

# Modeling Emotion Influence in Image Social Networks

Xiaohui Wang, Jia Jia, Jie Tang, Boya Wu, Lianhong Cai and Lexing Xie

**Abstract**—We study emotion influence in large image social networks. We focus on users' emotions reflected by images that they have uploaded and social influence that plays a role in changing users' emotions. We first verify the existence of emotion influence in the image networks, and then propose a probabilistic factor graph based emotion influence model to answer the questions of "who influences whom". Employing a real network from Flickr as the basis in our empirical study, we evaluate the effectiveness of different factors in the proposed model with in-depth data analysis. The learned influence is fundamental for social network analysis and can be applied to many applications. We consider using the influence to help predict users' emotions and our experiments can significantly improve the prediction accuracy (3.0%-26.2%) over several alternative methods such as Naive Bayesian, SVM (Support Vector Machine) or traditional Graph Model. We further examine the behavior of the emotion influence model, and find that more social interactions correlate with higher emotion influence between two users, and the influence of negative emotions is stronger than positive ones.

**Index Terms**—Emotion influence, image, social networks

## 1 INTRODUCTION

Emotion stimulates the mind 3,000 times quicker than rational thought [1]. One may make a quick decision simply because of a particular feeling (e.g., happy, sad, or angry). Nowadays, people are used to express their daily feeling directly in the online networking space through texts and images. Understanding the substantial mechanism of emotion dynamics can give us insight into how people's emotions influence their activities and how the emotions spread in the social networks.

Recently, image social networks, in which images are used as the major media to express users' feeling, opinions and activities that they are invoked, have been constructed and attracted billions of users (Flickr<sup>1</sup>, Instagram<sup>2</sup>). In social networks, if one's emotional status is changed by her/his friends' emotional statuses, we say that there is an emotion influence. By

emotion influence, we mean that a user's emotional status is influenced by her/his friends, in particular through their uploaded images. And image emotion just refers to the emotion an image can bring to users. In this work, we try to quantify users' emotions reflected by images that they have uploaded and social emotion influence that plays a role in changing users' emotions.

In existing literature, several researches studied the problem of inferring emotions in social networks from users' attributes like posted blogs, locations and calling logs [1], [2]. These methods mainly considered the text information. A public report on the Facebook data suggests that images drive event engagement 100 times faster (e.g., clicking "like" or adding comment) than text. It is the same case for the individual emotions, as images stimulate emotions faster and thus image emotions are easier to affect people. The difference is that the expression is much more implicit compared to that using text. For example, in images, people often prefer to use warm colors like red or pink to express happiness and cold colors like blue or dark to express sadness. Recently, Machajdik and Hanbury [3] studied the problem of inferring emotions from images using visual features. Some work studied the emotions or visual preferences of Flickr images [4], [5]. However, these studies do not consider the interaction between users. In social networks, users' behaviors and opinions are strongly influenced by each other [6], [7]. In image social networks, this can be reflected as the emotion correlation between images. How to leverage the correlation to help understand users' emotions is a challenging question.

To clearly demonstrate the problem we are going to address in this paper, we give a general framework

- X. Wang was with Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, and is currently with School of Mechanical Engineering, University of Science and Technology Beijing, Beijing, 100083.  
E-mail: xiaohui0506@gmail.com
- J. Jia is with Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China, Key Laboratory of Pervasive Computing, Ministry of Education, Tsinghua National Laboratory for Information Science and Technology (TNList).  
E-mail: jjia@tsinghua.edu.cn
- J. Tang, B. Wu and L. Cai are with Department of Computer Science and Technology, Tsinghua University, Beijing, 100084.  
E-mail: jietang@tsinghua.edu.cn, stella.1991@163.com, clh-dcs@tsinghua.edu.cn
- L. Xie is with Research School of Computer Science at the Australian National University, Canberra, 00120C.  
E-mail: lexing.xie@anu.edu.au

1. <http://www.flickr.com/>  
2. <http://instagram.com/>

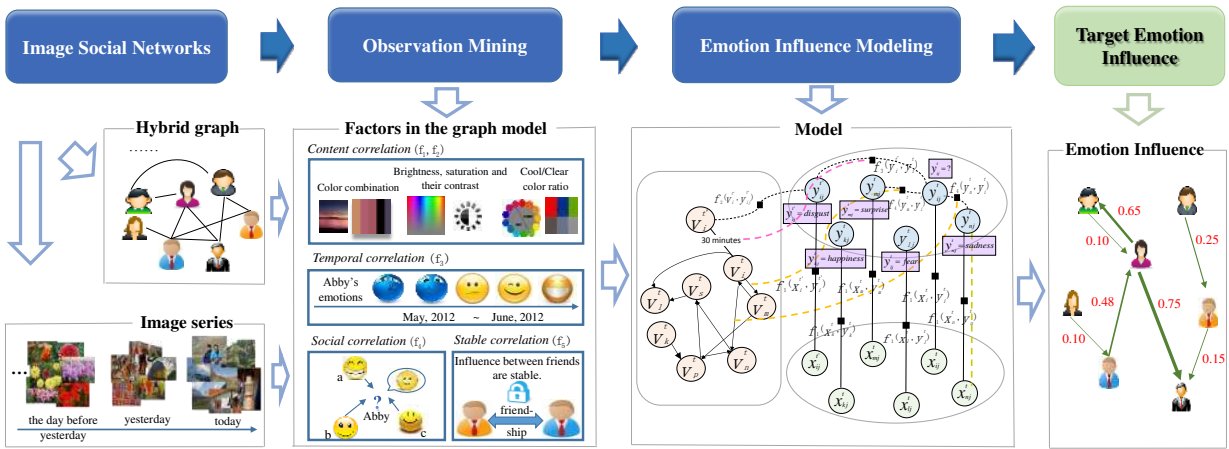


Fig. 1. Illustration of modeling emotion influence in image social networks.

in Figure 1. The input of our study is an image social network comprised of users and user relations. Each user is associated with a set of images that she/he uploaded to the social network. The expected output is the learned emotion influence between users in the network. The problem is non-trivial and has several challenges. First of all, it is unclear whether such an influence exists in image social networks. A more challenging question is how to identify the emotion influence patterns, as users' emotions are usually affected by various complex and subtle factors. Moreover, how to design a principled model to quantitatively describe the complex emotion influence among users and images?

In this paper, employing a networking data crawled from Flickr as our data source, we systematically study the problem of modeling emotion influence in image social networks. We first conduct a matched sampling test to verify the existence of influence in the image network. We further propose a probabilistic factor graph model to formalize the emotion influence learning problem. In particular, the model considers the following factors: (1) *content correlation*: how images' visual features reflect users' emotions; (2) *temporal correlation*: how a user's emotion is affected by her/his emotions in the recent past; (3) *social correlation*: how users' emotions influence (and are influenced by) their friends' emotions, and how stable the influence is. The learnt influence is fundamental for social network analysis and can be applied to many applications. We apply the learned influence to help predict users' emotions. Our experiments show that our model, by incorporating the emotion influence, can significantly improve the prediction accuracy (3.0%-26.2%) compared with several alternative methods such as Naive Bayesian, SVM (Support Vector Machine) or traditional Graph Model. We also demonstrate several interesting case studies on the behavior of the emotion influence model, which show that more social interactions correlate with higher

emotion influence between two users, and the influence of negative emotions is stronger than positive ones.

The rest of the paper is organized as follows: Section 2 gives related work. Section 3 proves the existence of emotion influence. Section 4 formally formulates the problem. Section 5 proposes the emotion influence model. Section 6 and Section 7 show the data observation and experimental results. Section 8 finally draws the conclusions.

## 2 RELATED WORK

**Existence validation of emotion influence.** The existence of emotion influence in image social networks is the basis of our study [8]. Fortunately, the phenomenon of emotion influence (a users' emotion can be affected by his friends) has also been reported by both psychology and computer science literature [9]. S. Hareli et al found that emotions of an individual influence the emotions, thoughts and behaviors of others and proposed that organizational dyads and groups inhabit emotion cycles [10]. Fowler et al. found that people's happiness depends on the happiness of others with whom they are connected [2]. Some affective prediction methods have considered the emotion influence among users in social networks [4]. Emotion influence has many applications. For example, J. Tang et al. found that using social emotion influence can improve emotion forecasting by 8% [1]. In this paper, we adopt Flickr data to validate the existence of emotion influence in image social networks.

**Text-based social influence analysis.** Previous researches studied the influence in social networks based on users' actions [11], [12], opinions [13], blogs and news articles [14], twitter [15]. However, they only use the text based information or actions (e.g. posting a photo). In image social networks, such as Flickr and Instagram, the text data or user actions are rather limited, and such methods can hardly work.

Besides the differences in media (text/image), none of the above literatures presents such a comprehensive optimization framework. For example, most of them do not consider time variation and just uses a static network [15]. None of them can model the dynamic influence in such an explicit way and give optimal possibilities for influence [1], [12], [15]. In contrast, our framework will consider the time variation in a dynamic image network, and model the influence explicitly. We even consider the change of influence over time, which makes our model more realistic.

The problem of predicting “observed emotions” from a history is analogous to the well-known problems of social action prediction or recommendation [16], no matter if the observed action or preference truly reflect users’ interests. So in this paper, we focus on measuring and predicting “observed emotional statement” (in text and images) rather than the “internal” emotion of users. The bridge between “observed emotional statement” and internal emotion state was covered in [10]. A more recent Facebook study [17] also confirms the relations between user emotions and their social network activities as well as the emotional influences in social networks. Their user studies serve as the user-centric foundation to our work. Hereafter, the user emotions all refer to the “observed emotions” – the emotions depicted in his uploading behaviors.

**Affective image classification.** Emotional image classification, also called affective image classification, is an important but hard problem [18], [19]. This is because emotions are highly subjective, and difficult to quantitatively measure [20]. Previous research focused on two crucial aspects for classification accuracy: training data and models. For the training data, the ground-truth emotions are usually manually labeled, therefore it is accurate but few in data number. Facing with the massive amounts of images in social networks, the above methods are powerless. Recently, some work has predicted emotions from images in social networks with images’ tags and comments [21], [22] or from users’ actions [1]. In this paper, we also adopt images’ tags and comments to obtain the emotion labels as the ground truth. For models, the commonly used methods are based on the machine learning such as SVM [19], Naive Bayesian [3], and probabilistic models [23]. Although these methods can achieve pretty good accuracies on affective image classification, they are difficult to incorporate different social influence factors. We try to integrate the traditional feature based image emotion prediction with social network information. The fusion of the two powerful and essentially different approaches makes it possible to further improve the performance.

**Factor graph model in social network analysis.** The structure of the graphic model is similar as that of social networks, so it is commonly used for social

analysis [24], [25]. A factor graph is a particular type of graph model that enables efficient computation of marginal distributions through the sum-product algorithm [26]. Tang et al. analyzed individual emotional states using the factor graph model [1]. In this paper, we adopt factor graph to build the emotion influence model, and the key problem is that which factors should be taken into account.

### 3 IMAGE-BASED EMOTION INFLUENCE STUDY

#### 3.1 Data Collection

We randomly download a data set from Flickr. The data set contains 4,725 users and their uploaded images (in total 1,254,640 images). The users are randomly chosen from active users of different user groups. Each user is associated with her/his shared images, contact list and personal information. Each image has timestamp, url, tags given by its owner and comments given by viewers. We define that user *A* is a friend of user *B* if user *A* has connection (follow action in Flickr) with user *B*. In the data set, on average, one user has 32.9 friends (one-way follow), and 18.6 among those friends also follow the user (two-way follow).

How to measure emotions is a key question in affective computing [27]. In our problem, the emotions are defined as six basic categories according to Ekman’s theory [28], *happiness*, *surprise*, *anger*, *disgust*, *fear* and *sadness*. Facing the vast scale of social images, manually labeling is powerless. Instead, we use tags and comments for automatic image labeling, which is the common method in previous work [29], [4], [30]. We use WordNet [31] and HowNet [32] dictionaries to obtain averagely more than 200 synonyms for each emotion category, and manually verify them. The numbers and examples of synonyms for all the emotion categories are shown in Table 1. For an image, we count the occurrences of each emotion synonym category in its tags and comments, and select the most frequent one (if exists) as the ground truth. Every image which has a significantly strong synonym category response (word count for the synonym category) in its tags or comments is labeled by this method. And we disregard images that do not have any emotion synonym or whose best category response is not significantly stronger than the second. This way further improves the labelling accuracy.

This has left us a labeled data set containing 354,192 images uploaded by 4066 users. The network status of the labeled data set is: one user have 33.4 friends (one-way follow) on average, and 18.8 among them also follow the user (two-way follow), which is consistent with the network status of the whole data set. So we use the labeled data set for data observation and experiments. We directly model the emotions reflected in the Flickr images. The correlation between these

TABLE 1  
Examples and number of synonyms per emotion category.

| Category  | #   | Synonym examples   |
|-----------|-----|--|
| Happiness | 177 | pleasure, charming, joyous, brilliant, glad, delight, sweet, enjoyable, satisfying |
| Surprise  | 39  | shock, startle, surprising, amazed, astonish, astonished                           |
| Anger     | 442 | annoyance, fury, ire, irritation, rage, wrath, storming, aggravate                 |
| Disgust   | 241 | nauseate, offend, repel, revolt, sicken, nasty                                     |
| Fear      | 102 | horror, terror, frightful, dread, scare, dismay, direful                           |
| Sadness   | 432 | depressed, unhappy, miserable, discourage, moody, cried, displeasing               |

images and the users' true emotions is beyond the scope of this study.

### 3.2 Sampling Test

We first need to validate the existence of emotion influence in image networks. The basic validation method is the sampling test [33]. The users are divided into two groups: the friend-related group  $G_R$  and the friend-independent group  $G_I$ . Taking the "happy" emotion as an example, a user is said to be "happy" if most of her/his uploaded images are labeled "happy" at a specified time.  $G_R$  contains users who uploaded images at time  $t$  and has one or more friends with the "happy" emotion at time  $t - \Delta t$ , while  $G_I$  contains users who uploaded images at time  $t$  and has no friend with the "happy" emotion at time  $t - \Delta t$ . Finally, we compare the "happy" ratio of users at time  $t$ . The "happy" ratio is defined as the ratio of "happy" users in each group, e.g.

$$Ratio = \frac{\#happy\_users}{\#all\_users} \quad (1)$$

**Experimental setting.** For each group, 50 users are randomly chosen from our data set. We set  $\Delta t = 1, 2, 3, 4$  weeks and  $t$  is randomly chosen from March, 2006 to June, 2012 without overlap. The time interval we choose here is one week. As in our experiment, we mainly model the long-term emotions, just as in [1] (over days) and [34] (over weeks). We repeat the experiment 10 times to calculate the average ratio. Intuitively, the more friends have the same emotion, the greater is their influence. To investigate the effect of the number of friends, we further divide the friend-related group  $G_R$  into two subgroups: subgroup with one or two friends and subgroup with at least three friends (with the "happy" emotion at time  $t - \Delta t$ ).

**Results.** Figure 2 shows the results. The average ratio of the friend-related group  $G_R$  is higher than that of the friend-independent group  $G_I$ , which confirms the existence of friend influence through images. The results also indicate more friends with the same emotion usually have bigger impact. And the downward trend

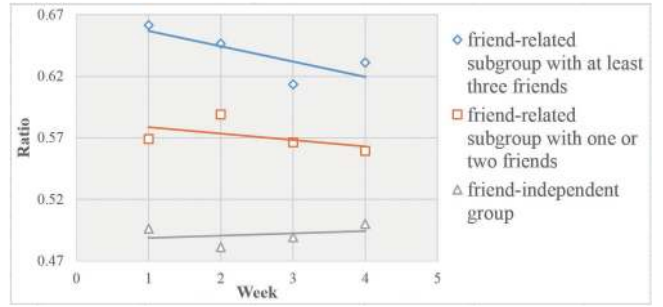


Fig. 2. The result of sampling test for image-based emotion influence.

shows that the influence becomes weakened as time passed, which will be learned and incorporated in our model. Note that although the downward trend is very obvious, it may be not as significant as one may expect. Actually as we know, emotion influence between friends, especially in social networks, cannot be as strong as events in our real life. So the slope here is meaningful and acceptable.

## 4 FORMULATION AND SETTINGS

Our goal is to derive the emotion influence in social networks from analyzing the emotional states from users' uploaded images. We have the following intuitions:

- 1) Users' emotions are expressed by their uploaded images in image-based social networks. Note that here we refer to the "observed emotion" introduced in Section 2.
- 2) Users' emotions at time  $t$  are influenced by their friends' emotions at the recent past before  $t$ .

With the above intuition, we present the problem formulation of modeling the emotion influence from images in social networks. The social network can be defined as a graph  $G = (V, E)$ , where  $V$  is the set of  $|V| = N$  users,  $E \subset V \times V$  is the set of relationships among users, e.g. friendships. Notation  $e_{ij} \in E$  indicate user  $v_i$  and  $v_j$  are friends in the social network (In unsymmetrical network such as Flickr,  $e_{ij}$  means user  $v_j$  follows user  $v_i$ ). In this work, we transform continuous time to discrete time slices and thus we use "at time  $t$ " to refer to the  $t$ -th time slice. We use  $E^t$  to denote relationship information at time  $t$ .

A user  $v_i$  uploaded a set of images at time  $t$ , denoted as  $\mathbf{X}_i^t$ .  $\mathbf{x}_{ij}^t$  is an element of  $\mathbf{X}_i^t$  indicating the  $j$ th image uploaded by user  $v_i$  at time  $t$ , which is instantiated as its visual features. Table 2 summarizes the notations used throughout this paper.

In our problem, the emotion space  $\mathcal{C} = \{c_p\}$  is defined as six most common emotional states, *happiness*, *surprise*, *anger*, *disgust*, *fear* and *sadness* [28].

Without loss of generality, we treat it as a combination of several binary classification problems. Given

an emotion  $c_p \in \mathcal{C}$ , for each image  $\mathbf{x}_{ik}^t$  we have a binary variable  $y_{ik}^t \in \{-1, 1\}$  indicating whether the image has the emotion  $c_p$ , and for each user  $v_i$  we have a binary variable  $y_i^t \in \{-1, 1\}$  to denote whether the user has the emotion  $c_p$  at time  $t$ . We can train a model for each emotion  $c_p \in \mathcal{C}$ , and combine them to infer the emotions. Our approach to handle each emotion separately is flexible. When all models suggest negative values for an image/a user at some time, it means that the image/user has a neutral emotion at the time. When multiple models suggest positive values, we can either select the emotion with the highest probability to be the major emotion of the image/user at the time, or output all the positive emotions, based on the requirement. This flexibility allows modeling multiply emotions at the same time, as emotion is rather complex and sometimes we may have more than one emotion at the same time. We now formally define the time-varying social network and emotion influence, and formulate the leaning task.

**Definition 1. Time-varying social network.** The time-varying social network is denoted as  $G = (V, \{E^t\}, \{\mathbf{X}_i^t\})$ , where  $V$  is the set of users,  $t, i$  iterates over all time slices and users,  $e_{ij}^t \in E^t$  denotes user  $v_i$  and user  $v_j$  are friends at time  $t$ ,  $\mathbf{X}_i^t$  denotes images uploaded by user  $v_i$  at  $t$ .

**Definition 2. Emotion influence.** Social influence from user  $v_i$  to user  $v_j$  at time  $t$  is denoted as  $\mu_{ij}^t$ . In our implementation, we treat it as a binary variable  $\{0, 1\}$ , and use the inferred possibility as the output weight, indicating the extent to which user emotion  $y_j^t$  is influenced by user emotion  $y_i^t$ , where  $t' < t$ .

**Learning task:** Given a time-varying social network  $G$ , let  $\mathbf{Q} = (\{y_i^t\}, \{y_{ik}^t\}, \{\mu_{ij}^t\})$  represents the collection of all the variables, including user/image emotion variables and social influence variables. And the target is to find a function for predicting emotions of all the unlabeled images, users' emotions and influence in different time slices:

$$f : (G, \text{labeled data}) \rightarrow \mathbf{Q} \quad (2)$$

## 5 EMOTION INFLUENCE MODEL

We propose a factor graph model to infer emotion influence from images in social networks (Figure 1). We consider the following aspects: (1) *content correlation*: a user's emotion is induced by visual features of their uploaded images; (2) *temporal correlation*: the user's current emotion has correlations with her/his emotion in the recent past; (3) *social correlation*: the user's emotion may be largely affected by his friends, and the influence relationship does not change frequently within a short time. By leveraging these aspects, we formulate the emotion influence process as a dynamic factor graph model. All the correlations

TABLE 2  
Notations

| Symbol              | description  |
|---------------------|--|
| $G$                 | the whole network containing uploaded image and connection information               |
| $V$                 | the set of users in the social network   |
| $E^t$               | the set of edges at time $t$   |
| $N$                 | number of users  |
| $\mathcal{T}$       | the collections of all the successive time slices, namely 1,2,3...                   |
| $\mathbf{x}_{ij}^t$ | the vector of visual features of the $j$ th image uploaded by user $v_i$ at time $t$ |
| $\mathbf{X}_i^t$    | the set of images uploaded by user $v_i$ at time $t$                                 |
| $NB^t(v_i)$         | the set of $v_i$ 's friends at time $t$  |
| $y_{ij}^t$          | binary variable indicating whether image $\mathbf{x}_{ij}^t$ has a specified emotion |
| $y_i^t$             | binary variable indicating the emotion of user $v_i$ at time $t$                     |
| $\mu_{ij}^t$        | binary variable indicating the influence of user $v_i$ on $v_j$ at time $t$          |

are expressed as correlation functions. A correlation function is actually a factor function in the graph model, mapping values of some random variables to non-negative real numbers. In our model, we choose exponential functions as our correlation functions for the convenience of calculations.

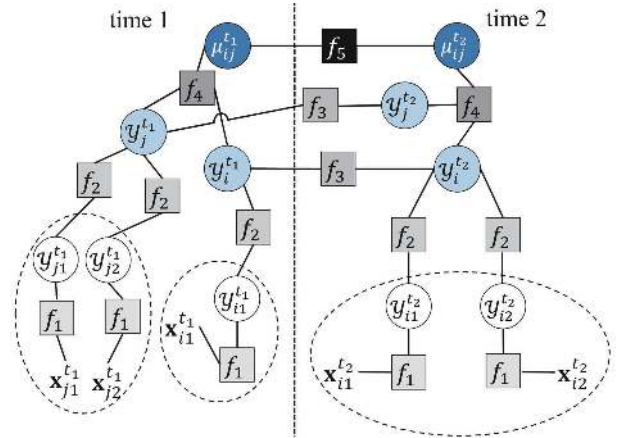


Fig. 3. Graphical representation of the emotion influence model.

### 5.1 The Predictive Model

The graphical representation of the proposed model is illustrated in Figure 3. In the model, the above aspects are quantized as different types of factor functions:

- **Content correlation function.** It contains two parts, visual feature factor  $f_1(\mathbf{x}_{ij}^t, y_{ij}^t)$ , representing the correlation between image emotion  $y_{ij}^t$  (see Section 4,  $y_{ij}^t \in \{-1, 1\}$  indicating whether an image has the emotion or not) and its corresponding visual feature  $\mathbf{x}_{ij}^t$ . The other part is image induction factor  $f_2(y_{ij}^t, y_i^t)$ , denoting the emotions of uploaded images reflect their owners' emotions.

- **Temporal correlation function**  $f_3(y_i^{t'}, y_i^t), t' < t$ . It represents the influence of user  $v_i$ 's emotions in the recent past  $t'$  on her/his current emotion at time  $t$ .
- **Social correlation function**  $f_4(y_i^t, y_j^t, \mu_{ij}^t)$ . It encodes the dependence of user  $v_i, v_j$ 's emotion at time  $t$ , and their influence variable  $\mu_{ij}^t$ . To better constrain the influence, we assume that the influence between two users is stable over a short time, encoded by stable correlation factor  $f_5(\mu_{ij}^{t'}, \mu_{ij}^t), t' < t$ .

Visual feature factor  $f_1$  is instantiated as exponential-linear function:

$$f_1(\mathbf{x}_{ij}^t, y_{ij}^t) = \exp\{\alpha^T \cdot \mathbf{x}_{ij}^t y_{ij}^t\} \quad (3)$$

where  $y_{ij}^t \in \{-1, 1\}$  is the binary emotion indicator.  $\mathbf{x}_{ij}^t$  is the visual feature,  $\alpha$  is the parameter vector for all the images. Note that we add a dimension with constant "1" in the feature vector  $\mathbf{x}_{ij}^t$ , and  $\alpha^T \cdot \mathbf{x}_{ij}^t$  is actually a linear classification.  $f_1$  can be thought of a weak predictor, working directly on images. We adopt the interpretable aesthetic features [35] as our visual features, as summarized in Table 3. These features contain color features, figure-ground relationship, shape and composition, which are all emotion related. For example, the well known five-color combination [36] is included, which has direct impact on image emotions [37].

The basic assumption in our study is that a user's emotion can be reflected by emotions of the images she/he uploaded around that time. Thus we could define the following image induction factor function:

$$f_2(y_{ij}^t, y_i^t) = \exp\{-\beta_i |y_{ij}^t - y_i^t|\} \quad (4)$$

Usually, a user's emotion does not change rapidly and her/his current emotion is highly dependent on her/his emotions in the recent past. So the temporal correlation factor function is used to model this phenomenon:

$$f_3(y_i^{t'}, y_i^t) = \exp\{-\xi_i \cdot h(t, t') \cdot |y_i^{t'} - y_i^t|\} \quad (5)$$

where  $h(t, t') = e^{-\delta \cdot (t-t')}$  describes the exponential decrease of influence over time,  $\delta$  is a manually defined parameter, and  $\xi_i$  is the per-user weight parameter.

As validated in data observation in Section 6, the emotion can be affected by his friends. And the social correlation factor function is defined to model this effect:

$$f_4(y_i^t, y_j^t, \mu_{ij}^t) = \exp\{-\lambda_{ij} |2(1 - \mu_{ij}^t) - |y_i^t - y_j^t||\} \quad (6)$$

where  $\mu_{ij}^t \in \{0, 1\}$  is a binary variable, representing at time  $t$ , whether user  $v_i$  has influence on user  $v_j$ .  $\lambda_{ij}$  is the per user pair weight parameter.

For stable correlation factor function, we assume that the social influence is stable, which means that if a friend has a strong influence on you before, the impact

is likely to be also strong afterwards. This constraint is defined as follows:

$$f_5(\mu_{ij}^{t'}, \mu_{ij}^t) = \exp\{-\eta_{ij} g(t, t') |\mu_{ij}^{t'} - \mu_{ij}^t|\} \quad (7)$$

where  $g(t, t') = e^{-\tau \cdot (t-t')}$  with a manually defined parameter  $\tau$  indicates the decay effect over time, and  $\eta_{ij}$  is the per user pair weight parameter.

As introduced in Section 4, there are several types of variables in the graph factors, e.g.  $\{y_i^t\}, \{y_{ij}^t\}, \{u_{ij}^t\}$ , and we pack them into one variable vector  $\mathbf{Q}$  to facilitate explanations. There are quite a lot of free parameters in the formulation, including  $\alpha, \{\beta_i\}, \{\xi_i\}, \{\lambda_{ij}\}, \{\mu_{ij}\}$ , where  $\xi_i$  iterate over users, and  $\lambda_{ij}, \mu_{ij}$  iterate over user pairs. To avoid possible over-fittings, we transform these user pair variables to one user variables. That is, we assume  $\lambda_{ij} = \lambda_i, \eta_{ij} = \eta_i, \forall j \in V$ . We denote  $\theta = (\alpha, \{\beta_i\}, \{\xi_i\}, \{\lambda_i\}, \{\eta_i\})$ .

Given the definitions of the above factor functions, we define the probability of  $P(\mathbf{Q}|G, \theta)$  as follows:

$$\begin{aligned} P(\mathbf{Q}|G, \theta) = & \frac{1}{Z} \times \prod_{v_i \in V, t \in \mathcal{T}} \prod_{\mathbf{x}_{ij}^t \in \mathbf{X}_i^t} f_1(\mathbf{x}_{ij}^t, y_{ij}^t) \times \prod_{v_i \in V, t \in \mathcal{T}} \prod_{\mathbf{x}_{ij}^t \in \mathbf{X}_i^t} f_2(y_{ij}^t, y_i^t) \\ & \times \prod_{v_i \in V, t, t' \in \mathcal{T}} f_3(y_i^{t'}, y_i^t) \times \prod_{v_i \in V, t \in \mathcal{T}} \prod_{e_{ij}^t \in E^t} f_4(y_i^t, y_j^t, \mu_{ij}^t) \\ & \times \prod_{v_i \in V, t, t' \in \mathcal{T}} \prod_{e_{ij}^{t'} \in E^{t'}, e_{ij}^t \in E^t} f_5(\mu_{ij}^{t'}, \mu_{ij}^t) \end{aligned} \quad (8)$$

where  $Z$  is a normalization term. We then define the objective function as

$$\mathcal{O}(\mathbf{Q}, \theta) = \log P(\mathbf{Q}|G, \theta) \quad (9)$$

Challenges of the proposed emotion influence model are mainly two-fold. The first challenge is how to model the complex interdependency and correlation among several factors, e.g., user emotions are closely related to images emotions, and the emotions of users will influence one another and spread in a time-varying social network. The second challenge is due to the unobservability of latent factors. According to our definition in Section 4, both user emotion and emotion influence are unobservable. Therefore, it is not straightforward for model learning and inference.

## 5.2 Model Learning and Inference

Our goal is to estimate an optimal parameter  $\theta$  and obtain the appropriate variables  $\mathbf{Q}$  from the input graph to maximize the objective function  $\mathcal{O}(\mathbf{Q}, \theta)$  defined in Eq. 9, i.e.,

$$\mathbf{Q}^*, \theta^* = \arg \max_{\mathbf{Q}, \theta} \mathcal{O}(\mathbf{Q}, \theta) \quad (10)$$

TABLE 3  
Summary of visual features. The column ‘#’ represents the dimension of each feature.

| Name                                      | #  | Short description   |
|---|----|---|
| Five-color combinations                   | 15 | five dominant colors in the HSV color space   |
| Brightness, saturation and their contrast | 4  | mean brightness, mean saturation and their average contrast   |
| Cool color ratio                          | 1  | ratio of color colors. Colors can be divided into cool colors with hue ([0,360]) in the HSV space between 30 and 110 and warm colors. |
| Clear color ratio                         | 1  | ratio of clear colors with brightness ([0,1]) greater than 0.7. The colors whose brightness less than 0.7 are dull colors.            |

---

**Algorithm 1: The learning and inference algorithm**


---

**Input:**  $G$ , learning rate  $\epsilon$   
**Output:**  $\theta = (\alpha, \{\beta_i\}, \{\xi_i\}, \{\lambda_i\}, \{\eta_i\})$ .  
 $\mathbf{Q} = (\{y_i^t\}, \{y_{ij}^t\}, \{u_{ij}^t\})$   
Initialize parameters randomly  
**repeat**  
  %Use max-product belief propagation to update  $\mathbf{Q}$   
   $\mathbf{Q} \leftarrow \arg \max_{\mathbf{Q}} \mathcal{O}(\mathbf{Q}, \theta)$   
  %Update parameter  $\theta$   
  **repeat**  
    Use Gibbs Sampling to obtain a sample collection of  $\mathbf{Q}$ , according to  $\theta$ , e.g.  $\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_C$ , and let  
     $E_{\mathbf{Q}|\theta}(\Phi(\mathbf{Q})) \approx \frac{1}{C} \sum_{i=1}^C \Phi(\mathbf{Q}_i)$   
    Calculate  $\frac{\partial \mathcal{O}}{\partial \theta} = \Phi(\mathbf{Q}) - E_{\mathbf{Q}|\theta}(\Phi(\mathbf{Q}))$   
    Update  $\theta \leftarrow \theta + \epsilon \frac{\partial \mathcal{O}}{\partial \alpha}$   
  **until** convergence;  
**until** convergence;  
%Optional  
Run general belief propagation to get marginal distributions for individual variable elements of  $\mathbf{Q}$

---

distribution calculation can be thought of the “inference” counterpart.

Given a fixed set of model parameters, we find the optimal latent variables which maximize the target objective function, and for a given set of latent variables, we optimize the model parameters for maximization. The two steps are iterated until convergence.

The main challenges lie in the second step. First, the irregular connections of latent variables  $\{y_{ij}^t\}$  and  $\{\mu_{ij}^t\}$  greatly increase the objective function’s complexity. Second, the normalization term  $Z$  is hard to compute as it needs to sum over the exponential combinations of variable assignments. These all make the objective function’s gradients with respect to model parameters very hard to calculate, preventing us to easily employ a gradient based algorithm. To address the problems, we utilize the graph structure and use a Gibbs sampling based approach to approximate the gradient. Note that some variables ( $y_{ij}^t$ ) are already labeled, and treated as constants during optimization.

It is clear that by introducing some variable function  $\Phi(\mathbf{Q})$ , the objective function (Eq. 11) can be rewritten as  $\mathcal{O} = \frac{1}{Z} e^{\theta^T \Phi(\mathbf{Q})}$ . To make it easier to follow, we list the components of  $\Phi(\mathbf{Q})$  as follows:

In this subsection, we present a variant of the hard EM algorithm to address the optimization. For clarity, we rewrite the objective function as

$$\begin{aligned}
\mathcal{O} = & \sum_{v_i \in V, t \in \mathcal{T}} \sum_{\mathbf{x}_{ij}^t \in \mathbf{X}_i^t} \alpha^T \cdot \mathbf{x}_{ij}^t y_{ij}^t + \sum_{v_i \in V, t \in \mathcal{T}} \sum_{\mathbf{x}_{ij}^t \in \mathbf{X}_i^t} -\beta_i |y_i^t - y_{ij}^t| \\
& + \sum_{v_i \in V, t, t' \in \mathcal{T}} -\xi_i e^{-\delta \cdot |t' - t|} |y_i^t - y_i^{t'}| \\
& + \sum_{v_i \in V, t \in \mathcal{T}} \sum_{e_{ij}^t \in E^t} -\lambda_{ij} |1 - \mu_{ij}^t - |y_i^t - y_j^t|| \\
& + \sum_{v_i \in V, t, t' \in \mathcal{T}} \sum_{e_{ij}^{t'} \in E^{t'}, e_{ij}^t \in E^t} -\eta_{ij} e^{-\tau \cdot |t' - t|} |\mu_{ij}^t - \mu_{ij}^{t'}| - \log Z
\end{aligned} \tag{11}$$

The general idea can be classified to the hard Expectation Maximization (EM) algorithm, and is summarized in Algorithm 1. Note that in our algorithm, there are no explicit separate learning and inference. Instead, the parameters are tightly correlated to the data (e.g. per user parameters). The optional marginal

| Component in $\theta$ | Corresponding part in $\Phi(\mathbf{Q})$  |
|-----------------------|---|
| $\alpha$              | $\sum_{v_i \in V, t \in \mathcal{T}} \sum_{\mathbf{x}_{ij}^t \in \mathbf{X}_i^t} \mathbf{x}_{ij}^t y_{ij}^t$                          |
| $\beta_i$             | $\sum_{t \in \mathcal{T}} \sum_{\mathbf{x}_{ij}^t \in \mathbf{X}_i^t} - y_i^t - y_{ij}^t $  |
| $\xi_i$               | $\sum_{t, t' \in \mathcal{T}} -e^{-\delta \cdot  t' - t }  y_i^t - y_i^{t'} $   |
| $\lambda_i$           | $\sum_{t \in \mathcal{T}} \sum_{e_{ij}^t \in E^t} - 1 - \mu_{ij}^t -  y_i^t - y_j^t  $  |
| $\eta_i$              | $\sum_{t, t' \in \mathcal{T}} \sum_{e_{ij}^{t'} \in E^{t'}, e_{ij}^t \in E^t} -e^{-\tau \cdot  t' - t }  \mu_{ij}^t - \mu_{ij}^{t'} $ |

**Update  $\mathbf{Q}$ .** This step is to obtain an optimal  $\mathbf{Q}$  given a fixed set of parameter  $\theta$ , e.g.  $\mathbf{Q}^* = \arg \max_{\mathbf{Q}} \mathcal{O}(\mathbf{Q}, \theta)$ . This is actually the inference on a given graph model, and we use a standard max-product belief propagation algorithm [38] to infer the best value given the current parameter configuration.

**Update  $\theta$ .** For our objective form  $\mathcal{O} = \frac{1}{Z} e^{\theta^T \Phi(\mathbf{Q})}$  the derivative is  $\frac{\partial \mathcal{O}}{\partial \theta} = \Phi(\mathbf{Q}) - E_{\mathbf{Q}|\theta}(\Phi(\mathbf{Q}))$  [39].

The derivative obviously corresponds to the original variables, e.g.

$$\frac{\partial \mathcal{O}}{\partial \alpha} = \sum_{v_i \in V, t \in \mathcal{T}} \sum_{\mathbf{x}_{ij}^t \in \mathbf{X}_i^t} \mathbf{x}_{ij}^t y_{ij}^t - E\left(\sum_{v_i \in V, t \in \mathcal{T}} \sum_{\mathbf{x}_{ij}^t \in \mathbf{X}_i^t} \mathbf{x}_{ij}^t y_{ij}^t \mid \alpha, \beta_i, \xi_i, \lambda_i, \eta_i\right) \quad (12)$$

The above shows that in order to obtain the gradient,  $E(\Phi(\mathbf{Q}))$  is also required, which is however intractable. We find that Gibbs Sampling is a great tool for our problem. The basic idea is, given  $\theta$ , we use the Gibbs Sampling to obtain a collection of  $\mathbf{Q}$ . The samples are used to obtain a collection of  $\Phi(\mathbf{Q})$ , which are further used to give an estimate of  $E(\Phi(\mathbf{Q}))$ . The Gibbs Sampling process works on a sequence of variables. We denote  $\mathbf{Q} = \{q_1, q_2, \dots, q_n\}$ . We first pick some initial value  $\mathbf{Q}^{(0)}$ . For each iteration  $i > 0$ , we sample variable  $q_j^{(i)}$  from the conditional distribution

$$\hat{p} = p(q_j \mid q_1^{(i)}, \dots, q_{j-1}^{(i)}, q_{j+1}^{(i-1)}, \dots, q_n^{(i-1)}) \quad (13)$$

And the only remaining question is that how to calculate  $\hat{p}$ . To simplify the description, we again rewrite our objective function as  $\mathcal{O}(\mathbf{Q}, \theta) = \frac{\sum_i \theta_i \Phi_i(\mathbf{Q}_i)}{Z}$ , where each  $\mathbf{Q}_i \subset \mathbf{Q}$  is the elements of a factor node. The problem is equivalent to the calculation of

$$p(q \mid \mathbf{Q} - q, \theta) = \frac{e^{\sum_{i \in I} \theta_i \Phi_i(\mathbf{Q}_i)} \cdot e^{\sum_{i \notin I} \theta_i \Phi_i(\mathbf{Q}_i)}}{Z \cdot p((\mathbf{Q}) - q \mid \theta)} \quad (14)$$

where  $I$  is the index set for factors containing  $q$ . Note that our variables are all binary variables, and we have

$$W = \frac{p(q=0 \mid \mathbf{Q} - q, \theta)}{p(q=1 \mid \mathbf{Q} - q, \theta)} = \frac{e^{\sum_{i \in I} \theta_i \Phi_i(q=0, \mathbf{Q}_i - \{q\})}}{e^{\sum_{i \in I} \theta_i \Phi_i(q=1, \mathbf{Q}_i - \{q\})}} \quad (15)$$

$$= e^{\sum_{i \in I} \theta_i (\Phi_i(q=0, \mathbf{Q}_i - \{q\}) - \Phi_i(q=1, \mathbf{Q}_i - \{q\}))}$$

Finally we have  $p(q=1 \mid \mathbf{Q} - q, \theta) = \frac{1}{W+1}$ . Note that in our sparse graph,  $I$  is usually a small set, and each update of the sampling process only involves a small set of variables.

Here we can get the gradient with respect to parameters  $\theta$ , and then we use a standard gradient method to obtain an optimal  $\theta$ .

## 6 DATA OBSERVATION

In this section, we evaluate the rationality of the model factors by statistical data observations on our labeled data set of 354,192 images uploaded by 4066 users.

### 6.1 Content Correlation

Factor  $f_1$  is determined by visual features of images. We test whether there exist explicit correlations between the visual features and emotions by Canonical Correlation Analysis (CCA). CCA is a statistical approach suitable for multidimensional data [40], which can explicitly show that how related two random vectors are, to the maximal possible extent compared with simple correlation methods like Pearson analysis. To comply with the requirements of CCA, emotions should be quantitatively described. Here we adopt the image-scale space, composed of two dimensions *warm-cool* and *hard-soft* [36], to represent emotions. Each labeled image in our data set is assigned by the image scale according to its tags and comments using the method in [41]. We have two sets of variables,  $X = \{x_1, x_2, \dots, x_{21}\}$  representing visual features and  $Y = \{wc, hs\}$  indicating the image scale. CCA will find linear combinations of the  $x$ 's and the  $y$ 's that have maximum correlation with each other. The final result on our data set shows the maximum correlation coefficient 0.74, indicating that there exists considerable correlation between the visual features and emotions.

### 6.2 Temporal Correlation

To better illustrate the temporal factor  $f_3$ , we track the users' emotion over time. Figure 4(a) shows the temporal correlation rate  $Rate_T$  of randomly selected 2500 users ( $n = 2500$ ) during a month ( $\delta = 1, 2, \dots, 29$ ).  $Rate_T$  is the average rate of users with the same emotion between time  $t$  and  $t + \delta$ , defined as:

$$Rate_T = \frac{1}{n} \sum_{i=1}^n \frac{\#Same\_emotion(v_i^t, v_i^{t+\delta})}{\#\{t\}} \quad (16)$$

$\#\{t\}$  is number of tested  $t$  and  $t + \delta$  pairs for user  $v_i$  (user  $v_i$  uploaded images at the two time points). The overall tendency is downward, showing that the emotion similarity drops for longer time intervals. This confirms the correlation between a user's current emotion and her/his emotions in the recent past.

### 6.3 Social Correlation

To better examine the factors  $f_4$  and  $f_5$ , we generalize the statistical view of the social correlation in Section 3.2, by counting general user pairs and test whether they have the same emotions. This test further tell us whether one's emotions could influence her/his friends. We randomly choose  $n = 1000$  users and their uploaded images. Figure 4(b) shows the similarity rate  $Rate_I$  for friends and non-friends, defined as:

$$Rate_I = \frac{1}{n} \sum_{i=1}^n \frac{\#Same\_emotion(v_i^t, NB(v_i)^{t+\delta})}{\#NB(v_i)} \quad (17)$$



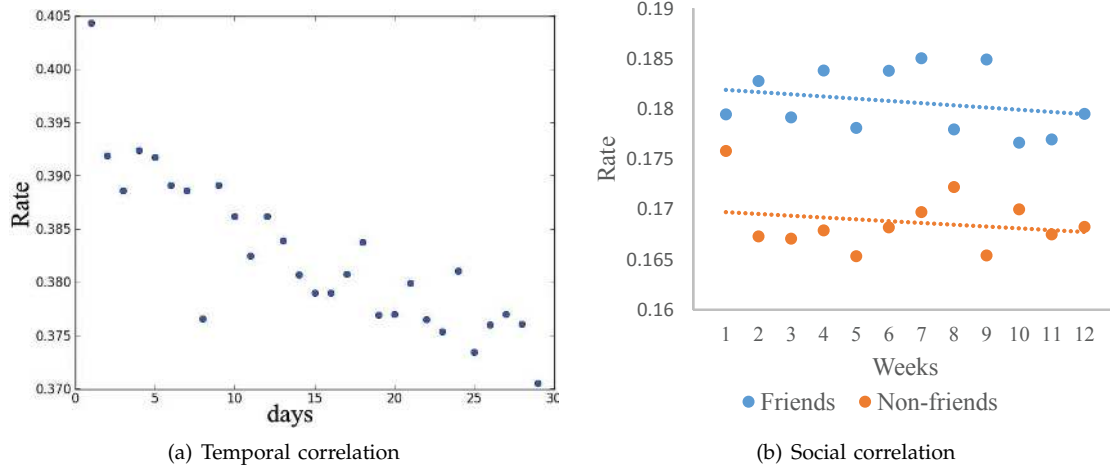


Fig. 4. Data observation results.

TABLE 4  
Number of images per emotion category.

| Happiness | Surprise | Anger  | Disgust | Fear   | Sadness |
|-----------|----------|--------|---------|--------|---------|
| 166,482   | 35,735   | 28,798 | 15,645  | 57,321 | 50,211  |

where  $Same\_emotion(v_i^t, NB(v_i)^{t+\delta})$  is a function counting the number of users of group  $NB(v_i)$  at time  $t + \delta$  with the same emotion as user  $v_i$  at time  $t$ . Note that we do not require members of  $NB(v_i)$  to have an emotion at  $t + \delta$ .  $\delta$  is the time interval, ranging from 1 to 12 weeks. For test of friends,  $NB(v_i)$  is defined as friends of  $v_i$ , and for test of non-friends,  $NB(v_i)$  is chosen randomly. The result shows the influence rate of friends is greater than that of non-friends, indicating significant emotion influence among friends.

## 7 EXPERIMENTS AND RESULTS

In this section, we evaluate the emotion prediction accuracy of images using the proposed model and then analyze how social factors help improve the inferring performance. To make the optimization easier, we use a subnet containing 954 users and 50,000 labeled images. For 40% of the labeled images, we ignore their labels during model calculation, and their labels are only used for testing. Finally, we give a qualitative case study to further demonstrate the effectiveness of the method.

### 7.1 Model Performance

We compare the proposed model with alternative baseline methods for emotion prediction of images. The data configuration for the baseline algorithms is the same as our model, e.g. 60% for training and 40% for testing. The three baseline methods are as follows:

- **Naive Bayesian (NB).** The method directly uses the same visual features as our model to train

a classifier. This method is used for traditional affective image classification and achieves good performance [3].

- **Support Vector Machine (SVM).** SVM directly uses the visual features as our model to train a predictive model. This method is adopted for content-based affective image classification [19].
- **Graph Model (GM).** The method is proposed for learning to infer emotions from images in social networks [4]. We combine their factors of social correlations and the visual features used in our model.

We evaluate the performance by Accuracy. Accuracy is the proportion of true results (both true positives and true negatives), where true positive is the number of positive data correctly predicted as positive; and true negative is the number of negative data correctly predicted as negative.

Table 5 shows the comparison performance results of our model and the above baseline methods. For all the six emotion categories, the average accuracy of the proposed model achieves 75.00%. Using the same features, we find that the Naive Bayesian method is totally invalid. The proposed model can achieve better performance with an average 26.2% accuracy improvement. For SVM and Graph Model, our model can achieve an average 3.5% improvement on Accuracy than SVM and 3.0% improvement than Graph Model.

Specially, the performances of our model and Graph Model show a marked advance when classifying images with negative emotions. For example, in case of anger category, the accuracy of our model is 4.1% higher than SVM. It can be seen from Table 4 that the negative images are relatively rare. Besides, the visual features of those images are very similar. Let's take anger and fear as an example. Anger images are dull and cool, while fear images are dull and cool as well. So the classification of negative images

TABLE 5  
Performance comparison between our model and baseline methods (%).

| Emotional categories | Accuracy     |              |              |              |
|----------------------|--------------|--------------|--------------|--------------|
|                      | NB           | SVM          | GM           | Our Model    |
| Happiness            | 55.35        | 63.74        | 56.90        | 61.40        |
| Surprise             | 42.55        | 72.80        | 79.67        | 77.78        |
| Anger                | 43.39        | 73.09        | 70.80        | 77.19        |
| Disgust              | 45.34        | 78.94        | 75.52        | 83.63        |
| Fear                 | 50.08        | 69.29        | 79.35        | 73.98        |
| Sadness              | 56.17        | 71.05        | 68.76        | 76.02        |
| <b>Average</b>       | <b>48.81</b> | <b>71.49</b> | <b>72.04</b> | <b>75.00</b> |

is rather challenging. Our model and Graph Model show a noticeable improvement in classifying negative images, indicating the effectiveness of temporal and social correlations.

Furthermore, our model achieves higher performance on average than Graph Model. The social correlations of Graph Model are only the uploaded time and owner of an image, which our model captures and leverages more, such as temporal correlation ( $f_3$ ), social correlation ( $f_4$ ) and stable correlation ( $f_5$ ). These factors have made contributions to improve prediction performance.

However, when classifying images labeled with happiness, the performance of our method is marginally lower than SVM. We assume that this is because the happy images hold a rather large proportion of training samples. So it is natural that the SVM method turns out better. However, when it comes negative emotion categories with few training samples, the performance of SVM hurts obviously. So our model is capable of handling the unbalance of the number of different emotional images.

## 7.2 Factor Contribution Analysis

In the predictive model, we incorporate content, temporal and social correlation factors. Since the content correlation are the basic factors for inferring image emotions, here we investigate how temporal correlation ( $f_3$ ), social correlation ( $f_4$ ) and stable correlation ( $f_5$ ) help improve the performance. We test the contribution for each factor function by removing it from the model in turn and compare the prediction performance.

Table 6 shows the results of factor contribution analysis. For most emotion categories, we can see the descending on Accuracy when removing each of the three factors. Previous researches have revealed that social information plays an important role in text-based prediction task in social networks [14], [8]. Our results further verify that the social information can also help improve the prediction performance of image-based tasks. Among the three factors, stable correlation ( $f_5$ ) has the most contribution, which further indicates the users' emotion influence is not

TABLE 6  
Contribution of different factor functions by performance comparison of emotion prediction (%). Model- $f_3$  represents removing temporal correlation factor, Model- $f_4$  represents removing social correlation factor, Model- $f_5$  represents removing stable correlation factor.

| Emotional categories | Accuracy     |              |              |              |
|----------------------|--------------|--------------|--------------|--------------|
|                      | Model        | Model- $f_3$ | Model- $f_4$ | Model- $f_5$ |
| Happiness            | 61.40        | 59.65        | 60.23        | 59.94        |
| Surprise             | 77.78        | 77.49        | 75.73        | 75.73        |
| Anger                | 77.19        | 76.02        | 75.73        | 75.73        |
| Disgust              | 83.63        | 85.09        | 83.63        | 83.63        |
| Fear                 | 73.98        | 72.81        | 72.81        | 72.51        |
| Sadness              | 76.02        | 73.98        | 74.85        | 74.85        |
| <b>Average</b>       | <b>75.00</b> | <b>74.17</b> | <b>73.83</b> | <b>73.73</b> |

short-lived but last for some time; temporal correlation ( $f_3$ ) has the most contribution for happiness and sadness, indicating that these two emotions have a lasting temporal influence; social correlation ( $f_4$ ) has the most contribution for surprise and anger, which means these two emotions have strong influence among social friends.

## 7.3 Case Study

Since it is very hard to find an objective way to evaluate the whole influence results, we would like to show some interesting cases to demonstrate the effectiveness of the model.

We randomly pick up 412 users from our data set. Then we collect the images they uploaded and their behaviors on Flickr, such as publishing comments. By leveraging this information, we discover several interesting as well as important phenomena.

**More social interactions lead to higher emotion influence between two users.** To better explore the correlation between emotion influence and user' social interactions, 412 users' behaviors are analyzed. If user  $v_i$  comments an image published by user  $v_j$ , we call that there is a social interaction between user  $v_i$  and user  $v_j$ . Thus the average interacting frequency of different emotion influence  $v$  can be defined as:

$$Frequency_v = \sum_{i,j} = \frac{\#interaction\_times(i,j)}{\#all\_user\_pairs(v)} \quad (18)$$

Using the proposed emotion influence model, we first get the predicted emotion influence between every two users. Then, we classify the prediction results to different emotion categories. We further compute the average interacting frequency of every two users according to the above equation. Given these results, we investigate the correlation between emotion influence and social interaction for different emotion categories. We clearly depict our observation in Figure 5. For the bar graphs in Figure 5, their horizontal axes represent the different value intervals of the predicted

emotion influence, while their vertical axes represent the average frequency of social interactions. It can be found that for most emotion categories, with the increase of the emotion influence, the overall trend of the average interacting frequency is growing too. This phenomenon indicates that more social interactions correlate with higher emotion influence between two users. The fact corroborates our common sense and validates the effectiveness of our method.

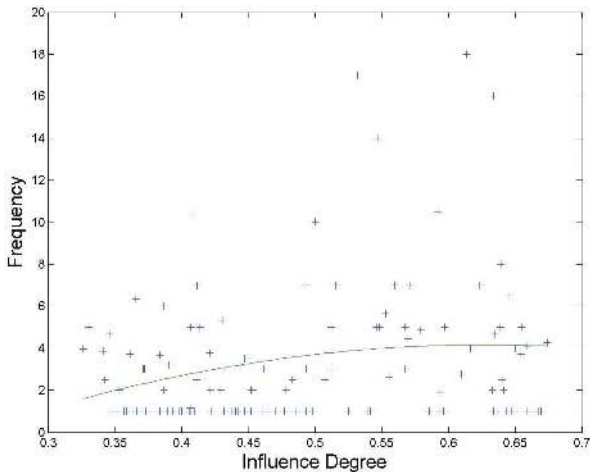


Fig. 5. The correlation between emotion influence and social interaction.

**Disgust is a special category with an opposite trend.** Taking that a step further, Table 7 shows the correlation varies. We can see that for happiness, surprise, anger, fear and sadness, with the increase of the emotion influence, the average users' interacting frequency grows. But another interesting phenomenon is that in the case of disgust, the rule works contrariwise. Disgust is one prototypic emotion with multiple reaction patterns which are linked to a set of aversive experiences and a unique mode of experiencing morality [42]. Thus the feeling of disgust is highly subjective. Some people feel disgusted when seeing dense patterns, while some people feel nothing about it, but are extremely disgusted seeing fluffy things. That is to say, even if our closest friend feels extremely disgusted, we may not be influenced at all. The second reason we assume to explain the phenomenon is that expressing one's disgust may hurt the others' feelings so that close friends are less likely to leave comments to reveal their opinions, while strangers can relieve their feelings thoughtlessly.

**The influence of negative emotions is stronger than positive ones.** Based on the discussion above, we can further explore the correlation on higher dimensions. We classify the emotion categories into positive emotions and negative emotions. Positive emotions mainly contain happiness, while negative emotions contain anger, fear, disgust and sadness. Although positive emotions only have one type, half of the

labeled images fall into the happiness category. So the amount of positive and negative images are nearly the same. Deriving from Table 5, we compare the prediction results of our model with traditional GM. It seems that when taking the proposed temporal and social correlation factors into consideration, compared to the improvement of classifying positive images, the performance of classifying negative images improves more significantly. That means temporal and social correlation plays a more vital role in classifying negative images, indicating that the influence of negative emotions is stronger.

## 7.4 Error Analysis

At last, we would like to give our analysis on the possible sources of the errors based on the prediction results of the proposed model.


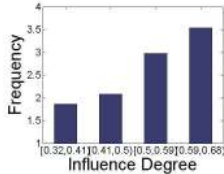

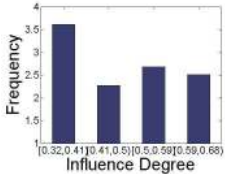

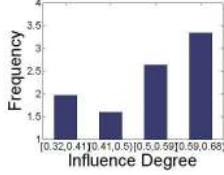

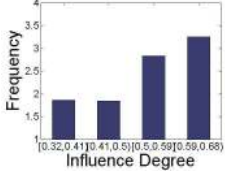

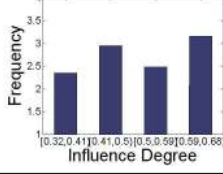

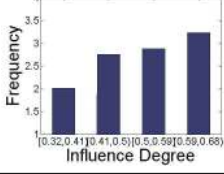
- 1) *Noise and missing data.* The data preparation by using the user-given tags to build the ground truth is widely used in existing research on big data analysis. This is a practical idea under the situation when there is no existing ground truth and manually labelling the ground truth is mission impossible for a large-scale data set. But it also may introduce some noise. Not all the users are active, or the user may wrongly label her/his mood.
- 2) *Other factors.* In this paper, we assume that in image social networks, people's emotions are expressed and spread by their uploading and sharing images. Quantitative analysis of causal influence is a very difficult challenge, which is not the focus of this paper. We select a task of emotion prediction of images to quantitative evaluation of the proposed influence model. The model incorporates various pivotal factors including image content, temporal correlation and social correlation with their weights. And the experiments validate the effectiveness of these factors.

## 8 CONCLUSIONS

In this paper, we study a novel problem of modeling emotion influence in image social networks. We first examine the existence of the image-based emotion influence among friends. Then, an emotion influence model is proposed to combine uploaded images, users' emotion and their influence. We evaluate the proposed model on a large image data set. The experiments show that our model is much more effective on inferring emotions from social images than traditional classifiers. A case study further demonstrates that the prediction of the emotion influence is consistent with the reality.

Emotion is one of the most important user behaviors and presents a fundamental research direction.

TABLE 7  
The correlation varies according to different emotion categories.

| Categories   | Social interaction  | Trend    | Categories   | Social interaction  | Trend    |
|--|---|----------|--|---|----------|
| Happiness<br> |  | Increase | Disgust<br> |  | Decrease |
| Surprise<br>  |  | Increase | Fear<br>    |  | Increase |
| Anger<br>     |  | Increase | Sadness<br> |  | Increase |

Explanation for the emotions of some example images: the majestic landscape in the surprise category makes us amazed and astonished; lots of things clustered randomly in the disgust category make us nauseated; natural disasters make us afraid and building a lonely climate under the big scene lets us scared in the fear category.

As the future work, it is intriguing to extend the peer influence study to group conformity behavior, such as how a user's emotion conforms to the community she/he belongs to. It is also interesting to further study how the users' emotions correlate with their social actions. There are many real applications of emotion analysis such as recommendation and social advertising.

## ACKNOWLEDGMENT

This work is supported by the National Basic Research Program of China (2012CB316401), Natural Science Foundation of China (61370023), and is partially supported by the major project of the National Social Science Foundation of China (13&ZD189), the National Basic Research Program of China (2011CB302201). Jie Tang is supported by the National High-tech R&D Program (2014AA015103), National Basic Research Program of China (2014CB340506), and the Natural Science Foundation of China (61222212). Lexing Xie is supported by Australian Research Council Discovery Project DP140102185. We would also like to thank the Microsoft Research Asia-Tsinghua University Joint Laboratory FY14-RES-SPONSOR-111 for its support.

## REFERENCES

- [1] J. Tang, Y. Zhang, J. Sun, J. Rao, W. Yu, Y. Chen, and A. Fong, "Quantitative study of individual emotional states in social networks," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 132–144, 2012.
- [2] J. H. Fowler and N. A. Christakis, "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study," in *British Medical Journal*, 2008.
- [3] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *ACM Multimedia*, 2010, pp. 83–92.
- [4] J. Jia, S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang, "Can we understand van gogh's mood? learning to infer affects from images in social networks," in *ACM Multimedia*, Japan, 2012.
- [5] P. Lovato, A. Perina, N. Sebe, O. Zandonà, A. Montagnini, M. Bicego, and M. Cristani, "Tell me what you like and i'll tell you what you are: Discriminating visual preferences on flickr data," in *Computer Vision-ACCV*. Springer, 2013, pp. 45–56.
- [6] E. Bakshy, D. Eckles, R. Yan, and I. Rosenn, "Social influence in social advertising: evidence from field experiments," in *EC'12*, 2012, pp. 146–161.
- [7] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler, "A 61-million-person experiment in social influence and political mobilization," *Nature*, vol. 489, pp. 295–298, 2012.
- [8] Y. Yang, P. Cui, W. Zhu, and S. Yang, "User interest and social influence based emotion prediction for individuals," in *ACM Multimedia*, 2013, pp. 785–788.
- [9] S. James H. Fowler, Nicholas A. Christakis and D. Roux, "Dynamic spread of happiness in a large social network: Longitudinal analysis of the framingham heart study social network," *BMJ: British Medical Journal*, vol. 338, no. 7685, pp. 23–27, 2009.
- [10] S. Hareli and A. Rafaeli, "Emotion cycles: On the social influence of emotion in organizations," *Research in Organizational Behavior*, vol. 28, no. 0, pp. 35 – 59, 2008.
- [11] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *ACM SIGKDD*, New York, NY, USA, 2008, pp. 7–15.
- [12] J. Tang, S. Wu, and J. Sun, "Confluence: Conformity influence in large social networks," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD'13. New York, NY, USA: ACM, 2013, pp. 347–355.
- [13] T. La Fond and J. Neville, "Randomization tests for distinguishing social influence and homophily effects," in *international Conference on World Wide Web*, New York, USA, 2010, pp. 601–610.
- [14] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring

- networks of diffusion and influence," *ACM Trans. Knowl. Discov. Data*, vol. 5, no. 4, pp. 21:1–21:37, 2012.
- [15] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, "User-level sentiment analysis incorporating social networks," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '11. New York, NY, USA: ACM, 2011, pp. 1397–1405.
- [16] H. Ma, H. Yang, M. R. Lyu, and I. King, "Sorec: Social recommendation using probabilistic matrix factorization," in *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, ser. CIKM '08. New York, NY, USA: ACM, 2008, pp. 931–940.
- [17] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. 24, pp. 8788–8790, 2014. [Online]. Available: <http://www.pnas.org/content/111/24/8788.abstract>
- [18] B. Li, S. Feng, W. Xiong, and W. Hu, "Scaring or pleasing: exploit emotional impact of an image," in *ACM Multimedia*, 2012, pp. 365–366.
- [19] B. Li, W. Xiong, W. Hu, and X. Ding, "Context-aware affective images classification based on bilayer sparse representation," in *ACM Multimedia*, 2012, pp. 721–724.
- [20] M. Cristani, A. Vinciarelli, C. Segalin, and A. Perina, "Unveiling the multimedia unconscious: implicit cognitive processes and multimedia content analysis," in *ACM Multimedia*, 2013, pp. 213–222.
- [21] S. Siersdorfer, E. Minack, F. Deng, and J. Hare, "Analyzing and predicting sentiment of images on the social web," in *ACM Multimedia*, 2010, pp. 715–718.
- [22] T. Chen, D. Lu, M.-Y. Kan, and P. Cui, "Analyzing and predicting sentiment of images on the social web," in *ACM Multimedia*, 2013, pp. 781–784.
- [23] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *ICML*, 2001, pp. 282–289.
- [24] Y. Liu, B. Wu, H. Wang, and P. Ma, "Bpgm: A big graph mining tool," *Tsinghua Science and Technology*, vol. 19, no. 1, pp. 33–38, 2014.
- [25] R. Bredereck, J. Chen, P. Faliszewski, J. Guo, R. Niedermeier, and G. J. Woeginger, "Parameterized algorithmics for computational social choice: Nine research challenges," *Tsinghua Science and Technology*, vol. 19, no. 4, pp. 358–373, 2014.
- [26] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.
- [27] K. R. Scherer, "What are emotions? and how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005.
- [28] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.
- [29] L. Xie and X. He, "Picture tags and world knowledge: Learning tag relations from visual semantic sources," in *ACM Multimedia*, 2013, pp. 967–976.
- [30] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, "Large-scale visual sentiment ontology and detectors using adjective noun pairs," in *ACM Multimedia*, 2013, pp. 223–232.
- [31] C. Fellbaum, "Wordnet," in *Theory and Applications of Ontology: Computer Applications*, 2010, pp. 231–243.
- [32] Z. Dong and Q. Dong, *HowNet And The Computation Of Meaning*. World Scientific, 2006.
- [33] J. Zhang, B. Liu, J. Tang, T. Chen, and J. Li, "Social influence locality for modeling retweeting behaviors," in *IJCAI*, 2013, pp. 2761–2767.
- [34] M. De Choudhury, S. Counts, and E. Horvitz, "Predicting post-partum changes in emotion and behavior via social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: ACM, 2013, pp. 3267–3276.
- [35] X. Wang, J. Jia, J. Yin, and L. Cai, "Interpretable aesthetic features for affective image classification," in *ICIP*, Melbourne, Australia, 2013, pp. 3230–3234.
- [36] S. Kobayashi, *Art of Color Combinations*. K. International, 1995.
- [37] J. Itten, *The art of color: the subjective experience and objective rationale of color*. Wiley, 1974.
- [38] J. S. Yedidia, W. T. Freeman, Y. Weiss *et al.*, "Generalized belief propagation," in *NIPS*, vol. 13, 2000, pp. 689–695.
- [39] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, Aug. 2012.
- [40] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.
- [41] X. Wang, J. Jia, and L. Cai, "Affective image adjustment with a single word," *The Visual Computer*, vol. 29, no. 11, pp. 1121–1133, 2013.
- [42] J. Moll, R. Zahn, R. de Oliveira-Souza, F. Krueger, and J. Grafman, "The neural basis of human moral cognition," *Nature Reviews Neuroscience*, vol. 6, no. 10, pp. 799–809, 2005.

**Xiaohui Wang** is a postdoctor in University of Science and Technology Beijing. Her research interests are affective computing, multimedia and big data.

**Jia Jia** is an associate professor in Tsinghua University. Her research interests are affective computing and computational speech perception.

**Jie Tang** is an associate professor in Tsinghua University. His research interests are social network analysis, data mining, and machine learning.

**Boya Wu** is a undergraduate student in Tsinghua University. Her research interests are social network analysis and affective computing of images.

**Lianhong Cai** is a professor in Tsinghua University. Her major research interests include human computer speech interactive, speech synthesis and text-to-speech (TTS), speech corpus, and multimedia technology.

**Lexing Xie** is Senior Lecturer in the Research School of Computer Science at the Australian National University. Her research interests are in applied machine learning, multimedia, social media.