



# Anger Is More Influential than Joy: Sentiment Correlation in Weibo

Rui Fan<sup>1</sup>, Jichang Zhao<sup>2</sup>, Yan Chen<sup>1</sup>, Ke Xu<sup>1\*</sup>

**1** State Key Laboratory of Software Development Environment, Beihang University, Beijing, P. R. China, **2** School of Economics and Management, Beihang University, Beijing, P. R. China

## Abstract

Recent years have witnessed the tremendous growth of the online social media. In China, Weibo, a Twitter-like service, has attracted more than 500 million users in less than five years. Connected by online social ties, different users might share similar affective states. We find that the correlation of anger among users is significantly higher than that of joy. While the correlation of sadness is surprisingly low. Moreover, there is a stronger sentiment correlation between a pair of users if they share more interactions. And users with larger number of friends possess more significant sentiment correlation with their neighborhoods. Our findings could provide insights for modeling sentiment influence and propagation in online social networks.

**Citation:** Fan R, Zhao J, Chen Y, Xu K (2014) Anger Is More Influential than Joy: Sentiment Correlation in Weibo. PLoS ONE 9(10): e110184. doi:10.1371/journal.pone.0110184

**Editor:** Rodrigo Huerta-Quintanilla, Cinvestav-Merida, Mexico

**Received:** September 10, 2013; **Accepted:** July 21, 2014; **Published:** October 15, 2014

**Copyright:** © 2014 Fan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by 863 Program (grant no. 2012AA011005), SKLSDE (grant no. SKLSDE-2013ZX-06), and Research Fund for the Doctoral Program of Higher Education of China (grant no. 20111102110019). Jichang Zhao was partially supported by the Fundamental Research Funds for the Central Universities (grant nos. YWF-14-RSC-109 and YWF-14-JGXY-001). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* Email: kexu@nlsde.buaa.edu.cn

## Introduction

From the view of conventional social theory, *homophily* leads to connections in social networks, as the saying “Birds of a feather flock together” states [1]. Even in the online social network, more and more evidence indicates that the users with similar properties would be connected in the future with high probabilities [2,3]. It is clear that *homophily* could affect user behavior both online and offline [4,5], while the records in online social networks are relatively easier to be tracked and collected. Moreover, the continuous growth of the online social media attracts a vast number of internet users and produces many huge social networks. Twitter(www.twitter.com), a microblogging website launched in 2006, has over 300 million active users, with over 500 million microblog posts, known as tweets, being posted everyday. In China, Weibo(www.weibo.com), a Twitter-like service launched in 2009, has accumulated more than 500 million registered users in less than five years. Every day there will be more than 100 million Chinese tweets published. The high-dimension content generated by millions of global users is a “big data” window [6] to investigate the online social networks. That is to say, these large-scale online social networks provide an unprecedented opportunity for the study of human behavior.

Beyond typical demographic features such as ages, races, hometowns, common friends and interests, *homophily* also includes psychological states, like loneliness and happiness [1,4,7]. Previous studies also show that the computer-mediated emotional communication is similar to the traditional face-to-face communication, which means there is no evident indication that human communication in online social media is less emotional or less personally [8]. Each user in the online social network could be a

social sensor and the huge amount of tweets convey complicated signals about the users and the real-world events, among which the sentiments are an essential part. Therefore, emotion states of the users play a key role in understanding the user behaviors in social networks, whether from an individual or group perspective [9–11]. Meanwhile, users’ mood states are significantly affected by the real-world events [12], which could be employed to predict the stock market [13] or to detect the abnormal event [14]. Recent study [4] shows that happiness is assortative in Twitter network and [6] finds that the average happiness scores are positively correlated between the Twitter users connected by one, two or three social ties. An interesting phenomenon of emotion synchronization is also unraveled in [15]. While in these studies, the human emotion is simplified to two classes of positive and negative or just a score of general happiness, neglecting the detailed aspects of human sentiment, especially the negative emotion. Because of oversimplification of the emotion classification, it is hard for the previous literature to disclose the different correlations of different sentiments and then make comparisons. However, the negative emotions, like anger, sadness or disgust, are more applicable in real world scenarios such as abnormal event detection or emergency tracking [14]. In [5], the authors also find that negative emotion could boost user activity in BBC forum. In fact, figuring out the correlation of these emotions might shed light on understanding why people participate in the diffusion of abnormal event in the network and how the large-scale collective behavior could form across the entire network. On the other hand, the investigation of how the local structure affects the emotion correlation is not systematically performed yet, while which is essential to studying the mechanism of sentiment influence and contagion.

Aiming at filling these vital gaps, we divide the sentiment of an individual into four categories, including *anger*, *joy*, *sadness* and *disgust*, and investigate the emotion correlation between connected users in the interaction network obtained from Weibo. Out of our expectation, it is found that *anger* has a stronger correlation between different users than that of *joy*, while *sadness*'s correlation is trivial. Further analysis demonstrates that *anger* in Weibo is related with the real-world events about food security, government bribery or demolition scandal, which are always the hot trends in Internet of China. Moreover, node degree, node clustering and tie strength all could positively boost the emotion correlation in online social networks, especially for the mood of *anger*. Finally, we make our data set in this paper publicly available to the research community.

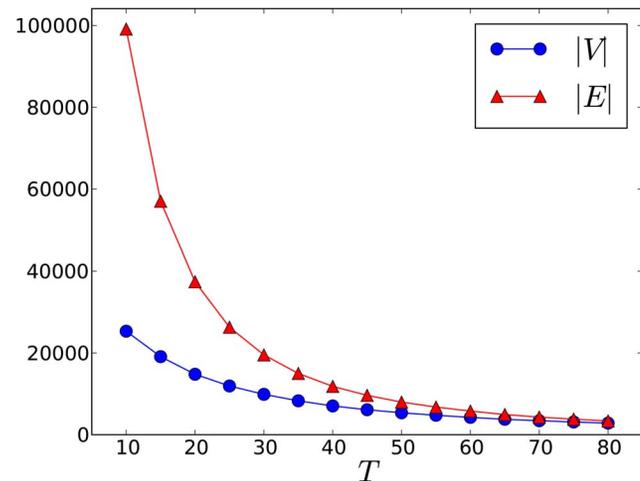
## Materials and Methods

### Weibo Dataset

As pointed out in [4], the following relationship in Twitter-like social networks does not stand for the social interaction, while if two users reply, retweet or mention each other in their tweets for certain times, the online social tie between them is sufficient to present an alternative means of deriving a conventional social network [6]. Starting from several influential seeds (like the users verified by Weibo), we adapt a typical Breadth-First-Search strategy to crawl tweets from Weibo through its open APIs. For each user we get, we first save all its tweets into the database and then add its followers (users that follow it) into the candidate queue for further explorations. Finally from December 2010 to February 2011, we accumulated around 70 million tweets posted by 278,654 users. While here we only construct an interaction network from the tweets posted during April 2010 to September 2010, where interaction means the number that two users retweet or mention each other is larger than a threshold  $T$ . From around 26 million tweets posted during the period we select and 140,000 users we crawled, an undirected but weighted graph  $G(V, E, T)$  is constructed, in which  $V$  is the set of users (the ones without links are omitted),  $E$  represents the set of interactive links among  $V$ , and  $T$  is the minimum number of interactions on each link. For each link in  $E$ , its weight is the sum of retweeting or mentioning times between its two ends in the specified time period. Specifically, to exclude occasional users that are not truly involved in the Weibo social network, we only reserve those active users in our interaction network that posted more than one tweet every two days on average over the six months. And to guarantee the validity of users' social interaction, if the number of two users retweet or mention each other is less than  $T$ , we would omit the connection between them. As shown in Figure 1, by tuning  $T$  we can obtain networks of different scales. Generally we set  $T=30$  and then the interaction network  $G$  contains 9,868 nodes and 19,517 links. We also make our entire dataset publicly available (<http://www.datatang.com/data/44650>, <http://goo.gl/iXzoXm>). Note that here we collect tweets from Weibo through its open APIs (<http://open.weibo.com>) under the authority granted by Weibo and we have also anonymized user IDs and names in the published data set to protect users' privacy.

### Emotion classification

The content in online social media like Twitter or Weibo is mainly recorded in the form of short text. Many approaches have been presented to mine sentiment from these texts in recent years. One of them is the lexicon based method, in which the sentiment of a tweet is determined by counting the number of sentimental words, i.e., positive terms and negative terms. For example, Dodds



**Figure 1. The number of nodes or edges varies for different interaction threshold  $T$ .** In the following part of the present work, we set  $T=30$  to extract a large enough network with convincing interaction strength.

doi:10.1371/journal.pone.0110184.g001

and Danforth measured the happiness of songs, blogs and presidents [16]. They also employed Amazon Mechanical Turk to score over 10,000 unique English words on an integer scale from 1 to 9, where 1 represents sadness and 9 represents happiness [9]. Golder and Macy collected 509 million English tweets from 2.4 million users in Twitter, then measured the positive and negative affects using Linguistic Inquiry and Word Count (LIWC) (<http://www.liwc.net>). While another one is the machine learning based solution, in which different features are considered to perform the task of classification, including terms, smileys, emoticons and etc. The first step was taken by Pang et al. in [17], they treated the sentiment classification of movie reviews simply as a text categorization task and investigated several typical classification algorithms. According to the experimental results, machine learning based classifiers outperform the baseline method based on human words list [18–20]. Different from most work which just categorized the emotion into negative and positive, our previous work [14] divided the sentiment into four classes, then presented a framework based on emoticons without manually-labelled training tweets and achieved a convincing precision. Because of the ability of multi-emotions classification, we employ this framework in the present paper.

A vast number of training samples is necessary for handling the extremely short text in social media. To avoid intensive labor, we use emoticon to label tweets into different emotions. It has been found that both smiley and emoticon are strongly related with typical sentiment words and could be convincing indicators of different emotions [21]. They help the users to express their moods when post the tweet [22]. Tossell et al. also confirm that emoticon usage is contextual [23]. Hence, we could treat these emoticons as sentiment labels of the tweets. In fact, it is a kind of crowdsourcing, i.e., the users label the tweet with emoticons to express their emotions themselves [14]. In the labeling stage, first we manually label the emotion of the emoticon. We select the most popular 95 emoticons and several students are working separately to label their emotions. Their judgements are based on the image of the emoticon and around 50 frequent words occurring together with the emoticon. Finally we find that most of the emoticons are labelled by four sentiments, including *anger*, *joy*, *disgust* and *sadness*. For other emotions like *fear* or *surprise*, we do not find

enough votes. So we split the emotion into these four classes. In fact it is also sort of consistent with the traditional Chinese culture that the human emotion are mainly constituted by four elements, including pleasure, anger, sorrow and joy.

From around 70 million tweets, 3.5 million tweets with valid emoticons are extracted and labeled. Using this data set as a training corpus  $C_{tr}$ , a simple but fast Bayesian classifier is built in the second stage to mine the sentiment of the tweets without emoticons, which are about 95% in Weibo. Be specific, for each tweet  $t$  in  $C_{tr}$ , we converts it into a sequence of words  $\{w_i\}$ , where  $w_i$  is a word and  $i$  is its position in  $t$ . From the labeled tweets, we could obtain the word  $w_i$ 's prior probability of belonging to the sentiment category  $m$  is  $P(w_i||m) = \frac{n^m(w_i)+1}{\sum_q (n^m(w_q)+1)}$ , where

$m=1,2,3$  or  $4$ ,  $n^m(w_i)$  is the times that  $w_i$  appears in all the tweets in the category  $m$  and *Laplace smoothing* is used to avoid the problem of zero probability. Then for an unlabeled tweet  $t$  with word sequence  $\{w_i\}$ , its category could be obtained as  $m^*(t) = \arg \max_m P(m) \prod_i P(w_i||m)$ , where  $P(m)$  is the prior probability of  $m$ . The averaged precision of this classifier is 64.3% and particularly the large amount of tweets we employ in the experiment can guarantee its accuracy further. For example, in the applications like *MoodLens*([http://gana.nlsde.buaa.edu.cn/hourly\\_happy/moodlens.html](http://gana.nlsde.buaa.edu.cn/hourly_happy/moodlens.html)) and *Sentiment Search*(<http://xinqings.nlsde.buaa.edu.cn/>), it can be used to detect abnormal events effectively in real-time tracking. Moreover, the mechanism of incremental learning in this classifier can tackle the problems like sentiment drift of terms [9] or emergence of new features [14].

Based on this framework, we demonstrate a sampled snapshot of the interaction network with  $T=30$ . As shown in Figure 2b, in which each user is colored by its emotion. We can roughly find that closely connected nodes generally share the same color, indicating emotion correlations in Weibo network. Besides, different colors show different clusterings. For example, the color of red, which represents *anger*, shows more evident clustering. These preliminary findings inspire us that different emotions might have different correlations and a deep investigation is indeed necessary.

### Emotion correlation

Emotion correlation is a metric to quantify the strength of sentiment similarity between connected users. For a fixed  $T$ , we first extract an interaction network  $G$  and all the tweets posted by the nodes in  $G$ . Then by employing the classifier established in the former section, the tweets for each user are divided into four categories, in which  $f_1, f_2, f_3$  and  $f_4$  represent the fraction of angry, joyful, sad and disgusting tweets, respectively. Hence we can use emotion vector  $e_i(f_1^i, f_2^i, f_3^i, f_4^i)$  to denote user  $i$ 's sentiment status. Based on this, we define the pairwise sentiment correlation as follows. Given a certain hop distance  $h$ , we collect all user pairs with distance  $h$  from  $G$ . For one of the four emotions  $m(m=1,2,3,4)$  and a user pair  $(j, q)$ , we put the source user  $j$ 's  $f_m^j$  into a sequence  $S_m$ , and the target user  $q$ 's  $f_m^q$  to another sequence  $T_m$ . Then the pairwise correlation could be calculated by Pearson correlation as

$$C_p^m = \frac{1}{l-1} \sum_{i=1}^l \left( \frac{S_i - \langle S_m \rangle}{\sigma_{S_m}} \right) \left( \frac{T_i - \langle T_m \rangle}{\sigma_{T_m}} \right),$$

where  $\langle S_m \rangle = \frac{1}{l} \sum_{i=1}^l S_i$  is the mean,  $\sigma_{S_m} = \sqrt{\frac{1}{l-1} \sum_{i=1}^l (S_i - \langle S_m \rangle)^2}$  is the standard deviation and  $l$  is the

length of  $S_m$  or  $T_m$ . Or it can also be obtained from Spearman correlation as

$$C_s^m = 1 - \frac{6 \sum_{i=1}^l d_i^2}{l(l^2-1)},$$

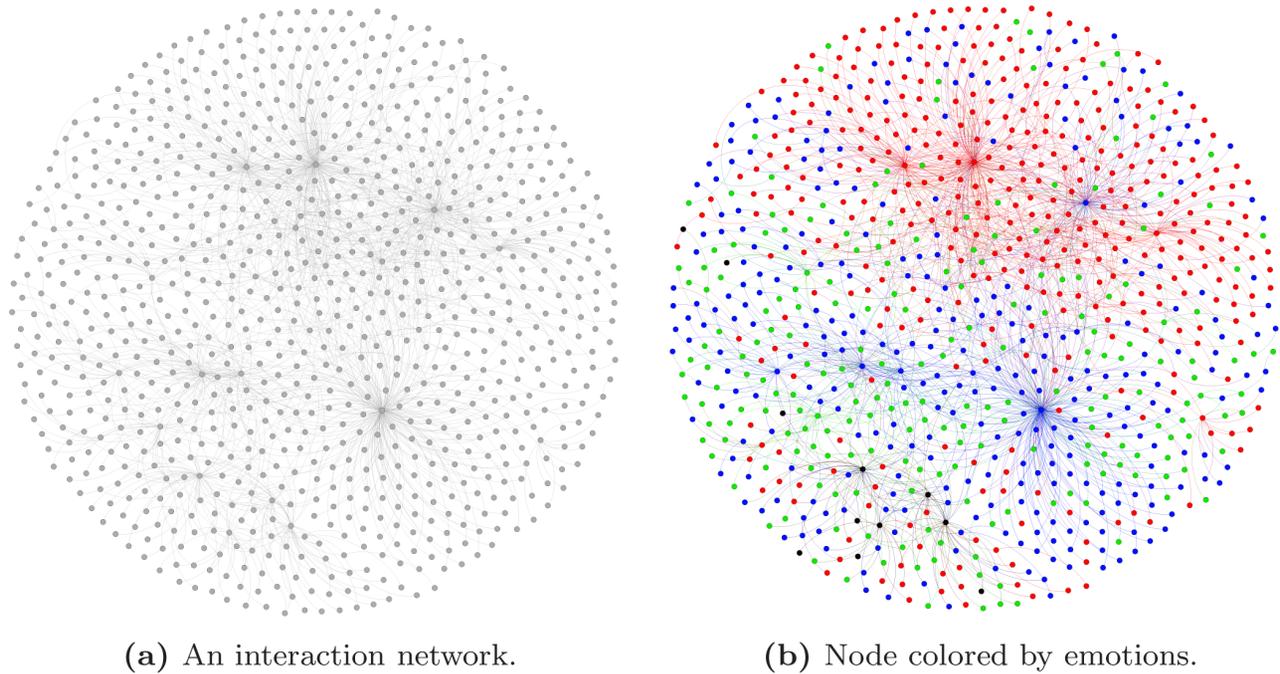
where  $d_i$  is the rank difference between  $S_i$  in  $S_m$  and  $T_i$  in  $T_m$ . Intuitively larger  $C_p^m$  and  $C_s^m$  both suggest a more positive correlation for sentiment  $m$ . In order to investigate fluctuations in the sentiment correlation, we also use the approach of *bootstrap* [24] to perform the error analysis. For instance, given two emotion sequences of length  $x > 1$ , denoted as  $S_m = \{s_1, s_2, \dots, s_x\}$  and  $T_m = \{t_1, t_2, \dots, t_x\}$ , we do not calculate their correlation directly. Contrarily, we first uniformly sample  $x$  integer indexes from the range of  $[1, x]$  with replacement and then put them into an index sequence defined as  $R_m = \{r_1, r_2, \dots, r_x\}$ . Through traversing each index  $r_i \in R_m$ , we can construct two new lists by putting  $s_{r_i}$  into  $S'_m$  and  $t_{r_i}$  into  $T'_m$ , respectively. Obviously after this we can generate two sampled sequences  $S'_m = \{s'_1, s'_2, \dots, s'_x\}$  and  $T'_m = \{t'_1, t'_2, \dots, t'_x\}$ . Finally for each round of index-sampling we can obtain a correlation value between  $S'_m$  and  $T'_m$ , and through  $z=1000$  times of repetitions we would obtain an averaged correlation and a standard deviation (the error) for  $m$ . Apparently lower errors stand for more significant correlations. Note that in the rest of the paper, averaged correlation would be presented as correlation if there is no conflict in the context.

Based on the dataset and classifier, interaction networks could be built and tweets of each user in the network would be emotionally labelled. Using the definition of correlations, we can then present the comparison of emotion correlations and the impact of local structures in the following section.

### Results

First we compare the correlation of different emotions based on the graph of  $T=30$ , which ensures enough number of ties and users, and at the same time guarantees relatively strong social tie strength. As shown in Figure 3, both Pearson correlation and Spearman correlation indicate that different sentiments have different correlations and *anger* has a surprisingly higher correlation than other emotions. In addition, the standard deviations of all sentiments' correlations are extremely small, which indicates that the sentiment correlation in online social networks is indeed significant and only shows trivial fluctuations. Although the previous studies [4,6] show that happiness is assortative in online social networks, but Figure 3 further demonstrates that the correlation of *anger* is much stronger than that of happiness, especially as  $h < 3$ . While for *sadness* and *disgust*, they both have an unexpected low correlation even for small  $h$ . For instance, the correlation of *sadness* is less than 0.15 as  $h=1$ . The results are also consistent with the previous findings that strength of the emotion correlation decreases as  $h$  grows, especially after  $h > 6$  [6]. In fact, as  $h > 3$ , the emotion correlation becomes weak for all the sentiments, which means that the correlation of the sentiment in the social network is limited significantly by the social distance. For example, for strong assortative emotions like *anger* and *joy*, their correlations just fluctuate around 0 as  $h > 3$ .

In order to test the above correlation further, we also shuffle  $S_m$  and  $T_m$  randomly for sentiment  $m$  and recalculate its correlation. As shown in Figure 4, for the shuffled emotion sequence, there is no correlation existing for all the sentiments. It indicates that the former correlation we get is truly significant and for random pairs of users in the social network, there is no *emotion homophily*. It

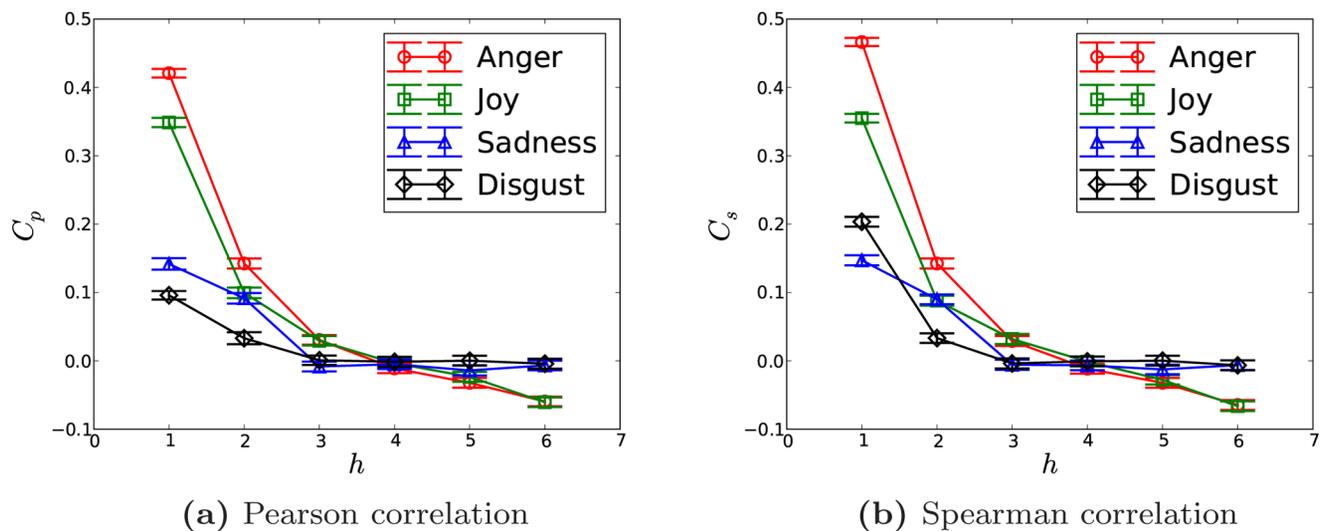


**Figure 2. (Color online) The giant connected cluster of a network sample with  $T=30$ .** (a) is the network structure, in which each node stands for a user and the link between two users represents the interaction between them. Based on this topology, we color each node by its emotion, i.e., the sentiment with the maximum tweets published by this node in the sampling period. In (b), the red stands for *anger*, the green represents *joy*, the blue stands for *sadness* and the black represents *disgust*. The regions of same color indicate that closely connected nodes share the same sentiment.  
doi:10.1371/journal.pone.0110184.g002

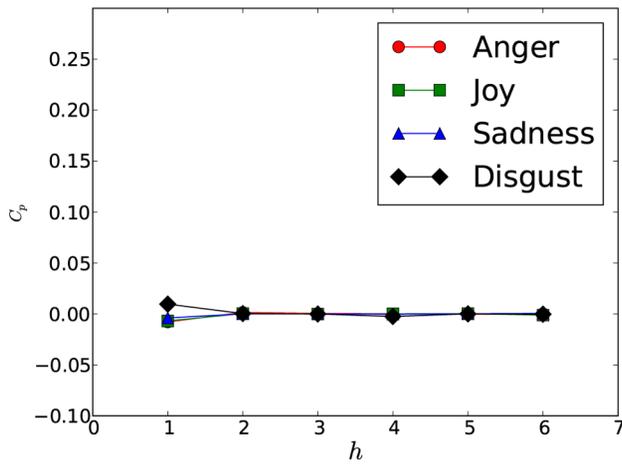
further justifies that through social ties, closely connected friends indeed share similar affective states.

Investigating to what extent the local structure, like tie strength, node degree and node clustering, could affect the emotion correlation and its error is of importance for modeling sentiment influence and propagation in future work. As shown in Figure 5, we first disclose how the interaction threshold  $T$  affects the sentiment correlations. As discussed in Section *Weibo Dataset*,

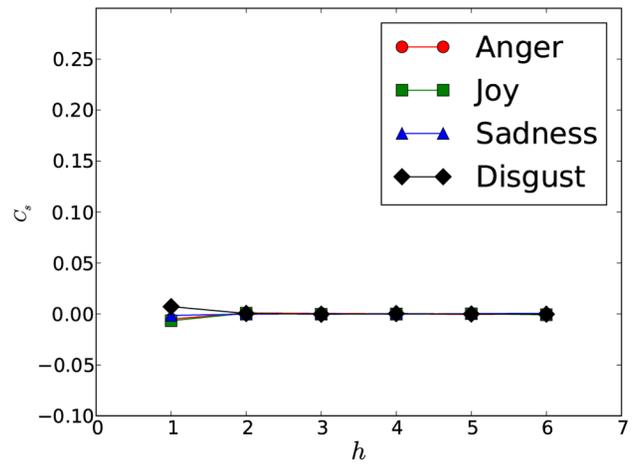
larger  $T$  produces smaller networks but with closer social relations and more frequent online interactions. It is also intuitive that frequent interactions in online social networks are positively related with strong social ties and convincing social bonds. Because of this, we can see in Figure 5 that for all the four emotions, their correlations inside two hops continue a steady increasing trend with  $T$ 's growth. Particularly for *anger*, its Pearson correlation could rise to around 0.5. For weakly correlated emotions like



**Figure 3. Correlations with error-bar for different emotions as the hop distance varies.** Large  $h$  means a pair of users are far away from each other in the social network we build. Here  $T=30$  is fixed.  
doi:10.1371/journal.pone.0110184.g003



(a) Pearson correlation



(b) Spearman correlation

**Figure 4. The emotion sequence is randomly shuffled to test the correlation significance.**  
doi:10.1371/journal.pone.01110184.g004

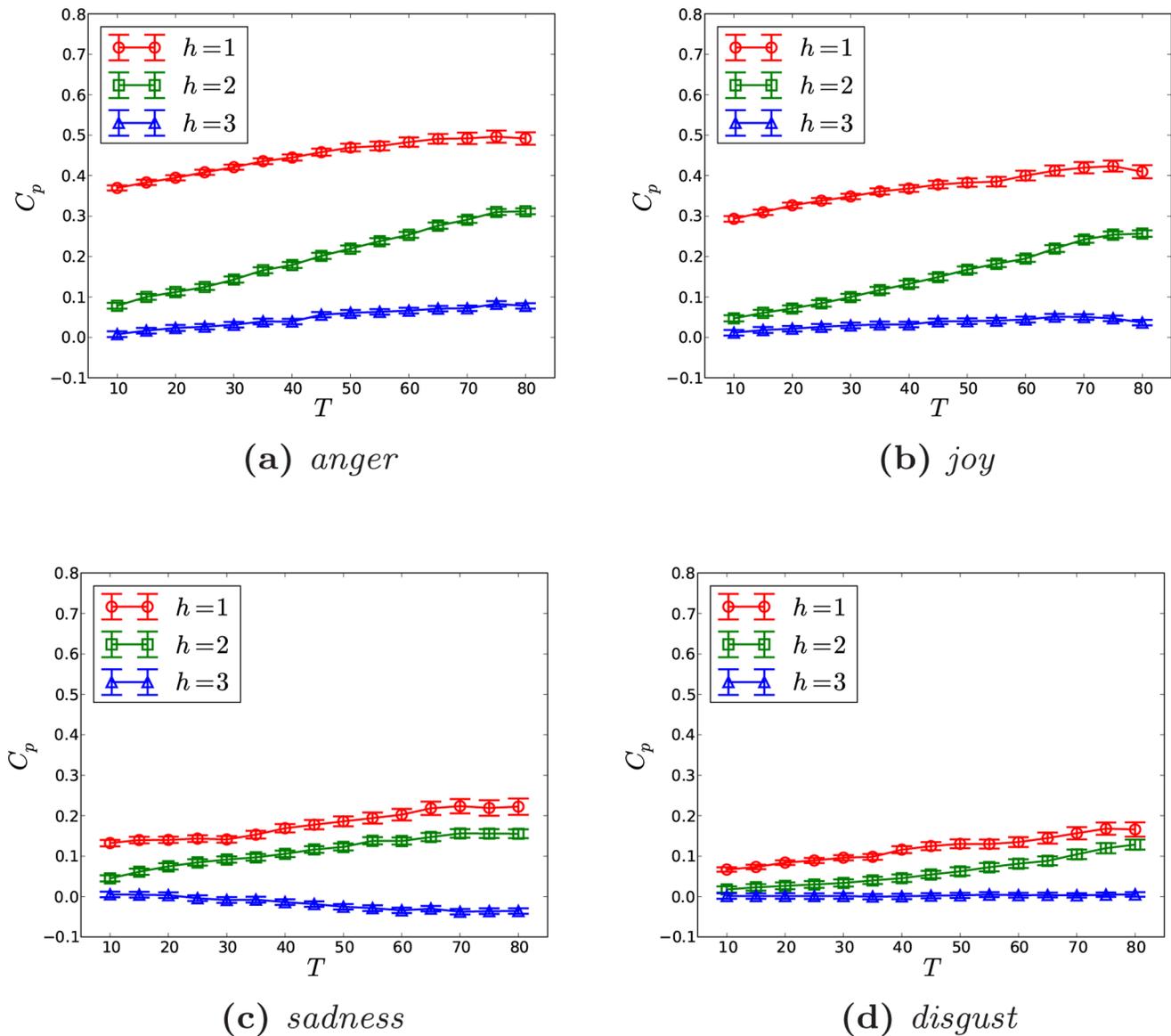
*sadness* and *disgust*, although the correlation shows a slow growth for  $h = 1$  and  $h = 2$ , while the maximum value of the correlation is still lower than 0.25. As  $h = 3$ , the increment of the sentiment correlation is trivial, especially for *sadness* and *disgust*. It illustrates that the primary factor of controlling the emotion correlation is still the social distance and the social tie strength just functions for close neighbors in the scope of two hops. Note that as  $T$  grows, the size of the network is reduced and the length of emotion sequences would be shortened accordingly, which might import more noise and produce larger errors. As can be seen in Figure 5, errors of correlations grow with  $T$ , especially as  $h = 1$ . Secondly, we check the effect of users' degrees to the sentiment correlation. Given a random node  $i$  with degree  $k$ , we have a sequence of the number of tweets with sentiment  $m$  for its friends, which is denoted as  $\{\eta_m^i\}_i$ , and  $\eta_m^i$  is the number of tweets with emotion  $m$  posted by an arbitrary neighbor  $j$ . Then the number of tweets with emotion  $m$  posted by  $i$ 's neighborhood is  $\sum_j \eta_m^i$ . Ultimately the sentiment vector for  $i$ 's neighborhood could be defined as  $e_i^{nei}$  ( $\frac{\sum_j \eta_1^i}{\sum_m \sum_j \eta_m^i}, \frac{\sum_j \eta_2^i}{\sum_m \sum_j \eta_m^i}, \frac{\sum_j \eta_3^i}{\sum_m \sum_j \eta_m^i}, \frac{\sum_j \eta_4^i}{\sum_m \sum_j \eta_m^i}$ ). Through adding  $f_m^i$  into  $S_m$  and  $\frac{\sum_j \eta_m^i}{\sum_m \sum_j \eta_m^i}$  into  $T_m$ , we could get the correlation of sentiment  $m$  for the users with degree  $k$ . As can be seen in Figure 6a, the sentiment correlation grows with  $k$ , especially for *anger* and *joy*, which illustrates that nodes with higher degrees in online social networks possess stronger emotion correlation with their neighborhoods. That is to say, having more friends in the social media indicates more significant sentimental correlation with the neighborhood. Specifically, the correlation of *anger* and *joy* are almost same for very small degrees, but later *anger* shows a significant jump for large degrees and enlarges the gap as compared to *joy*. As  $k$  raises to 30, the correlation of *anger* grows almost to 0.85. While the correlation of *sadness* and *disgust* do not demonstrate an obvious increasing trend and just fluctuate around 0.2 or even lower. Similarly, here errors also grow with increasing  $k$ , because nodes with large degrees only occupy a little fraction in the online social network [25], which would reduce the length of the emotion sequence and import noise, especially for *sadness* and *disgust* as shown in Figure 6a. However, *anger*'s

correlation is still significantly higher than that of *joy* as  $k > 15$ . It is also worthy emphasizing that because the network size is small and we only have the maximum degree around 30, which is far below the Dunbar's Number [25,26]. We suspect that the correlation might stop rising if the degree is larger than Dunbar's Number. Thirdly, we investigate the impact of the clustering of a node  $i$ , which is defined as  $\frac{2|E_i|}{k_i(k_i - 1)}$ , where  $E_i$  is the set of ties among  $i$ 's friends and  $k_i$  is the degree of  $i$ . For  $k_i = 1$ , we set  $i$ 's clustering to zero. Then similar to the case of degree, we could get the correlation variation as clustering grows for different emotions. As shown in Figure 6b, correlations of *anger* and *joy* grow very slowly with the clustering, while *disgust* and *sadness* just demonstrate fluctuations without obvious increment. The correlation of *anger* is still stronger than that of *joy*. However, different from the case of degree, even the correlations of *anger* and *joy* fluctuate as the clustering rises. While with respect to the error, it again shows a rising trend with the growth of clustering, since nodes with highly clustered neighborhoods take a trivial fraction in the social network [25]. Generally the above observations indicate that for *anger* and *joy*, the emotion correlation between a node and its neighborhood would be a little bit stronger as its neighbors are more closely clustered. The results of Spearman correlation are similar and not reported here.

To sum up, different emotions have different correlations in the social media. Compared to other sentiments, *anger* has the most positive correlation. Local structures can affect the sentiment correlation in near neighborhoods, from which we can learn that tie strength, node degree and node clustering could enhance the sentiment correlation, especially for *anger* and *joy*, and their contributions to *sadness* and *disgust* are greatly limited.

### Discussion

Users with similar demographics have high probabilities to get connected in both online and offline social networks. Recent studies reveal that even the psychological states like happiness are assortative, which means the happiness or well-being is strongly correlated between connected users in online social media like Twitter. Considering the oversimplification of the sentiment classification in the previous literature, we divide the emotion



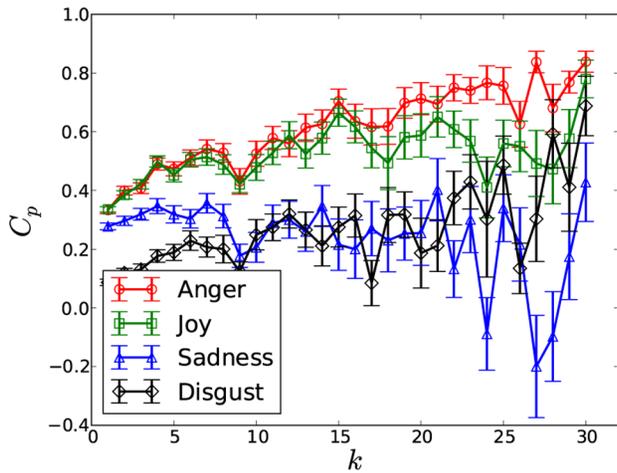
**Figure 5. Pearson correlations of different  $h$  for different networks extracted by varying  $T$ .** The case of  $h > 3$  is not considered here because of the weak sentiment correlation found in Figure 3. doi:10.1371/journal.pone.0110184.g005

into four categories and discuss their different correlations in details based on the tweets collected from Weibo of China, and the dataset has been publicly available to research community. Our results show that *anger* is more significantly correlated than other emotions like *joy*. While out of our expectation, the correlation of *sadness* is low.

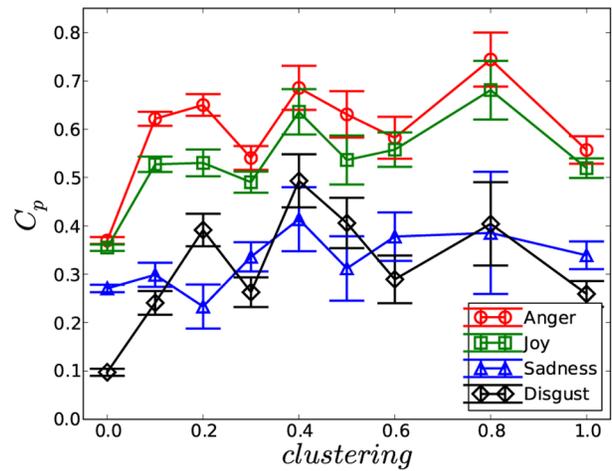
We try to unravel the underlying reason of why *anger* has a surprisingly high correlation but the correlation of *sadness* is weak from the view of keywords the corresponding tweets present. For a certain emotion  $m$ , we collect all the retweeted tweets (usually contain phrase like “@” or “retweet”) with this sentiment in a specified time period to combine into a long text document. Focusing only on retweeted tweets could help reduce the impact of external media and just consider the contagion from the social ties in Weibo. Several typical techniques are employed to mine the keywords or topic phrases from the documents [27], which are reported in Figure 7. Based on the keywords or topics we find, the

real-world events or social issues could be summarized to understand the sentiment correlation in detail.

With respect to *anger*, we find two kinds of social events are apt to trigger the angry mood of users in Weibo. First one is the domestic social problems like food security, government bribery and demolition for resettlement. The “shrimp washing powder” which results in muscle degeneration and the self-burning event in Fenggang Yihuang County of Jiangxi province represent this category. These events reflect that people living in China are dissatisfied about some aspects of the current society and this type of event can spread quickly as the users want to show their sympathy to the victims by retweeting tweets and criticizing the criminals or the government. Frequently appearing phrases like “government”, “bribery”, “demolition” and so on are strongly related with these events. The second type is about the diplomatic issues, such as the conflict between China and foreign countries. For instances, in August 2010, United States and South Korea



(a) Node Degree



(b) Node Clustering

**Figure 6.** Here  $T$  is fixed to 10 to reduce the data sparsity. Because the network is relatively small, the largest degree we get is only 30. Therefore, the results in Figure 6a just demonstrate that when the degree is small, how the sentiments' correlations vary with node degrees. While regarding to Figure 6b, the linear bin is used to get emotion sequences for nodes with clusterings within the same bin. doi:10.1371/journal.pone.01110184.g006

held a drill on the Yellow Sea, which locates in the east of China. In September 2010, the ship collision of China and Japan also made users in Weibo extremely rageful. Actually, these events could arouse patriotism and stimulate the angry mood. Keywords like “Diaoyu Island”, “ship collision” and “Philippines” show the popularity of these events at that time. To sum up, Weibo is a convenient and ubiquitously channel for Chinese to share their concern about the continuous social problems and diplomatic issues. Pushed by the real-world events, these users tend to retweet tweets, express their anger and hope to get resonance from neighborhoods in online social networks. While regarding to *sadness*, we find its strength of correlation is strongly affected by the real-world natural disasters like earthquake, as shown in Figure 7(right). Because the natural disaster happens occasionally and then the averaged correlation of the sadness is very low and the strength of its correlation might be highly fluctuated.

With the continuous growth, online social media in China like Weibo have been becoming the primary channel of information

exchange. In Weibo, the messages do not only deliver the factual information but also propagate the users' opinions about the social event or individual affairs. Real-world society issues are easy to get attention from the public and people tend to express their feelings towards these issues through posting and retweeting tweets in online social media. Through keywords and topics mining in retweeted angry tweets, we find the public opinion towards social problems and diplomatic issues are always angry and this extreme mental status also possesses the strongest correlation between connected users in Weibo. We conjecture that *anger* plays a non-ignorable role in massive propagations of the negative news about the society, which are always hot trends in today's Internet of China. This might be the origin of large scale online collective behavior in Weibo about society problems such as food security and demolition for resettlement in recent years. It is also consistent with a finding that good news never goes beyond the gate while bad news spread far and wide by ancient Chinese people more than one thousand years ago [28]. It should also be mentioned that



**Figure 7.** The example Chinese keywords extracted for *anger*(left) and *sadness*(right), respectively. The top 20 keywords are also translated into English, which could be found through <http://goo.gl/tl4q45>. doi:10.1371/journal.pone.01110184.g007

the study of this paper has received great attention from many media like MIT Technology Review, CNN, BBC and the Washington Post (just to name a few of them) after a preprint of this paper was posted on the Internet [29]. It is widely believed that similar results might also exist in the social media of other countries. In the future work we would investigate the role of emotion in the information diffusion and comprehensively understand how different sentiments function in the formation of the public opinion or the massive collective behavior. Meanwhile, our findings could inspire the modeling of emotion contagion, like different emotions might diffuse with diverse strengths, local structures such as tie strengths, degrees and clusterings might affect the spread and the emotion might only function between individuals with social distance no more than three hops. Another interesting direction is to study how to make the social media more

neutral by introducing some new mechanisms, e.g. delaying the post of an angry tweet can give people additional time for consideration and so might be of use for reducing the number of angry tweets.

## Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable suggestions, which improve this paper greatly.

## Author Contributions

Performed the experiments: RF JZ YC. Analyzed the data: RF JZ YC. Wrote the paper: JZ KX. Designed the experiments: JZ RF KX. Prepared the figures: RF JZ YC.

## References

1. Miller M, Lynn SL, James CM (2001) Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27: 415–444.
2. Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B (2007) Measurement and analysis of online social networks. In: the 7th ACM SIGCOMM conference on Internet measurement. IMC '07, pp. 29–42.
3. Liben-Nowell D, Kleinberg J (2003) The link prediction problem for social networks. In: the twelfth international conference on Information and knowledge management. CIKM '03, pp. 556–559.
4. Bollen J, Gonçalves B, Ruan G, Mao H (2011) Happiness is assortative in online social networks. *Artif Life* 17: 237–251.
5. Chmiel A, Sobkowicz P, Sienkiewicz J, Paltoglou G, Buckley K, et al. (2011) Negative emotions boost user activity at BBC forum. *Physica A: Statistical Mechanics and its Applications* 390: 2936–2944.
6. Bliss CA, Kloumann IM, Harris KD, Danforth CM, Dodds PS (2012) Twitter reciprocal reply networks exhibit assortativity with respect to happiness. *Journal of Computational Science* 3: 388–397.
7. Traud AL, Kelsic ED, Mucha PJ, Porter MA (2008) Comparing community structure to characteristics in online collegiate social networks. *ArXiv e-prints*, arXiv:0809.0690.
8. Derks D, Fischer AH, Bos AE (2008) The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior* 24: 766–785.
9. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM (2011) Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PLoS ONE* 6: e26752.
10. Golder SA, Macy MW (2011) Diurnal and seasonal mood vary with work, sleep, and day length across diverse cultures. *Science* 333: 1878–1881.
11. Marchetti-Bowick M, Chambers N (2012) Learning for microblogs with distant supervision: Political forecasting with twitter. In: 13th EACL. pp. 603–612.
12. Bollen J, Pepe A, Mao H (2011) Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: Fifth ICWSM.
13. Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. *Journal of Computational Science* 2: 1–8.
14. Zhao J, Dong L, Wu J, Xu K (2012) Moodlens: an emoticon-based sentiment analysis system for Chinese tweets. In: KDD '12. pp. 1528–1531.
15. Xiong X, Zhou G, Huang Y, Haiyong C, Xu K (2013) Dynamic evolution of collective emotions in social networks: a case study of sina weibo. *Science China Information Sciences* 56: 1–18.
16. Dodds PS, Danforth CM (2010) Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies* 11: 441–456.
17. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up?: sentiment classification using machine learning techniques. In: EMNLP. pp. 79–86.
18. Parikh R, Movassate M (2009) Sentiment analysis of user-generated twitter updates using various classification techniques. Technical report.
19. Read J (2005) Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: ACLstudent. pp. 43–48.
20. Go A, Bhayani R, Huang L (2011) Twitter sentiment classification using distant supervision. Technical report, Stanford Digital Library Technologies Project.
21. Liu B (2010) Sentiment analysis and subjectivity. In: *Handbook of Natural Language Processing*, Second Edition. Taylor and Francis Group, Boca.
22. Aoki S, Uchida O (2011) A method for automatically generating the emotional vectors of emoticons using weblog articles. In: Proceedings of the 10th WSEAS international conference on Applied computer and applied computational science. ACACOS'11, pp. 132–136.
23. Tossell CC, Kortum P, Shepard C, Barg-Walkow LH, Rahmati A, et al. (2012) A longitudinal study of emoticon use in text messaging from smartphones. *Computers in Human Behavior* 28: 659–663.
24. Shalizi C (2010) The bootstrap. *American Scientist* 98: 186–190.
25. Zhao J, Wu J, Liu G, Xu K, Chen G (2010) Being rational or aggressive? A revisit to Dunbar's Number in online social networks. *ArXiv e-prints*, arXiv: 1011.1547.
26. Dunbar R (1998) *Grooming, Gossip, and the Evolution of Language*. Harvard University Press, Cambridge, MA.
27. Li C, Sun A, Datta A (2012) Twevent: Segment-based event detection from tweets. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management. CIKM '12, pp. 155–164.
28. Sun GX (2012) *Tang and Five Dynasties*. Bei meng suo yan. A book of short stories from the Tang and Five Dynasties.
29. Fan R, Zhao J, Chen Y, Xu K (2013) Anger is more influential than joy: Sentiment correlation in Weibo. *ArXiv e-prints*, arXiv: 1309.2402.