Microblog sentiment analysis using social and topic context

Xiaomei Zou, Jing Yang*, Jianpei Zhang

School of Computer Science and Technology, Harbin Engineering University, Harbin, Heilongjiang, China

* yangjing@hrbeu.edu.cn

Abstract

Analyzing massive user-generated microblogs is very crucial in many fields, attracting many researchers to study. However, it is very challenging to process such noisy and short microblogs. Most prior works only use texts to identify sentiment polarity and assume that microblogs are independent and identically distributed, which ignore microblogs are networked data. Therefore, their performance is not usually satisfactory. Inspired by two sociological theories (sentimental consistency and emotional contagion), in this paper, we propose a new method combining social context and topic context to analyze microblog sentiment. In particular, different from previous work using direct user relations, we introduce structure similarity context into social contexts and propose a method to measure structure similarity. In addition, we also introduce topic context to model the semantic relations between microblogs. Social context and topic context are combined by the Laplacian matrix of the graph built by these contexts and Laplacian regularization are added into the microblog sentiment analysis model. Experimental results on two real Twitter datasets demonstrate that our proposed model can outperform baseline methods consistently and significantly.

Introduction

It is a very challenging task to get users' real sentiment from large collections of short user-generated social media contents (e.g. microblogs). It is also of great value and has a wide range of application prospects to mining users' sentiment, such as customer relationship management, recommendation systems, and business intelligence [1–3]. The automatic sentiment analysis task usually requires the machine to have a deep understanding of natural language [4], which has achieved some satisfactory performances in long formal text sentiment analysis [5–8]. However, its performance drops sharply when it is applied to microblog sentiment analysis as it assumes that texts are independent and identically distributed (i.i.d.). Compared with long formal texts, microblogs are much shorter and have various expression style, e.g., 'lol' and 'It is so cooooooooool!', which exacerbates the problem of vocabulary sparsity. On the other hand, social media provides different types of metadata, such as user relations, which can be leveraged to improve the accuracy of microblog sentiment analysis.

Studying the influence of other metadata beyond texts (called social context) on microblog sentiment analysis has recently attracted much attention of many researchers, for example,
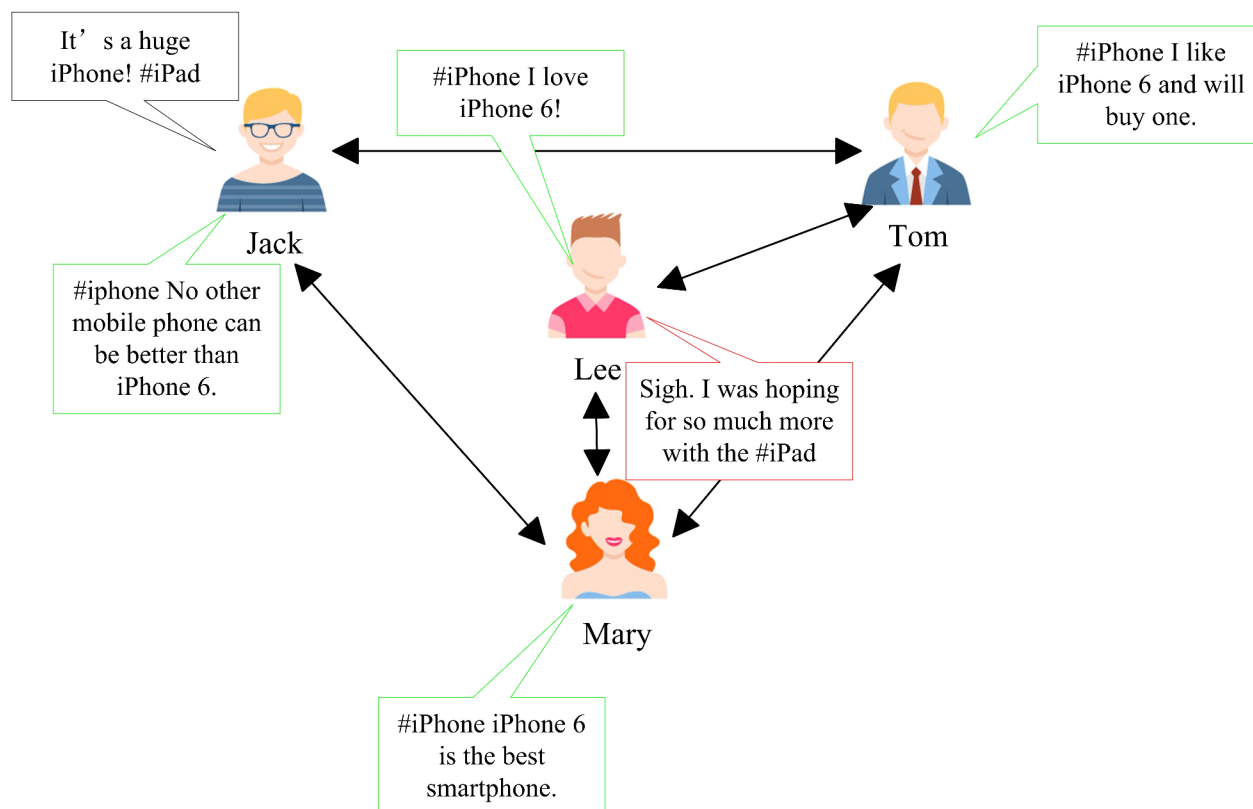
introducing user direct relations to microblog sentiment analysis models [9, 10]. There are two basic sociological theories: sentiment consistency [11] and emotional contagion [12] to support these methods. As an aspect of social context, sentiment consistency, which is called user context, indicates that microblogs posted by the same person tend to have the same sentiment label. Emotional contagion implies that similar people tend to have the same opinion, and it is usually called friends context, which is also an aspect of social context. Although there are already works [9, 10] which exploit social context for sentiment analysis in microblogging, they only take the effects of user direct relationships on sentiment analysis into account, ignoring the impact of user indirect relationships. However, connections in a social network are heterogeneous in nature [13–15], so it is not enough to analyze microblog sentiment by only using user direct relationships. Here is an example. In Fig 1, a green dialog box represents that the sentiment of its corresponding text is positive while the sentiment of the text in a red dialog box is negative. The text in the black dialog box represents the one needed to be classified. There are no direct connections between Jack and Lee, but they have two common friends (Mary and Tom). All users have a positive opinion for iPhone 6. Jack posts a tweet about iPad: "It's a huge iPhone!" which is a negative comment towards iPad. However, it is difficult to recognize its polarity for a machine from its literal meaning. Besides, if we use direct relationships between users in this graph to assist sentiment analysis, we still can not classify this text into the right class as Lee's friends (Mary and Tom) have no comments on iPad, which results in a classification error.

Recently, indirect relationships between users have been applied into recommendation systems [16, 17]. The basic idea of these works is that similar users have the same preferences or behavior habits. However, there is little literature that studies the usefulness of indirect relationships in sentiment analysis. At the same time, with the development of sociological theory, homophily [18, 19] has received much more attention [17, 20]. It is the principle that a contact between similar people occurs at a higher rate than among dissimilar people [20], which has a great influence on the formation of friendships. As a result of homophily, the information such as culture and behavior that flows through the network tends to be localized. In addition, [21] has found some evidence of both positive and negative sentiment homophily in social networks.

Inspired by these works, we propose our own method: using indirect relations, in particular, user structure similarity to analyze microblog sentiment. Our method is based on an assumption: opinions of similar users should be similar, and we experimentally verify this assumption. First, we find similar users through common friend relationships and establish a similarity context matrix. It is a common practice of finding similar users through common friends [17, 20, 22, 23] and similarity breeds new connection [19]. Further, two users who may have a new connection between them may share the same opinion [20]. The essence of our method is to look for potential user relationships that may be friends, and then take them into account in the sentiment analysis model. Second, topic factors are introduced and a topic context matrix is established. The phenomenon of homophily is more significant on the same topic [24], and in turn, the topic context can better exploit the theory of homophily. Finally, the structure similarity context and topic context are combined into a graph model, and the Laplacian matrix of this graph is used to analyze microblog sentiment. Go back to the example given in Fig 1 again. Jack and Lee have two common friends. According to our assumption, they have a certain probability of becoming friends, so they may share the same sentiment with a certain probability. Therefore, due to Lee's negative comment on #iPad, Jack may also have a negative comment on #iPad, then the accuracy of sentiment analysis is guaranteed.

Fig 1. An example. Green dialog boxes represent the corresponding texts are positive, while red dialog boxes represent the corresponding texts are negative.

The main contributions of this paper include:

1. Proposing a method using structure similarity to model homophily in social networks.

2. Introducing structure similarity into social context of microblogs as a substitute to user direct relations.

3. Introducing topic context to model the semantic relations between microblogs.

4. Proposing a novel microblog sentiment analysis model which incorporates user context, structure similarity context, topic context and text information.

5. Evaluating the proposed model extensively using real-world datasets to understand the working of the proposed model.

The remainder of this paper is organized as follows. In Section 2, several related works are introduced. In Section 3, we define the problem we study and propose our model. In Section 4, the experimental results are presented. In Section 5, we conclude the whole paper.

Related work

In this section, we review some related works about sentiment analysis and microblog sentiment analysis.

Sentiment analysis

Existing approaches to sentiment analysis fall into two main categories: lexicon-based methods and machine learning methods. Lexicon-based methods [25–32] usually utilize lexicon such as SentiWordNet [33], SenticNet [34] to tag words occurring in the sentence into positive and negative labels, then the sentiment of the whole document is judged by summarizing the tagged words. Lexicon-based methods are unsupervised which don't need datasets with polarity labels. However, these methods rely on lexicons too much and are domain-related for the sentiment polarity of words varies from domain to domain.

Machine learning methods regard sentiment analysis as a text classification problem [35–42]. In these methods, features such as unigrams, bigrams, word embeddings are extracted from the text firstly and then features are fed to classification models such as SVM, NB and deep neural networks (CNNs, RNNs) and so on. Machine learning methods are supervised and usually need lots of training data with polarity labels. The accuracy of sentiment classification is related to the size of training data.

Microblog sentiment analysis

Microblog sentiment analysis has become a hot research topic in these years [10, 43, 44]. Because microblogs are short and noisy, many methods are proposed to solve this problem. [45] used emoticons as features to analyze the sentiment of tweets. In [46], generalized emoticons, repeated punctuations, and repeated words were used to build a co-occurrence graph by label propagation algorithm and the co-occurrence graph was used to identify the sentiment polarities of tweets. [47] built a sentiment lexicon using the relations between words and emoticons, then they used the lexicon to extract sentiment features and analyze microblogs. All these methods mentioned above utilize text information only and ignore the extra information provided by the microblog media.

In recent years, there are more and more research works about how to utilize user information to analyze sentiment. [10] proposed a method using user follow relations and '@' information to identify the sentiment of users on Twitter. [48] took sentiment analysis of users to a specific topic as a problem of collaborative filtering, relations between users were applied to predict sentiment of users. Similarly, [49] also exploited user relations graph. The classification results of the maximum entropy model were used as labels and then the authors implemented label propagation algorithm to identify sentiment. These works are user-level or user-topic level sentiment classification methods, while our methods are microblog-level. In [9], Hu et al. proposed a framework named SANT (a **S**ociological **A**pproach to handling **N**oisy and short **T**exts) combining social context to classify sentiment of microblogs. On the basis of [9], [50] added content similarity to the framework of SANT and proposed a semi-supervised method to identify sentiment of tweets. [51] argued the framework proposed by [9] was a purely content-based approach so they proposed a **S**tructured **M**icroblog **S**entiment **C**lassification (SMSC) framework which used social context at the prediction stage. There are also works which introduced user relations into microblog retrieval [52, 53]. However, all these methods employ direct user relations and ignore user similarity. Base on the observation in Section 1, two users who have common friends may share the same sentiment with each other, which means using direct user relations only are not enough for sentiment analysis.

Model

Datasets

In this paper, our experiments are conducted on two Twitter sentiment analysis benchmark datasets: HCR and OMD. Many proposed works use these two datasets to evaluate the performance of using social relations for sentiment analysis. These two datasets include raw texts and sentiment labels which are labeled manually.

HCR: This dataset is collected by [49]. It includes tweets about health care reform of America in March 2010. It has three parts: a training set, a development set, and a test set. There are five kinds of labels in the dataset: positive, negative, neutral, irrelevant and unsure and this corpus was manually annotated by the authors. In this paper, we only use tweets with positive and negative labels. We use the complete follower graph built by [54] in 2009 to construct the user relations of HCR and take the graph as undirected. The dataset has 9 different topics, i.e. health care reform, Obama, Republicans, Democrats, conservatives, liberals, Tea Party, Stupak and other. Each microblog corresponds to one of these targets.

OMD: This dataset is built by [55]. It consists of tweets discussing the US Presidential Debates between Barack Obama and John McCain. This dataset is manually labeled by Amazon Mechanical Turk. Every tweet is tagged by at least three Turkers and its inter-annotator agreement is 0.655 reported in [55], which shows a relatively good agreement between annotators. Four kinds of labels appear in the dataset, they are positive, negative, mixed and irrelevant. We use majority voting to determine the final label of each tweet. The same as HCR, we only use tweets with positive and negative labels. The relation graph is also built by using the follower graph clawed by [54] in 2009. Microblogs in this dataset can be divided into three topics by using keywords, i.e., Obama (containing keyword "Obama" but no "McCain"), McCain (containing keyword "McCain" but no "Obama"), and debate (containing both "Obama" and "McCain" or none of them).

In this paper, we reserve users who have friends and delete those microblogs whose author has no friends. The information about the two datasets is shown in Table 1.

Notation

In this paper, uppercase letters like B are used to denote matrices, lowercase bold letters like $\mathbf{x}$ denote vectors. Lowercase letters like a denote scalar values. We use B_{i*} to denote the i-th row of matrix B and B_{*j} to denote the j-th column of matrix B. The entry at the i-th row and j-th column is denoted as B_{ij}. B^T is used as the transposition of matrix B. $\|B\|_F$ denotes the Frobenius norm of B. $tr(.)$ is the trace of a matrix.

The goal of this paper is by using the training set feature matrix $X \in R^{n \times m}$ (where n represents the number of microblogs in training set, m represents the number of features.) and label matrix $Y \in R^{n \times c}$ (where c is the number of sentiment polarities) to construct a classifier

Table 1. Statistics of datasets.

Emoticon	HCR	OMD
# of Tweets	1434	1184
# of users	806	636
# positive Tweets	387	475
Average Tweets per User	1.78	1.86
Average Friends per User	14.95	5.54

Table 2. Meaning of variables.

Variables	Meaning	Type
uppercase letter like B	matrix	–
lowercase letters like b	scalar	–
lowercase bold letters like $\mathbf{b}$	vector	–
B_{*i}	the i-th column of matrix B	
B_{i*}	the i-th row of matrix B	
X	feature matrix	$R^{n \times m}$
Y	ground truth label matrix	$R^{n \times c}$
$\hat{Y}$	fitted sentiment label matrix	$R^{n \times c}$
n	number of features	integer
m	number of training set	integer
c	number of sentiment classification	integer
t	number of topics	integer
W	classifier	$R^{m \times c}$
$\mathbf{x}$	feature vector of a microblog	R^m
U	user-microblog matrix	$R^{d \times n}$
d	number of users	integer
S	structure similarity matrix	$R^{d \times d}$
T	microblog-topic matrix	$R^{n \times t}$
M	microblog-microblog topic matrix	$R^{n \times n}$
A	microblog-microblog relation matrix	$R^{n \times n}$
D	diagonal matrix	$R^{n \times n}$
L	Laplacian matrix	$R^{n \times n}$
F	user-user direct relation matrix	$R^{d \times d}$
G	graph	
E	edges	
V	nodes	

$W \in R^{m \times c}$, and then classifier $W \in R^{m \times c}$ was used to predict an unseen microblogs $\mathbf{x}$. Y represents ground truth labels of microblogs. We use $\hat{Y} = XW \in R^{n \times c}$ to represent the fitted value of the ground truth label matrix Y. In this paper, we only consider binary classification of sentiments, that is, $c = 2$. Therefore, if a microblog is positive, then its ground truth label is $Y_{i*} = [+1\ -1]$. And if the sentiment of a microblog is negative, then its label is $Y_{i*} = [-1\ +1]$.

Given an undirected graph $G = (V, E)$, A represents its adjacency matrix, $L = D - A$ represents the Laplacian matrix of G [56], where D is diagonal matrix and D_{ii} indicates the degree of the i-th vertex.

To classify an unseen microblog, we use the prediction function in Eq (1). Variables and their meanings are shown in Table 2.

$$
g(\mathbf{x}) = \begin{cases} +1 & if \quad \mathbf{x}W_{*1} > \mathbf{x}W_{*2} \\ -1 & if \quad \mathbf{x}W_{*1} < \mathbf{x}W_{*2} \\ +1 \quad or \quad -1 \quad randomly & if \quad \mathbf{x}W_{*1} = \mathbf{x}W_{*2} \end{cases} \tag{1}
$$

Modeling microblog content

The popular method Least Squares is applied to fit the classification model for text information. In terms of multiclass classification tasks, the Least Squares aims to learn c classifiers by solving the Eq (2) optimization problem:

$$\min_W \frac{1}{2} \| XW - Y \|_F^2 \tag{2}$$

Unlike traditional text information, microblogs are short and have many noises which lead to a sparse matrix of unigrams. To handle this problem, we use sparse regularization L_1 norm to seek a sparse reconstruction of the feature space. To minimize the L_1 norm based linear reconstruction error can implement feature selection automatically and get a sparse representation of texts [57]. Thus, we add L_1 norm in our model to get a more robust model (see Eq (3)).

$$\min_W f(W; X, Y) = \min_W \frac{1}{2} \| XW - Y \|_F^2 + \beta \| W \|_1 \tag{3}$$

where β is the weight of regularization.

Context besides text

In this section, we will introduce the different contexts used in this paper and integrate them into the final model.

Topic context. In this section, we introduce the topic context. Hashtags are a type of mechanisms provided by microblogging services, by which users can insert topic information into microblogs conveniently. For example, in a tweet (Twitter microblog message), the symbol # is used to tag topics in a tweet. A tweet "I love #iPhone6" indicates that this tweet is about "iPhone6". Users post various microblogs towards different topics in social media as a way to express themselves. Although different users may have different opinions towards the same topic and a user may hold different opinions towards different topics, the opinions of a same person on the same topic usually consistent with each other. In addition, similar users tend to hold similar opinions towards the same topic. Topic context is used to indicate whether two microblog messages are related to the same topic. It is important to introduce topic information into microblog sentiment analysis as it models the semantic connections between microblogs. It is noted that we use topics not text similarity to model this semantic relation. This is because the data representation is very sparse in microblogging platform, and if we use text similarity the values of semantic similarity between microblogs will be very small which cannot model the semantic relation efficiently. We can get a microblog-microblog matrix M towards topics using Eq (4), where $T \in R^{n \times t}$ is the microblog-topic matrix and $T_{ij} = 1$ if and only if the i-th microblog is about the j-th topic.

$$M = T \times T^T \tag{4}$$

$M_{ij} = 1$ if and only if microblog p_i and p_j are about the same topic. The diagonal elements of M are set to zeros.

User context. User context is based on a sociological theory called sentiment consistency. Sentiment consistency suggests that the sentiments of two microblogs posted by the same user have a higher probability to be the same than those of two randomly selected microblogs, which has been verified in [9] and [10]. $A_{sc} \in R^{n \times n}$ represents microblog-microblog matrix for sentiment consistency. We can use Eq (5) to calculate A_{sc}. $U \in R^{d \times n}$ is a user-microblog matrix
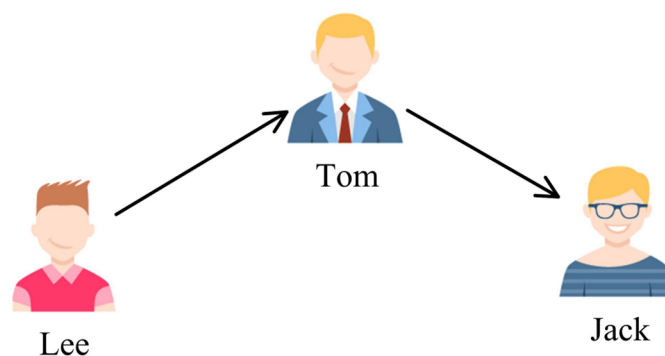
where $U_{ij} = 1$ if the i-th user posts the j-th microblog and d is the number of users.

$$A_{sc} = U^T \times U \qquad (5)$$

$A_{scij} = 1$ if and only if microblog p_i and p_j are posted by the same user.

Structure similarity context. This part is also based on a basic theory of sociology: emotional contagion, which discloses that the sentiments of two microblogs posted by similar users have a higher probability than those of two randomly selected microblogs. In previous work, if two microblogs are posted by two users connected with follower/friend relationships, a model is built to make the sentiments of these two microblogs as close as possible. It is called friends context, represented by $A_{ec} = U^T \times F \times U$, where $F \in R^{d \times d}$ represents a user-user matrix and $F_{ij} = 1$ if there exists a following/followee relation between the i-th user and the j-th user. However, previous works only use direct relations between users and ignore common friendships between users. As discussed in Section 1, a user may also share the same opinion with the user who is a friend of his friends, which is an expression of homophily. Therefore, in this part, we use structure similarity which takes common friendships into consideration to model the emotional contagion theory. Common friendships often induce new friendships [58]. In real life, if B and C have a common friend A, the probability of becoming friends between them increases. This principle is called "Triadic Closure" [59, 60]. One of the reasons for triadic closure is the fact that both B and C are friends of A (as they all know it) provides them with the basic trust that is lacked among strangers during the formation of friendships. The second reason is based on A's incentive: it can decrease the latent stress of A in two separate relationships to bring B and C together.

There are three cases that three users are connected by two following relations in Twitter. This is shown in Figs 2, 3 and 4, where the user pointed by the arrow is a followee and the user at the other end of the arrow is a follower. The first case (Fig 2) represents the process of the flow of information, an opinion may flow from Jack to Lee through Tom. The second one (Fig 3) describes a situation that two users share a common followee. Sharing a common followee emphasizes the establishment of friend-of-friend relationships, which means that the more common followees between two users, the easier it is to build a following relation between them. The third one (Fig 4) demonstrates the situation that two users have a common follower. This case reflects the similarity of two users' image and attractiveness. No matter what the case is, all the three cases are an expression of user similarity, which implies the

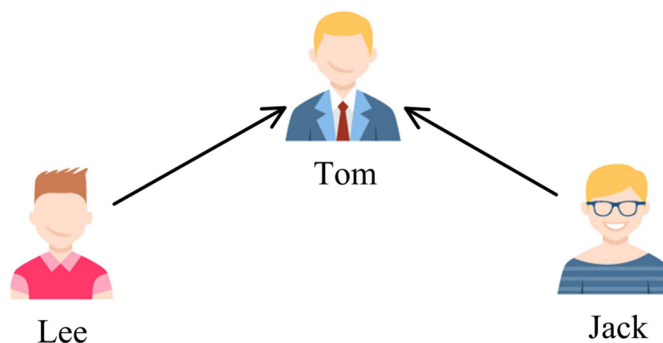**Fig 2. Different relation types in Twitter: The first case.**

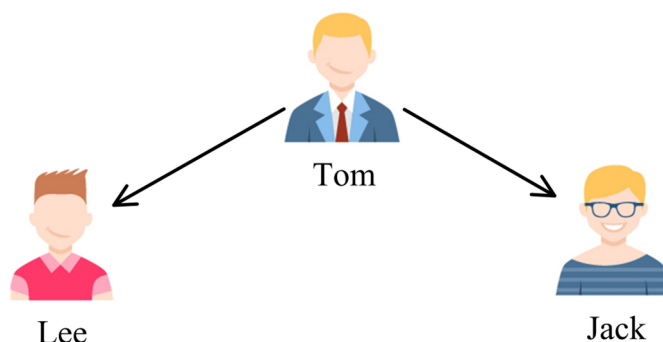**Fig 3. Different relation types in Twitter: The second case.**

Fig 4. Different relation types in Twitter: The third case.

possibility of the formation of a friendship between the two unconnected users. For this reason, we can take the follower graph as undirected.

Given two users u_i and u_j, their structure similarity can be calculated by Eq (6).

$$S_{ij} = Sim(u_i, u_j) = |N_{u_i} \bigcap N_{u_j}| \qquad (6)$$

The structure similarity is measured by two users' common friends. N_{u_i} represents neighbours of user u_i. $|N_{u_i} \bigcap N_{u_j}|$ represents the number of u_i and u_j's common friends. However, considering a condition, shown in Fig 5, user 2 and user 4 have two common friends 1 and 3. Compared with Fig 5, in Fig 6, user 2 has many more friends. If we use Eq (6) to compute the structure similarity between user 2 and user 4 of Fig 6, we will get the same similarity value as that in Fig 5. To handle this problem, we use Eq (7) in which all neighbors of two users are taken into consideration to compute the structure similarity.

$$S_{ij} = Sim(u_i, u_j) = \begin{cases} \dfrac{|N_{u_i} \bigcap N_{u_j}|}{|N_{u_i} \bigcup N_{u_j}|} & for \quad two \quad unconnected \quad users. \\[2ex] \dfrac{|N_{u_i} \bigcap N_{u_j}|}{|N_{u_i} \bigcup N_{u_j}|} + 1 & for \quad two \quad connected \quad users. \end{cases} \qquad (7)$$
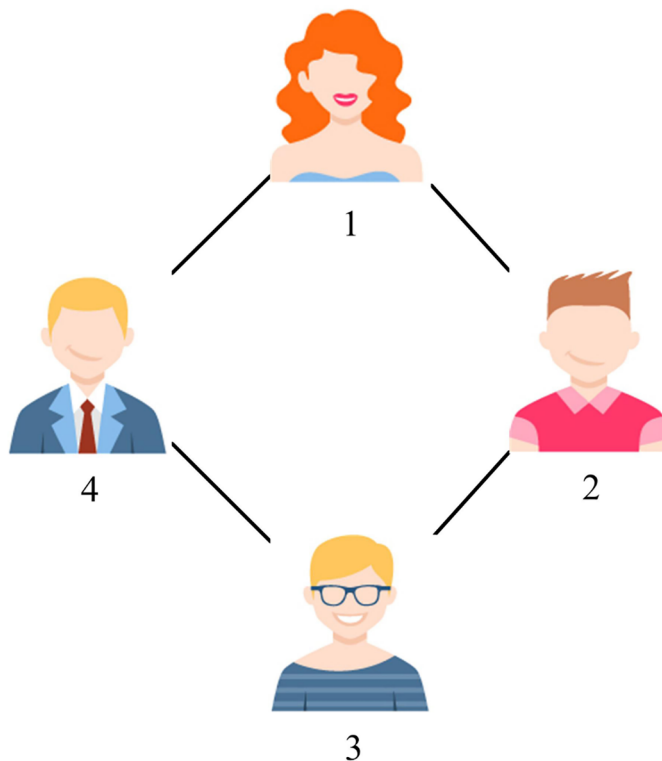
Fig 5. Example of similarity (a).

where $N_{u_i} \bigcup N_{u_j}$ represents the union set of friends of both user u_i and u_j and $|N_{u_i} \bigcup N_{u_j}|$ represents the number of users in the union set.

After getting user structure similarity matrix S, the emotional contagion matrix A_{ec} can be computed by Eq (8).

$$A_{ec} = U^T \times S \times U \tag{8}$$

Incorporating structure similarity context

In this section, we combine the three kinds of contexts into our framework. $A_1 \in R^{n \times n}$ is used to represent the combination of user context and stucture similarity context. It can be calculated by Eq (9). $A_2 \in R^{n \times n}$ represents the combination of user context, structure similarity context, and topic context. Eq (10) can be used to compute A_2. We set $\theta = 1$.

$$A_1 = A_{sc} + \theta * A_{ec} \tag{9}$$

$$A_2 = (A_{sc} + \theta * A_{ec}) \circ M \tag{10}$$

where $\circ$ represents Hadamard product.

We use the SANT framework proposed by [9]. Based on sentiment consistency and emotional contagion, to integrate sentiment relations between microblogs in sentiment classification, the basic idea is to make two microblogs as close as possible if they are posted by the
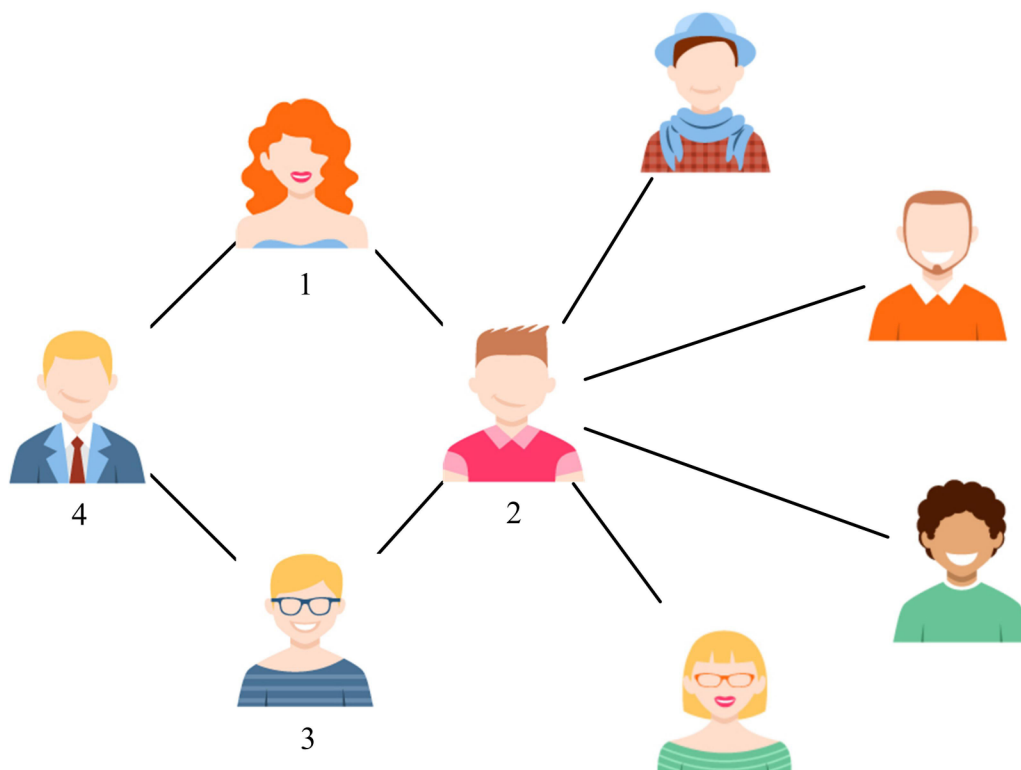
Fig 6. Example of similarity (b).

same user or two users are very similar with each other. In this situation, we can solve the problem by minimizing Eq (11).

$$\min_{W} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \| \hat{Y}_{i*} - \hat{Y}_{j*} \|^2$$

$$= \min_{W} \sum_{k=1}^{c} \hat{Y}_{*k}^{T} (D - A) \hat{Y}_{*k} \qquad (11)$$

$$= \min_{W} tr(W^T X^T L X W)$$

If we only use user context and structure similarity context, $A = A_1$. If topic context is used, $A = A_2$. So the final model which combines text information and social context can be represented by Eq (12).

$$f(W; X, Y) = \min_{W} \frac{1}{2} \| XW - Y \|_F^2 + \frac{\alpha}{2} tr(W^T X^T L X W) + \beta \| W \|_1 \qquad (12)$$

where α is the weight of social context in the model, β is the weight of regularization.

Learning

Motivated by [61], we solve the non-smooth optimization problem in Eq (12) by optimizing its equivalent smooth convex reformulations. Firstly, Eq (12) can be reformulated as Eq (13) as

a constrained smooth convex optimization problem.

$$\min_{W \in Z} L(W; X, Y) = \frac{1}{2} \| XW - Y \|_F^2 + \frac{\alpha}{2} tr(W^T X^T L X W),$$

$$where \quad Z = \{W | \| W \|_1 \leq z\} \tag{13}$$

$L(W;X, Y)$ is the differentiable part and Z is the non-differentiable part. $z \geq 0$ is the radius of the L_1-ball, and there is a one-to-one correspondence between β and z.

The smooth part of the optimization problem can be reformulated equivalently as a proximal regularization [62] of the linearized function $L(W;X, Y)$ at W_t, which is formally defined as:

$$W_{t+1} = \arg \min_W G_{\lambda_t, W_t}(W)$$

$$where \quad G_{\lambda_t, W_t}(W) = L(W_t; X, Y) + \tag{14}$$

$$< \bigtriangledown L(W_t; X, Y), W - W_t > + \frac{\lambda_t}{2} \| W - W_t \|_F^2$$

where λ_t is the step size in the t-th iteration. In this paper, the gradient of $L(W;X, Y)$ with respect to W can be computed using Eq (15).

$$\bigtriangledown L(W; X, Y) = X^T(XW - Y) + \alpha X^T L X W \tag{15}$$

When considering the constraints Z in Eq (13), and given β, the $(t+1)$-th W can be computed by Eq (16).

$$(W_{t+1})_{j*} = \begin{cases} \left(1 - \frac{\beta}{\lambda_t \| (U_t)_{j*} \|}\right)(U_t)_{j*}, if \; \| (U_t)_{j*} \| \geq \frac{\beta}{\lambda_t} \\ 0, otherwise \end{cases} \tag{16}$$

where $U_t = W_t - \frac{1}{\lambda_t} \bigtriangledown L(W_t; X, Y)$. As discussed in [61], to achieve the optimal convergence, we can further accelerate our constrained smooth convex optimization problem. In particular, two sequences W_t and V_t are used in this accelerated algorithm. W_t is the sequence of approximate solutions, and V_t, an affine combination of W_t and W_{t-1}, is the sequence of search points. V_t can be computed by Eq (17).

$$V_t = W_t + \gamma_t(W_t - W_{t-1}) \tag{17}$$

Where γ_t is the combination coefficient. The approximate solution W_{t+1} is computed as a "gradient" step of V_t through $G_{\lambda_t, Vt}$. We use Nesterov's method [63] to solve the optimization problem. The details are shown in Algorithm 1 in which η_t is set according to [61].

Algorithm 1 SASS: Sentiment analysis using structure similarity

```
Input: X, Y, L, α, β
Output: W
1. Initialize W₀ by random
2. Set η₀ = 0, η₁ = 1, W₁ = W₀, t = 1
3. while not convergent do
4.     Compute Vₜ = Wₜ + (ηₜ₋₁-1)/ηₜ (Wₜ - Wₜ₋₁)
5.     Compute ▽L(Wₜ;X, Y)
6.     while True do
7.         Compute Uₜ = Vₜ - (1/λₜ) ▽L(Wₜ;X, Y)
8.         Compute Wₜ₊₁ according to Eq (16)
```

```
9.         if L(W_{t+1}, X, Y)≤G_{λ_t, vt}(W_{t+1}) then
10.            Set λ_{t+1} = λ_t
11.            Break
12.        end if
13.         Set λ_t = 2 × λ_t
14.     end while
15.     if t > MaxIter then
16.         Return W_{t+1}
17.     end if
18.     Set η_{t+1} = (1+√(1+4η_t))/2
19.     Set t = t+1
20. end while
```

Experiments

In this section, we present empirical evaluation results to assess the effectiveness of our proposed framework. In particular, we evaluate the proposed method on the two datasets introduced in Section 3. Impacts brought by different contexts and parameters are further discussed.

Correlation between structure similarity and sentiment

The positive relation between friends context and microblogs sentiment labels are verified in [9] and [10]. In this paper, we also engage in a statistical study of the degree to which structure similarity and microblogs sentiment labels correlate. Given the unweighted graph $G = (V, E)$ built on microblog-microblog relations, we compute the ratio of edges whose corresponding nodes have the same sentiment labels to all edges in E, denoted by $\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}(Y_{i*}=Y_{j*}, e_{ij} \in E)}{\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}(e_{ij} \in E)}$ where $\mathbf{1}(.)$ is the indicator function. Given an weighted graph G, we can also compute this ratio by taking its weights matrix into consiration, so the ratio can be computed by Eq (18). In Eq (18), weights are regarded as a degree to what the two microblogs have the same sentiment label. [50] also use the index p to evaluate the correlation between text similarity and sentiment labels.

$$p = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}(Y_{i*} = Y_{j*}, e_{ij} \in E) \cdot A_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}(e_{ij} \in E) \cdot A_{ij}} \tag{18}$$

where $\mathbf{1}(.)$ is the indicator function, A represents the weights matrix of G.

Fig 7 clearly shows the ratio of different methods on both HCR and OMD. SS denotes the microblog-microblog graph constructed by structure similarity, SS-T denotes the microblog-microblog graph built by structure similarity and topic context. We find that the ratio of SS and SS-T is much higher than chance on both HCR and OMD in Fig 7, that is, there is a positive relation between structure similarity and sentiment labels, which paves the way for our next study: how to exploit and model structure similarity into the microblog sentiment analysis system. It is noted that the ratio of SS-T method is higher than SS method. This is because that homophily is more obvious on the same topic which has been verified in [24] and similar people tend to have the same opinion on the same topic. Adding topic context can better exploit the heterogeneous relations between microblogs.
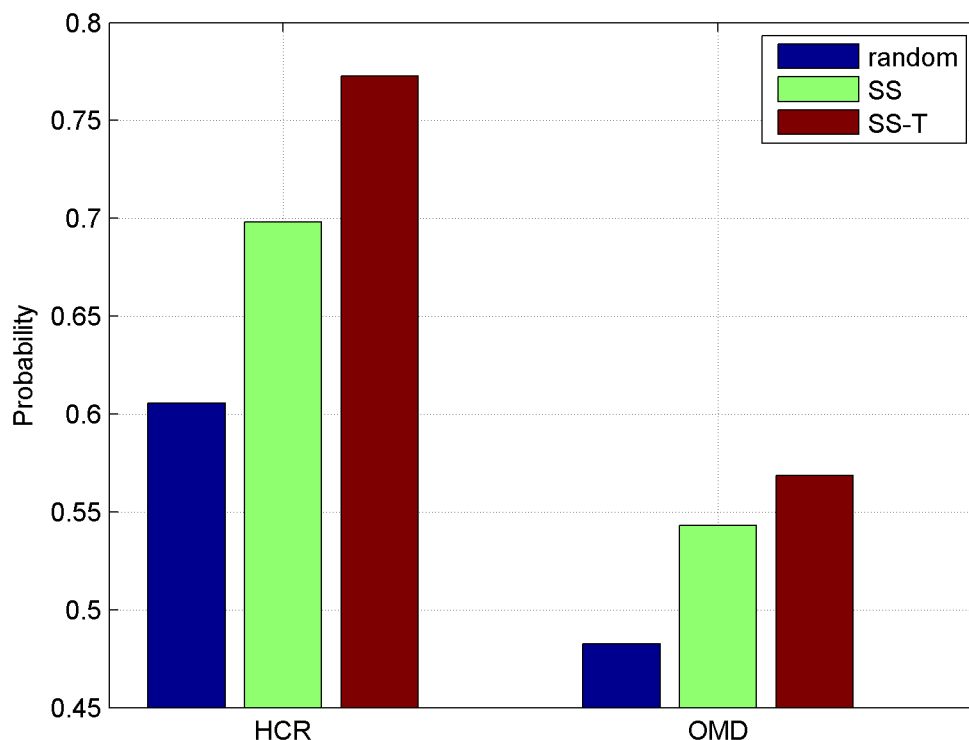
Fig 7. Shared sentiment probability conditioned on structure similarity.

Usefulness of social context

In this section, we perform experiments to assess the validity of different contexts whether they could improve the accuracy of sentiment classification. We use 90% microblogs for training. "TC" represents the method of only using **t**ext **c**ontext (TC), "UC" represents the method using **u**ser **c**ontext (UC) and texts. Similarly, "SSC" denotes the method combining **s**tructure **s**imilarity **c**ontext (SSC) and texts, "FC" is the method using **f**riends **c**ontext (FC) and text information. We use accuracy, which is the proportion of true results (both true positives and true negatives) among the total number of cases examined, as a metric to measure the performance of different algorithms. It can be computed by $accuracy = (TP + TN)/(num)$, where num represents the number of both positive samples and negative samples in the training set. TP and TN represent the number of items correctly labeled as belonging to the positive class and the negative class respectively. The result is shown in Table 3.

Table 3. Performance of different context.

	TC	FC	UC	SSC
HCR	0.631944	0.784722	0.770833	0.791667
OMD	0.647058	0.781513	0.773109	0.789916

From this table, we can conclude following observations.

1. Using social context can improve the performance of sentiment analysis on both HCR and OMD datasets. The accuracy of the methods using social context is higher than the accuracy of using text only, which validates the usefulness of user context, friends context and structure similarity context. The performance of social context reveals that sentiment consistency and emotional contagion hold true in microblogging platform and this can be an experimental basis for the two theory.

2. User context gets the lower improvement than the other social context. This is mainly due to the fact that the average number of friends is larger than the average number of microblogs one user posted, which lead to a sparser sentiment consistency matrix. For example, according to Table 1, each user in HCR dataset only has 1.78 tweets on average, while the average number of friends is 14.95.

3. It is also noted that the method using structure similarity context gets the best performance among all social context. Structure similarity can get more information than direct relation such as common friends and weights about whose influence are larger on users, which is the reason behind its better performance than others.

Performance evaluation

In this section, we use random sampling method to test the accuracy of different methods in different size training set. The methods we use in this paper are listed below.

Least Square (LS): Least Square method [64] is a widely used supervised classifier. Its goal is to find W which minimize the function $f(X) = \frac{1}{2} \| XW - Y \|_F^2$.

Lasso: Lasso [64] only use texts to identify sentiment. Comparing with Least Square method, Lasso adds $\| W \|_1$ to handle the sparse problem of classifier W.

Support Vector Machine (SVM): SVM [45] is a widely used classifier in the fields of text and hypertext categorization, images classification and so on.

Naive Bayes (NB): Like SVM, NB [45] is also a supervised classifier in many fields.

Logistic Regression (LR): LR [45] denotes L_2-norm regularized Logistic Regression, a popular sentiment classification method.

SANT: A method proposed by [9] which combined sentiment consistency and emotional contagion.

SMSC: A method proposed by [51] which use graph information at the prediction stage.

SASS: Our method of Sentiment Analysis based on Structure Similarity, which uses structure similarity and user context to analyze sentiment.

In our method, there are two import parameters: α, β. The two parameters are all nonnegative. In this section, we set $\alpha = 0.0005$, $\beta = 1$ which are tuned by cross-validation. α is the parameter that controls the contribution of social context information, β is the sparse regularization parameter. The training set and the test set are selected randomly from the original dataset to test our method. p% represents the percentage of the training set, and the rest is used for testing. Experimental results of HCR and OMD are shown in Figs 8 and 9 respectively.

Via comparing the results of different methods, we can draw the following observations.

1. Methods only using texts achieve lower improvement than methods using social context. Two-sample one-tail t tests are conducted and the results show that methods using social context can get improved sentiment classification accuracy with a significance level 0.01. Text information in microblogging platform is very noisy, irony and sarcasm are always
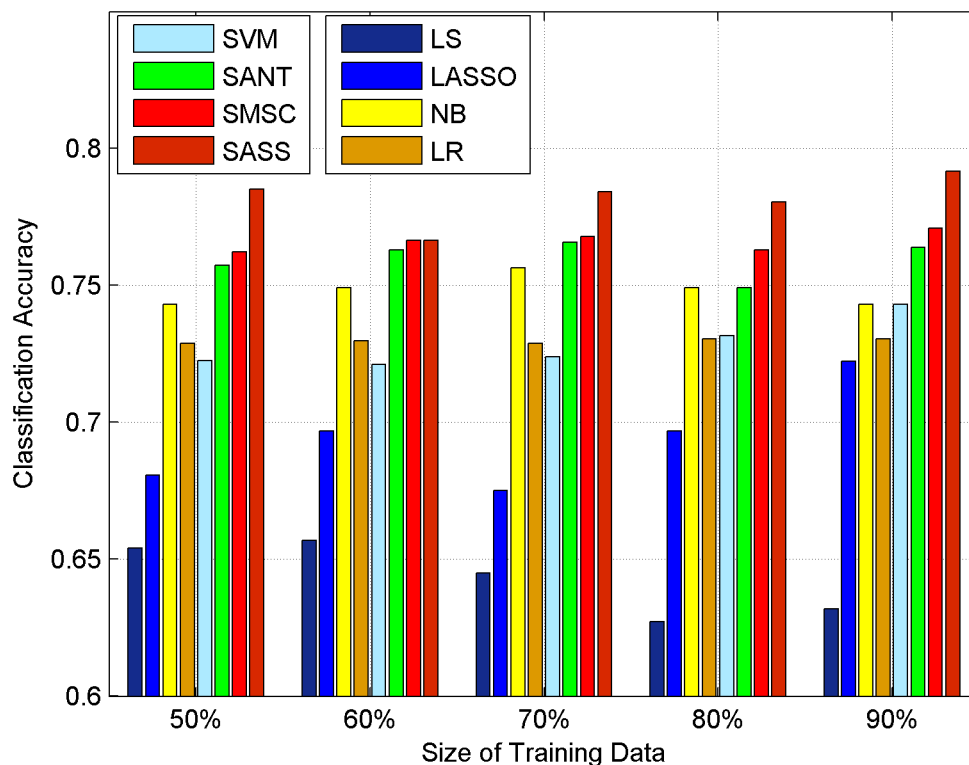
Fig 8. The performance of our method and baseline methods on HCR without topic context.

used to express the negative feelings of users. Methods such as SVM, LS, LR, and NB cannot handle this situation, while using social context can solve the problem to some extent as they take microblogs which are connected to them into consideration and lead to a better performance.

2. SASS outperforms SANT and SMSC and get the best performance among all methods on both datasets with different sizes of training data consistently and significantly. Compared with the traditional LS method, our proposed method SASS can get the average percentage improvement ($\frac{the\ accuracy\ of\ SASS\ -\ the\ accuracy\ of\ LS)}{the\ accuracy\ of\ LS}$) of microblog sentiment analysis accuracy by about [21.61108% and 15.9152%] in HCR and OMD respectively, which outperforms SMSC's [19.18897% and 12.3313%] and SANT's [18.19744% and 11.0669%] improvement respectively. This improvement is continuous and significant in both datasets. SANT and SMSC only use user context and friends context. In contrast, our method SASS which uses structure similarity can deeply explore the relations between microblogs. In our method, every microblog has different contributions to the sentiments of other microblogs while in SANT and SMSC all microblogs are regarded to have the same contribution to others. Besides, our method takes potential friendships into consideration. This is the reason that our method can achieve a better performance.
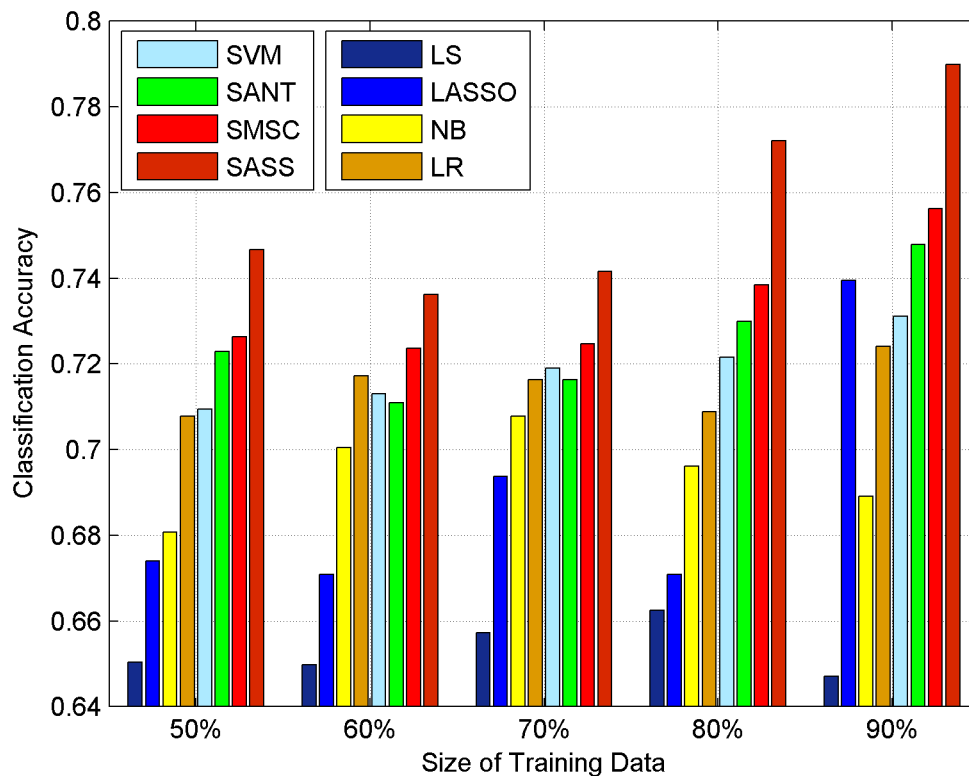
Fig 9. The performance of our method and baseline methods on OMD without topic context.

3. Lasso achieves a better performance than LS, this implies using a sparse solution is an effective way to handle noisy microblog texts as it can select features in an automatic way.

4. When there is only 50% data for training, our method still outperforms other methods on OMD and HCR and the performance of SASS is not sensitive to the changes of the size of training data. This demonstrates we can save a lot of cost in labeling, which has its significance to solve the problem of "lack of manually labeled training data".

Usefulness of topic context

In this subsection, we introduce topic context into our model and compare SASS with topic context (SASS-T) and SASS in different sizes of training data. Classification results are plotted in Figs 10 and 11 for HCR and OMD respectively. From the figures, we can see that adding topic context can improve the accuracy of microblog sentiment analysis to some extent. Compared with the traditional LS method, SASS-T can get the average percentage improvement by [22.4206% and 17.7539%] in HCR and OMD respectively, which is larger than SASS's [21.61108% and 15.9152%]. T-tests are also applied in this subsection and there is also a
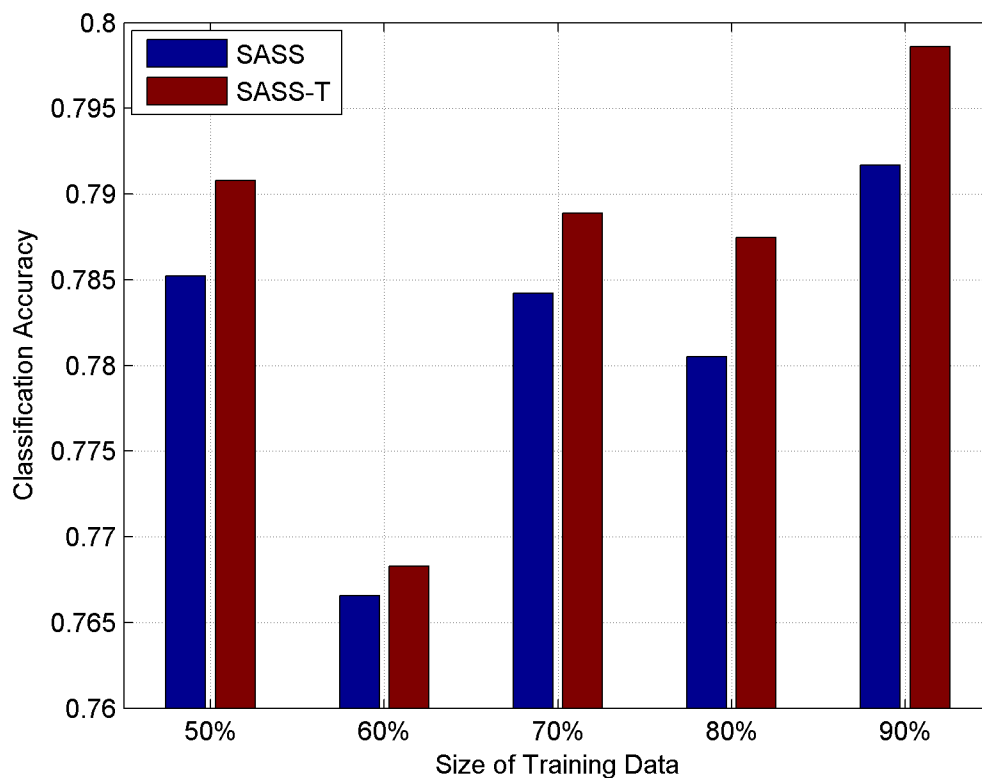
Fig 10. Classification accuracy on HCR with topic context.

significant improvement with the significance level 0.01 on both datasets. The results indicate the positive effect of using topic context to model the semantic relations between microblogs in microblog sentiment analysis. The reason why adding topics can improve the accuracy of sentiment analysis is that the opinions of a same person and similar users on the same topic usually consistent with each other.

Parameter analysis

In this subsection, we evaluate the effects of parameter selection of α and β on our method. We use 90% of data on both datasets which are randomly selected, and the left data are used for test. Fig 12 shows the effect of α in detail when $\beta = 1$. Obviously, the performance of SASS is not sensitive to the variation of α. When α is too small, social context is not fully used in the sentiment analysis. Thus, the performance increases as α increases from 0. However, when α is too large, the performance of the model mainly depends on social context so it becomes worse. Fig 13 shows the performance of SASS with the variation of β when $\alpha = 0.0005$.
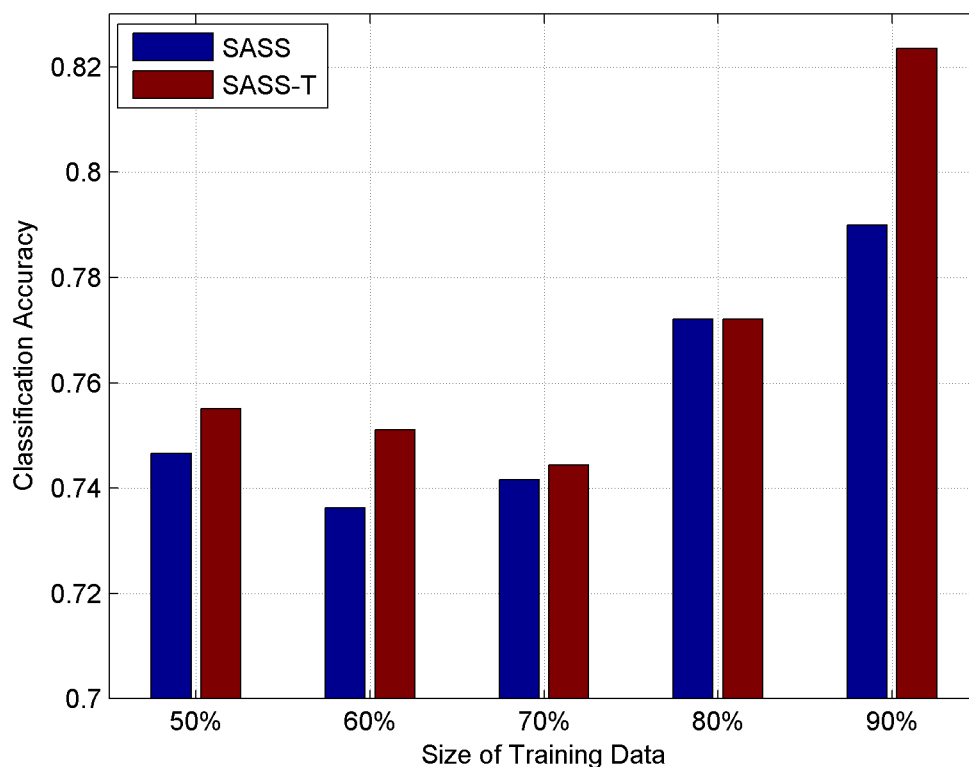
Fig 11. Classification accuracy on OMD with topic context.

https://doi.org/10.1371/journal.pone.0191163.g011

It is noted that when β is too large, the performance of our model goes down as it mainly relies on the sparse regularization and many features are filtered by the regularization. When β is too small, the sparse regularization is not fully used and many noises remain in the training set, so the accuracy of sentiment analysis also increases from 0. Besides, it is clear that the model is not very sensitive to the variation of α and β and it is an appealing property as it can save a lot of time to tune parameters.

Conclusion and discussion

In this paper, we propose a new method which using social context to identify sentiment polarity. Inspired by sentimental consistency and emotional contagion, we take three kinds of context into account: user context, structure similarity context, and topic context. We introduce a measure to structure similarity, build structure similarity matrix. We also introduce topic context and build a topic context matrix. We add all these contexts into the model by using the Laplacian matrix of the graph constructed by the contexts. Experimental results show that
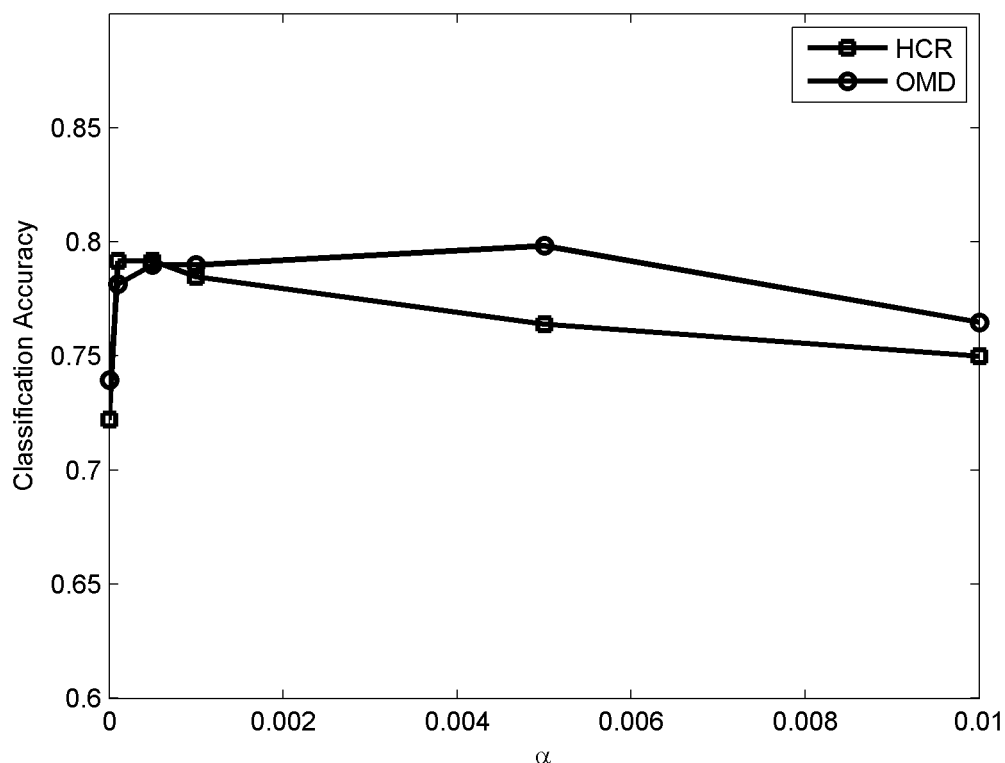
Fig 12. Sensitivity to hyper-parameter α.

https://doi.org/10.1371/journal.pone.0191163.g012

structure similarity has a better performance than user direct relations. Besides, adding topic context is helpful for improving the accuracy of sentiment classification. Meanwhile, our method can be easily extended to other models such as semi-supervised classification model proposed by [50] and the structured model proposed by [51].

In this paper, we use Least Squares to model text information of microblogs. In future, we also want to extend Laplacian regularization to support vector machine (SVM) and maximum entropy model to see the differences between them. Deep learning methods have obtained a very good performance across many different NLP tasks recently, so we also want to study how to combine social context with deep learning models.

Fig 13. Sensitivity to hyper-parameter β.

https://doi.org/10.1371/journal.pone.0191163.g013

Acknowledgments

Author Contributions

Formal analysis: Xiaomei Zou.

Funding acquisition: Jing Yang, Jianpei Zhang.

Methodology: Xiaomei Zou.

Supervision: Jing Yang, Jianpei Zhang.

Writing – original draft: Xiaomei Zou.

Writing – review & editing: Jing Yang.

References

1. Bollen J, Mao H, Zeng X, 2011. Twitter mood predicts the stock market. Journal of Computational Science 2(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

2. Yang D, Zhang D, Yu Z, Wang Z, 2013. A sentiment-enhanced personalized location recommendation system. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media. ACM, pp. 119–128.

3. Cambria E, Mar 2016. Affective computing and sentiment analysis. IEEE Intelligent Systems 31(2), 102–107. https://doi.org/10.1109/MIS.2016.31

4. Cambria E, Schuller B, Xia Y, White B, Sep 2016. New avenues in knowledge bases for natural language processing. Know.-Based Syst. 108(C), 1–4. https://doi.org/10.1016/j.knosys.2016.07.025

5. Turney PD, 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, pp. 417–424.

6. Hu M, Liu B, 2004. Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 168–177.

7. Godbole N, Srinivasaiah M, Skiena S, 2007. Large-scale sentiment analysis for news and blogs. ICWSM 7(21), 219–222.

8. Mei Q, Ling X, Wondra M, Su H, Zhai C, 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In: Proceedings of the 16th international conference on World Wide Web. ACM, pp. 171–180.

9. Hu X, Tang L, Tang J, Liu H, 2013. Exploiting social relations for sentiment analysis in microblogging. In: Proceedings of the sixth ACM international conference on Web search and data mining. ACM, pp. 537–546.

10. Tan C, Lee L, Tang J, Jiang L, Zhou M, Li P, 2011. User-level sentiment analysis incorporating social networks. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 1397–1405.

11. Abelson RP, 1983. Whatever became of consistency theory? Personality and Social Psychology Bulletin.

12. Hatfield E, Cacioppo JT, Rapson RL, 1994. Emotional contagion. Cambridge university press.

13. Tang L, Liu H, 2009. Relational learning via latent social dimensions. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data Mining. KDD'09. ACM, New York, NY, USA, pp. 817–826.

14. Tang J, Hu X, Gao H, Liu H, 2013. Exploiting local and global social context for recommendation. In: Ijcai. pp. 2712–2718.

15. Sun Y, Han J, 2012. Mining heterogeneous information networks: principles and methodologies. Synthesis Lectures on Data Mining and Knowledge Discovery 3(2), 1–159. https://doi.org/10.2200/S00433ED1V01Y201207DMK005

16. Tang J, Wang S, Hu X, Yin D, Bi Y, Chang Y, Liu H, 2016. Recommendation with social dimensions. In: Proceedings of the thirtieth AAAI conference on Artificial Intelligence. AAAI'16. AAAI Press, pp. 251–257.

17. Carullo Giuliana and Castiglione Aniello and De Santis Alfredo and Palmieri Francesco, 2015. A triadic closure and homophily-based recommendation system for online social networks. World Wide Web 18 (6), 1579–1601. https://doi.org/10.1007/s11280-015-0333-5

18. Wasserman S, Faust K, 1994. Social network analysis: Methods and applications. Vol. 8. Cambridge university press.

19. Mcpherson Miller and Smithlovin Lynn and Cook James M, 2001. BIRDS OF A FEATHER: Homophily in Social Networks. Review of Sociology 27(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

20. Crimaldi Irene and Vicario Michela Del and Morrison Greg and Quattrociocchi Walter and Riccaboni Massimo, 2015. Homophily and Triadic Closure in Evolving Social Networks. arXiv: Social and Information Networks.

21. Thelwall Mike. 2010. Emotion Homophily in Social Network Site Messages. First Monday 15(4).

22. Liang Y, Li Q, 2011. Incorporating interest preference and social proximity into collaborative filtering for folk recommendation. In: SWSM 2011 (SIGIR workshop)

23. Xie Yan Bo and Zhou Tao and Wang Bing Hong, 2008. Scale-free networks without growth. Physica A Statistical Mechanics & Its Applications 387(7), 1683–1688. https://doi.org/10.1016/j.physa.2007.11.005

24. Kang JH, Lerman K. 2012. Using lists to measure homophily on twitter. In AAAI workshop on Intelligent techniques for web personalization and recommendation.

25. Neviarouskaya A, Prendinger H, Ishizuka M, 2009. Sentiful: Generating a reliable lexicon for sentiment analysis. In: 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops. IEEE, pp. 1–6.

26. Qiu G, Liu B, Bu J, Chen C, 2009. Expanding domain sentiment lexicon through double propagation. In: IJCAI. Vol. 9. pp. 1199–1204.

27. Deng Shuyuan and Sinha Atish P and Zhao Huimin, 2017. Adapting sentiment lexicons to domain-specific social media texts. decision support systems, 2017: 65–76. https://doi.org/10.1016/j.dss.2016.11.001

28. Al-Twairesh Nora and Al-Khalifa Hend and Alsalman Abdulmalik, 2016. AraSenTi: Large-Scale Twitter-Specific Arabic Sentiment Lexicons. Meeting of the Association for Computational Linguistics, 697–705.

29. Bandhakavi Anil and Wiratunga Nirmalie and Padmanabhan Deepak and Massie Stewart, 2016. Lexicon based feature extraction for emotion text classification. Pattern Recognition Letters 93(SI),133–142.

30. Fu Xianghua and Liu Wangwang and Xu Yingying and Cui Laizhong, 2017.Combine HowNet Lexicon to Train Phrase Recursive Autoencoder for Sentence-Level Sentiment Analysis. Neurocomputing 241 (3),851–872.

31. Khan Farhan Hassan and Qamar Usman and Bashir Saba, 2017. Lexicon based semantic detection of sentiments using expected likelihood estimate smoothed odds ratio. Artificial Intelligence Review 48, 1–26.

32. Khan Farhan Hassan and Qamar Usman and Bashir Saba, 2017. A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet. Knowledge & Information Systems 51(3),851–872. https://doi.org/10.1007/s10115-016-0993-1

33. Baccianella S, Esuli A, Sebastiani F, 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: LREC. Vol. 10. pp. 2200–2204.

34. Cambria E, Poria S, Bajpai R, Schuller B, 2016. Senticnet 4: A semantic resource for sentiment analysis based on conceptual primitives. In: the 26th International Conference on Computational Linguistics (COLING), Osaka.

35. Ortigosa-Hernández J, Rodríguez JD, Alzate L, Lucania M, Inza I, Lozano JA, 2012. Approaching sentiment analysis by using semi-supervised learning of multi-dimensional classifiers. Neurocomputing 92, 98–115. https://doi.org/10.1016/j.neucom.2012.01.030

36. Pang B, Lee L, Vaithyanathan S, 2002. Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, pp. 79–86.

37. Read J, 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: Proceedings of the ACL student research workshop. Association for Computational Linguistics, pp. 43–48.

38. Severyn A, Moschitti A, 2015. Twitter sentiment analysis with deep convolutional neural networks. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, pp. 959–962.

39. Ren Y, Zhang Y, Zhang M, Ji D, 2016. Context-sensitive twitter sentiment classification using neural network. In: AAAI. pp. 215–221.

40. Wang Y, Huang M, Zhu X, Zhao L, 2016. Attention-based lstm for aspect-level sentiment classification. In: EMNLP. pp. 606–615.

41. Poria S, Chaturvedi I, Cambria E, Hussain A, 2016. Convolutional mkl based multimodal emotion recognition and sentiment analysis. In: Data Mining (ICDM), 2016 IEEE 16th International Conference on. IEEE, pp. 439–448.

42. Chen Tao and Xu Ruifeng and He Yulan and Wang Xuan, 2017. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Systems With Applications 72,221–230. https://doi.org/10.1016/j.eswa.2016.10.065

43. Pandarachalil R, Sendhilkumar S, Mahalakshmi G, 2015. Twitter sentiment analysis for large-scale data: an unsupervised approach. Cognitive Computation 7(2), 254–262. https://doi.org/10.1007/s12559-014-9310-z

44. Liu K-L, Li W-J, Guo M, 2012. Emoticon smoothed language models for twitter sentiment analysis. In: AAAI.

45. Go A, Bhayani R, Huang L, 2009. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford 1, 12.

46. Cui A, Zhang M, Liu Y, Ma S, 2011. Emotion tokens: Bridging the gap among multilingual twitter sentiment analysis. In: Asia Information Retrieval Symposium. Springer, pp. 238–249.

47. Kiritchenko S, Zhu X, Mohammad SM, 2014. Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research 50, 723–762.

48. Ren F, Wu Y, 2013. Predicting user-topic opinions in twitter with social and topical context. IEEE Transactions on Affective Computing 4(4), 412–424. https://doi.org/10.1109/T-AFFC.2013.22

49. Speriosu M, Sudan N, Upadhyay S, Baldridge J, 2011. Twitter polarity classification with label propagation over lexical links and the follower graph. In: Proceedings of the First workshop on Unsupervised Learning in NLP. Association for Computational Linguistics, pp. 53–63.

50. Lu T-J, 2015. Semi-supervised microblog sentiment analysis using social relation and text similarity. In: 2015 International Conference on Big Data and Smart Computing (BigComp). IEEE, pp. 194–201.

51. Wu F, Huang Y, Song Y, 2016. Structured microblog sentiment classification via social context regularization. Neurocomputing 175, 599–609. https://doi.org/10.1016/j.neucom.2015.10.101

52. Vosecky J, Leung KW, Ng W, 2014. Collaborative personalized Twitter search with topic-language models. international acm sigir conference on research and development in information retrieval, 53–62.

53. Kotov A, Agichtein E, 2013. The importance of being socially-savvy: quantifying the influence of social networks on microblog retrieval. Conference on Information and Knowledge Management, 1905–1908.

54. Kwak H, Lee C, Park H, Moon S, 2010. What is Twitter, a social network or a news media? In: WWW'10: Proceedings of the 19th international conference on World wide web. ACM, New York, NY, USA, pp. 591–600.

55. Shamma DA, Kennedy L, Churchill EF, 2009. Tweet the debates: understanding community annotation of uncollected sources. In: Proceedings of the first SIGMM workshop on Social media. ACM, pp. 3–10.

56. Chung FR, 1997. Spectral graph theory. No. 92. American Mathematical Soc.

57. Simon N, Friedman J, Hastie T, Tibshirani R, 2013. A sparse-group lasso. Journal of Computational and Graphical Statistics 22(2), 231–245. https://doi.org/10.1080/10618600.2012.681250

58. Easley D, Kleinberg J, 2010. Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press.

59. Jackson MO, Rogers BW. 2007. Meeting Strangers and Friends of Friends: How Random Are Social Networks? American Economic Review 97(3), 890–915. https://doi.org/10.1257/aer.97.3.890

60. Palla G, Vicsek T. 2007. Quantifying social group evolution. Nature 446 (7136), 664. https://doi.org/10.1038/nature05670 PMID: 17410175

61. Liu J, Ji S, Ye J. 2009. Multi-task feature learning via efficient l 2, 1 -norm minimization. Conference on Uncertainty in Artificial Intelligence. AUAI Press, 339–348.

62. Kernighan BW, Lin S, Feb 1970. An efficient heuristic procedure for partitioning graphs. The Bell System Technical Journal 49(2), 291–307. https://doi.org/10.1002/j.1538-7305.1970.tb01770.x

63. Nesterov Y and Nesterov I. Introductory lectures on convex optimization: A basic course. 2004.

64. Friedman J, Hastie T, Tibshirani R, 2001. The elements of statistical learning. Vol. 1. Springer series in statistics Springer, Berlin.