Neural Temporal Opinion Modelling for Opinion Prediction on Twitter

Lixing Zhu[†]   **Yulan He**[†*]   **Deyu Zhou**[§]

[†]Department of Computer Science, University of Warwick, UK
[§]School of Computer Science and Engineering, Key Laboratory of Computer Network
and Information Integration, Ministry of Education, Southeast University, China
{Lixing.Zhu,Yulan.He}@warwick.ac.uk   d.zhou@seu.edu.cn

Abstract

Opinion prediction on Twitter is challenging due to the transient nature of tweet content and neighbourhood context. In this paper, we model users' tweet posting behaviour as a temporal point process to jointly predict the posting time and the stance label of the next tweet given a user's historical tweet sequence and tweets posted by their neighbours. We design a topic-driven attention mechanism to capture the dynamic topic shifts in the neighbourhood context. Experimental results show that the proposed model predicts both the posting time and the stance labels of future tweets more accurately compared to a number of competitive baselines.

1 Introduction

Social media platforms allow users to express their opinions online towards various subject matters. Despite much progress in sentiment analysis in social media, the prediction of opinions, however, remains challenging. Opinion formation is a complex process. An individual's opinion could be influenced by their own prior belief, their social circles and external factors. Existing studies often assume that socially connected users hold similar opinions. Social network information is integrated with user representations via weighted links and encoded using neural networks with attentions or more recently Graphical Convolutional Networks (GCNs) (Chen et al., 2016; Li and Goldwasser, 2019). This strand of work, including (Chen et al., 2018; Zhu et al., 2020; Del Tredici et al., 2019), leverages both the chronological tweet sequence and social networks to predict users' opinions.

The majority of previous work requires a manual segmentation of a tweet sequence into equally-spaced intervals based on either tweet counts or time duration. Models trained on the current interval are used to predict users' opinions in the next interval. However, we argue that such a manual segmentation may not be appropriate since users post tweets at different frequency. Also, the time interval between two consecutively published tweets by a user is important to study the underlying opinion dynamics system and hence should be treated as a random variable.

Inspired by the multivariate Hawkes process (Aalen et al., 2008; Du et al., 2016), we propose to model a user's posting behaviour by a temporal point process that when user u posts a tweet d at time t, they need to decide on whether they want to post a new topic/opinion, or post a topic/opinion influenced by past tweets either posted by other users or by themselves. We thus propose a neural temporal opinion model to jointly predict the time when the new post will be published and its associated stance. Instead of using the fixed formulation of the multivariate Hawkes process, the intensity function of the point process is automatically learned by a gated recurrent neural network. In addition, one's neighbourhood context and the topics of their previously published tweets are also taken into account for the prediction of both the posting time and stance of the next tweet.

To the best of our knowledge, this is the first work to exploit the temporal point process for opinion prediction on Twitter. Experimental results on the two Twitter datasets relating to Brexit and US general election show that our proposed model outperforms existing approaches on both stance and posting time prediction.

2 Methodology

We present in Figure 1 the overall architecture of our proposed Neural Temporal Opinion Model (NTOM). The input to the model at time step i

---

[*]Corresponding author
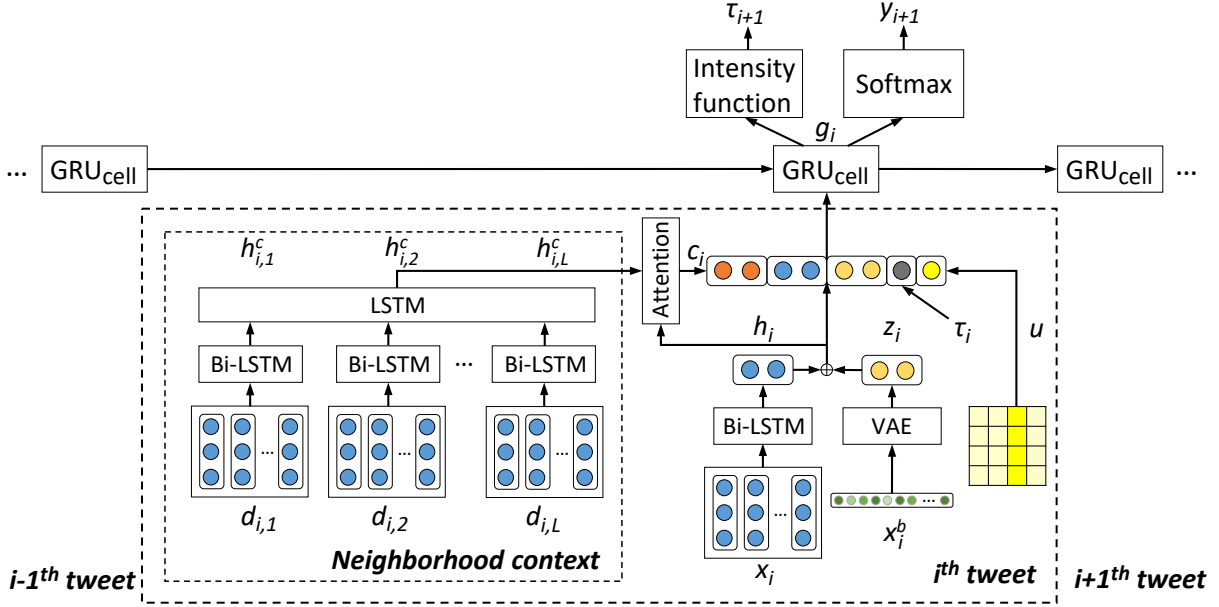
Figure 1: Overview of the Neural Temporal Opinion Model.

consists of user's own tweet x_i, bag-of-word representation x_i^b, time interval τ_i between the $i-1^{th}$ tweet and the ith tweet, user embedding u, and neighbours' tweet queue $\{d_{i,1}, d_{i,2}, \ldots, d_{i,L}\}$. At first, a Bi-LSTM layer is applied to extract features from input tweets. Then the neighborhood tweets are processed by a stacked Bi-LSTM/LSTM layer for the extraction of neighborhood context, which is fed into an attention module queried by the user's own tweet h_i and topic z_i. The output of attention module is concatenated with tweet representation, time interval τ_i, user representation u, and topic representation z_i, which is encoded from x_i^b via a Variational Autoencoder (VAE). Finally, the combined representation is sent to a GRU cell, whose hidden state participates in computing the intensity function and the softmax function, for the prediction of the posting time interval and the stance label of the next tweet. In the following, we elaborate the model in more details:

Tweet representation: Words in tweets are mapped to pre-trained word embeddings (Baziotis et al., 2017)[1], which is specially trained for tweets. Then Bi-LSTM is used to generate the tweet representation.

Topic extraction: The topic representation z_i in Figure 1 captures the topic focus of the ith tweet. It is learned by VAE (Kingma and Welling, 2014), which approximates the intractable true posterior

---

[1] https://github.com/cbaziotis/
datastories-semeval2017-task4

by optimising the reconstruction error between the generated tweet and the original tweet. Specifically, we convert each tweet to the bag-of-word format weighted by term frequency, x_i^b, and feed it to two inference neural networks defined as f_{μ_ϕ} and f_{Σ_ϕ}. These generate mean and variance of a Gaussian distribution from which the latent topic vector z_i is sampled. Then the approximated posterior would be $q_\phi(z_i|x_i^b) = \mathcal{N}(z_i|f_{\mu_\phi}(x_i^b), f_{\Sigma_\phi}(x_i^b))$. To generate the observation $\tilde{x}_i^b$ conditional on the latent topic vector z_i, we define the generative network as $p_\varphi(x_i^b|z_i) = \mathcal{N}(x_i^b|f_{\mu_\varphi}(z_i), f_{\Sigma_\varphi}(z_i))$. The reconstruction loss for the tweet x_i^b is then:

$$\mathcal{L}_x = \mathbb{E}_{q_\phi(z_i|x_i^b)}[\log p_\varphi(x_i^b|z_i)] - \text{KL}(q_\phi(z_i|x_i^b)||p(z_i)) \quad (1)$$

Neighbourhood Context Attention: To capture the influence from the neighbourhood context, we first input the neighbours' recent L tweets to an LSTM in a temporal ascending order. The output of the LSTM is weighed by the attention signals queried by the user's i^{th} tweet and topic:

$$c_i = \sum_{l=1}^{L} \alpha_l h_{i,l}^c \quad (2)$$

$$\alpha_l \propto \exp([h_i^\mathsf{T}, z_i^\mathsf{T}]\tanh(W_h h_{i,l}^c + W_z z_{i,l}^c)) \quad (3)$$

where $\{h_{i,1}^c, h_{i,2}^c, \ldots, h_{i,L}^c\}$ denotes the hidden state output of each tweet $d_{i,l}$ in the neighbourhood context, $z_{i,l}^c$ denotes the associated topic, h_i is the representation of the user's own tweet at time step i, and both W_h and W_z are weight matrices.

We use this attention mechanism to align the user's tweet to the most relevant part in the neighbourhood context. Our rationale is that a user would attend to their neighbours' tweets that discuss similar topics. The attention output c_i is then concatenated with a user's own tweet h_i and the extracted topic z_i. We further enrich the representation with the elapsed time τ_i between the posting time of the current tweet and the last posted tweet, and add a randomly initialised user vector u to distinguish the user from others. The final representation is passed to a GRU cell for the joint prediction of the posting time and stance label of the next tweet.

Temporal Point Process: The goal of NTOM is to forecast the time gap till the next post, together with the stance label. Instead of modelling the time interval value based on regression analysis, we use the GRU (Cho et al., 2014) to simulate the temporal point process.

At each time step, the combined representation $[c_i, h_i, z_i, \tau_i, u]$ is input to the GRU cell to iteratively update the hidden state taking into account the influence of previous tweets:

$$g_i = f_{GRU}(g_{i-1}, c_i, h_i, z_i, \tau_i, u) \qquad (4)$$

where g_i is the hidden state of GRU cell. Given g_i, the intensity function is formulated as:

$$\lambda^*(t) = \lambda(t|\mathcal{H}_i) = \exp(b_\lambda + v_\lambda^\mathsf{T} g_i + w_\lambda t) \quad (5)$$

Here, $\mathcal{H}_i$ summarises all the tweet histories up to tweet i, b_λ denotes the base density level, the term $v_\lambda^\mathsf{T} g_i$ captures the influence from all previous tweets and $w_\lambda t$ denotes the influence from the instant interval. The likelihood that the next tweet will be posted at the next interval τ given the history is:

$$f^*(\tau) = \lambda^*(\tau) \exp\left(-\int_0^\tau \lambda^*(t)dt\right) \qquad (6)$$

The expectation for the occurrence of the next tweet can be estimated using:

$$\hat{\tau}_{i+1} = \int_0^\infty \tau \cdot f^*(\tau)d\tau \qquad (7)$$

Loss: We expect the predicted interval to be close to the actual interval as much as possible by minimising the Gaussian penalty function:

$$\mathcal{L}_{time} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(\tau_{i+1} - \hat{\tau}_{i+1})^2}{2\sigma^2}\right) \qquad (8)$$
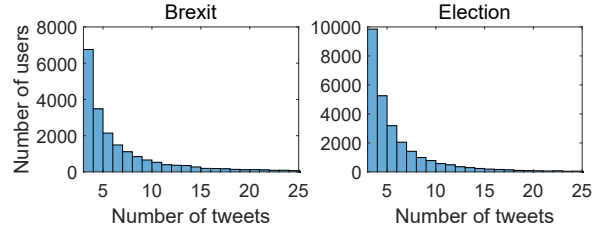

Figure 2: Number of users versus number of tweets.

For the stance prediction we employ the cross-entropy loss denoted as $\mathcal{L}_{stan}$. The final objective function is computed as:

$$\mathcal{L} = \eta\mathcal{L}_x + \beta\mathcal{L}_{time} + \gamma\mathcal{L}_{stan} \qquad (9)$$

where η, β and γ are coefficients determining the contribution of various loss functions.

3 Experiments

3.1 Setup

We perform experiments on two publicly available Twitter datasets[2] (Zhu et al., 2020) on Brexit and US election. The Brexit dataset consists of 363k tweets with 31.6%/29.3%/39.1% supporting/opposing/neutral tweets towards Brexit. The Election dataset consists of 452k tweets with 74.2%/20.4%/5.4% supporting/opposing/neutral tweets towards Trump. We filter out users who posted less than 3 tweets and are left with $20,914$ users in Brexit and $26,965$ users in Election. We plot in Figure 2 the number of users versus the number of tweets and found that over 81.6% users have published fewer than 7 tweets, we therefore set the maximum length of the tweet sequence of each user to 7. For users who have published more than 7 tweets, we split their tweet sequence into multiple training sequences of length 7 with an overlapping window size of 1. For each user, we use 90% of their tweets for training and 10% (round up) for testing.

Our settings are $\eta = 0.2$, $\beta = 0.4$ and $\gamma = 0.4$. We set the topic number to 50 and the vocabulary size to 3k for the tweet bag-of-words input to VAE. The mini-batch size is 16. We use Adam optimizer with learning rate 0.0005 and learning rate decay 0.9. The evaluation metrics are accuracy for stance prediction and Mean Squared Error (MSE) for posting time prediction. The results are compared against the following baselines:

Model	Brexit		Election	
	Acc.	MSE	Acc.	MSE
CSIM_W	0.653	–	0.656	–
NOD	0.675	–	0.690	–
LING+GAT	0.692	–	0.704	–
RNN	0.636	7.81	0.659	9.62
LSTM	0.677	3.37	0.683	4.51
GRU	0.691	2.80	0.693	3.92
NTOM$_{-VAE}$	0.697	2.67	0.705	4.01
NTOM$_{-context}$	0.665	3.34	0.682	4.78
NTOM$_{-GCN}$	0.680	2.65	0.706	4.29
NTOM	**0.713**	**2.59**	**0.715**	**3.70**

Table 1: Stance prediction accuracy and Mean Squared Errors of predicted posting time on the Brexit and Election datasets.

- CSIM_W (Chen et al., 2018) gauges the social influence by an attention mechanism for the prediction of the user sentiment of the next tweet.
- NOD (Zhu et al., 2020) takes into account the neighborhood context and pre-extracted topics for tweet stance prediction.
- LING+GAT (Del Tredici et al., 2019) places a GCN variant over linguistic features to extract node representations. Tweets are aggregated by users for user-level prediction.

We also perform ablation study on our model by removing the topic extraction component (NTOM$_{-VAE}$) or removing the neighbourhood context component (NTOM$_{-context}$). In addition, to validate that NTOM does benefit from point process modelling and can better forecast the time and stance of the next tweet, we remove the intensity function (i.e. no Eq. (5)-(7)) and directly use vanilla RNN and its variants including LSTM and GRU to predict the true time interval. Furthermore, to investigate if is is more beneficial to use GCN to encode the neighbourhood context, we learn tweet representation using GCN[3] (Hamilton et al., 2017), which preserves high-order influence in social networks through convolution. As in (Li and Goldwasser, 2019), we use a 2-hop GCN and denote the variant as NTOM$_{-GCN}$. For the Brexit dataset, MSE is measured in hours, while for the Election dataset it is measured in minutes due to the intensive tweets published within two days.

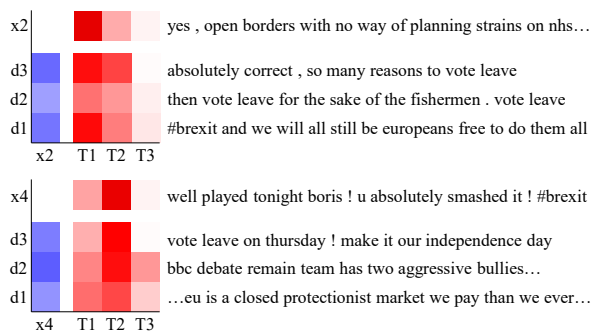[3] https://github.com/williamleif/GraphSAGE

3.2 Results

We report in Table 1 the stance prediction accuracy and MSE scores of predicted posting time. Compared to baselines, NTOM consistently achieves better performance on both datasets, showing the benefit of modelling the tweet posting sequence as a temporal point process. In the second set of experiments, we study the effect of temporal process modelling. The results verify the benefit of using the intensity function, with at least a 2% increase in accuracy and 0.2 decrease in MSE compared with vanilla RNN and its variants. In the ablation study, the removal of neighbourhood context component caused the largest performance decline compared to other components, verifying the importance of social influence in opinion prediction. Removing either VAE (for topic extraction) or intensity function (using only GRU) results in slight drops in stance prediction and more noticeable performance gaps in time prediction. It can be also observed that using GCN to model higher-order influence in social networks does not bring any benefits, possibly due to extra noise introduced to the model.

3.3 Visualisation of Topical Attention

To investigate the effectiveness of the context attention that is queried by topics, we first select some example topics from the topic-word matrix in VAE. The label of each topic is manually assigned based on its associated top 10 words. Then we display a tweet's topic distribution together with its neighborhood tweets' topic distribution. We also visualize the attention weights assigned to the 3 neighborhood tweets.

Figure 3 illustrates the example topics, topic distribution and attention signals towards context tweets. Here, x_2 and x_4 denote a user's 2nd and 4th tweets respectively. The most recent 3 neighborhood tweets are denoted as d_1, d_2, d_3. Blue in the leftmost separate column denotes the attention weights, and each row on top of $T1, T2$ and $T3$ denotes the topic distribution. It can be observed that the user's concerned topic shifts from *immigration* to *Boris Johnson* in 2 time steps. The drift also appears in the neighbour's tweets. Higher attention weights are assigned to the neighbour's tweets which share similar topical distribution as the user. We can thus infer that the topic vector does help select the most relevant neighborhood tweet.

x2				yes , open borders with no way of planning strains on nhs…
d3				absolutely correct , so many reasons to vote leave
d2				then vote leave for the sake of the fishermen . vote leave
d1				#brexit and we will all still be europeans free to do them all
	x2	T1	T2	T3

x4				well played tonight boris ! u absolutely smashed it ! #brexit
d3				vote leave on thursday ! make it our independence day
d2				bbc debate remain team has two aggressive bullies…
d1				…eu is a closed protectionist market we pay than we ever…
	x4	T1	T2	T3

	Topic	Top words
T1	immigration	immigration, stop, free, work, change, countries, immigrants, migrants, migration, open
T2	Boris Johnson	Boris, live, Johnson, politics, sturgeon, TV, Nicola, morning, takebackcontrol, guy
T3	vote remain	voteremain, strongerin, Cameron, eureferendum, David, inorout, pm, eudebate, osborne, positive

Figure 3: Distribution over 3 topics and attention signals on 3 neighborhood tweets, respectively in 2 time steps. Topics are labelled based on the top 10 words.

4 Related Work

The prediction of real-time stances on social media is challenging, partly caused by the diversity and fickleness of users (Andrews and Bishop, 2019). A line of work mitigated the problem by taking into account the homophily that users are similar to their friends (McPherson et al., 2001; Halberstam and Knight, 2016). For example, Chen et al. (2016) gauged a user's opinion as an aggregated stance of their neighborhood users. Linmei et al. (2019) took a step further by exploiting the extracted topics, which discern a user's focus on neighborhood tweets. Recent advances in this strand also include the application of GCNs, with which the social relationships are leveraged to enrich the user representations (Li and Goldwasser, 2019; Del Tredici et al., 2019).

On the other hand, several work has utilized the chronological order of tweets. Chen et al. (2018) presented an opinion tracker that predicts a stance every time a user publishes a tweet, whereas (Zhu et al., 2020) extended the previous work by introducing a topic-dependent attention. Shrestha et al. (2019) considered diverse social behaviors and jointly forecast them through a hierarchical neural network. However, the aforementioned work requires a manual segmentation of a tweet sequence. Furthermore, they are unable to predict when a user will next publish a tweet and what its associated stance is. These problems can be addressed using the Hawkes process (Hawkes, 1971), which has been successfully applied to event tracking (Sri-

jith et al., 2017), rumor detection (Lukasik et al., 2016; Zubiaga et al., 2016; Alvari and Shakarian, 2019) and retweet prediction (Kobayashi and Lambiotte, 2016). A combination of the Hawkes process with recurrent neural networks, called Recurrent Marked Temporal Pointed Process (RMTPP), was proposed to automatically capture the influence of the past events on future events, which shows promising results on geolocation prediction (Du et al., 2016). Benefiting from the flexibility and scalability of neural networks, several work has been done in this vein including event sequence prediction (Mei and Eisner, 2017) and failure prediction (Xiao et al., 2017). Our work is partly inspired by RMTPP, but departs from the previous work by jointly considering users' social relations and topical attentions for stance prediction on social media.

5 Conclusion

In this paper, we propose a novel Neural Temporal Opinion Model (NTOM) to address users' changing interest and dynamic social context. We model users' tweet posting behaviour based on a temporal point process for the joint prediction of the posting time and stance label of the next tweet. Experimental results verify the effectiveness of the model. Furthermore, visualisation of the topics and attention signals shows that NTOM captures the dynamics in the focused topics and contextual attention.

Acknowledgments

References

Odd Aalen, Ornulf Borgan, and Hakon Gjessing. 2008. *Survival and event history analysis: a process point of view*. Springer Science & Business Media.

Hamidreza Alvari and Paulo Shakarian. 2019. Hawkes process for understanding the influence of pathogenic social media accounts. In *2019 2nd*

International Conference on Data Intelligence and Security (ICDIS), pages 36–42. IEEE.

Nicholas Andrews and Marcus Bishop. 2019. Learning invariant representations of social media users. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1684–1695.

Christos Baziotis, Nikos Pelekis, and Christos Doulkeridis. 2017. DataStories at SemEval-2017 task 4: Deep LSTM with attention for message-level and topic-based sentiment analysis. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 747–754, Vancouver, Canada. Association for Computational Linguistics.

Chengyao Chen, Zhitao Wang, Yu Lei, and Wenjie Li. 2016. Content-based influence modeling for opinion behavior prediction. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2207–2216.

Chengyao Chen, Zhitao Wang, and Wenjie Li. 2018. Tracking dynamics of opinion behaviors with a content-based sequential opinion influence model. *IEEE Transactions on Affective Computing*.

KyungHyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. *CoRR*, abs/1409.1259.

Marco Del Tredici, Diego Marcheggiani, Sabine Schulte im Walde, and Raquel Fernández. 2019. You shall know a user by the company it keeps: Dynamic representations for social media users in nlp. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4701–4711.

Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. 2016. Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1555–1564. ACM.

Yosh Halberstam and Brian Knight. 2016. Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter. *Journal of Public Economics*, 143:73–88.

Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034.

Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.

Diederik P Kingma and Max Welling. 2014. Auto-encoding variational bayes. *stat*, 1050:1.

Ryota Kobayashi and Renaud Lambiotte. 2016. Tideh: Time-dependent hawkes process for predicting retweet dynamics. In *Tenth International AAAI Conference on Web and Social Media*.

Chang Li and Dan Goldwasser. 2019. Encoding social information with graph convolutional networks forpolitical perspective detection in news media. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2594–2604.

Hu Linmei, Tianchi Yang, Chuan Shi, Houye Ji, and Xiaoli Li. 2019. Heterogeneous graph attention networks for semi-supervised short text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4823–4832.

Michal Lukasik, PK Srijith, Duy Vu, Kalina Bontcheva, Arkaitz Zubiaga, and Trevor Cohn. 2016. Hawkes processes for continuous time sequence classification: an application to rumour stance classification in twitter. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 393–398.

Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.

Hongyuan Mei and Jason M Eisner. 2017. The neural hawkes process: A neurally self-modulating multivariate point process. In *Advances in Neural Information Processing Systems*, pages 6754–6764.

Prasha Shrestha, Suraj Maharjan, Dustin Arendt, and Svitlana Volkova. 2019. Learning from dynamic user interaction graphs to forecast diverse social behavior. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 2033–2042. ACM.

PK Srijith, Michal Lukasik, Kalina Bontcheva, and Trevor Cohn. 2017. Longitudinal modeling of social media with hawkes process based on users and networks. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 195–202.

Shuai Xiao, Junchi Yan, Xiaokang Yang, Hongyuan Zha, and Stephen M Chu. 2017. Modeling the intensity function of point process via recurrent neural networks. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Lixing Zhu, Yulan He, and Deyu Zhou. 2020. Neural opinion dynamics model for the prediction of user-level stance dynamics. *Information Processing & Management*, 57(2):102031.

Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, and Michal Lukasik. 2016. Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2438–2448.