

# Approval network: a novel approach for sentiment analysis in social networks

E. Fersini<sup>1</sup> · F. A. Pozzi<sup>1</sup> · E. Messina<sup>1</sup>

Received: 4 March 2015 / Revised: 19 September 2016 /  
Accepted: 22 September 2016 / Published online: 11 October 2016  
© Springer Science+Business Media New York 2016

**Abstract** The data-centric impetus and the development of online social networks has led to a significant amount of research that is nowadays more flexible in demonstrating several sociological hypotheses, such as the sentiment influence and transfer among users. Most of the works regarding sentiment classification usually consider text as unique source of information, do not taking into account that social networks are actually networked environments. To overcome this limitation, two main sociological theories should be accounted for addressing any sentiment analysis tasks: homophily and constructivism. In this paper, we propose Approval Network as a novel graph representation to jointly model homophily and constructivism, which is intended to better represent the contagion on social networks. To show the potentiality of the proposed representation, two novel sentiment analysis models have been proposed. The first one, related to user-level polarity classification, is approached by presenting a semi-supervised framework grounded on a Markov-based probabilistic model. The second task, aimed at simultaneously extracting aspects and sentiment at message level, is addressed by proposing a novel fully unsupervised generative model. The experimental results show that the proposed sentiment analysis models grounded on Approval Networks are able to outperform not only the traditional models where the relationships are disregarded, but also those computational approaches based on traditional friendship connections.

**Keywords** Approval networks · User-level sentiment analysis · Aspect-level sentiment analysis · Social networks

---

✉ E. Fersini  
fersini@disco.unimib.it  
F. A. Pozzi  
federico.pozzi@disco.unimib.it  
E. Messina  
messina@disco.unimib.it

<sup>1</sup> Department of Informatics, Systems and Communication (DISCo), University of Milano-Bicocca, Viale Sarca, 336 - 20126 Milan, Italy

## 1 Introduction and motivation

Opinions are central to almost all human activities and are key influencers of our behaviours [4]. When people need to make a decision, such as which restaurant is the best one in the city or who would people vote for a candidate in the next election, they often listen to the opinions of others. With the development and spread of Social Networks, user-generated opinions are getting richer and ubiquitously available. The automatic extraction and quantification of subjective information from natural language text is therefore becoming a fundamental task of Sentiment Analysis. This research area has been extensively investigated with several approaches, such as lexicon- [56] and learning- based methods [51] to deal with explicit opinions (see Example 1).

*Example 1* [**explicit, negative**]: The latest menswear campaign is on fire: it's disgusting!

Most of the works usually take into account content as unique information to infer sentiment [3, 22, 42, 63]. For instance, [22] presented the results of machine learning algorithms for classifying the sentiments of Twitter messages using distant supervision (i.e. emoticons are considered as ground truth for tweet labelling), while [3] explored the linguistic characteristics of how tweets are written and the meta-information of words for sentiment classification. Less work has been done on implicit opinions [70] (see Example 2).

*Example 2* [**implicit, positive**]: "Saving Soldier Ryan"...I can't wait to watch it!

An implicit opinion is a statement that implies a regular or comparative opinion that usually expresses a desirable or undesirable fact. Example 2 shows that there is some good expectation about the movie, although it is not encoded in words. In this cases the textual features do not provide any explicit information about the intended sentiment, making its recognition of the sentiment difficult even for humans.

When dealing with both explicit and implicit opinions, we should take into account their real nature: they are expressed as **texts** interconnected by several heterogeneous **relationships** (e.g. a user-generated opinion can be linked to its author, it can be related to one or more topics, it can be referenced by several readers). Considering the real characteristics of opinions in online social networks, there are two sources of information that can be used to address the sentiment analysis tasks. Textual information can be exploited to model the "similarity" of opinions, while the relational information can both complement and smooth the evidence given by the text. Considering only one of these two information sources, may lead to define biased sentiment analysis models do not able to account for important structure in the data [66]. To this purpose, we propose a novel network modelling approach ables to encapsulate the heterogeneity of the information sources (contents and relationships) by capturing two sociological processes underlying the social network interactions: homophily and constructuralism.

The paper is organised as follows. In Section 2, a literature review on sentiment analysis is presented. In Section 3, homophily and constructuralism underlying social network interactions are discussed. In Section 4, Approval Networks for modelling the user interactions are introduced. In Section 5, two novel sentiment analysis models are presented to take advantage of Approval Networks at user-level and aspect-level. Finally, in Section 6 conclusions are derived.

## 2 Literature review

During the last decade, sentiment analysis has been mainly focused on well-formed text at different granularity such as document level [1, 62, 69], sentence level [41, 68] and aspect level [16, 20, 72]. In the recent years, the informal texts on social networks become one of the major forms of online communications, enabling a quasi real-time diffusion of contents provided by people and organisations. Although the recent studies are moving towards informal text, most of them are only based on textual contents [3, 19, 29, 42, 53]. In those cases, sentiment analysis is based on the assumption that user-generated opinions are independent and identically distributed (i.i.d). These methods are aimed at handling the complex characteristics of natural language, do not considering the networked context of the data: a user-generated opinion is linked to its author and can be explicitly referenced by several readers. The debating and informal nature of user-generated opinions have led therefore to additional challenges for addressing the sentiment analysis tasks: it is impractical to exploit traditional supervised machine learning techniques that require annotated and non-relational training data.

Recently, some approaches [11, 13, 21, 37, 67] have been proposed to exploit some structural network information for sentiment analysis purposes.<sup>1</sup> In [21] a “friendship” network is used to simulate and analyse the diffusion model of opinion across the network, while in [37] the hyperlink structure of blogs is exploited to track how users join and construct opinions and how some of these opinions around a specific topic spread. In [11], the authors exploited the network structure to group the users with respect to the sentiment, without considering the textual contents. Although these approaches provide a fundamental contribution to the sentiment analysis research field, their focus is on structural properties of the social networks disregarding the information embedded within the text. Recently, an investigation about the emotional contagion in large social networks has been presented [13]. The authors have modelled the emotions of the users as dependent not only on the endogenous and exogenous factors (e.g. being always happy and rainfall effect), but also on contagion of groups of friends.

Some recent investigations, which have attempted to combine both text and network information for sentiment analysis purposes, are focused on user-user relationships [55, 58, 61, 71]. Tan et al. [61] proposed a semi-supervised approach to predict the user-sentiment by introducing explicitly available undirected user-user relationships (“friendships”) into a text based factor-graph model. In [58], Speriosu et al. proposed to enrich the content representation by including directed user-user relationships as additional features to the text ones. The same kind of directed user-user social relationships (e.g. “following” and “follower” in Twitter) has been exploited in [55] and [71] to predict the sentiment orientation by means of supervised and unsupervised relational approaches.

Although the above mentioned approaches represent a fertile ground for sentiment analysis in online social networks, they strongly assume that the explicitly available user-user relationships (friendships and following/follower connections) unconditionally represent the sentiment agreement between connected users. However, this assumption does not reflect what happens in the real world, where two structurally connected users (e.g. friends)

---

<sup>1</sup>For a detailed survey about diffusion models for social networks please refer to [59].

can have divergent opinions on a given topic. In order to better capture the sentiment agreement among users, in the following section two main sociological theories are presented and analysed to finally converge to a novel network modelling approach for sentiment analysis purposes.

### 3 Homophily and constructuralism: two sociological processes

People with different characteristics (e. g., genders, races, ages, class backgrounds, etc.) usually show very different personalities: educated people are tolerant, women are sensitive, and gang members are violent [48]. Since people generally have significant contacts with others who tend to be like themselves, any personal characteristic tends to converge. **Homophily** is the principle stating that *a contact among similar people occurs at a higher rate than among dissimilar people*. Homophily implies that differences in terms of social characteristics translate into network distance, i.e. the number of relationships through which a piece of information must travel to connect two individuals [48].

The concept of homophily is very ancient. In Aristotle's *Rhetoric* and *Nichomachean Ethics*, he noted that people "love those who are like themselves" [2]. Plato observed in *Phaedrus* that "similarity begets friendship" [52]. However, social scientists began systematic observations of group formation and network ties only in the 1920s [6, 31, 64]. They noted that school children formed friendships and play groups at higher rates if they were similar on demographic characteristics. The classic and most famous work in sociology is [39], where the friendship process is studied. They also quoted the proverbial expression of homophily, "*birds of a feather flock together*", which is often used to summarise this sociological process. Researchers have studied homophily ranging from the strong relationships of "discussing important matters" [43, 44] to the more circumscribed relationships of "knowing about" someone [23] or appearing with them in a public place [47].

In particular, [39] distinguished two types of homophily: **status homophily**, in which similarity is based on informal, formal, or ascribed status, and **value homophily**, which is based on values, attitudes, and beliefs [48]. Status homophily includes the major sociodemographic dimensions like race, ethnicity, sex, or age, and acquired characteristics like religion, education, occupation, or behaviour patterns. Value homophily includes the wide range of internal states presumed to shape our orientation towards future behaviours: attitude, belief, and value similarity lead to attraction and interaction [32]. Value homophily is the homophily facet that has been considered as assumption in this paper, where interactions are preferred compared to static user attributes.

Besides homophily, [8] has developed a sociological theory called **constructuralism**, whose core assumption is that *people who share knowledge are more likely to interact* (i.e. form ties). In particular, constructuralism argues that individual learning from interactions takes place on two levels. First, social interactions allow us to collect over time new knowledge that represents similarity among users better than static sociodemographic dimensions like race, ethnicity, sex, or age (i.e. status homophily). Second, as humans receive and share knowledge with interaction partners, we "learn" a perception of what we expect them to know. Paired with the assumption of homophily that people tend to interact with others similar to them, constructuralism explains how social relationships evolve via interactions as the knowledge that two actors share increases [35]. This approach to the coevolution of knowledge and social relationships has a considerable explanatory power over the dynamics of social networks and has shown to be an effective tool for social simulation [9, 24]. Since text does not always provide explicit or sufficient information about sentiment, early studies

on sentiment classification [30, 61] overcome this limitation by exploiting the principle of homophily, which is usually modelled through friendships. However, considering the similarity among users on the basis of constructivism appears to be much more powerful than interpersonal influence within the friendship network [10, 36].

Considering friendship connections as proxy of homophily is therefore a strong assumption: (1) being friends/followers does not necessarily mean agreeing on a particular topic (e.g. there are often opposite political views among friends), (2) dynamic interactions are preferred compared to static attributes (value homophily): once friendship is established in online social networks, it is rare that it could be interrupted, or even when it occurs it changes slowly over time, (3) social interactions allow us to collect over time new knowledge that represents similarity among users better than the static sociodemographic dimensions, such as friendship (constructivism).

For these reasons, we propose a novel paradigm called **Approval Network** to jointly model homophily and constructivism. The general idea behind Approval Network is that a user who approves (e.g., by ‘likes’ on Facebook, ‘+1’ on Google+ and ‘retweets’ on Twitter) a given message is likely to hold the same opinion of the author. This is because an approval tool does not allow the user to add a comment against the original message.<sup>2</sup> Thus, the main underlying principle is that **approval relationships** could be a strong indication about the sentiment agreement between two users: the higher is the number of approvals between two users on a given topic, more likely will be their agreement on that topic.

For instance, information can spread in Twitter in the form of *retweets*, which are tweets that have been forwarded by a user to his or her followers. A retweet is identified by the pattern “RT @” followed by the name of the tweet’s author and the original tweet (e.g. John tweets “I like the new iPhone” and Mary retweets the John’s tweet: “RT @John: I like the new iPhone”, i.e. John and Mary positively agrees about iPhone). While approving can simply be perceived as the act of copying and rebroadcasting, the practice contributes to a conversational ecology in which conversations are composed of a public interplay of voices that give rise to an emotional sense of shared conversational context [7].

## 4 Approval networks

Approving contents is a well adopted practice in online social networks. Among the practical motivations, it has been highlighted that the agreement with the original user message is one of the main motivations [7, 33]. In order to model the two underlying social theories and capture the practical behaviour of the users, we have formally defined what we have called Approval Networks.

**Definition 1** Given a topic of interest  $q$ , a **Directed Approval Graph** is a quadruple  $\text{DAG}_q = \{V_q, E_q, \mathbf{X}_q^V, \mathbf{X}_q^E\}$ , where  $V_q = \{v_1, \dots, v_n\}$  represents the set of active users on  $q$ ;  $E_q = \{(v_i, v_j) | v_i, v_j \in V_q\}$  is the set of approval edges, meaning that the extent that  $v_i$  approved  $v_j$ ’s messages;  $\mathbf{X}_q^E = \{w_{i,j} | (v_i, v_j) \in E_q\}$  is the set of weights assigned to approval edges, indicating that  $v_i$  approved  $w_{i,j}$  messages of  $v_j$  on  $q$ ;  $\mathbf{X}_q^V = \{k_i | v_i \in V_q\}$  is the set of coefficients related to nodes, where  $k_i$  represents the total number of messages of  $v_i$  on  $q$ .

<sup>2</sup>Note that for this reason the Facebook’s ‘Share’ does not belong to approval tools.

Given a  $DAG_q$ , we can define a network representation that takes into account the real usage of social network approval tools. In particular, we defined an Augmented Directed Approval Graph as follows:

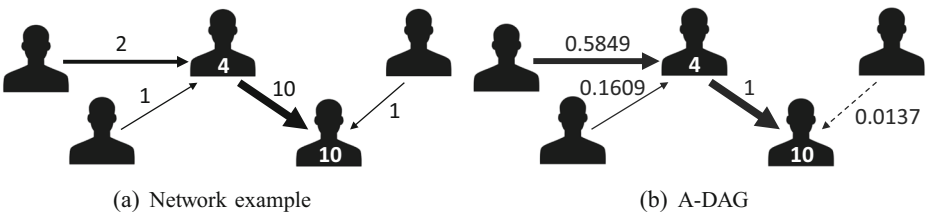
**Definition 2** Given a  $DAG_q = \{V_q, E_q, X_q^V, X_q^E\}$ , an **Augmented Directed Approval Graph** is derived as a triple **A-DAG** $_q = \{V_q, E_q, C_q^E\}$ , where  $C_q^E = \{c_{i,j}\}$  is the set of normalised weights of approval edges, and  $c_{i,j}$  is computed as

$$c_{i,j} = \frac{w_{i,j}}{\max_i w_{i,j}} \log_2 \left( 1 + \frac{w_{i,j}}{k_j} \right) \tag{1}$$

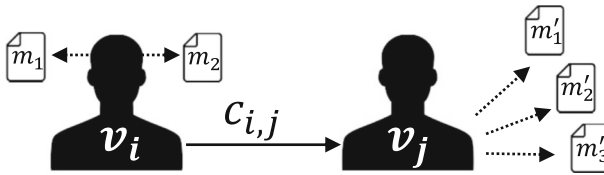
The measure presented in (1) tries to capture the behaviour of users on social networks. First, the common characteristic of an approval network is that most of the users usually approve only one message of a target user, and very few users approve two or more messages [38]. Thus, a logarithmic scale should be used instead of a linear one. Second, the number of approvals between two users does not necessarily indicate how much they agree with a particular topic. It could be influenced by the interest and originality of the target user’s messages. For example, a user *A* could completely agree with user *B* but approves it one or two times only because the weak originality of *B*’s messages. For this reason, the number of approvals from user *A* to *B* has been normalised by considering the maximum number of approvals from any user connected to *B*. Finally, this network representation penalises users who approve few messages of a particular target user if there are other users who approve many posts of the same target user. A toy example of A-DAG is reported in Figure 1.

In order to take into account both texts and relationships available in online social networks, A-DAG $_q$  has been extended by defining an heterogeneous graph as a unique representation of both user-user and user-message relationships.

**Definition 3** Given an A-DAG $_q$ , let  $M_q = \{m_1, \dots, m_m\}$  be the set of nodes representing messages about *q* and  $A_q^M = \{(v_i, m_t) | v_i \in V_q, m_t \in M_q\}$  be the set of arcs that connect the user  $v_i$  and the message  $m_t$ . A **Heterogeneous Directed Approval Graph** is a quintuple **H-DAG** $_q = \{V_q, E_q, C_q^E, M_q, A_q^M\}$ .



**Figure 1** Example of A-DAG $_q$  modeling approval relations. White numbers represent the number of messages provided by every user. **a** The weight on the edges corresponds to the number of approvals that a source user makes with respect to the messages of the target user. **b** The score on the edges is the weight of approval edges computed according (1)



**Figure 2** H-DAG representing user-message and user-user relationships

A graphical representation of  $H-DAG_q$  is reported in Figure 2. In the following, topic  $q$  is intended to be fixed and therefore omitted.

## 5 Sentiment analysis with approval networks

In the following sections, we will present two tasks where the proposed Heterogeneous Directed Approval Graph has been exploited for defining two novel sentiment analysis models. The first task relates to user-level sentiment analysis addressed by a semi-supervised model, while the second one is concerned with aspect-level sentiment analysis solved by presenting a fully unsupervised approach.

### 5.1 User-level sentiment analysis

Dealing with sentiment classification on social networks usually requires a fully supervised learning paradigm, where the sentiment orientation of users must be known a priori to derive suitable predictive models. However, this does not reflect the real settings of social networks, where the polarity on a given topic is explicitly available only for some users (black nodes) while for others could be derived from their posts and relations with other users (white nodes). As black nodes we considered those users whose bio (description on Twitter) or name clearly state a positive or negative opinion about the topic 'Obama'. For instance, a positive user's bio could report "*I love football, TV series and Obama!*" and/or the name could be "*ObamaSupporter*".

In this context, a semi-supervised learning paradigm better represents the real setting. To this purpose, we introduce a semi-supervised sentiment learning approach named  $S^2-LAN$  [54] ables to deal both with text and Approval Network: given a small proportion of users already labelled in terms of polarity, it predicts the sentiments of the remaining unlabelled users by combining textual information and Approval Network directly.

#### 5.1.1 Semi-supervised sentiment learning by approval network ( $S^2-LAN$ )

Given a H-DAG denoted by  $\phi$ , two vectors need to be introduced to tackle the sentiment classification problem at user-level: a vector of labels  $\mathbf{L}^V = \{l(v_i) \in \{+, -\} | v_i \in V\}$  that defines each user as either "positive" (+) or "negative" (-) and an analogous vector of labels  $\mathbf{L}^M = \{l_{v_i}(m_t) \in \{+, -\} | v_i \in V, m_t \in M\}$  that represents the polarity label of each message  $m_t$  written by the user  $v_i$ .

In our model, we assume that the sentiment label  $l(v_i)$  of the user  $v_i$  can be derived by considering the sentiment labels  $l_{v_i}(m_t)$  of his/her messages and influenced by the sentiment labels of the directly connected neighbours  $N(v_i)$ . The user-message and user-user (approval) relations are combined in the following probabilistic model:

$$\log P(\mathbf{L}^V | \phi) = \left( \sum_{v_i \in V} \left[ \sum_{m_t \in M} \sum_{\alpha} \sum_{\beta} \mu_{\alpha, \beta} f_{\alpha, \beta}(l(v_i), l_{v_i}(m_t)) + \sum_{\substack{v_j \in N(v_i), \\ (v_i, v_j) \in \phi}} \sum_{\alpha} \sum_{\beta} \lambda_{\alpha, \beta} g_{\alpha, \beta}(l(v_i), l(v_j)) \right] \right) - \log Z \tag{2}$$

where  $\alpha, \beta \in \{+, -\}$  denote the polarity labels,  $f_{\alpha, \beta}(\cdot, \cdot)$  and  $g_{\alpha, \beta}(\cdot, \cdot)$  are feature functions used to evaluate the user-message and the user-user relations respectively, and the weights  $\mu_{\alpha, \beta}, \lambda_{\alpha, \beta}$  are parameters to be estimated.  $Z$  is a normalisation factor that enables a coherent probability distribution of  $P(\mathbf{L}^V)$ . Regarding the estimation of  $\mu, \lambda$  and the assignment of user sentiment labels which maximises  $\log P(\mathbf{L}^V)$  refer to [54], where a modified version of the SampleRank algorithm [65] is presented.

**User-message feature function** A user-message feature function evaluates whether the message polarity agrees (or disagrees) with respect to the user sentiment. Formally,  $f_{\alpha, \beta}(l(v_i), l_{v_i}(m_t))$  is defined as:

$$f_{\alpha, \beta}(l(v_i), l_{v_i}(m_t)) = \begin{cases} \frac{\rho_{T-black}}{|M_{v_i}|} & l(v_i) = \alpha, l_{v_i}(m_t) = \beta, v_i \in \text{black} \\ \frac{\rho_{T-white}}{|M_{v_i}|} & l(v_i) = \alpha, l_{v_i}(m_t) = \beta, v_i \in \text{white} \\ 0 & otherwise \end{cases} \tag{3}$$

where “ $v_i \in \text{black}$ ” means that the user  $v_i$  is initially labeled (i. e. its polarity label is known a priori), and “ $v_i \in \text{white}$ ” means that  $v_i$  is unlabelled (i.e. its polarity label is unknown a priori). The parameters  $\rho_{T-black}$  and  $\rho_{T-white}$  represent the different level of confidence related to black and white users, and  $M_{v_i} \subset M$  denotes the set of messages written by user  $v_i$ . Since the user-message feature function  $f_{\alpha, \beta}$  assumes that every message  $m_t \in M$  has a polarity label, a sentiment classification methodology for messages is required. To this purpose, an ensemble method based on Bayesian Model Averaging (BMA) has been used [19, 53]. This choice is motivated by the fact that an ensemble of different models could be less sensitive to noise and could provide a more accurate prediction than single learners [14].

**User-user feature function** A user-user feature function evaluates whether the polarity of a given user agrees (or disagrees) with its neighbour’s sentiment. Given a H-DAG,  $g_{\alpha, \beta}(l(v_i), l(v_j))$  is formally defined as follows:

$$g_{\alpha, \beta}(l(v_i), l(v_j)) = \begin{cases} \frac{\rho_{neigh} \cdot c_{i,j}}{\sum_{v_k \in N(v_i)} c_{i,k}} & l(v_i) = \alpha, l(v_j) = \beta \\ 0 & otherwise \end{cases} \tag{4}$$

where  $\rho_{neigh}$  represents the confidence level of the relationships among users<sup>3</sup> and  $c_{i,j}$  denotes the normalised weights of approval edges in  $\phi$ .

<sup>3</sup>Note that  $\rho_{T-black}, \rho_{T-white}$  and  $\rho_{neigh}$  are empirically estimated.



### 5.1.2 Experiments

In this section, we present a case study to validate the semi-supervised model  $S^2$ -LAN presented above. The experimental investigation is based on connections and messages obtained from Twitter and presents a comparison between several approaches able to consider textual information, structural information and their combination.

**Dataset** In order to evaluate the proposed model, two datasets have been considered:

- The first dataset, named **Obama**, comprises 62 users posting about the President of U.S. Barack Obama, whose tweets have been monitored during the period 8-10 May 2013. The resulting tweets and authors have been manually labeled as positive or negative by three annotators. This dataset consists also of 270 relationships and 160 posts related to the 62 users. In order to validate the proposed modelling on a more complex scenario, a larger dataset has been collected.
- The second dataset, named **Superman**, comprises 3835 users posting about the movie entitled “*Man of Steel*”, whose posts have been monitored during the period 11-12 June 2013. Also in this case, the resulting tweets and authors have been manually labelled. This dataset consists also of 4032 relationships and 5167 posts related to the 3835 users.

Some statistics about the considered dataset are reported in Table 1.

**Compared models** In order to evaluate the proposed model,  $S^2$ -LAN has been compared with traditional approaches based on text, i.e. Dictionary-based Classifier (DIC) [26], Naive Bayes (NB) [45], Maximum Entropy (ME) [46], Support Vector Machines (SVM) [12], Conditional Random Fields (CRF) [60] and Bayesian Model Averaging (BMA) [19].

The classical state-of-the-art measures for classification have been employed, i.e. Precision (P), Recall (R), F1-measure (F1) distinguished in positive (+) and negative (-), together with the well known global Accuracy (Acc) measure. Moreover, in order to show the importance of textual and social network information,  $S^2$ -LAN has been investigated according to the following experimental conditions:

- $S^2$ -LAN (T+A): this corresponds to the proposed model in (2), where both *text* and *approval* relationship are evaluated by the user-message and user-user feature functions;
- $S^2$ -LAN (T+F): this model estimates the likelihood reported in (2) considering *text* and *friendship* relationships (i.e. following/follower) in their respective feature functions;
- $S^2$ -LAN (T): the model is estimated only evaluating the user-message feature functions based on *text*;
- $S^2$ -LAN (A): the model is estimated only evaluating the user-user feature functions based on *approval* relationships;
- $S^2$ -LAN (F): the model is estimated only evaluating the user-user feature functions based on *friendship* (i.e. following/follower) connections.

**Table 1** Characteristics of the Twitter datasets

Dataset	# Users	# Posts	# Relationships	Avg. Neighbours	Net. Diameter
<b>Obama</b>	62	160	267	7.1	3
<b>Superman</b>	3835	5167	4032	2.1	5

**Table 2** Performance on *Obama*: best baseline based on text (BMA) vs the proposed  $S^2$ -LAN (T+A) based on relations and text

	$P_-$	$R_-$	$F1_-$	$P_+$	$R_+$	$F1_+$	Acc
<b>BMA</b>	0.655	0.826	0.731	0.692	0.474	0.563	0.666
$S^2$ -LAN (T+A)	0.933	0.963	0.945	0.958	0.907	0.925	0.938
<b>Gain</b>	+27 %	+13 %	+21 %	+26 %	+43 %	35 %	27 %

Considering that the estimation of the parameters  $\lambda$  and  $\mu$  of  $S^2$ -LAN models is performed by the gradient-based approach named SampleRank, we fixed its maximum number of steps equal to 10000 and we assumed that convergence is achieved when the results are persistent for 500 steps. Moreover, since is based on a sampling function, we performed  $k = \{1, 5, 11, 15, 21, 101\}$  runs to get  $k$  predictions (votes) and take a majority vote among the  $k$  possible labels for each user. For each  $k$ , we performed 500 experiments to compute the average performance.

**Results** The first evaluation relates to the ability of classifiers to detect the polarity of tweets on both dataset. By comparing the state of the art approaches, it emerges that BMA on *Obama* achieves 60.37 % of Accuracy compared with 58.49 % obtained by the best single classifier (CRF), 58.04 % of SVM, 56.12 % of NB, 55.35 % of ME and 55.01 % of DIC. Analogous results can be observed on *Superman*, where BMA obtains 53.1 % of accuracy followed by SVM (52.6 %), CRF (52.0 %), NB (51.7 %), ME (51.3 %) and DIC (51.1 %). The experimental comparison between BMA and the other baseline classifiers has shown that the adopted solution is particularly effective and efficient, thanks to its ability to define a strategic combination of different classifiers through an accurate and computationally efficient heuristic. According to these results, BMA has been selected as text-based baseline at user-level for the comparison with  $S^2$ -LAN. To this purpose, the polarity of a user has been derived by aggregating the polarity prediction on tweets given by BMA through a majority voting mechanism. For instance, if BMA detects three positive and two negative tweets for a given user, the final user label will be positive.

Tables 2 and 3 summarise the performance on *Obama* and *Superman* achieved by  $S^2$ -LAN (T+A) with respect to the BMA based heuristics. The reported results refer to the  $S^2$ -LAN (T+A) configuration based on  $k = 21$ , which represents a good trade-off between running time and accuracy. It can be easily noted that the approval relations enclosed in  $S^2$ -LAN (T+A) ensure a global improvement of 27 % with respect to the text-only method on *Obama* and 14 % on *Superman*. Since BMA does not take into account any kind of relationship, the correct prediction of a user does not have any effect on adjoining users. Considering the network structure, the prediction of each user has impact on all the other

**Table 3** Performance on *Superman*: best baseline based on text (BMA) vs the proposed  $S^2$ -LAN (T+A) based on relations and text

	$P_-$	$R_-$	$F1_-$	$P_+$	$R_+$	$F1_+$	Acc
<b>BMA</b>	0.701	0.681	0.690	0.452	0.403	0.426	0.542
$S^2$ -LAN (T+A)	0.758	0.768	0.762	0.480	0.579	0.524	0.671
<b>Gain</b>	+5 %	+8 %	+7 %	+3 %	17 %	10 %	+14 %

**Table 4** Accuracy comparison: importance of textual and social network information on the Obama dataset

OBAMA					
$k$	S <sup>2</sup> -LAN (T+A)	S <sup>2</sup> -LAN (T+F)	S <sup>2</sup> -LAN (T)	S <sup>2</sup> -LAN (A)	S <sup>2</sup> -LAN (F)
1	<b>0.777</b>	0.692	0.592	0.746	0.735
5	<b>0.856</b>	0.744	0.597	0.839	0.827
11	<b>0.916</b>	0.801	0.596	0.900	0.867
15	<b>0.921</b>	0.799	0.595	0.901	0.880
21	<b>0.938</b>	0.829	0.595	0.919	0.895
101	<b>0.971</b>	0.892	0.595	0.952	0.931

nodes by a “propagation” effect, smoothing each predicted label according to the connected nodes.

In order to better understand the contribution that text and relationships provide to the proposed model, several additional results are reported in the following. We notice that the results achieved on *Obama* are extremely higher compared to the ones obtained on *Superman*. The main reason can be found in the characteristics of the two networks: the *Obama* network structure is more dense and compact than the *Superman* one. This can be deduced by comparing the average number of neighbours and the diameter of the two graphs: while the *Obama* dataset is characterised by an average number of edges per user equal to 7.1 and a diameter coefficient equal to 3, the *Superman* dataset has 2.1 and 5 respectively (see Table 1). In this scenario, a higher average number of neighbours positively affects the diffusion model because more information (from adjoining nodes) is available when processing each user. Additionally a compact network denoted by a low diameter, i.e. the longest distance in the graph, makes the propagation effect less sensible to the noise.

Several additional considerations can be derived when comparing the five variants of S<sup>2</sup>-LAN (See Table 4 and 5). First of all, it is easy to note that in both datasets, the model based only on text (S<sup>2</sup>-LAN (T)) is extremely poor. When considering either only friendship or approval relationships, S<sup>2</sup>-LAN (F) and S<sup>2</sup>-LAN (A) have two different behaviours in the two datasets. While in *Obama* only considering the relational information produces a positive effect on the user-level sentiment prediction, on the *Superman* dataset there is no significant improvement with respect to the text information.

**Table 5** Accuracy comparison: importance of textual and social network information on the Superman dataset

SUPERMAN					
$k$	S <sup>2</sup> -LAN (T+A)	S <sup>2</sup> -LAN (T+F)	S <sup>2</sup> -LAN (T)	S <sup>2</sup> -LAN (A)	S <sup>2</sup> -LAN (F)
1	<b>0.559</b>	0.541	0.551	0.508	0.500
5	<b>0.597</b>	0.590	0.551	0.502	0.497
11	<b>0.634</b>	0.608	0.551	0.533	0.495
15	<b>0.648</b>	0.621	0.553	0.541	0.510
21	<b>0.671</b>	0.639	0.553	0.554	0.514
101	<b>0.687</b>	0.640	0.551	0.562	0.530

These two different behaviours can be again explained by analysing the dataset characteristics. While in *Obama* we can roughly estimate a number of edges per user close to 4.3, for the *Superman* dataset this estimation is about 1.05. According to these statistics, both friendships and approvals are able to provide a good improvement with respect to text in the dense network, while no significant variations into the sparse one. In this context, the denser is the network and the higher is the improvement.

Finally, looking at the results of those models that jointly consider text and relationships, i.e.  $S^2$ -LAN (T+A) and  $S^2$ -LAN (T+F), we can derive two considerations. First of all, both kinds of relationships (when combined with text) are able to improve the user-level sentiment predictions with respect to consider text only. Second, the proposed model  $S^2$ -LAN (T+A) is able to better combine the two ingredients outperforming  $S^2$ -LAN (T+F) in both datasets. This investigation not only confirms that the inclusion of relationships in predictive models, as suggested in other studies [17, 18, 57], leads to improve recognition performance when dealing with non-propositional environments, but also that approval relationships better capture the concepts of homophily and costructuralism than simple friendships.

## 5.2 Aspect-level sentiment analysis

This section is focused on simultaneously extracting aspects and classifying sentiments from textual messages. Most of the works [42, 63] tackle sentiment analysis in social networks at document level. However, social networks contain highly diverse aspects in the same message (e.g., “*iOS7 is very good, but improvements on battery and screen are required*”). The above mentioned works would classify the entire message as globally positive, but the company Apple would be probably more interested in opinions about the aspect ‘*battery*’ and ‘*screen*’ in order to understand how ‘*iOS7*’ could be improved. In order to deal with aspect-level sentiment analysis [27], two tasks need to be addressed:

- **Aspect Extraction:** Aspect extraction consists of two sub-tasks: (1) extracting all aspect terms (e.g., ‘*screen*’) from the corpus, and (2) clustering aspect terms with similar meanings (e.g., group ‘*screen*’ and ‘*display*’ into one aspect category in the domain ‘*iPhone*’). Topic models, such as Latent Dirichlet Allocation (LDA) [5] and Probabilistic Latent Semantic Analysis (pLSA) [25], have been successfully applied to perform both sub-tasks simultaneously.
- **Sentiment Classification:** Once aspects have been extracted, sentiment analysis techniques can be applied to discover the underlying sentiment orientations. In the previous example, the polarity on the aspects *battery* and *screen* is negative.

Several works which deal with sentiment classification and topic modelling have been proposed in the literature. *Topic Sentiment Mixture (TSM)* [49] separates topic and sentiment words using an extended pLSA model. Further models based on the LDA principle can be found in [40] and [34], where *Joint Sentiment/Topic (JST)* and *Aspect and Sentiment Unification Model (ASUM)* have been proposed respectively. The main advantage of joint modelling both tasks derives from its ability to reciprocally reduce their noise.

However, these techniques consider only textual information, disregarding that some relationships can provide useful information to infer the latent aspects and sentiment orientations. To this issue, an unsupervised probabilistic model called **NAS** (Networked Aspect-Sentiment) is proposed as an extension of JST. With NAS, we intend to extract both aspects and sentiments from microblog messages by simultaneously incorporating all the information made available by the proposed approval networks.

5.2.1 Networked Aspect-Sentiment (NAS) model

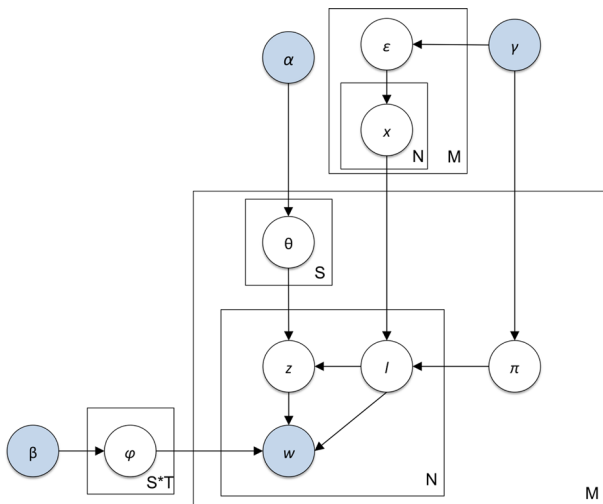
JST (in its original form) is characterised by a good ability to jointly model aspects and sentiment orientations. However, it is not able to use some additional information given by the network structure underlying the user messages. Taking into account the relationships based on approvals could enhance the model to maintain coherent sentiment labels of strong connected users posting about the same topic. The main idea to overcome the limitation of the JST model is to consider (into the generative process and therefore during the inference phase) not only the sentiment label related to specific topic-message-words, but also the sentiment label of messages provided by adjoining users.

Based on the above intuition, we propose a novel generative model called **Network Aspect Sentiment (NAS)**, which drops the sentiment independence assumption that characterise the JST model. As depicted in Figure 3, several dependencies have been modelled in NAS. In particular, analogously to the JST model, the topic  $z$  is conditioned on the sentiment-topic-message distribution  $\theta$ . However, concerning the sentiment label  $l$ , the NAS model not only consider its conditional dependency on the sentiment-message distribution  $\pi$  but also the dependencies on adjoining sentiment labels modelled through the variable  $x$  (regulated by a sentiment-approval-distribution  $\epsilon$ ). The generation of each word  $w$  finally depends on its sentiment-topic-word distribution  $\varphi$ , the topic  $z$  and the sentiment label  $l$ . The entire process takes advantage of hyper-parameters  $\alpha, \beta, \gamma$  to regulate the prior of their respective discrete distributions.

More formally, let's assume that we have a corpus composed of  $M$  messages, where each message  $m$  is a sequence of  $N_m$  words, i.e.  $d = (w_1, w_2, \dots, w_{N_m})$  and each unique word  $w_i$  contributes to create a vocabulary of  $V$  distinct terms. Let  $S$  be the number of sentiment labels and  $T$  the total number of topics.

The joint distribution of the NAS model can be written as:

$$p(w, z, l, x, \varphi, \omega, \epsilon, \pi \mid \alpha, \beta, \gamma) = P(\varphi \mid \beta) P(\omega \mid \alpha) P(\pi \mid \gamma) P(\epsilon \mid \gamma) P(z \mid s, \theta) P(l \mid x, \pi) P(x \mid \epsilon) P(w \mid z, l, \varphi) \quad (5)$$



**Figure 3** Graphical representation of NAS through plate notation. Nodes are random variables, edges are dependencies, and plates are replications. Only shaded nodes are observable

$$\begin{aligned}
 &= \prod_{s=1}^S \prod_{t=1}^T P(\varphi_{st}|\beta) \times \prod_{s=1}^S \prod_{m=1}^M P(\theta_{sm}|\alpha) \times \prod_{m=1}^M P(\pi_m|\gamma) \\
 &\times \prod_{m'=1}^M P(\epsilon_{m'}|\gamma) \times \prod_{m=1}^M \prod_{n=1}^{N_m} P(z_{nm}|\theta_{sm}) \times \prod_{m'=1}^M \prod_{n=1}^{N_{m'}} P(x_{nm'}|\epsilon_{m'}) \\
 &\times \prod_{m=1}^M \prod_{n=1}^{N_m} P(w_{md}|\varphi_{st}) \times \prod_{m=1}^M \prod_{n=1}^{N_m} P(l_{nm}|\pi_m) \tag{6}
 \end{aligned}$$

$$\begin{aligned}
 &= \prod_{s=1}^S \prod_{t=1}^T Dir(\varphi_{st}|\beta) \times \prod_{s=1}^S \prod_{m=1}^M Dir(\theta_{sm}|\alpha) \times \prod_{m=1}^M Dir(\pi_m|\gamma) \\
 &\times \prod_{m=1}^M Dir(\epsilon_{m'}|\gamma) \times \prod_{m=1}^M \prod_{n=1}^{N_m} Disc(z_{nm}|\theta_{sm}) \times \prod_{m=1}^M \prod_{n=1}^{N_m} Disc(x_{nm'}|\epsilon_{m'}) \\
 &\times \prod_{m=1}^M \prod_{n=1}^{N_m} Disc(w_{nm}|\varphi_{st}) \times \prod_{m=1}^M \prod_{n=1}^{N_m} Disc(l_{nm}|\pi_m) \tag{7}
 \end{aligned}$$

where  $Dir(\cdot)$  and  $Disc(\cdot)$  denote Dirichlet and Discrete distributions respectively.

Given the  $n$ -th word of the  $m$ -th message, the collapsed sampler needs to compute the probability of topic  $z_{nm}$  and label  $l_{nm}$  being assigned to the word  $w_{nm}$ , given all the other topic assignments ( $z_{nm}^-$ ) and the sentiment realisations ( $l_{nm}^-$ ) of all the other words, which are further conditioned on the sentiment realisations  $x_{nm'}$  of messages  $m'$  of adjoining users. The sampler needs therefore to estimate the joint probability of topic and sentiment as follows:

$$P(z_{nm}, l_{nm} | z_{nm}^-, l_{nm}^-, w_{nm}, x_{nm}, \alpha, \beta, \gamma) \tag{8}$$

By definition of conditional probability, we can derive that (8) corresponds to:

$$\frac{P(z_{nm}, l_{nm}, z_{nm}^-, l_{nm}^-, w_{nm}, x_{nm} | \alpha, \beta, \gamma)}{P(z_{nm}^-, l_{nm}^-, w_{nm}, x_{nm} | \alpha, \beta, \gamma)} \tag{9}$$

Since the denominator does not depend on  $z_{nm}$  and  $l_{nm}$ , it can be removed obtaining:

$$\propto P(z_{nm}, l_{nm}, z_{nm}^-, l_{nm}^-, w_{nm}, x_{nm} | \alpha, \beta, \gamma) \tag{10}$$

Now, considering that  $z_{nm}$  together with  $z_{nm}^-$  is just  $z$ , and  $l_{nm}$  together with  $l_{nm}^-$  corresponds to  $l$ , (10) can be written in a compact form as:

$$P(z, l, w, x | \alpha, \beta, \gamma) \tag{11}$$

Using the rule of total probability, the topic distribution  $\theta$ , the sentiment distributions  $\pi$  and  $\epsilon$ , and the word distribution  $\varphi$ , can be integrated out as follows:

$$\begin{aligned}
 &P(z, l, w, x|\alpha, \beta, \gamma) \\
 &= \int \int \int \int P(z, l, w, x, \varphi, \theta, \epsilon, \pi|\alpha, \beta, \gamma) d\varphi d\theta d\epsilon d\pi \tag{12}
 \end{aligned}$$

Expanding the integrand given the model defined in (5), we can obtain:

$$\begin{aligned}
 &P(z, l, w, x|\alpha, \beta, \gamma) \\
 &= \int \int \int \int P(\varphi|\beta)P(\theta|\alpha)P(\pi|\gamma)P(\epsilon|\gamma)P(z|l, \theta)P(l|x, \pi)P(x|\epsilon)P(w|z, l, \varphi) d\varphi d\theta d\epsilon d\pi \\
 &= \int P(\varphi|\beta)P(w|z, l, \varphi)d\varphi \times \int P(\theta|\alpha)P(z|l, \theta)d\theta \\
 &\times \int P(\pi|\gamma)P(l|x, \pi)d\pi \times \int P(\epsilon|\gamma)P(x|\epsilon)d\epsilon \tag{13}
 \end{aligned}$$

Expanding out the terms according to the independence assumption in (6), we obtain the following form of the joint probability distribution:

$$\begin{aligned}
 P(z, l, w, x|\alpha, \beta, \gamma) &= \int \prod_{s=1}^S \prod_{t=1}^T P(\varphi_{st}|\beta) \prod_{m=1}^M \prod_{n=1}^{N_m} P(w_{nm}|\varphi_{stn_m}) d\varphi \\
 &\times \int \prod_{s=1}^S \prod_{m=1}^M P(\theta_{sm}|\alpha) \prod_{m=1}^M \prod_{n=1}^{N_m} P(z_{nm}|\theta_{sm_t}) d\theta \\
 &\times \int \prod_{m=1}^M P(\pi_m|\gamma) \prod_{m=1}^M \prod_{n=1}^{N_m} P(l_{nm}|\pi_{m_s}) d\pi \\
 &\times \int \prod_{m'=1}^M P(\epsilon_{m'}|\gamma) \prod_{m'=1}^M \prod_{n=1}^{N_m} P(x_{nm'}|\epsilon_{m'}) d\epsilon \tag{14}
 \end{aligned}$$

Expanding out the Dirichlet priors and the discrete distributions according to their definitions, we can derive:

$$\begin{aligned}
 P(z, l, w, x|\alpha, \beta, \gamma) &= \left( \frac{\Gamma(\sum_{i=1}^V \beta_i)}{\prod_{i=1}^V \Gamma(\beta_i)} \right)^{ST} \times \prod_{t=1}^T \prod_{s=1}^S \frac{\prod_{i=1}^V \Gamma(N_{its} + \beta)}{\Gamma(N_{ts} + V\beta)} \\
 &\times \left( \frac{\Gamma(\sum_{i=1}^T \alpha_i)}{\prod_{i=1}^T \Gamma(\alpha_i)} \right)^{SM} \times \prod_{s=1}^S \prod_{m=1}^M \frac{\prod_{t=1}^T \Gamma(N_{tsm} + \alpha)}{\Gamma(N_{ts} + T\alpha)} \\
 &\times \left( \frac{\Gamma(\sum_{i=1}^S \gamma_i)}{\prod_{i=1}^S \Gamma(\gamma_i)} \right)^M \times \prod_{m=1}^M \frac{\prod_{s=1}^S \Gamma(N_{sm} + \gamma)}{\Gamma(N_m + S\gamma)} \\
 &\times \left( \frac{\Gamma(\sum_{i=1}^S \gamma_i)}{\prod_{i=1}^S \Gamma(\gamma_i)} \right)^M \times \prod_{m'=1}^M \frac{\prod_{s=1}^S \Gamma(N_{sm'} + \gamma)}{\Gamma(N_{m'} + S\gamma)} \tag{15}
 \end{aligned}$$

**Table 6** Characteristics of the Twitter datasets

Dataset	# Users	# Posts	# Relationships
<b>iOS7</b>	62343	80001	25932
<b>Lone Survivor</b>	69389	59825	18591

A sample obtained from the Markov chain of the Gibbs sampler can be used to approximate the distribution of words in topic and sentiment labels, obtaining therefore the following joint probability distribution:

$$p(z, l, w, x | \alpha, \beta, \gamma) \propto \underbrace{\frac{N_{its} + \beta}{N_{ts} + V\beta} \times \frac{N_{ism} + \alpha}{N_{sm} + T\alpha} \times \frac{N_{sm} + \gamma}{N_m + S\gamma}}_{\mathbf{A} \text{ (i. e. JST)}} \times \underbrace{\sum_{(v_i, v_j) \in E} c_{i,j} \frac{1}{|M(v_j)|} \frac{\sum_{m' \in M(v_j)} \frac{N_{sm'} + \gamma}{N'_m + S\gamma}}{\sum_{(v_i, v_j) \in E} c_{i,j}}}_{\mathbf{B} \text{ (i. e. Approval Network)} \times \mathbf{C}} \tag{16}$$

The last constituent (**B**) highlights the contribution provided by the Approval Network. The rationale behind **B** is to flatten the relationships between messages of adjoining users ( $m'$ ) by aggregating their sentiment label  $s$  through their average (**C**). This contribution is weighted by considering the edge coefficients  $c_{ij}$  encoded by the HN-DAG and representing the strength of the existing relationships between users.

### 5.2.2 Experiments

In this section, we present the experimental results on real data to demonstrate that the inclusion of the approval network in the proposed NAS model outperforms the state-of-the-art baseline methods for both sentiment classification and aspect extraction tasks.

**Dataset and Settings** Two datasets from Twitter have been collected by monitoring users (tweets and retweets) posting on “*iOS7*” and “*Lone Survivor*”. *iOS7* contains positive, negative and neutral tweets, while *Lone Survivor* is composed of positive and neutral tweets. In Table 6, some basic statistics are reported for both datasets.

In order to transform microblog messages to a more canonical language, URLs, mention tags (@), hashtags (#) and retweet (RT) symbols have been removed and misspelled tokens have been automatically corrected using the Google’s Spell Checker API.<sup>4</sup> Emoticons, initialisms for emphatic expressions (e.g., ‘ROFL’, ‘LMAOL’, ‘ahahah’, ‘ehh’, ...) and onomatopoeic expressions (e.g., ‘bleh’, ‘wow’, ...) have been replaced by POS\_EXP, NEU\_EXP and NEG\_EXPR, according to their sentiment.

For a direct comparison of the proposed model with the state-of-the-art, a comparative analysis has been performed with the following approaches:

<sup>4</sup><https://code.google.com/p/google-api-spelling-java/>.



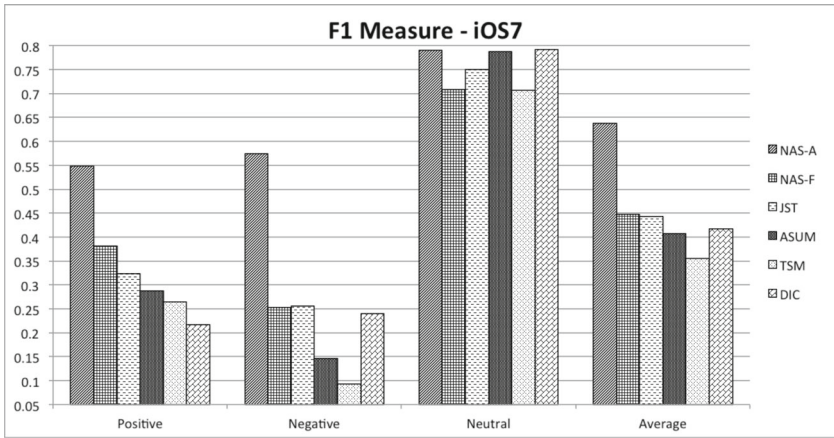
- **DIC**: the first baseline for sentiment classification purposes is the dictionary-based classifier, where the overall message polarity is determined by first checking whether each term belongs to the positive, negative or neutral lexicon [28] and then using the majority voting strategy;
- **ASUM**: this model, based on the Aspect Sentiment Unification Model presented in [34], computes the joint probability  $P(z, l, w)$  using 1000 Gibbs iterations with 100 burn-in iterations, disregarding any relational information encoded by the social network.
- **JST**: it corresponds to the Joint Sentiment Topic model presented in [40] and works analogously to the ASUM model for sampling the joint probability  $P(z, l, w)$ . No relational information is included.
- **TSM**: this model, based on the Topic Sentiment Model presented in [49], computes the joint probability  $P(z, l, w)$  using the Kullback-Leibler divergence measure over the sentiment coverage in order to determine the model convergence. Also in this case, no relational information is included.
- **NAS-A**: it corresponds to the model presented in Section 5.2.1, where the joint probability  $P(z, l, w, x)$  is estimated using 1000 Gibbs iterations with 100 burn-in iterations. This model includes the relational information encoded by the *Approval Network*.
- **NAS-F**: this approach works similarly to NAS-A, but it considers the *Following/Follower* relationships instead of the approval ones when computing the joint probability distribution.

Supported by a detailed investigation of the main aspects into the data, the number of aspects has been set  $T = 10$  for the *iOS7* dataset, while  $T = 5$  for *Lone Survivor*. Concerning the sentiment associated to each aspect, the SentiWordNet resource [15] has been exploited for all the investigated models. This allowed us to automatically derive dynamic and context-dependent sentiment labelling.

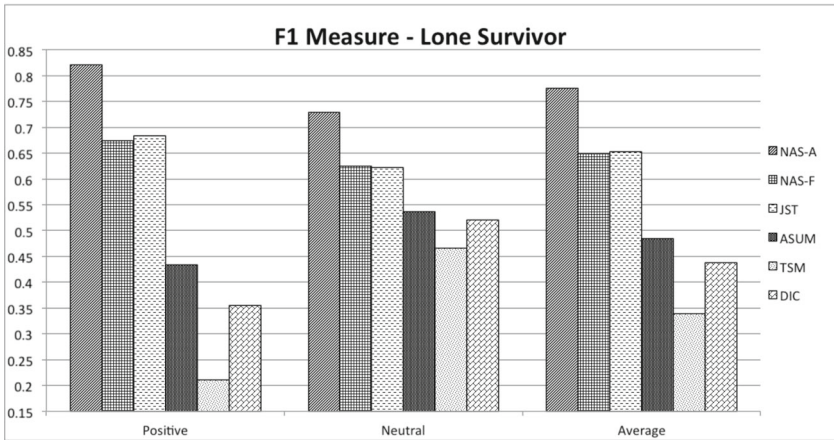
**Results on Sentiment classification** In order to evaluate the performance of the considered models for the sentiment classification task, F1-measure (which aggregates both Precision and Recall) has been computed over all the sentiments.

If we focus on *iOS7* (Figure 4a), we can note the outperforming results of NAS-A with respect to the other approaches. The first consideration relates to the direct comparison of NAS-A with the popular baseline DIC. As expected, DIC achieves low performance of F1-measure (0.41 on *iOS7* and 0.43 on *Lone Survivor*). Conversely, NAS-A is able to achieve valuable improvements showing macro-average F1-measure of 0.63 on *iOS7* and 0.77 on *Lone Survivor*. Looking at the performance of NAS-F, it can be noted that although it is able to provide a small relative gain with respect to JST, TSM and ASUM, the use of friendship relationships is not able to outperform the NAS-A model. NAS-A is able to achieve the best results on *iOS7*, leading to a gain of 23 % compared to ASUM, 19 % compared to JST and NAS-F, 28 % compared to TSM. Similar results can be observed in Figure 4b on *Lone Survivor*, where NAS-A achieves a gain of 29 % over ASUM, 12 % over JST and NAS-F and 42 % over TSM.

The outperforming results of the proposed model are due to several reasons. First, NAS-A (and NAS-F) overcomes the ASUM assumption stating that the sentiment of all the words of a given sentence must be consistent to each other. Second, it works well with positive, negative and neutral opinions because it is able to deal with the co-occurrence of different sentiment seed words. For instance, consider the sentence “*I bought the new iPhone and the screen is very nice.*”, where a neutral word (‘new’) co-occurs with a positive word (‘nice’)



(a) F1-measure on iOS7



(b) F1-measure on Lone Survivor

**Figure 4** Results on sentiment classification

in the same sentence. Finally, the proposed model shows that including the information encoded by the approval network leads to tackle those situations where words are not sufficient or their sentiment orientations are contradictory with respect to the overall message polarity. According to its ability to take into account the sentiment value of adjoining users, the proposed model is capable to detect positive (or negative) messages when none of their words are positive (or negative). This ability can be noted by looking at some examples reported in Table 7, where tweets are correctly classified only by NAS-A.

For instance, the tweet *'iOS 7 looks like a child's coloring book!!'* is correctly classified as negative by NAS-A, even if it does not contain negative words. Conversely, the tweet *'If Lone Survivor didn't change your life, you're fucking insane'* as been classified by

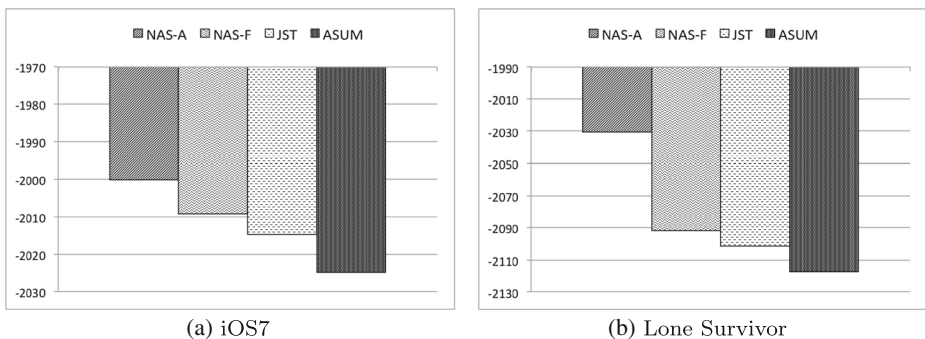
**Table 7** Examples of tweets whose sentiment is correctly captured by the NAS model

NAS-A	JST	ASUM	Tweet
POS	NEG	NEG	If Lone Survivor didn't change your life, you're fucking insane.
NEU	NEG	NEG	So I know how to get iOS 7 but I'm to stupid to figure out what I'm doing
POS	NEU	NEU	Lone survivor time!!
NEG	POS	NEU	Why are you thinking of switching? iOS7 looks great, just like Android
NEG	NEU	NEU	iOS 7 looks like a child's coloring book!!

NAS-A as positive towards *Lone Survivor*, even if it contains only neutral and negative words. In these two examples, NAS-A correctly classifies them as negative/positive because their authors retweet other users who emit negative/positive tweets.

**Results on Aspect Discovery** In order to automatically measure the quality of the aspects extracted by all the considered models, the *Topic Coherence* measure [50] has been adopted. This measure depends on the corpus without using any other resource and is computed according to the co-document frequency ratio among top topical terms. Note that, given the initial number of aspects  $T$  and sentiments  $S$ , NAS-A, NAS-F, JST and ASUM produce  $T \times S$  aspects. Since TSM performs the worst in sentiment classification and it has  $T$  aspects as output, Topic Coherence has not been computed because it is sensitive to the number of aspects [50]. If we focus on Figure 5, we can note that the inclusion of the network through NAS-A leads to significant improvements on both datasets. ASUM achieves the lowest performance, followed by JST and finally by NAS-A.

Even if Topic Coherence is a good measure to compare aspects extracted by different models, it is not able to provide a clear idea of their pros and cons. For this reason, we discuss some examples reported in Tables 8 and 9 related to two main aspects, i.e. *Battery Life* and *Security*. The aspect *Security* has been weakly identified by ASUM only through the words ‘privacy’ and ‘unlocked’, because the model tends to relate nouns with neutral aspects and adjectives to positive/negative aspects, making their characterisation very difficult. Moreover, inconsistent adjectives for positive and negative aspects (e.g., ‘love’ for a negative aspect) and several non-related words are chosen by ASUM.



**Figure 5** Comparison of Topic Coherence

**Table 8** Aspects related to “Battery Life” in iOS7

<b>BATTERY LIFE</b>			
ASUM (neu)	JST (neu)	NAS-F (n)	NAS-A (n)
screen	battery	<b>battery</b>	<b>battery</b>
<b>battery</b>	<b>life</b>	<b>life</b>	<b>life</b>
lock	<b>os</b>	<del>fun</del>	<b>half</b>
panoramas	<b>half</b>	<u>bad</u>	<u>bad</u>
<del>cool</del>	<b>problem</b>	<del>posexp</del>	<u>negexp</u>
video	unauthorized	<b>hours</b>	<u>fast</u>
panoramic	toy	lock	<b>problem</b>
background	drop	app	<b>night</b>
wallpaper	update	lock-screen	<b>day</b>
lock-screen	app	<u>fast</u>	<b>hours</b>

Red/~~strike~~ indicates incoherent adjectives; green/underline indicates coherent adjectives; **bold** denotes coherent nouns; *italic* represents incoherent nouns

Conversely, JST, NAS-F and NAS-A have a good balance of adjectives and nouns for positive and negative aspects that allow us to easily characterise the aspect and simultaneously understand its perceived sentiment. Thanks to a manual investigation of the data, we have discovered that “Battery life” is negatively perceived because of its short life, while “Gaming” is positively perceived thanks to the introduction of controllers’ support and kits. Unlike neutral opinions, positive and negative opinions are mostly identifiable by the presence of opinionated words (e.g., adjectives). According to this consideration, only NAS-A and NAS-F lead to the correct identification of the negative aspect *Battery Life*. Conversely, JST has a more clear behaviour in excluding opinionated words that generally leads the model to infer the neutral sentiment.

**Table 9** Aspects related to “Security” in iOS7

<b>SECURITY</b>			
ASUM (n)	JST (n)	NAS-F (n)	NAS-A (n)
<u>shit</u>	<u>bad</u>	<b>lock</b>	<b>lock</b>
<u>ugly</u>	<b>security</b>	spent	hidden
<u>hate</u>	<b>major</b>	mind	<b>flaw</b>
<del>love</del>	true	<b>security</b>	<b>privacy</b>
<u>negexp</u>	<b>flaw</b>	stop	<b>security</b>
<u>bad</u>	<u>ugly</u>	gamepad	<u>annoying</u>
<u>horrible</u>	<u>negexp</u>	<u>scared</u>	<u>scared</u>
<u>tough</u>	privacy	sitting	<u>ugly</u>
<b>unlocked</b>	<u>annoying</u>	<del>love</del>	<u>negexp</u>
<b>privacy</b>	unauthorized	entire	<u>hate</u>

Red/~~strike~~ indicates incoherent adjectives; green/underline indicates coherent adjectives; **bold** denotes coherent nouns; *italic* represents incoherent nouns

As general remark, the experimental comparison suggests that the introduction of network information may not only help the sentiment classification step, but also the ability of identifying aspects. A more important observation is concerned with NAS-A when compared to NAS-F. The proposed model based on approval relationships is able to better capture the aspects underlying the microblog posts thanks to its ability to model the interest of users on the same topic.

A final remark related to the proposed NAS model, relates to its ability to deal with short and noisy text. The fact that social network text is composed of few words poses considerable problems when applying traditional topic/sentiment models. These models typically suffer from data sparsity to estimate robust word co-occurrence statistics when dealing with short and ill-formed text. The proposed model is able to reduce the negative impact of short and noisy text thanks to the contagion of connected users. The relational information enclosed into the generative model is able to compensate the biased statistics when considering independent text.

## 6 Conclusion

Text does not always provide explicit or sufficient information about the sentiment orientation of a short text. Early studies have tried to overcome this limitation by modelling the user-user similarity through friendships. However, being friends does not necessarily mean agreeing on all issues. Supported by the theory behind two sociological processes (homophily and constructivism), this paper has proposed to model user-user relationships through Approval Networks. To finally capture the real heterogeneity of social network data, a Heterogeneous Directed Approval Graph (H-DAG) has been presented to model both textual contents and user relationships. Two novel sentiment analysis models based on H-DAG have been introduced, confirming that the inclusion of approval relationships in predictive models leads to significant improvement in terms of effectiveness.

## References

1. Agarwal, B., Mittal, N., Bansal, P., Garg, S.: Sentiment analysis using common-sense and context information. *Intell. Neuroscience*
2. Aristotle: Rhetoric. nichomachean ethics. In: Aristotle, vol. 23, p. 1371. Rackman Translation Cambridge: Harvard University Press (1934)
3. Barbosa, L., Feng, J.: Robust sentiment detection on twitter from biased and noisy data. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL '10 (2010)
4. Beigi, G., Hu, X., Maciejewski, R., Liu, H.: An overview of sentiment analysis in social media and its applications in disaster relief. In: Sentiment Analysis and Ontology Engineering, pp. 313–340. Springer (2016)
5. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003)
6. Bott, H.: Observation of play activities in a nursery school. *Genet. Psychol. Monogr.* **4**, 44–88 (1928)
7. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: Proceedings of the 2010 43rd Hawaii International Conference on System Sciences, HICSS '10, pp. 1–10. IEEE Computer Society (2010)
8. Carley, K.: A theory of group stability. *Am. Sociol. Rev.* **56**(3), 331–354 (1991)
9. Carley, K.M., Hill, V.: Structural change and learning within organizations. *Dynamics of organizations: Computational modeling and organizational theories* 63–92 (2001)
10. Cohen, J.M.: Sources of peer group homogeneity. *Sociol. Educ.* **50**(4), 227–241 (1977)
11. Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., Flammini, A.: Political polarization on twitter. In: ICWSM (2011)

12. Cortes, C., Vapnik, V.: Support-vector networks. *ML* **20**(3), 273–297 (1995)
13. Coviello, L., Sohn, Y., Kramer, A.D., Marlow, C., Franceschetti, M., Christakis, N.A., Fowler, J.H.: Detecting emotional contagion in massive social networks. *PLoS one* **9**(3), e90, 315 (2014)
14. Dietterich, T.G.: Ensemble learning. In: *The Handbook of Brain Theory and Neural Networks*, pp. 405–508. MIT Pr (2002)
15. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: *Proceedings of LREC*, vol. 6, pp. 417–422. Citeseer (2006)
16. Fang, L., Huang, M., Zhu, X.: Exploring weakly supervised latent sentiment explanations for aspect-level review analysis. In: *Proceedings of the 22nd ACM International Conference on Conference on Information and Knowledge management*, pp. 1057–1066. ACM (2013)
17. Fersini, E., Messina, E.: Web page classification through probabilistic relational models *International Journal of Pattern Recognition and Artificial Intelligence* (2013)
18. Fersini, E., Messina, E., Archetti, F.: A probabilistic relational approach for Web document clustering. *Inf. Process. Manag.* **46**(2), 117–130 (2010)
19. Fersini, E., Messina, E., Pozzi, F.: Sentiment analysis: Bayesian ensemble learning. *Decis. Support Syst.* **68**, 26–38 (2014)
20. Garcia-Pablos, A., Cuadros, M., Gaines, S., Rigau, G.: V3: Unsupervised generation of domain aspect terms for aspect based sentiment analysis. *SemEval* **2014**, 833 (2014)
21. Gatti, M., Cavalin, P., Neto, S., Pinhanez, C., dos Santos, C., Gribel, D., Appel, A.: Large-scale multi-agent-based modeling and simulation of microblogging-based online social network. In: *Multi-Agent-Based Simulation XIV*, pp. 17–33. Springer, Berlin Heidelberg (2014)
22. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. Technical report, Stanford (2009)
23. Hampton, K.N., Wellman, B.: Examining community in the digital neighborhood: Early results from Canada’s wired suburb. In: *Digital Cities*, vol. 1765, pp. 194–208. Springer, Berlin Heidelberg (2000)
24. Hirshman, B.R., Charles, J., Carley, K.M.: Leaving us in tiers: Can homophily be used to generate tiering effects. *Comput. Math. Organ. Theory* **17**(4), 318–343 (2011)
25. Hofmann, T.: Probabilistic latent semantic indexing. In: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 50–57. ACM (1999)
26. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and DM*, pp. 168–177 (2004)
27. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’04*, pp. 168–177. ACM (2004)
28. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: *Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining*, pp. 168–177 (2004)
29. Hu, X., Tang, J., Gao, H., Liu, H.: Unsupervised sentiment analysis with emotional signals. In: *Proceedings of the 22nd International Conference on World Wide Web*, pp. 607–618. International World Wide Web Conferences Steering Committee (2013)
30. Hu, X., Tang, L., Tang, J., Liu, H.: Exploiting social relations for sentiment analysis in microblogging. In: *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, WSDM ’13*, pp. 537–546. ACM (2013)
31. Hubbard, R.M.: A method of studying spontaneous group formation. *Some New Techniques for Studying Social Behavior* pp 76–85 (1929)
32. Huston, T.L., Levinger, G.: Interpersonal attraction and relationships. *Ann. Rev. Psychol.* **29**, 115–156 (1978)
33. Jin, X., Wang, C., Luo, J., Yu, X., Han, J.: Likeminer: A system for mining the power of ‘like’ in social media networks. In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 753–756. ACM, NY, USA (2011)
34. Jo, Y., Oh, A.H.: Aspect and sentiment unification model for online review analysis. In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM ’11*, pp. 815–824. ACM (2011)
35. Joseph, K., Morgan, G.P., Martin, M.K., Carley, K.M.: On the coevolution of stereotype, culture, and social relationships: An agent-based model *Social Science Computer Review* (2013)
36. Kandel, D.B.: Homophily, selection, and socialization in adolescent friendships. *Am. J. Sociol.* **84**(2), 427–436 (1978)
37. Kascheky, M., Riedl, R.: Tracing opinion-formation on political issues on the internet: A model and methodology for qualitative analysis and results. In: *2011 44th Hawaii International Conference on System Sciences (HICSS)*, pp. 1–10 (2011)

38. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World wide Web, WWW '10, pp. 591–600 (2010)
39. Lazarsfeld, P.F., Merton, R.K.: Friendship as a social process: A substantive and methodological analysis. *Freedom and Control in Modern Society* pp 18–66 (1954)
40. Lin, C., He, Y.: Joint sentiment/topic model for sentiment analysis. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management, pp. 375–384. ACM (2009)
41. Liu, Y., Yu, X., Liu, B., Chen, Z.: Sentence-level sentiment analysis in the presence of modalities. In: Computational Linguistics and Intelligent Text Processing, pp. 1–16. Springer (2014)
42. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, HLT '11, vol. 1, pp. 142–150 (2011)
43. Marsden, P.V.: Core discussion networks of americans. *Am. Sociol. Rev.* **52**(1), 122–131 (1987)
44. Marsden, P.V.: Homogeneity in confiding relations. *Soc. Netw.* **10**, 57–76 (1988)
45. McCallum, A., Nigam, K.: A comparison of event models for naive bayes text classification. In: AAAI-98 workshop on Learning for Text Category, pp. 41–48 (1998)
46. McCallum, A., Pal, C., Druck, G., Wang, X.: Multi-conditional learning: Generative/discriminative training for clustering and classification, pp. 433–439. AAAI (2006)
47. McPherson, M., Mayhew, B.H., Rotolo, T., Smith-Lovin, L.: Sex and ethnic heterogeneity in face-to-face groups in public places: An ecological perspective on social interaction. *Soc. Forces* **74**, 15–52 (1995)
48. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. *Annual review of sociology* pp 415–444 (2001)
49. Mei, Q., Ling, X., Wondra, M., Su, H., Zhai, C.: Topic sentiment mixture: Modeling facets and opinions in Weblogs. In: Proceedings of the 16th International Conference on World Wide Web, WWW '07, pp. 171–180. ACM (2007)
50. Mimno, D., Wallach, H.M., Talley, E., Leenders, M., McCallum, A.: Optimizing semantic coherence in topic models. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 262–272. Association for Computational Linguistics (2011)
51. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* **2**, 1–135 (2008)
52. Plato: *Laws*. In: Plato in Twelve Volumes, vol. 11, p. 837. Bury Translator. Cambridge: Harvard University Press (1968)
53. Pozzi, F.A., Fersini, E., Messina, E.: Bayesian model averaging and model selection for polarity classification. In: Proceedings of the 18th International Conference on Application of Natural Language to Information Systems, LNCS, pp. 189–200 (2013)
54. Pozzi, F.A., Maccagnola, D., Fersini, E., Messina, E.: Enhance user-level sentiment analysis on microblogs with approval relations. In: AI\*IA 2013: Advances in Artificial Intelligence, vol. 8249, pp. 133–144. Springer International Publishing (2013)
55. Rabelo, J.C., Prudêncio, R.C., Barros, F.A.: Leveraging relationships in social networks for sentiment analysis. In: Proceedings of the 18th Brazilian Symposium on Multimedia and the Web, WebMedia '12, pp. 181–188. ACM (2012)
56. Scharl, A., Weichselbraun, A.: An automated approach to investigating the online media coverage of us presidential elections. *J. Inf. Technol. Polit.* **5**(1), 121–132 (2008)
57. Sharara, H., Getoor, L., Norton, M.: Active surveying: A probabilistic approach for identifying key opinion leaders. In: IJCAI, pp. 1485–1490 (2011)
58. Speriosu, M., Sudan, N., Upadhyay, S., Baldrige, J.: Twitter polarity classification with label propagation over lexical links and the follower graph. In: Proceedings of the First Workshop on Unsupervised Learning in NLP, EMNLP '11 (2011)
59. Sun, J., Tang, J.: A survey of models and algorithms for social influence analysis. In: *Social Network Data Analytics*, pp. 177–214. Springer (2011)
60. Sutton, C.A., McCallum, A.: An introduction to conditional random fields. *Found. Trends ML* **4**(4), 267–373 (2012)
61. Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., Li, P.: User-level sentiment analysis incorporating social networks. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '11, pp. 1397–1405 (2011)
62. Tang, D.: Sentiment-specific representation learning for document-level sentiment analysis. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, pp. 447–452. ACM (2015)
63. Wang, S., Manning, C.D.: Baselines and bigrams: simple, good sentiment and topic classification. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, ACL '12, vol. 2, pp. 90–94 (2012)
64. Wellman, B.: The school child's choice of companions. *J. Educ. Res.* **14**, 126–32 (1929)

65. Wick, M., Rohanimanesh, K., Culotta, A., McCallum, A.: Samplerank: Learning preferences from atomic gradients. In: NIPS Workshop on Advances in Ranking (2009)
66. Wu, L., Hu, X., Liu, H.: Relational learning with social status analysis. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, pp. 513–522. ACM (2016)
67. Xiong, X., Ma, J., Wang, M., Zhou, G., Xu, K.: Information diffusion model in modular microblogging networks. *World Wide Web* **18**(4), 1051–1069 (2015)
68. Yang, B., Cardie, C.: Context-aware learning for sentence-level sentiment analysis with posterior regularization. In: Proceedings of ACL (2014)
69. Yessenalina, A., Yue, Y., Cardie, C.: Multi-level structured models for document-level sentiment classification. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 1046–1056. Association for Computational Linguistics (2010)
70. Zhang, L., Liu, B.: Identifying noun product features that imply opinions. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2, HLT '11, pp. 575–580. Association for Computational Linguistics (2011)
71. Zhou, X., Wan, X., Xiao, J.: Collective opinion target extraction in chinese microblogs, vol. 13. Citeseer (2013)
72. Zhu, C., Zhu, H., Ge, Y., Chen, E., Liu, Q.: Tracking the evolution of social emotions: A time-aware topic modeling perspective. In: 2014 IEEE International Conference on Data Mining (ICDM), pp. 697–706. IEEE (2014)