, z | \gamma, \phi) = \prod_u q_u(\theta_u | \gamma_u) \prod_n q_z(z_{u,n} | \phi_{u,n}) \quad (6)$$

is used to approximate the posterior p of θ and z. And the goal is to minimize the Kullback-Leibler (KL) divergence between the approximation and the exact posterior. Equivalently, we try to maximize the variational free energy, which is the evidence lower bound of the log marginal probability of the observations:

$$\mathcal{L} = \sum_u \sum_n E_q[\log p(x_{u,n} | z_{u,n}, \{\mu, \Sigma\}_{1:K})]$$

$$+ \sum_u \sum_n E_q[\log p(z_{u,n} | \theta_u)] + \sum_u E_q[\log p(\theta_u | \alpha)]$$

$$- \sum_u E_q[\log q_\theta(\theta_u | \gamma_u)] - \sum_u \sum_n E_q[\log q_z(z_{u,n} | \phi_{u,n})]$$

$$+ \sum_{(u,v)} E_q[\log p(y_{u,v} | z_{u,v}, \eta, \nu)].$$

$$(7)$$

B. Estimation

Having the evidence lower bound of the log likelihood, the maximization of the lower bound can be achieved through coordinate ascent, where each variable is iteratively updated assuming all others are fixed, until convergence. By taking the derivative of the evidence lower bound in Eq. 7, subject to the sum-to-one constraint, the update of the free parameter ϕ is

$$\phi_{u,n,k} \propto \exp(\log N(x_{u,n} | \mu_k, \Sigma_k) + \Psi(\gamma_i) - \Psi(\sum_{j=1}^K \gamma_j)$$
$$+ \sum_{v | y_{u,v} = 1} \frac{\eta \circ \bar{\phi}_v}{N_u}),$$

$$(8)$$

where $\bar{\phi}_v = E_q[\bar{z}_v] = \frac{1}{N_v} \sum_n \phi_{v,n}$. And the normalization term is the sum over all K.

With a similar method, the update of the free parameter γ is

$$\gamma_{u,k} = \alpha_k + \sum_{n=1}^{N_u} \phi_{u,n,k}. \quad (9)$$

Isolating the terms involving $\{\mu, \Sigma\}_{1:K}$ in the lower bound and taking the derivative, the update of the topic parameters $\{\mu, \Sigma\}_{1:K}$ is

$$\mu_k = \frac{\sum_u \sum_n \phi_{u,n,k} x_{u,n}}{\sum_u \sum_n \phi_{u,n,k}}, \quad (10)$$

$$\Sigma_k = \frac{\sum_u \sum_n \phi_{u,n,k}(x_{u,n} - \mu_k)(x_{u,n} - \mu_k)^T}{\sum_u \sum_n \phi_{u,n,k}}. \quad (11)$$

It would be inappropriate to regard all links except observed positive links as negative training examples since there might be positive but unobserved links between users, which is expected to predict. Therefore, following [7], we use a regularization penalty parameterized by ρ for the negative observations. And the updates of the link function are conducted analytically by

$$\nu \leftarrow \log(M - \mathbf{1}^T \bar{\Pi}) - \log(\rho(1 - \frac{1}{K}) + M - \mathbf{1}^T \bar{\Pi}), \quad (12)$$

$$\eta \leftarrow \log(\bar{\Pi}) - \log(\bar{\Pi} + \frac{\rho}{K^2} \mathbf{1} - \mathbf{1}\nu), \quad (13)$$

where $M = \sum_{u,v} \mathbb{1}(y_{u,v} = 1)$, $\bar{\Pi} = \sum_{u,v} \bar{\pi}_{u,v} \mathbb{1}(y_{u,v} = 1)$ and $\bar{\pi}_{u,v} = \frac{1}{N_u} \sum_n \phi_{u,n} \circ \frac{1}{N_v} \sum_n \phi_{v,n}$.

C. Prediction

With all the variational parameters estimated, the link between two users given their shared collection of images can be predicted. The variational Bayesian prediction is given by:

$$p(y_{u,v} | x_u, x_v) = E_q[p(y_{u,v} | \bar{z}_u, \bar{z}_v)]. \quad (14)$$

However, it still involves complicated summations over all possible z for all the images considered. Instead, we perform plug-in approximation and substitute all z with its variational expectation. Hence, the predictive probability is approximated with

$$p(y_{u,v} | x_u, x_v) = \exp(\eta^T \bar{\pi}_{u,v} + \nu), \quad (15)$$

where $\bar{\pi}_{u,v} = \frac{1}{N_u} \sum_n \phi_{u,n} \circ \frac{1}{N_v} \sum_n \phi_{v,n}$.

V. Experimental Results

In this section, experiments are conducted to investigate the effectiveness of the proposed GRTM for connection discovery using user shared images.

A. Dataset and Experimental Setup

A set of 201,006 images uploaded by 542 users are scraped from Flickr, an online social network for image sharing, with millions of images uploaded. The 542 users are selected randomly from images under the same tag query page to provide diversity. The average number of shared images for each user is 370, covering diverse content, and there are 902 connections among the 542 users. The images are processed by GoogLeNet [14], pretrained using the ILSVRC 2014 dataset, for image representation, so that rich semantic information can be extracted. The links between users are divided into two parts: 60% of the links are used for training and 40% of the links are used for testing. Using the user shared images and training links, the latent variables and model parameters of the GRTM are estimated through an iterative process, as described in the previous section. Using Eq. 15, the testing process is then undertaken to predict the probability of the links existing between users , excluding the observed training links, and the results are then compared with the groundtruth of the testing links. The Dirichlet hyperparameter α is set to 2.0, and the number of topics is set to be 100 in our experiment. Though the number of topics has some influence to some extent, the overall performance does not vary much.

As a comparison with previous works on similar applications, the Mean method from [4] where users' profiles are constructed through their shared images by taking the mean features of all the images, and the BoFT method [2] where users' profiles are obtained by counting the occurrences of cluster labels in users' collections obtained through image clustering, are implemented. The prediction of user connections is thus conducted through computing the similarity between users. And instead of using the scale invariant feature transform (SIFT) for image feature extraction, CNN (GoogLeNet) features are used.

B. Results

The prediction performance of the Mean method, BoFT method and proposed GRTM is shown in Fig. 4, where both the receiver operating characteristic (ROC) curve and precision-recall curve are plotted. The ROC curve is the function of true positive rate (TPR) and false positive rate (FPR) and is used to measure how well the model can distinguish whether two users are friends or not. As it is shown, the ROC curve of the GRTM clearly dominates the other two methods. For example, for the point in the ROC curve with TPR of 0.8, given a pair of users with a true friendship, GRTM has 80% probability of predicting it correctly, while given a pair of users without a friendship,
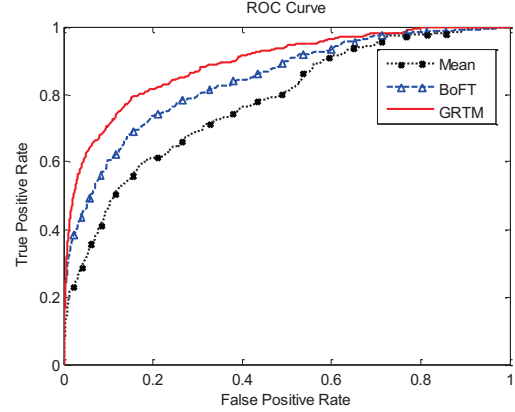

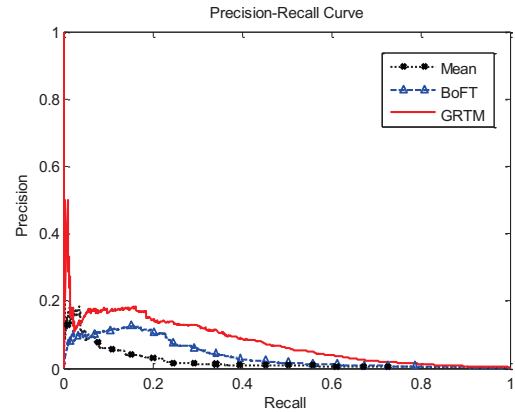Figure 4: ROC curve for Mean method [4], BoFT method [2] and proposed GRTM.

Figure 5: Precision-Recall curve for Mean method [4], BoFT method [2] and proposed GRTM.

it has 20% probability of predicting it incorrectly. For the same TPR, however, the BoFT method has 35% probability of incorrect prediction for users without a friendship, and Mean method has 45%. Overall, the area under curve (AUC) for ROC curve of the proposed GRTM is 0.89, while that of the BoFT method is 0.84 and that of Mean method is 0.78. As an alternative measure, the precision-recall curve, as shown in Fig. 5, is the function of the precision rate and recall rate. It measures what fraction of the recommended candidates are the user's true friends and what fraction of true friends are recommended. As with the ROC curve, the precision-recall curve of the proposed GRTM also dominates the others. For example, for the same recall rate of 10%, the precision rate of the GRTM is 18%, while that of the BoFT method is 11% and that of the Mean method is 5.9%. Overall, the precision-recall AUC of the proposed GRTM is 0.05, while that of the BoFT method is 0.039 and that of the Mean method is 0.021.

In order to illustrate how the proposed GRTM interprets the user shared images, some example images are shown
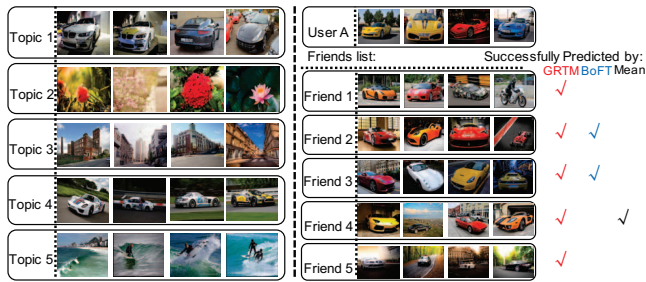
Figure 6: Left: examples of Gaussian topics; Right: examples of friendship prediction by Mean method, BoFT method and the proposed GRTM.

in Fig. 6. Since each Gaussian topic is defined by μ_k and Σ_k and each image has a probability under the Gaussian topic, images that achieve the highest probabilities under that topic are selected as representatives. As shown in Fig. 6, the semantic meaning of the Gaussian topics, such as cars, flowers, buildings, race cars and surfing, can be readily distinguished. The Gaussian topics cover a variety of objects, scenes, sports, etc. Those images that contain more than one topic could also be revealed through the probabilities under different Gaussian topics. It is also noted that ordinary cars and race cars are separated as different topics, which is also beneficial to reflect users' preferences for subsequent connection discovery. Fig. 6 shows examples of the friendship prediction results obtained by the three methods. The example user, A, shares a lot of car images. For recommendation purposes, all three methods are directed to recommend 10 friends. As shown in the figure, out of the 10 recommendations, the proposed GRTM successfully predicts 5 of them, whereas the BoFT only successfully predicts 2 and the Mean method successfully predicts 1. Furthmore, the successful predictions of the other two methods are in fact a subset of those of the proposed GRTM.

VI. CONCLUSION

This paper proposes a Gaussian relational topic model (GRTM) for connection discovery using user shared images in social media. The GRTM not only models users' interests as latent variables through user shared image content but also models the connections between users as a result of their shared images. It explicitly relates user shared images to the connections between users in a hierarchical, systematic and supervisory way and provides an end-to-end model for connection discovery using shared images. It is demonstrated by experiment that the proposed model significantly outperforms the methods in previous works where the modeling of users' interests and connection discovery are separated.

ACKNOWLEDGMENT

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] M. Cheung, J. She, and Z. Jie, "Connection discovery using big data of user-shared images in social media," *Multimedia, IEEE Transactions on*, vol. 17, no. 9, pp. 1417–1428, 2015.

[3] L. Yang, C.-K. Hsieh, and D. Estrin, "Beyond classification: Latent user interests profiling from visual contents analysis," in *International Conference on Data Mining Workshop*. IEEE, 2015, pp. 1410–1416.

[4] J. Zhuang, T. Mei, S. C. Hoi, X.-S. Hua, and S. Li, "Modeling social strength in social media community via kernel-based learning," in *Proceedings of the 19th International Conference on Multimedia*. ACM, 2011, pp. 113–122.

[5] X. Geng, H. Zhang, Z. Song, Y. Yang, H. Luan, and T.-S. Chua, "One of a kind: User profiling by social curation," in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 567–576.

[6] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[7] J. Chang and D. M. Blei, "Relational topic models for document networks." in *AIStats*, vol. 9, 2009, pp. 81–88.

[8] 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, International Conference on*. IEEE, 2009, pp. 1452–1455.

[9] X. Zhang, X. Zhao, Z. Li, J. Xia, R. Jain, and W. Chao, "Social image tagging using graph-based reinforcement on multi-type interrelated objects," *Signal Processing*, vol. 93, no. 8, pp. 2178–2189, 2013.

[10] M. Cheung, J. She, and X. Li, "Non-user generated annotation on user shared images for connection discovery," in *Cyber, Physical and Social Computing, the 8th IEEE Conference on*. IEEE, 2015.

[11] J. D. Mcauliffe and D. M. Blei, "Supervised topic models," in *Advances in Neural Information Processing Systems*, 2008, pp. 121–128.

[12] P. Xie, Y. Pei, Y. Xie, and E. P. Xing, "Mining user interests from personal photos." in *AAAI*. Citeseer, 2015, pp. 1896–1902.

[13] J. Sang and C. Xu, "Right buddy makes the difference: An early exploration of social relation analysis in multimedia applications," in *Proceedings of the 20th ACM International Conference on Multimedia*. ACM, 2012, pp. 19–28.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.