# Sentiment Analysis of Tweets using Heterogeneous Multi-layer Network Representation and Embedding

**Loitongbam Gyanendro Singh**[*], **Anasua Mitra**[*], **Sanasam Ranbir Singh**
Department of Computer Science and Engineering
Indian Institute of Technology Guwahati, Assam, India
{gyanendrol9, anasua.mitra, ranbir}@iitg.ac.in

## Abstract

Sentiment classification on Twitter text often needs to deal with the problems of under-specificity, noise, and multilingual content. This study proposes a heterogeneous multi-layer network-based representation of tweets to generate multiple representations of a tweet and address the above issues. The generated representations are further ensembled and classified using a neural-based early fusion approach. Further, we propose a centrality aware random-walk for node embedding and tweet representations suitable for the multi-layer network. From various experimental analyses, it is evident that the proposed method can address the problem of under-specificity, noisy text, and multilingual content present in a tweet and provides better classification performance than the text-based counterpart. Further, the proposed centrality aware based random walk provides better representations than unbiased and other biased counterparts.

## 1 Introduction

With the growing popularity of Twitter, sentiment analysis of tweets has drawn the attention of several researchers from both academia and industry in recent times. Unlike other regular texts, sentiment analysis on Twitter text poses plenty of challenges because of various characteristics such as (i) under-specificity due to text limit, (ii) nonstandard writing format due to the presence of texts, hashtags, mentions in any order as the user wishes to write, (iii) noisy text due to the presence of short-form, long-form, multilingual, transliterated text, misspelling, etc. Researchers tried to address these problems by adopting various methods like task-specific representation learning (Singh et al., 2020; Pham and Le, 2018; Fu et al., 2018; Tang et al., 2016; Kim, 2014), incorporating additional information such as hashtags (Alfina et al., 2017; Qadir and Riloff, 2014), relationship between users (Zhao et al., 2017), multi-source information (Zhou and Huang, 2017), ensembling (Al-Twairesh and Al-Negheimish, 2019; Araque et al., 2017; Wang et al., 2014), etc.

This paper proposes a novel approach to handle the above issues using a heterogeneous multi-layer network representation of a tweet. A multi-layer network is a network formulated by connecting different layers of networks. For example, a heterogeneous multi-layer network can be formed by connecting layers of mention's network, hashtag's network, keyword's co-occurrence networks. Multi-layer networks have shown to provide promising performance in the other tasks like community detection and clustering (Hanteer and Rossi, 2019; Luo et al., 2020), node classification (Li et al., 2018; Zitnik and Leskovec, 2017; Ghorbani et al., 2019), representation learning in graphs (Cen et al., 2019; Zhang et al., 2018; Ni et al., 2018), etc. A tweet or a collection of tweets can be represented by a multi-layer network. An advantage of using network-based representation is that a network can be expanded by adding nodes or shrunk by removing nodes. The motivations of using a multi-layer network in this paper are as follows. (i) The semantic relation between keywords, hashtags, and mentions can be captured by applying an effective network embedding method. (ii) The noise and under-specificity can be reduced by expanding the network with related nodes or shrinking the network after removing unrelated nodes. Further, the co-occurring keywords, hashtags, and mentions often share semantic relationship (Wang et al., 2016; Weston et al., 2014; Qadir and Riloff, 2013; Wang et al., 2011).

This paper has three major contributions. First, it proposes a centrality[1] aware random walk based

[*]Equal contributions

[1]Prominence of a node in a network

network embedding from a large multi-layer network of a large tweet collection. Second, it can represent a tweet as a multi-layer network, extracts multiple forms of representation from the network, applies early fusion to merge different representations, and classifies the tweet. Third, it reduces noise in the tweet by expanding the network with related nodes or shrinking network after removing noisy nodes.

Sentiment classification is a domain-dependent task (Karamibekr and Ghorbani, 2012). Therefore, we evaluate the proposed method over datasets in different domains. From extensive experimental evaluations, the proposed method is found to outperform its counterparts in majority of the cases. To the best of the authors' knowledge, this study is the first of its kind to investigate sentiment classification task by transforming tweet into a heterogeneous multi-layer network.

Rest part of the paper is organized as follows. Section 2 present the literature related to this study. Section 3 present our proposed framework. The experimental setup of the study is described in Section 4. We discuss the results and observations in Section 5. Finally, we conclude the study in Section 6.

## 2 Related studies

Sentiment analysis is an old research area. Initial work on sentiment classification can be traced back as early as 2000 (Turney, 2002; Pang et al., 2002; Turney and Littman, 2003). There have been several paradigm shifts in sentiment analysis methods from statistical methods (Turney, 2002; Pang et al., 2002; Turney and Littman, 2003) to rule-based (Prabowo and Thelwall, 2009), to lexicon-based (Taboada et al., 2011; Balamurali et al., 2011; Mohammad et al., 2009), to feature-based (Kouloumpis et al., 2011; Barbosa and Feng, 2010), to deep neural network (Kim, 2014; Severyn and Moschitti, 2015). Majority of the recent studies focus on the application of neural network models. Therefore, this section briefly reviews a few of the recent and related studies which have exploited graph and neural models.

Authors in (Violos et al., 2016) use a homogeneous network known as *word graph* to represent a document by connecting co-occurring words in the document. Three different networks are created for positive, negative, and neutral classes using the documents in respective classes. Using these networks, a document is represented by a three-dimensional vector defined by the three sentiment classes). The elements of the vector correspond to the similarity of the word graph of the document and the word graph of the respective sentiment class. The vector thus obtained is used for classifying the document. Similarly, authors in (Bijari et al., 2020) construct co-occurrence word-graph of a document collection and generate word embedding using Node2Vec (Grover and Leskovec, 2016). The embeddings thus obtained are used to represent words in the text and build a classifier using the Convolution Neural Network (CNN) model. Further in the studies (Gui et al., 2017; Zhao et al., 2017), the advantages of exploiting the relationship between keywords, sentiment, products and users etc. have also been evident in sentiment analysis.

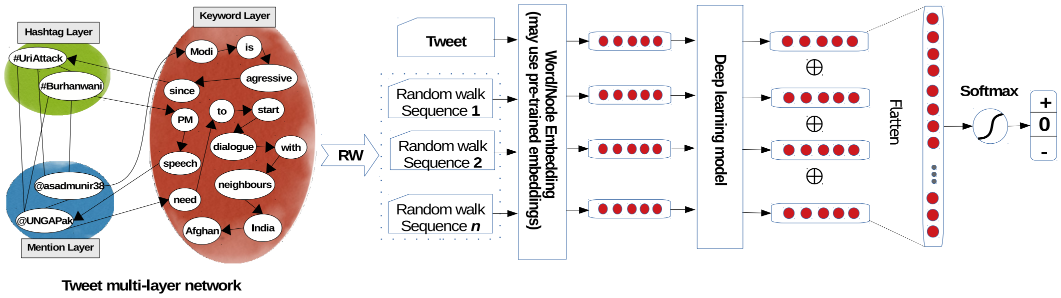
In recent times, deep learning based models are extensively used for sentiment classification. To mention few of them, authors in (Jianqiang et al., 2018; Dahou et al., 2016; Severyn and Moschitti, 2015; dos Santos and Gatti, 2014; Kim, 2014) use CNN, (Xu et al., 2019; Liu and Zhang, 2017) use Long Short Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), (Nguyen and Nguyen, 2018; Chen et al., 2017) use a combination of CNN and reccurrent based models. Further, studies (Al-Twairesh and Al-Negheimish, 2019; Araque et al., 2017) use neural ensembled models to combine different representation of text.

## 3 Proposed framework

As mentioned earlier, the proposed method has four distinct components; (i) representation of a tweet or collection of a tweet using multi-layer network, (ii) embedding of the multi-layer network using centrality aware random walk, (iii) tweet classification using multiple representations generated from the multi-layer network of a tweet, and (iv) reduction of noise in a tweet by expanding or shrinking network. This section discusses the details of these components. Figure 1 shows a high-level schematic diagram of the proposed model using a heterogeneous multi-layer network.

### 3.1 Representation of Tweets using Multi-layer Network

A $L$-layer network $\mathbf{G}$ is defined by $(\mathbf{V}, \mathbf{E}, \mathcal{L})$ where $\mathcal{L}$ denotes the set of layer indices $\{1, 2, ..., L\}$, $\mathbf{V} = \{\mathbf{V}^1 \cup \mathbf{V}^2 \cup ... \cup \mathbf{V}^L\}$, $\mathbf{V}^i$ denotes the set of vertices in layer $i$ of the network,




**Tweet**: *@asadmunir38 Modi is agressive since #UriAttack, #BurhanWani & PM speech @UNGAPak needs to start dialogue with neighbours India, Afghan*

Figure 1: Proposed heterogeneous multi-layer network based tweet sentiment classification framework

$\mathbf{E}$ denotes the set of edges. Considering three important components of a tweet, the proposed multi-layer network is formed with three layers i.e., *hashtag, mention* and *keyword* as $\{H, M, K\}$. To capture both the co-occurrence and sequential characteristics of keywords, hashtags and mentions in a tweet, the proposed network consists of both directed and un-directed edges. An edge $e_{x,y} \in \mathbf{E}$ is directed if $x$ and $y$ occur sequentially next to other in a tweet where, i) $x, y \in V^K$ or ii) $x \in V^K$ and $y \in \{V^H \cup V^M\}$ or iii) $x \in \{V^H \cup V^M\}$ and $y \in V^K$. Whereas, an edge $e_{x,y} \in \mathbf{E}$ is un-directed if $x, y \in \{V^H \cup V^M\}$ co-occur in a tweet. An example of the proposed multi layer network for the tweet *"@asadmunir38 Modi is agressive since #UriAttack, #BurhanWani & PM speech @UNGAPak needs to start dialogue with neighbours India, Afghan"* is shown in Figure 1(a). Edge set $\mathbf{E} = \{\mathbf{A} \cup \mathbf{B}\}$ which comprises of a set of intra-layer adjacency matrices $\mathbf{A} = \{\mathbf{A}^1, \mathbf{A}^2, ..., \mathbf{A}^L\}$ with matrix $\mathbf{A}^i \in \mathcal{R}^{N^i \times N^i}$ in each layer $i$. A set of bipartite matrices $\mathbf{B}^{i,j} \in \mathcal{R}^{N^i \times N^j}$ represents cross-layer association between layer $i$ and layer $j$. For our tweet-multilayer network, we have three layers $\mathbf{A} = \{\mathbf{A}^H, \mathbf{A}^M, \mathbf{A}^K\}$ and five types of bipartite associations $\mathbf{B} = \{\mathbf{B}^{HM}, \mathbf{B}^{MK}, \mathbf{B}^{HK}, \mathbf{B}^{KM}, \mathbf{B}^{KH}\}$. This kind of complex networks also have one flattened representation in form of supra-adjacency matrix $S$, with total nodes $N = |\mathbf{V}^H| + |\mathbf{V}^M| + |\mathbf{V}^K|$,

$$\mathbf{S}_{N \times N} = \begin{bmatrix} \mathbf{A}^H & \mathbf{B}^{HM} & \mathbf{B}^{HK} \\ \mathbf{B}^{MH} & \mathbf{A}^M & \mathbf{B}^{MK} \\ \mathbf{B}^{KH} & \mathbf{B}^{KM} & \mathbf{A}^K \end{bmatrix} \tag{1}$$

The intra-layer associations $\mathbf{A}$s are on the main-diagonal of $\mathbf{S}$ and the cross-layer connections $\mathbf{B}$ are the off-diagonal elements. Further, $\mathbf{A}^K, \mathbf{B}^{HK}, \mathbf{B}^{KH}, \mathbf{B}^{MK}, \mathbf{B}^{KM}$ are asymmetric matrices and other matrices of $\mathbf{S}$ are symmetric. A tweet or a collection of tweets can be represented as a multi-layer network as discussed above.

### 3.2 Centrality aware random-walk with restart for heterogeneous multi-layer network

To generate network embedding from the proposed multi-layer tweet network, we extend the random walk followed in PageRank (Brin and Page, 1998) algorithm. Given a row stochastic adjacency matrix $\mathbf{A}$ of a network, the PageRank of the nodes in the network can be defined as the following vector.

$$\vec{\pi}_{t+1} = (1 - \delta)\mathbf{A}\vec{\pi}_t + \delta\vec{\pi}_0 \tag{2}$$

where $\vec{\pi}_t$ is the stationary probability distribution vector that depicts the probability with which a random walker would stay in a particular node at time $t$. The restart probability $\delta \in [0, 1]$ denotes the probability of jumping to a random node and $\vec{\pi}_0$ is the initial stationary probability vector.

As in (Li and Patra, 2010), the above random-walk can be extended to our tweet multi-layer heterogeneous network in the following manner. If $\lambda \in (0, 1)$ is the probability that a random-walker jumps to a different layer while surfing, in presence of $L$ number of layers and considering jumping to any of the remaining layers is equiprobable, the transition probability $\mathbf{M}$ aka column-normalized supra-adjacency matrix $\mathbf{S}$ in equation 1, is modified as,

$$\mathbf{M} = \begin{bmatrix} (1-\lambda)A^H & \frac{\lambda}{L-1}B^{HM} & \frac{\lambda}{L-1}B^{HK} \\ \frac{\lambda}{L-1}B^{MH} & (1-\lambda)A^M & \frac{\lambda}{L-1}B^{MK} \\ \frac{\lambda}{L-1}B^{KH} & \frac{\lambda}{L-1}B^{KM} & (1-\lambda)A^K \end{bmatrix} \tag{3}$$

That is, for a node, if its bipartite association exists, a random-surfer can stay in the same layer with probability $(1 - \lambda)$ or transit to a different layer with probability $(\frac{\lambda}{L-1})$. Now, equation 2 can be re-written as follows,

$$\vec{\pi}_{t+1} = (1-\delta)\mathbf{M}\vec{\pi}_t + \delta\vec{\pi}_{rs} \tag{4}$$

where $\vec{\pi}_{rs} = \begin{bmatrix} \eta_H.\vec{\pi}_0^H \\ \eta_M.\vec{\pi}_0^M \\ \eta_K.\vec{\pi}_0^K \end{bmatrix}$, $\eta_i$ denotes the importance of layer $i$, $\vec{\pi}_0^i$ denotes the initial stationary distribution of nodes in layer $i$ and $\sum_{i\in\{H,M,K\}} \eta_i = 1$. And, $\vec{\pi}_t \in \mathcal{R}^{(N^H+N^M+N^K)}$ is the stationary probability distribution of a random surfer on the heterogeneous multi-layer network at time $t$.

In this study, we propose to personalize the above PageRank algorithm using global importance of nodes in the proposed heterogeneous multi-layer network. In Equation 4, $\vec{\pi}_{rs}$ the initial stationary probability vector can be interpreted as layer importance weighted initial stationary probabilities of nodes based on node-importances. This interpretation needs not only the node centrality scores, but also the layer importances. MultiRank (Rahmede et al., 2018), a centrality estimate for multiplex networks[2] formulated using a modified version of PageRank, can estimate both the node centrality scores as well as the layer influences. MultiRank uses one layer-influence weighted aggregated adjacency matrix and one weighted bipartite matrix that relates nodes with layers to determine the node and layer centrality scores simultaneously. We specifically change the definition of these two matrices in accordance with the heterogeneous tri-layer network structure to customize MultiRank. As we calculate the centrality scores, we modify $\vec{\pi}_{rs}$ of Equation 4 by replacing each $\eta_i$ with respective influence score of layer $i$ and each initial steady-state vector $\vec{\pi}_0^i$ with node centrality scores in layer $i$.

In the Heterogeneous MultiRank algorithm, we have tuned free-parameters i) to suppress/enhance the contribution of low-centrality nodes, ii) to take into account elite layers that contain a few highly central nodes, iii) to/ not to normalize layer influences by weighted layer in-strength – while calculating the centrality scores. We have tuned the restart parameter in MultiRank and in multi-layer random walks in the range $\in (0.5, 0.7, 0.85)$. We set the walk-length at 30 and the number of walks as 10. All the free parameters are tuned based on the end-task performance.

[2]Multiplex network (Kivelä et al., 2014) is a special case of a multi-layer network that has the same set of nodes exhibiting distinct relations in different layers.

| Node embedding methods |
|---|
| FastText Embedding (FT) (Bojanowski et al., 2017) |
| Multi-View Embedding (MVE) (Qu et al., 2017) |
| Multiplex Network Embedding (MNE) (Zhang et al., 2018) |
| Sentiment Hashtag Embedding (SHE) (Singh et al., 2020) |

* The embedding dimension is of 128 sizes. Same hyper-parameter as define in the literature.

| Deep-learning models | Hyper-parameter |
|---|---|
| Convolution Neural Network (CNN) | 3 Kernels, 128 #Filters, *ReLu* Activation Function |
| Bidirectional Long Short Term Memory (Bi-LSTM) | 64 LSTM Units, *ReLu* Activation Function |

Table 1: Different embedding and neural methods

### 3.3 Classification of tweets represented with a multi-layer network

Let $G_i$ be the multi-layer network representing a tweet $T_i$. Over this network, we generate $n$ number of node sequences $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_n\}$ by using the above proposed random walk. Each node sequence is maintained to have a length of $m$ nodes. With $n$ number of random sequences and the original tweet, we have $(n+1)$ sentences to represent the tweet $T_i$. Each word in these sentences can be represented using a vector obtained from an appropriate embedding method. This paper has considered different embedding methods, as listed in Table 1, over a large collection of tweets.

For each node sequence $\mathcal{S}_i$, we apply a neural model (Bi-LSTM (Chen et al., 2017) and CNN (Kim, 2014) in our study) to generate a representation of the sequence. While using Bi-LSTM, the hidden state obtained after applying the last node in the sequence is considered to represent the node sequence $\mathcal{S}_i$. Similarly, the vector that we obtained after applying the pooling step in CNN represents the sequence. Thus, we obtained $(n+1)$ vectors for each tweet. We concatenate these $(n+1)$ vectors and feed it to a feed-forward dense layer with three neurons (each for positive, negative, and neutral) and classify the sentiment of the tweet using softmax activation function in the output layer as shown in Figure 1(b). We use Keras[3] deep learning framework for building our proposed model.

We calculate the error loss ($\Delta$) for the classifier using the well-known cross-entropy loss as,

$$\Delta = -\frac{1}{l}\sum_{i=1}^{l}\sum_{c}\mathbf{t}_{ic}log(s_{ic}) \tag{5}$$

where $c$ is the number of sentiment classes, $\mathbf{t}_{ic}$ is the $c^{th}$ ground truth class for the tweet, $l$ is the total number of training samples, and $\mathbf{s}_{ic}$ is the predicted probability on sample $i$ for the $c^{th}$ class.

[3]https://keras.io

### 3.4 Network expansion and shrinking

One of the motivations of using the multi-layer network for representing a tweet lies in its flexibility to expand or shrink the network. Given a set of existing nodes in a tweet-network as query nodes, the idea is to identify the most related nodes or most noisy nodes by exploiting a multi-layer network of a global tweet collection. We consider the most central and most similar neighboring nodes of the query nodes as potential expansion candidates. To reduce the search space, we first select top query nodes from a tweet's network view ranked by their centrality scores in the network calculated from the tweet collection. We then find neighbours of the selected nodes and ranked them using weighted combination of similarity and centrality score using the scoring function defined below: $Score(v) = \sum_{u \in N_v} \alpha.sim(v, u) + (1 - \alpha).centrality(u)$ where $N_v$ denotes neighbouring nodes of $v$, $sim(v, u)$ denotes cosine similarity using node embeddings of node $v$ and node $u$ , and $centrality(u)$ denotes centrality score of node $u$ in global network. In this study, we take equal weights of cosine and centrality score by setting $\alpha = 0.5$. Top neighbouring nodes are selected using the above scoring function and added to the network in their respective layers using the edge policy discussed in Section 3.1.

The above node expansion method finds new nodes having semantic relation with the query nodes. However, for the sentiment analysis task, we are interested in adding only sentiment carrying nodes. We use Sentiment Hashtag Embedding (SHE) proposed in (Singh et al., 2020) to estimate the sentiment orientation of a node. For building the SHE classifier, we have used the same experimental setup as described in (Singh et al., 2020). Among the newly added nodes, we determine the domination of sentiment classes and keep only those nodes carrying the dominant sentiment class. The rest of the newly selected nodes are removed from the network of the tweet under consideration.

## 4 Experimental Setup

### 4.1 Dataset

This paper considers a locally annotated dataset named as *Societal*. We have collected $50,300$ tweets using Twitter Streaming API[4] over four events that happened in India during August-December 2016, namely *Uri Attack, Surgical Strike, GST Amendment Bill, and Demonetization*. Two annotators with strong command on English and Hindi are engaged to annotate the tweets with *positive, negative,* and *neutral* sentiments. We have selected $42,422$ tweets where the two annotators have agreed on the same sentiment, which is of 85% agreement having 82.35 Kappa coefficient scores. Majority of the disagreements among the annotators are with the tweets of stance, sarcastic nature. A similar observation is also reported in (Karamibekr and Ghorbani, 2012). The `Societal` dataset contains 18% non-English tweets (i.e., Hindi and it code-mix with English), of which $1,626$ code-mix tweets and $1,505$ tweets with less than five keywords are kept unseen for evaluation of our proposed model. Meanwhile, the hashtags and mentions cover 11% and 15% of the total $39,428$ unique vocabulary of the `Societal` dataset. This tweet collection is used to build sentiment classifiers and construct a multi-layer network to generate node embeddings. Details of the dataset is shown in Table 2.

| Heterogeneous Multi-layer Tweet Network | | | |
|---|---|---|---|
| **Relation** | **#Nodes** | **#Edges*** | **Edge-type** |
| Hashtag-Hashtag, $A^H$ | 3552 | 10776 | Undirected |
| Mention-Mention, $A^M$ | 4243 | 12277 | Undirected |
| Keyword-Keyword, $A^K$ | 28962 | 181849 | Directed |
| Hashtag-Mention, $B^{HM}$ | 6446 | 13765 | Undirected |
| Hashtag-Keyword, $B^{HK}$ | 4782 | 6648 | Directed |
| Mention-Keyword, $B^{MK}$ | 7958 | 14790 | Directed |
| Keyword-Hashtag, $B^{KH}$ | 6824 | 11825 | Directed |
| Keyword-Mention, $B^{KM}$ | 4018 | 5813 | Directed |

* The edges are weighted by normalized co-occurrence frequency.

| Tweet Corpus | | | | |
|---|---|---|---|---|
| **Dataset** | **#Positive** | **#Negative** | **#Neutral** | **Total Tweets** |
| *Societal* | 16375 | 17047 | 9000 | 42422 |

Table 2: Statistical characteristics of the dataset

### 4.2 Embedding method

We investigate the efficacy of our proposed multi-layer network using four different types of node embedding methods namely Multiplex Network Embedding (MNE) (Zhang et al., 2018), Multi-View Embedding (MVE) (Qu et al., 2017), FastText (FT) (Bojanowski et al., 2017), and Sentiment Hashtag Embedding (SHE) (Singh et al., 2020) (listed in Table 1). These embedding methods need a collection of node sequences. Therefore, we represent the tweet corpus into a large multi-layer network by combining the whole tweet networks. For experimental comparison, we investigate three random walk methods to generate the node sequences, namely *Unbiased* random walk used in MNE, biased random walk used in

[4] `http://docs.tweepy.org`

| Types of tweet representation | | Accuracy (in %) | | | | | | | | F-Macro (in %) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **CNN** | | | | **Bi-LSTM** | | | | **CNN** | | | | **Bi-LSTM** | | | |
| | RW | BFT | MNE | MVE | SHE | BFT | MNE | MVE | SHE | BFT | MNE | MVE | SHE | BFT | MNE | MVE | SHE |
| Original Tweet | – | 77.92 | 75.53 | 77.01 | 76.89 | 75.22 | 74.53 | 73.64 | 76.05 | 76.62 | 73.52 | 75.33 | 75.38 | 72.43 | 72.59 | 71.60 | 74.39 |
| [A] **T+MLN** | Unbiased | 73.96 | 74.90 | 75.10 | 76.51 | 74.83 | 74.38 | 73.90 | 75.70 | 70.99 | 72.14 | 72.68 | 73.49 | 71.88 | 72.62 | 71.69 | 72.67 |
| | N2V | 75.61 | 75.45 | 75.02 | 74.15 | 74.65 | 72.57 | 72.84 | 73.84 | 72.56 | 73.03 | 72.83 | 71.68 | 72.09 | 70.51 | 70.70 | 70.82 |
| | Biased | 77.88 | 74.30 | 74.39 | 77.27 | 75.89 | 74.70 | 74.37 | 75.63 | 75.07 | 71.34 | 72.83 | 74.85 | 73.35 | 72.58 | 72.73 | 73.00 |
| [B] **T+MLN+NE** | Unbiased | 76.20 | 75.30 | 75.08 | 77.18 | 75.31 | 74.96 | 74.53 | 75.51 | 73.85 | 72.93 | 73.04 | 74.48 | 72.87 | 71.63 | 72.17 | 73.08 |
| | N2V | 75.30 | 73.80 | 72.67 | 73.84 | 74.54 | 74.77 | 72.49 | 73.84 | 72.46 | 71.47 | 70.91 | 72.13 | 72.25 | 72.50 | 70.75 | 71.85 |
| | Biased | 78.33 | 76.57 | 76.54 | 77.88 | 76.33 | 75.08 | 75.05 | 76.53 | 76.84 | 74.15 | 73.01 | 75.05 | 74.92 | 73.41 | 73.32 | 74.44 |
| [C] **T+MLN+SNE** | Unbiased | 78.72 | 76.20 | 77.17 | 79.37 | 76.97 | 74.87 | 75.73 | 76.79 | 77.39 | 76.43 | 75.52 | 78.09 | 75.73 | 73.08 | 74.32 | 73.84 |
| | N2V | 77.77 | 76.66 | 77.38 | 76.87 | 76.72 | 72.45 | 76.47 | 76.11 | 76.68 | 75.50 | 76.13 | 74.65 | 75.30 | 70.86 | 73.41 | 73.69 |
| | Biased | 79.23 | 77.97 | 78.14 | **80.78** | 78.95 | 77.11 | 78.16 | 79.33 | 77.33 | 76.73 | 76.90 | **79.79** | 77.39 | 75.79 | 76.66 | 78.22 |
| [D] **T**+Shuffle | Unfiltered | 73.86 | 76.66 | 76.26 | 77.49 | 74.98 | 75.05 | 76.26 | 76.33 | 73.04 | 75.15 | 74.20 | 75.04 | 72.91 | 73.29 | 74.54 | 73.93 |
| | Filtered | 77.54 | 77.17 | 77.84 | 77.89 | 76.21 | 76.84 | 76.98 | 77.78 | 76.48 | 75.95 | 76.43 | 75.07 | 75.07 | 75.32 | 75.18 | 76.17 |

* **T**: Tweet, **MLN**: Multi-layer Network, **NE**: Node Expansion, **SNE**: Sentiment polarized Node Expansion

Table 3: Performance of sentiment classifiers across different embedding and representations. Blue: Embedding method that performs best for each tweet representations. Red: Best performing tweet representation for each embedding models. Purple: Best performing classifier across different representation of tweet and embedding models. **Purple bold**: Overall best.

Node2Vec (*N2V*) (Grover and Leskovec, 2016) and the proposed *Biased* random walk. Moreover, to investigate the efficacy of our proposed random walk (RW), we modeled the generated biased RW sequences using the FastText embedding model – which we refer to as Biased FT (BFT) in Table 3.

### 4.3 Selection of $n$ random walks

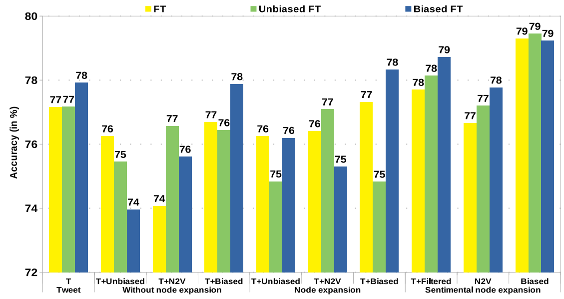
A random walker can generate various node sequences from a given network starting from a node. However, all of the sequences are not useful. To identify the nodes sequences of our interest, we consider a simple second-order Markov chain based language model (Lafferty and Zhai, 2001) by calculating the probability of generating a node sequence given a tweet network. This study considers the best three-node sequences.[5] [6]

## 5 Results and observations

In Table 3, we show the performance of two sentiment classifiers CNN (Nguyen and Nguyen, 2018) and Bi-LSTM (Xu et al., 2019) in terms of accuracy and F-Macro scores over the *Societal* dataset using 10-fold cross validation approach for four embedding models of our choice namely Multiplex Network Embedding (MNE) (Zhang et al., 2018), Multi-View Embedding (MVE) (Qu et al., 2017), FastText (FT) (Bojanowski et al., 2017), and Sentiment Hashtag Embedding (SHE) (Singh et al., 2020). We consider the work of (Nguyen and Nguyen, 2018; Xu et al., 2019) as the baseline models for text-based sentiment classification of tweet. Along the rows of Table 3, we have three groups namely [A], [B] and [C] pertaining to the three types of tweet-network representations, where we compare three different types of random-walks (RWs) – *Unbiased*, *Node2Vec*(N2V) (Grover and Leskovec, 2016) and the proposed *Biased RW* to generate node sequences required as inputs for the above embedding methods. From the table, we can see that the network representation of tweets helps the sentiment classification task. Though the tweet-text only classification (in the first row) is hard to beat with multi-layer network representation of a tweet without node expansion, but for Bi-LSTM based classifier, the biased RW in the group [A] beats text only prediction in $75\%$ of the cases with a maximum of $1.13\%$ in terms of F-macro using Biased FT embeddings. For CNN, biased RW in [A] beats original tweet prediction using SHE embeddings. Although group [B] gave a competitive performance as compared to text-only classifiers, sentiment polarized node expansion (SNE) methods in [C] beats tweet-text based prediction by a huge margin of $1.4\%$, and $1.9\%$ (on average) for CNN and Bi-LSTM classifiers respectively – indicating the network representation of tweets, especially when augmented with informative nodes, are useful and complements the text in tweets. Among the RW based methods for node sequence generation, the proposed *Biased* RW performs the best followed by *Unbiased* and *N2V*. The proposed *Biased* RW outperforms *Unbiased* RW decently – can be seen with prominence in [A] *Biased* vs *Unbiased* using Biased FT embeddings in CNN. Even the best performances in both the metrics pertain to [C] *Biased* RW with SHE embeddings using both the classifiers. We feel the *N2V* style global topology-based biasing is not that useful for sentiment prediction than our biased approach, which uses centrality scores intuitively. Among the embedding models, we observe that Biased FT and SHE give competitive performances.

[5] We have considered only the top few walks (3, 5, and 7) with the highest probability. Experiments show that considering the top 3 walks provide the best results.

[6] Our code is available at: https://github.com/gloitongbam/SA_Hetero_Net




* The plot shows different scale but of same value due to round-off error.

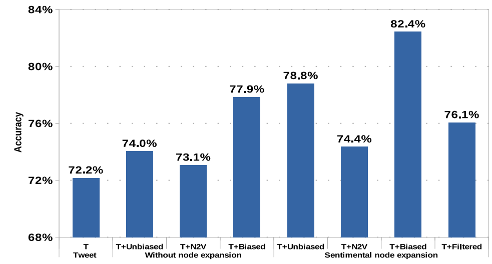
Figure 2: Performance of CNN classifier using different types of node embedding generated via FastText algorithm

We believe Biased FT performs competitively as it is trained on centrality-aware random-walks, additionally augmented with sentiment polarized nodes. Whereas, SHE systematically embeds sentiment information and also aided by biased tweet graph view – this makes it an unbeatable performer for sentiment classification.

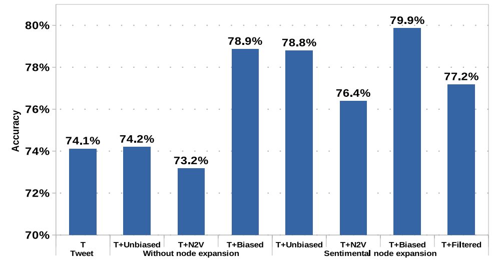
To realize the importance of generating node sequences with an effective RW method over the proposed network, we investigate another experimental setup by randomly mixing nodes selected for expansion and the tweet text. We call it as T+Shuffle–*Unfiltered* (*Filtered*) methods for plain(sentiment polarized node) expansion respectively in [D]. For Bi-LSTM, we can see [D] *Unfiltered* beats text-only prediction, which signifies that the list of selected nodes, though randomly shuffled, but are informative enough. For both the classifiers, [D] *Filtered* outperforms text-only prediction on average by $0.8\%, 2.4\%$, respectively, signifying selected nodes by SNE method aids in performance. Here we shall also showcase the novelty of node sequences over a randomly shuffled list of the same nodes. [D] *Unfiltered* is comparable with [B] view – biased RWs are seen to improve upon the prior. Whereas walks in the [C] view, which is comparable to [D] *Filtered* are seen to improve the performance of the latter. [C] *Biased* beats [D] *Filtered* by $1.6\%, 1.5\%$ points on average for CNN and Bi-LSTM.

### 5.1 Novelty of biased random-walk sequences

It is evident from the already-shown results that our proposed biased random-walks are useful for the effective representation of tweets. One may be further interested in knowing how far these biased RW sequences can improve the performance of any embedding models. We conduct a pilot study by creating three versions of the FastText algorithm – a word embedding based original version (FT), an unbiased RW sequence-based version (Unbiased FT), and a biased RW sequence-based version (Biased FT) as summarized in Figure 2. Biased FT beats original FT in 6 out of 10 cases by an average of $1.11\%$. Biased FT also beats Unbiased FT in $6/10$ cases by an average of $1.37\%$. Although Unbiased FT seems to perform poorer as compared to the original FT in general, in the case of sentiment polarized node expansion, it consistently outperformed the FT – which again proves the effectiveness of the SNE method.




(a) Tweets with keywords $< 5$



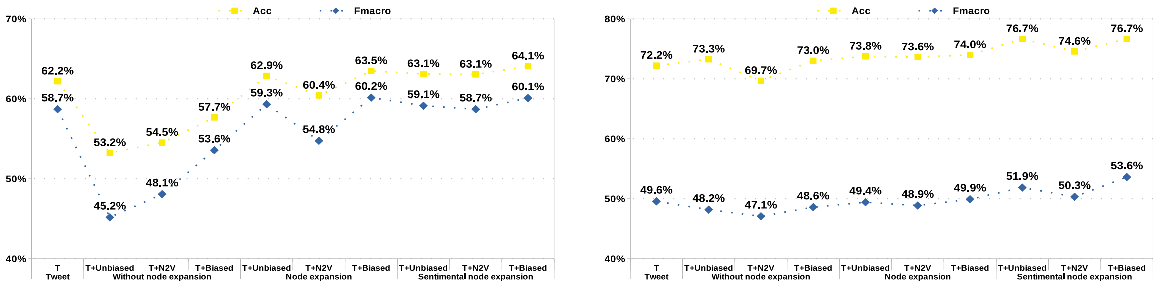

(b) Multilingual tweets

Figure 3: Performance of CNN classifier for different under-specified tweet categories. Inputs to classifier are 5 different tweet representations; i.e. (i) tweet-text only, and node expansion over the actual tweet using random walkers based on (ii) MNE (Unbiased), (iii) Node2Vec (N2V), and (iv) centrality biased node expansions (Biased), and (v) random shuffled of the selected sentiment polarized nodes (Filtered).

### 5.2 Response on under-specified Tweets

We consider tweets having less than five keywords[7] as an under-specified tweet. Tweets with fewer keywords, although informative, can pose challenges to sentiment classifiers due to under-specificity. We considered the CNN-based classifiers trained using Biased FT embedding to classify the under-specified tweets for this study. Figure 3(a) shows the CNN-based classifiers' performance based on the different types of tweet representations. From the figure, we observed that the sentiment classifier trained without any node expansion performs better

[7]Including hashtags and mentions




(a) SemEval-2013 (b) SemEval-2016

* (The figures are shown in the dotted line plots instead of bar plots due to space constraints.)

Figure 4: Performance of CNN classifier across Twitter challenge datasets

than the classifier trained with tweet-text only. This observation shows the power of optimally selected $n$ random-walk sequences as an alternative representation of tweets. Among no expansion methods, biased RW sequences give the best performance – beat tweet-text only prediction by $5.7\%$ and unbiased RW by $3.82\%$. We can see similar trends of performance for RW based sequences in case of sentiment polarized node expansion also. However, SNE expansion strategically mitigates the problem under-representation of lesser-keyword tweets by extending the tweet-network view to include less-noisy informative nodes so that the generated walks are more diverse and discriminating. The last pair of columns is one special scenario where we give the original tweet-text + list of randomly-shuffled sentiment polarized nodes to the sentiment classifier. This combination (T+Filtered) outperforms the tweet only prediction by $3.9\%$ – depicting nodes selected for expansion are important for inference. However, as T+Biased w/o node expansion (NE), T+Unbiased with SNE and T+Biased with SNE beat this T+Filtered by a large margin of $1.8\%, 2.7\%$ & $6.4\%$ accuracy respectively, it proves the veracity of this fact that random-walk sequences are a stronger representation of tweets as compared to mere inclusion of a shuffled-list of semantically related words to the tweet-text.

### 5.3 Response on Multilingual tweet

Figure 3(b) shows sentiment classification performance over the multilingual tweets – tweet-text written in the code-mixed language. This plot also follows similar trends, as reflected in Figure 3(a), but we have two striking observations this time. In the case of multilingual tweets, since the co-occurrence of multilingual words is rare, our proposed node expansion methods are useful to retrieve semantically related co-occurred English words that can aid in inference. We verify the same intuition with this plot. We can see the jump in prediction results where T+Unbiased with SNE, T+N2V with SNE, and T+Biased with SNE beat their counterparts in the previous group (w/o node expansion) with a margin of $4.6\%, 3.2\%$ and $0.1\%$ accuracy respectively. It is interesting to see the huge performance improvement of T+Biased w/o NE over tweet only prediction by a margin of $4.75\%$ accuracy – which we believe is due to the power of interpretable, centrality-score aided, optimally biased RW sequences of multilingual words.

### 5.4 Evaluation on Twitter challenge datasets

We further investigate the performance of the proposed method with two popularly datasets used in Twitter challenges; SemEval-2013[8] and SemEval-2016[9]. For this study, we consider the train and test split provided in the datasets. Figure 4(a) and (b) shows the performance of the CNN classifier trained over different types of tweet representation using the SemEval-2013 and SemEval-2016 datasets, respectively. For training the CNN classifier, we use Biased FT embeddings trained using the challenge datasets. Our proposed centrality aware-based biased random walker through sentiment polarized node expansion has achieved best performance up to 64% accuracy and 60% F-macro score on SemEval-2013 and up to 77% accuracy and 54% F-macro score for SemEval-2016. Further, comparing the performance of tweet representation between text-based and network-based without node expansion, it is observed that for both datasets, the representation without node expansion could hardly beat text-based representation in F-macro measure. However, for the SemEval-2016 dataset, our proposed method beats text-based representa-

[8] https://www.cs.york.ac.uk/semeval-2013/task2/
[9] http://saifmohammad.com/WebPages/StanceDataset.htm

tion in both the evaluation measures. We see substantial performance gain for N2V RW (Node2Vec) in both the datasets when augmented with any kind of node expansion. For SemEval-2016, a fascinating thing to observe is – unbiased and biased RW-based sequences almost give a comparable performance with the same total points in terms of accuracy. However, the biased RW view consistently outperformed the unbiased view in F-macro measure in both datasets for each of the cases of node expansion. This points to the fact that our method consistently performs better across all sentiment classes than the unbiased method.

## 6 Conclusion

This study investigates the efficacy of transforming tweets to heterogeneous multi-layer network for the sentiment classification task. Our proposed centrality aware random walk based method can generate random-sequences that capture better semantic relations than unbiased and other biased random walk counterparts. From various experimental observations, it is evident that sentiment-oriented node expansion can reduce under-specificity, noise in a tweet and enhance the representation. The proposed method outperforms its text-based counterpart in a majority of the cases.

## 7 Acknowledgement

This work is partially funded by the Ministry of Electronics & Information Technology, Government of India.

## References

Nora Al-Twairesh and Hadeel Al-Negheimish. 2019. Surface and deep features ensemble for sentiment analysis of arabic tweets. *IEEE Access*, 7:84122–84131.

Ika Alfina, Dinda Sigmawaty, Fitriasari Nurhidayati, and Achmad Nizar Hidayanto. 2017. Utilizing hashtags for sentiment analysis of tweets in the political domain. In *Proceedings of the 9th International Conference on Machine Learning and Computing*, pages 43–47. ACM.

Oscar Araque, Ignacio Corcuera-Platas, J Fernando Sánchez-Rada, and Carlos A Iglesias. 2017. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, 77:236–246.

AR Balamurali, Aditya Joshi, and Pushpak Bhattacharyya. 2011. Harnessing wordnet senses for supervised sentiment classification. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1081–1091. Association for Computational Linguistics.

Luciano Barbosa and Junlan Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd international conference on computational linguistics: posters*, pages 36–44. Association for Computational Linguistics.

Kayvan Bijari, Hadi Zare, Emad Kebriaei, and Hadi Veisi. 2020. Leveraging deep graph-based text representation for sentiment polarity applications. *Expert Systems with Applications*, 144:113090.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30:107–117.

Yukuo Cen, Xu Zou, Jianwei Zhang, Hongxia Yang, Jingren Zhou, and Jie Tang. 2019. Representation learning for attributed multiplex heterogeneous network. In *Proceedings of the 25th ACM International Conference on Knowledge Discovery & Data Mining*, pages 1358–1368.

Tao Chen, Ruifeng Xu, Yulan He, and Xuan Wang. 2017. Improving sentiment analysis via sentence type classification using bilstm-crf and cnn. *Expert Systems with Applications*, 72:221–230.

Abdelghani Dahou, Shengwu Xiong, Junwei Zhou, Mohamed Houcine Haddoud, and Pengfei Duan. 2016. Word embeddings and convolutional neural network for arabic sentiment classification. In *Proceedings of coling 2016, the 26th international conference on computational linguistics: Technical papers*, pages 2418–2427.

Peng Fu, Zheng Lin, Fengcheng Yuan, Weiping Wang, and Dan Meng. 2018. Learning sentiment-specific word embedding via global sentiment representation. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18)*, pages 4808–4815.

Mahsa Ghorbani, Mahdieh Soleymani Baghshah, and Hamid R Rabiee. 2019. Mgcn: semi-supervised classification in multi-layer graphs with graph convolutional networks. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 208–211.

Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864.

Lin Gui, Yu Zhou, Ruifeng Xu, Yulan He, and Qin Lu. 2017. Learning representations from heterogeneous network for sentiment classification of product reviews. *Knowledge-Based Systems*, 124:34–45.

Obaida Hanteer and Luca Rossi. 2019. An innovative way to model twitter topic-driven interactions using multiplex networks. *Frontiers in Big Data*, 2:9.

Zhao Jianqiang, Gui Xiaolin, and Zhang Xuejun. 2018. Deep convolution neural networks for twitter sentiment analysis. *IEEE Access*, 6:23253–23260.

Mostafa Karamibekr and Ali A Ghorbani. 2012. Sentiment analysis of social issues. In *Proceedings of the International Conference on Social Informatics (SocialInformatics)*, pages 215–221.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1746–1751. Association for Computational Linguistics.

Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. 2014. Multilayer networks. *Journal of complex networks*, 2(3):203–271.

Efthymios Kouloumpis, Theresa Wilson, and Johanna D Moore. 2011. Twitter sentiment analysis: The good the bad and the omg! In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, volume 11, pages 538–541.

John Lafferty and Chengxiang Zhai. 2001. Document language models, query models, and risk minimization for information retrieval. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 111–119.

Jundong Li, Chen Chen, Hanghang Tong, and Huan Liu. 2018. Multi-layered network embedding. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 684–692. SIAM.

Yongjin Li and Jagdish C Patra. 2010. Genome-wide inferring gene–phenotype relationship by walking on the heterogeneous network. *Bioinformatics*, 26(9):1219–1224.

Jiangming Liu and Yue Zhang. 2017. Attention modeling for targeted sentiment. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 572–577.

Dongsheng Luo, Jingchao Ni, Suhang Wang, Yuchen Bian, Xiong Yu, and Xiang Zhang. 2020. Deep multi-graph clustering via attentive cross-graph association. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 393–401.

Saif Mohammad, Cody Dunne, and Bonnie Dorr. 2009. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pages 599–608. Association for Computational Linguistics.

Huy Nguyen and Minh-Le Nguyen. 2018. A deep neural architecture for sentence-level sentiment classification in twitter social networking. In *Proceedings of the Computational Linguistics*, pages 15–27.

Jingchao Ni, Shiyu Chang, Xiao Liu, Wei Cheng, Haifeng Chen, Dongkuan Xu, and Xiang Zhang. 2018. Co-regularized deep multi-network embedding. In *Proceedings of the 2018 World Wide Web Conference*, pages 469–478.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 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, pages 79–86.

Duc-Hong Pham and Anh-Cuong Le. 2018. Learning multiple layers of knowledge representation for aspect based sentiment analysis. *Data & Knowledge Engineering*, 114:26–39.

Rudy Prabowo and Mike Thelwall. 2009. Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2):143–157.

Ashequl Qadir and Ellen Riloff. 2013. Bootstrapped learning of emotion hashtags# hashtags4you. In *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 2–11.

Ashequl Qadir and Ellen Riloff. 2014. Learning emotion indicators from tweets: Hashtags, hashtag patterns, and phrases. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1203–1209.

Meng Qu, Jian Tang, Jingbo Shang, Xiang Ren, Ming Zhang, and Jiawei Han. 2017. An attention-based collaboration framework for multi-view network representation learning. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1767–1776.

Christoph Rahmede, Jacopo Iacovacci, Alex Arenas, and Ginestra Bianconi. 2018. Centralities of nodes and influences of layers in large multiplex networks. *Journal of Complex Networks*, 6(5):733–752.

Cicero dos Santos and Maira Gatti. 2014. Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING): Technical Papers*, pages 69–78.

Aliaksei Severyn and Alessandro Moschitti. 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*, pages 959–962.

L. G. Singh, A. Anil, and S. R. Singh. 2020. She: Sentiment hashtag embedding through multitask learning. *IEEE Transactions on Computational Social Systems*, 7(2):417–424.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.

Duyu Tang, Furu Wei, Bing Qin, Nan Yang, Ting Liu, and Ming Zhou. 2016. Sentiment embeddings with applications to sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(2):496–509.

Peter D Turney. 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*, pages 417–424. Association for Computational Linguistics.

Peter D Turney and Michael L Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4):315–346.

John Violos, Konstantinos Tserpes, Evangelos Psomakelis, Konstantinos Psychas, and Theodora Varvarigou. 2016. Sentiment analysis using word-graphs. In *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics*, pages 1–9.

Gang Wang, Jianshan Sun, Jian Ma, Kaiquan Xu, and Jibao Gu. 2014. Sentiment classification: The contribution of ensemble learning. *Decision support systems*, 57:77–93.

Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang. 2011. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1031–1040. ACM.

Yuan Wang, Jie Liu, Yalou Huang, and Xia Feng. 2016. Using hashtag graph-based topic model to connect semantically-related words without co-occurrence in microblogs. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1919–1933.

Jason Weston, Sumit Chopra, and Keith Adams. 2014. # tagspace: Semantic embeddings from hashtags. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1822–1827.

Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu, and Xu Wu. 2019. Sentiment analysis of comment texts based on bilstm. *Ieee Access*, 7:51522–51532.

Hongming Zhang, Liwei Qiu, Lingling Yi, and Yangqiu Song. 2018. Scalable multiplex network embedding. In *IJCAI*, volume 18, pages 3082–3088.

Zhou Zhao, Hanqing Lu, Deng Cai, Xiaofei He, and Yueting Zhuang. 2017. Microblog sentiment classification via recurrent random walk network learning. In *IJCAI*, volume 17, pages 3532–3538.

Guang-You Zhou and Jimmy Xiangji Huang. 2017. Modeling and mining domain shared knowledge for sentiment analysis. *ACM Transactions on Information Systems (TOIS)*, 36(2):1–36.

Marinka Zitnik and Jure Leskovec. 2017. Predicting multicellular function through multi-layer tissue networks. *Bioinformatics*, 33(14):i190–i198.