Eve2Sign: Creating Signed Networks of News Events

Roshni Chakraborty
Aalborg University, Denmark
roshnic@cs.aau.dk

Srishti Bhandari
NIT Durgapur, India
sb.16u10463@btech.nitdgp.ac.in

Nilotpal Chakraborty
Aalborg University, Denmark
nilotpalc@cs.aau.dk

Ritwika Das
NIT Durgapur, India
rd.16u10468@btech.nitdgp.ac.in

Abstract

Studying news events and user opinions towards the news events have several applications, like detection of the popularity of the news, understanding user preferences and stance towards particular news, determining the evolution of the news story, etc. One of the major ways to understand a news event is through the main characters (targets) of the event and the relationships among those targets with respect to the news event. Therefore, in this paper, we propose an approach to creating a signed network of the news event to visualize the targets and their interrelationships. Our experimental evaluations on 3 news events indicate that the proposed approach can detect a large number of relevant goals related to any news event without supervision and, further, create a signed network effectively.

1 Introduction

With the growth of Twitter as a social media platform [WVG+16], a wide range of users are coming to this platform for discussing and sharing their opinions on various events [KLPM10]. The impacts of the events vary across users [SAMA17]; for example, the immigration policy reformulation by an government might directly impact the people seeking refuge and indirectly affect the economy of industries. Therefore, a proper understanding of news events related opinions is very important and finds applications, such as in the identification of the popularity of a news article [CEHPS14, Cas13], identifying the stance of a user [LPRR18] towards a particular news, summarizing users' opinions [ZSAG12, OGZ+17], etc. However, there are inherent challenges in identifying the user opinions relevant to a news event and further utilizing these user opinions for other applications, such as the vast vocabulary gap between the news event and the tweets, short length of tweets, usage of informal vocabulary and abbreviations [SAC+18, MCHV18]. Moreover, the user might not explicitly mention the news event name in the tweet while expressing the opinion related to the news event, and understanding of the user opinions related to a news event requires considering both the positive and negative interactions of the users towards the news event [DWT18]. Thus, identifying and understanding of the users' opinions from Twitter towards a news event would be eased if prior knowledge of the news event [1] [SAMA17], possible targets related to the news event [2] [MKS+16] and the polarity of the targets towards the news event are known.

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia (eds.): Proceedings of the Text2Story'20 Workshop, Lisbon, Portugal, 14-April-2020, published at http://ceur-ws.org

[1] A news event is a collection of one or many news articles related to the same real-life event

[2] A target may be a person, an organization, a government policy, a movement, a product related to the news event

A signed network representation of the news event, where the nodes represent the possible targets related to the news event and the news event, the edges represent the polarity relationship among the targets with respect to the news event can provide as a prior knowledge base of the news event. This further can aid in several applications, like identifying the stance [MSK17, Moh16] and bias of a user [DWT18] from his tweet irrespective of the tweet text and absence of event name in the tweet, generating fair summaries of news events by considering both negative and positive opinions of the readers towards the news event [DSB+19, JA18], identifying the major aspects of the news event [VCLDD17, KWHR16], understanding of the evolution of a news story [AKT18], etc. However, the creation of a signed network of a news event requires the resolution of multiple sub-problems, like identifying the targets of the news event and the polarity relationship of the targets with the news event. Although few existing research works discuss the significance of creating signed networks from text [HAJR12, AKR13, SCM16], these works identify the polarity relationships between pre-specified targets [HAJR12] or identify targets through degree and eigenvector centrality score [AKR13]. However, identifying targets only through degree or eigenvector centrality score might not be able to cover the huge list of possible targets related to the news event or identify the targets which comprise of multiple words. Further, there is a significant difference in content in news event-related information from the literary text and online discussions. Therefore, it remains a challenge to build an approach to create a signed network of a news event that can inherently identify the possible targets of a news event irrespective of the type of news event and number of targets and determine the polarity relationships among the identified targets related to the news event.

In this paper, we propose a novel approach that utilizes the news articles of a news event to create a signed network of the event where the nodes represent the entity names or phrases, and the edges represent the polarity relationship between a pair of nodes with respect to the news event. Effectively, the proposed intends to resolve two sub-problems, i.e., identifying the targets and determining the polarity relationship between the targets towards the news event. Existing research works on target identification from the text can be segregated into *supervised* and *unsupervised* techniques based on their proposed methodology [HN14]. Although the existing *supervised* approaches might provide better performance than the existing *unsupervised* approaches, the requirement of human-annotated information related to a news event makes it challenging to be used [KMKB13] for news events. Existing *unsupervised* approaches can be segregated into graph-based and embedding based approaches. The graph-based approaches either utilize the sub-group structure [TMV16] through *k-truss*, *k-core*, *k-clique* or *community* substructures, centrality based techniques, like *textrank* [MT04] and *position based page rank* [FC17]. However, our preliminary analysis indicates that the significance of a target of a news event is independent of their position in the news article, and further, a target might comprise of multiple words. We observed that centrality based measures fail to identify targets comprising multiple words and sub-group based measures generate extensive noise. Therefore, in this paper, we propose a page rank score (PRS) based edge traversal approach to identify the relevant targets related to a news event. Our experimental evaluation indicates that the proposed approach can identify targets of a news event irrespective of the number of words in the target and the type of news event. After successful identification of the targets of the news event, the proposed approach relies on existing natural language processing tools to identify the polarity among the detected targets from the news articles. The identified targets, along with their detected polarity towards each other and the news event is used to create a signed network of the news event.

The rest of the paper is organized as follows: we present a formal definition of the problem along with a brief description of the proposed approach in Section 2. We discuss the experiments and observations in Section 3 and finally draw our conclusions in Section 4.

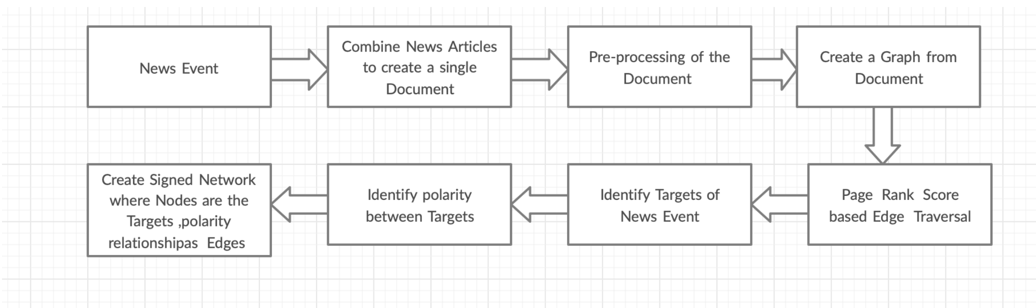

Figure 1: Overview of the Proposed Approach

2 Problem Definition

In this section, we provide a brief discussion of the problem followed by the details of the proposed approach. Given a news event $\mathcal{N}$, let $\mathcal{A} = \{a_1, a_2, \cdots\}$ be the set of news articles related to $\mathcal{N}$. We intend to create a signed network representation, $\mathcal{S} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ represents the targets, $\mathcal{T} = \{t_1, t_2, \cdots\}$ related to $\mathcal{N}$ and $\mathcal{E}$ represents the polarity relationship among $\mathcal{T}$, i.e., $\mathcal{E}_{ij} \in \{-1, +1\}$ indicates the polarity relationship between $\mathcal{T}_i$ and $\mathcal{T}_j$. An overview of the proposed approach is given in the figure 1, and we discuss the proposed approach in details next.

2.1 Proposed Approach

The proposed approach primarily identifies $\mathcal{T}$ related to $\mathcal{N}$ and thereafter, determines the polarity relationships between a pair of targets, say $\mathcal{T}_i$ and $\mathcal{T}_j$ from $\mathcal{A}$. We discuss each of these steps next.

2.1.1 Identification of Targets

In this section, we discuss the proposed approach to identify $\mathcal{T}$ related to $\mathcal{N}$. A target can comprise of single or multiple words. We primarily discuss the proposed procedure to identify single word $\mathcal{T}_i$ followed by multiple word $\mathcal{T}_i$. Given $\mathcal{N}$, we primarily create a document, $\mathcal{D}$ by combining the news articles, $\mathcal{A}$ related to $\mathcal{N}$. Thereafter, we create a graph $\mathcal{G}$, where the nodes, $\mathcal{Q}$ represent the words, w, from $\mathcal{D}$ after pre-processing (the pre-processing details is provided in section) and the edges, $\mathcal{R}$ represent the normalized co-occurrence score, $\mathcal{R}(w_i, w_j)$ between a pair of words, w_i and w_j calculated from $\mathcal{D}$ by:

$$\mathcal{R}(w_i, w_j) = \frac{coSc(w_i, w_j)}{\max_{w_r, w_s \in w}(coSc(w_r, w_s))}$$

where $coSc(w_i, w_j)$ represent the co-occurrence score of w_i and w_j in D. On $\mathcal{G}$, we calculate the page rank [PBMW99] centrality [3] score of each node, say w_k. We select a node into the set of targets, $\mathcal{T}$ if the PRS of w_k is greater than the threshold, i.e., $(pgR(w_k) \geq th_w)$. Based on two news events, the threshold was decided by a group of 2 manual annotators by considering that a minimum number of relevant targets is excluded, and a minimum number of irrelevant targets is included related to $\mathcal{N}$ into $\mathcal{T}$. We checked the effectiveness of the threshold on three different news events and found it relevant irrespective of the type of news event. Although the number of relevant targets excluded and the number of irrelevant targets included varies across news events, we found that the maximum number of irrelevant targets which was included was less than 4% of the total relevant targets and minimum number of relevant targets which was excluded was less than 9% of the total relevant targets for an event.

However, as previously discussed, a target might contain multiple words. Hence, we follow a path based PRS to identify targets of multiple words. For a word, w_i, we represent the score of his neighbours as $NgScore(w_i, w_k)$, where w_k is a neighbour of w_i, i.e., ($w_k \in$Ng(w_i)), as the weighted summation of the PRS of w_i, the PRS of w_k and normalized co-occurrence score of w_i and w_k, $\mathcal{R}(w_i, w_k)$. Therefore, $NgScore(w_i, w_k)$ can be formally defined as,

$$NgScore(w_i, w_k) = \alpha[pgR(w_i) + pgR(w_k) + \mathcal{R}(w_i, w_k)]$$

We have considered α as 0.33 to provide equal weightage to the page rank scores of w_i and w_k and the normalized co-occurrence score of w_i and w_k. Therefore, a phrase which comprises of w_i followed by w_k is added to the $\mathcal{T}$ if $(NgScore(w_i, w_k) \geq th_p)$. We further extend $NgScore$ to more than two words in a phrase by following a path of length 2, 3 and 4 starting from w_i to identify targets of length $3 - 5$ respectively. However, our experimental evaluations indicate that the number of phrases of size which are more than 4 words and is greater than the th_p is very few. Therefore, we consider targets that comprise of $1 - 4$ words in this paper. In table 1, we provide a brief overview of the graphs for each of the 3 events (details of the dataset of the news events are discussed in section 3.1) considered in this paper.

2.1.2 Identification of Polarity Relationships

After identification of $\mathcal{T}$ related to $\mathcal{N}$, we identify the polarity relationships among the targets from $\mathcal{D}$. We follow Fernandez et al. [FGÁLJM+16] to identify the polarity relationship by identifying the sentiment from the

[3] https://en.wikipedia.org/wiki/PageRank

Table 1: Overview of the event graph related E_1, E_2 and E_3, from which targets were detected.

Dataset	E_1	E_2	E_3
Nodes	3000	2351	1825
Edges	10673	14520	7831
Average Degree	7.12	12.35	8.58

consequent sentences, S of $\mathcal{D}$. We primarily discuss the approach proposed by Fernandez et al. [FGÁLJM+16] briefly followed by the details of our adopted methodology based on the same work [FGÁLJM+16]. Fernandez et al. [FGÁLJM+16] primarily identifies the dependency tree of a sentence and calculates the sentiment of the sentence based on the sentiment of the dependency sub-trees. The sentiment of each sub-tree is calculated through the sentiment score of the constituent lexicons along with the score due to the presence of intensifier, modifier, and negation in the subtree. The SO-CAL dictionary [TBT+11] is referred to calculate the sentiment score of a lexicon.

For our approach, we primarily identify the subject, s_t and object, o_t of s (for each $s \in S$) and thereafter, match if both s_t and o_t exist in $\mathcal{T}$, say as $\mathcal{T}_i$ and $\mathcal{T}_j$. We identify the polarity between $\mathcal{T}_i$ and $\mathcal{T}_j$ as the sentiment score calculated by following Fernandez et al. [FGÁLJM+16] on s. An iterative application of this procedure for each $s \in S$ can effectively identify a fraction of the existing links ($\mathcal{E}$) between the targets, ($\mathcal{T}$). Therefore, after this step, the polarity between $\mathcal{T}_i$ and $\mathcal{T}_j$ is identified as $+1$,-1 or 0. However, we observed that a fraction of the edges across the targets could not be identified by this procedure. We believe the absence of polarity information of all possible pairs of targets in S and the inherent complexity of the sentences are the two major limitations due to which we can not ensure identification of the polarity among all the targets in $\mathcal{T}$. In order to handle this issue, existing characteristics of signed networks, like structural balance theory and status theory, or different properties of signed networks could be used to predict the sign of a missing link [LHK10, DMT18] between the targets which we consider as one of the future directions of this work. On identification of all the possible polarity relationships among targets, we create a signed network of the event, $\mathcal{S} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ comprises of $\mathcal{T}$, $\mathcal{E}$ represents the identified polarity relationship between $\mathcal{V}$ as $\{+1, -1\}$.

3 Experiments and Discussion

In this section, we provide details of the datasets used, the pre-processing techniques followed in the experimental evaluation of the proposed approach, discuss the results obtained.

3.1 Datasets and Preprocessing

For our experimental evaluation, we selected 3 news events that belonged to *USA* [4], *Europe* [5] and *India* [6] respectively which has occurred during $2016 - 2017$ from the list of events provided by Wikipedia corresponding to each of the country or continent. We manually selected *USA*, *Europe* and *India* for our experiments to ensure the proposed approach is not biased by the location of the news event. For each of the these news events, we crawled the news articles from *Google News Search API* [7] by using *Python Newspaper API* [8]. We discuss the news events along with the news article datasets in details next.

1. *2016 Indian Banknote Demonetisation*, N_1 : In 2016, the Prime minister of India announced that the existing Rs 500 and Rs $1,000$ banknotes would not be allowed to be used and claimed that this action shall reduce the usage of counterfeit notes. This decision caused considerable debate in India. [9]. We crawled 86 news articles related to this event.

2. *Catalan Independence*, N_2 : The Parliament of Catalonia passed a resolution to declare the independence of Catalonia from Spain which led to considerable debate and discussions among the International community and Catalonia [10]. We crawled 47 news articles related to this event.

[4] https://en.wikipedia.org/wiki/2017_in_the_US
[5] https://en.wikipedia.org/wiki/2017
[6] https://en.wikipedia.org/wiki/2016_in_India
[7] https://news.google.com/?hl=en-IN&gl=IN&ceid=IN:en
[8] https://newspaper.readthedocs.io/en/latest/
[9] https://en.wikipedia.org/wiki/2016_Indian_banknote_demonetisation
[10] https://en.wikipedia.org/wiki/Catalan_independence_movement

3. *Allegations against Harvey Weinstein*, N_3 : Around 80 women accused Harvey Weinstein of sexual abuse and assault which led to severe discussions and debates [11]. We crawled 34 news articles related to this event.

Our dataset comprises of news events from different geographical locations, media coverage (the number of news articles related to an event range between $34 - 86$) to show that the proposed approach is not biased by the news event. We discuss the pre-processing details next.

(a) E_1 (b) E_2 (c) E_3

Figure 2: Word cloud representing a subset of the targets generated by the proposed approach.

Pre-processing Details We provide a brief overview of the pre-processing details hereby.

1. We combine $\mathcal{A}$ related to $\mathcal{N}$ into a single document, $\mathcal{D}$.

2. We use *SpaCy* [HM17] which provides an exhaustive list of around 325 stop-words (greater than the *NLTK library* [12]) to remove the stop words from $\mathcal{D}$. We also remove the punctuation marks from $\mathcal{D}$.

3. We segregrate $\mathcal{D}$ into it's consequent sentences, S. We use *SpaCy* [HM17] and *neuralcoref* [13] library for co-reference resolution for each sentence, i.e., $s \in S$. Further, we manually create a list to identify same words with different spellings into a single one, like the words *demonetization* and *demonetisation* was changed to the word *demonetization*.

4. We use *NLTK POS Tagger* [14] to identify the nouns, verbs, adjectives and prepositions of $\mathcal{D}$. Further, if a word has been tagged as both noun and adjective, then the word is considered to be of that tag which has higher occurrences. We only consider these words while creating $\mathcal{G}$ from $\mathcal{D}$ related to $\mathcal{N}$.

3.2 Experiments and Results

In this section, we describe our experiment details, followed by our observations. We compare the $\mathcal{T}$ identified by the proposed approach and the existing research works related to $\mathcal{N}$ (here, we consider $\mathcal{N}_1$, $\mathcal{N}_2$ and $\mathcal{N}_3$). We consider graph based approaches which identifies targets from text as baselines, like *TextRank*, *Clique* and *Community* sub-structure based approaches. To evaluate the performance of the proposed approach and the baselines in identifying targets related to $\mathcal{N}$, we measure the number of detected targets (D_{tar}) and the fraction of irrelevant targets (F_{irr}) detected by the proposed approach and the baselines related to $\mathcal{N}_1$, $\mathcal{N}_2$ and $\mathcal{N}_3$.

The decision of the relevance of a target was done by a group of 3 manual annotators who were given the set of targets, D_{tar} detected by the proposed approach related to an event, $\mathcal{N}_1$, $\mathcal{N}_2$ and $\mathcal{N}_3$. The manual annotators had no previous knowledge of the news events and was instructed to mark a target as relevant or irrelevant based on their knowledge of the news event based on $\mathcal{A}$ related to $\mathcal{N}$. A target was marked relevant or irrelevant based on the majority of the annotators. We repeated this for the targets detected by each of the baselines for each of the events, i.e., $\mathcal{N}_1$, $\mathcal{N}_2$ and $\mathcal{N}_3$, respectively. On comparing the results of the proposed approach with the baselines, as shown in table 2, we observe the proposed approach generates the maximum number of relevant targets followed by the baseline which utilizes *Clique* sub-structure. Although *TextRank* generates more targets,

[11] https://en.wikipedia.org/wiki/Harvey_Weinstein
[12] https://www.nltk.org
[13] https://pypi.org/project/neuralcoref/
[14] https://nlp.stanford.edu/software/tagger.shtml

it produces maximum noise and is only effective in generating single word targets. Further, we observe that *Clique* and *Community* based structure fails to indicate the order of the words in the phrase, thus requiring severe manual intervention to identify the relevant targets which comprises of multiple words. Therefore, as discussed in section 3.1 and highlighted in table 1, the proposed approach can identify targets of relevance as well as a large number of targets with respect to the existing works of target identification. In figure 2, we provide a word cloud representation of a subset of the targets identified by the proposed approach related to E_1, E_2 and E_3 respectively.

Table 2: Comparison of the proposed approach with the existing ones in terms of F_{irr} and T_{tar}.

Dataset	$Prop_{app}$		$TextRank$		$Clique$		$Community$	
	T_{tar}	F_{irr}	T_{tar}	F_{irr}	T_{tar}	F_{irr}	T_{tar}	F_{irr}
E_1	81	0.10	109	0.60	107	0.43	123	0.53
E_2	345	0.22	300	0.56	212	0.26	280	0.39
E_3	108	0.11	124	0.39	136	0.24	128	0.37

After detecting the targets, the proposed approach identifies the polarity relationship between a pair of targets following the approach discussed in section 2.1.2. We found the approach can effectively identify 0.63, 0.68, and 0.78 of the links for E_1, E_2 and E_3 respectively. On further analysis, we found that the fraction of edges which could not be resolved was either due to the inability of the proposed approach in identifying polarity relationship from a sentence or the absence of a target with any other target in any of the sentences. We further observed that the inability of the proposed approach to identifying the polarity relationship from a sentence was due to the presence of domain information and complex syntactic structure in the sentences. We intuitively believe utilizing the *structural balance theory* of signed networks along with the inherent features of signed networks that can effectively predict the sign of the missing links [LHK10, DMT18] can ensure identification of polarity relationship between a pair of targets which was not resolved by the approach discussed in section 2.1.2. In figure 3, we provide signed network representation of E_1 as created by the proposed approach, which comprises a subset of the detected targets (as the total number of detected targets is quite large) and the edges represent the polarity relationship between the targets towards the event. On visualizing the signed network for E_2, we found a similar distribution of targets on either side of the polarity towards the event. However, the signed network for E_3 indicates a large number of targets to be negatively connected to the event than the number of targets connected positively to the event.

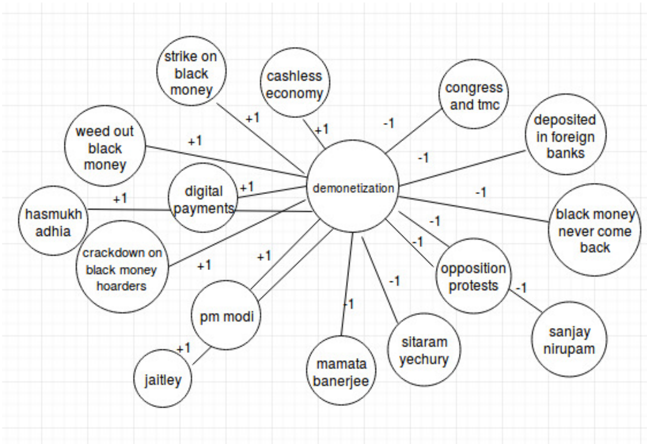

Figure 3: Signed network representing some of the targets along with their positive (or, negative) relation with the event E_1.

3.2.1 Analyzing Failures

Although the proposed approach effectively creates a signed network irrespective of the type of news event, we investigate more closely certain cases where our proposed approach fails. In the paper, the decision of the threshold for identifying relevant targets is done manually and is a fixed value irrespective of the event which might affect the effectiveness in identifying targets. Although we have tested the effectiveness of the threshold on

3 news events, we believe we require an exhaustive analysis on more news events to ensure it's effectiveness and applicability. Subsequently, the current version of the paper does not rigorously explore the determination of the polarity relationship between the targets. We believe the inclusion of signed network properties to predict the sign of a link as one of the immediate future directions of this paper can address the limitations of the polarity relationship detection by the proposed approach.

4 Conclusions

In this work, we propose an approach to creating the signed network representation of a news event, where the signed network comprises of the possible targets as nodes and the polarity relationship among the targets as edges. We propose a page rank score based edge traversal approach to identify the targets of a news article and rely on existing natural language processing tools to identify the polarity relationship between the identified targets from the articles of a news event. Our experimental evaluation on 3 events indicates that the proposed approach can detect a large number of relevant targets irrespective of the type of event with no supervision and, further, creates a signed network effectively. As one of our future directions, we believe the inclusion of information related to *semantic roles* of the words from the news articles and *structural information* from the created graph can aid in identifying more targets and further, reduce the noise in identified targets of the current proposed approach. We also intend to incorporate the attributes related to the signed network to identify the sign of the missing links.

References

[AKR13] Apoorv Agarwal, Anup Kotalwar, and Owen Rambow. Automatic extraction of social networks from literary text: A case study on alice in wonderland. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 1202–1208, 2013.

[AKT18] Omar Alonso, Vasileios Kandylas, and Serge-Eric Tremblay. How it happened: Discovering and archiving the evolution of a story using social signals. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, pages 193–202, 2018.

[Cas13] Carlos Castillo. Traffic prediction and discovery of news via news crowds. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 853–854. ACM, 2013.

[CEHPS14] Carlos Castillo, Mohammed El-Haddad, Jürgen Pfeffer, and Matt Stempeck. Characterizing the life cycle of online news stories using social media reactions. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 211–223. ACM, 2014.

[DMT18] Tyler Derr, Yao Ma, and Jiliang Tang. Signed graph convolutional networks. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 929–934. IEEE, 2018.

[DSB+19] Abhisek Dash, Anurag Shandilya, Arindam Biswas, Kripabandhu Ghosh, Saptarshi Ghosh, and Abhijnan Chakraborty. Summarizing user-generated textual content: Motivation and methods for fairness in algorithmic summaries. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–28, 2019.

[DWT18] Tyler Derr, Zhiwei Wang, and Jiliang Tang. Opinions power opinions: Joint link and interaction polarity predictions in signed networks. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 363–366. IEEE, 2018.

[FC17] Corina Florescu and Cornelia Caragea. Positionrank: An unsupervised approach to keyphrase extraction from scholarly documents. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1105–1115, 2017.

[FGÁLJM+16] Milagros Fernández-Gavilanes, Tamara Álvarez-López, Jonathan Juncal-Martínez, Enrique Costa-Montenegro, and Francisco Javier González-Castaño. Unsupervised method for sentiment analysis in online texts. *Expert Systems with Applications*, 58:57–75, 2016.

[HAJR12] Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev. Extracting signed social networks from text. In *Workshop Proceedings of TextGraphs-7 on Graph-based Methods for Natural Language Processing*, pages 6–14. Association for Computational Linguistics, 2012.

[HM17] Matthew Honnibal and Ines Montani. spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 7(1), 2017.

[HN14] Kazi Saidul Hasan and Vincent Ng. Automatic keyphrase extraction: A survey of the state of the art. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1262–1273, 2014.

[JA18] Myungha Jang and James Allan. Explaining controversy on social media via stance summarization. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1221–1224, 2018.

[KLPM10] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web*, pages 591–600. ACM, 2010.

[KMKB13] Su Nam Kim, Olena Medelyan, Min-Yen Kan, and Timothy Baldwin. Automatic keyphrase extraction from scientific articles. *Language resources and evaluation*, 47(3):723–742, 2013.

[KWHR16] Yaser Keneshloo, Shuguang Wang, Eui-Hong Han, and Naren Ramakrishnan. Predicting the popularity of news articles. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 441–449. SIAM, 2016.

[LHK10] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1361–1370. ACM, 2010.

[LPRR18] Mirko Lai, Viviana Patti, Giancarlo Ruffo, and Paolo Rosso. Stance evolution and twitter interactions in an italian political debate. In *International Conference on Applications of Natural Language to Information Systems*, pages 15–27. Springer, 2018.

[MCHV18] Béatrice Mazoyer, Julia Cagé, Céline Hudelot, and Marie-Luce Viaud. Real-time collection of reliable and representative tweets datasets related to news events. In *BroDyn@ ECIR*, pages 23–34, 2018.

[MKS+16] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, 2016.

[Moh16] Saif M Mohammad. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier, 2016.

[MSK17] Saif M Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3):26, 2017.

[MT04] Rada Mihalcea and Paul Tarau. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pages 404–411, 2004.

[OGZ+17] Yi Ouyang, Bin Guo, Jiafan Zhang, Zhiwen Yu, and Xingshe Zhou. Sentistory: multi-grained sentiment analysis and event summarization with crowdsourced social media data. *Personal and Ubiquitous Computing*, 21(1):97–111, 2017.

[PBMW99] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[SAC+18] Axel Suarez, Dyaa Albakour, David Corney, Miguel Martinez, and José Esquivel. A data collection for evaluating the retrieval of related tweets to news articles. In *European Conference on Information Retrieval*, pages 780–786. Springer, 2018.

[SAMA17] Vinay Setty, Abhijit Anand, Arunav Mishra, and Avishek Anand. Modeling event importance for ranking daily news events. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 231–240. ACM, 2017.

[SCM16] Shashank Srivastava, Snigdha Chaturvedi, and Tom Mitchell. Inferring interpersonal relations in narrative summaries. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[TBT+11] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.

[TMV16] Antoine Tixier, Fragkiskos Malliaros, and Michalis Vazirgiannis. A graph degeneracy-based approach to keyword extraction. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 1860–1870, 2016.

[VCLDD17] Steven Van Canneyt, Philip Leroux, Bart Dhoedt, and Thomas Demeester. Modeling and predicting the popularity of online news based on temporal and content-related features. *Multimedia Tools and Applications*, pages 1–28, 2017.

[WVG+16] Henning M Wold, Linn Vikre, Jon Atle Gulla, Özlem Özgöbek, and Xiaomeng Su. Twitter topic modeling for breaking news detection. In *WEBIST (2)*, pages 211–218, 2016.

[ZSAG12] Arkaitz Zubiaga, Damiano Spina, Enrique Amigó, and Julio Gonzalo. Towards real-time summarization of scheduled events from twitter streams. In *Proceedings of the 23rd ACM conference on Hypertext and social media*, pages 319–320. ACM, 2012.