Detecting and Forecasting Misinformation via Temporal and Geometric Propagation Patterns

Qiang Zhang[1(✉)], Jonathan Cook[1], and Emine Yilmaz[1,2]

[1] University College London, London, UK
qiang.zhang.16@ucl.ac.uk
[2] Alan Turing Institute, London, UK

Abstract. Misinformation takes the form of a false claim under the guise of fact. It is necessary to protect social media against misinformation by means of effective misinformation detection and analysis. To this end, we formulate misinformation propagation as a dynamic graph, then extract the temporal evolution patterns and geometric features of the propagation graph based on Temporal Point Processes (TPPs). TPPs provide the appropriate modelling framework for a list of stochastic, discrete events. In this context, that is a sequence of social user engagements. Furthermore, we forecast the cumulative number of engaged users based on a power law. Such forecasting capabilities can be useful in assessing the threat level of misinformation pieces. By jointly considering the geometric and temporal propagation patterns, our model has achieved comparable performance with state-of-the-art baselines on two well known datasets.

Keywords: Misinformation · Propagation graph · Point processes

1 Introduction

Social media has empowered human society in many ways. It is easier than ever to keep in touch with those we wish to, allowing an enormous variety of relationships to transcend physical isolation [19]. More so than ever before, social media has a responsibility for our mental wellbeing, as the arbiter of interactions between colleagues, friends and loved ones [13,24]. It is therefore a matter of the utmost importance that we make this platform a safe environment, protected against those wishing to corrupt the service with *fake news* [20].

Various methods have been used to tackle the misinformation problem. Content-based misinformation analysis models apply natural language processing tools to the text content of claims [23]. Alone, content-based models fail to trace the dynamics of spread for tasks such as early detection or spread forecasting. Recent misinformation analysis models use static graph neural networks to extract geometric propagation patterns; others leverage time-series analysis by treating misinformation spread as a temporal event sequence [4,15]. These

two approaches each neglect the alternative propagation structure with neither leveraging both geometric and temporal dissemination features.

Propagation-based misinformation analysis makes use of patterns that can be attributed to the dynamics of spread. Our principal goal is to utilise the maximum space of these spreading features, so as to make the most effective use of the available data. Specifically, we first formulate misinformation propagation as a dynamic graph, then we employ a continuous-time temporal point process to extract the temporal evolution patterns and geometric features. Furthermore, we use a power law to model the growth in the temporal network scale, so as to forecast the future rate of spread for a claim identified as misinformation. The contributions of this study can thus be summarised as follows. (i) We formulate misinformation propagation as a dynamic graph. (ii) We then design temporal point processes (TPPs) to utilize both temporal and geometric features of the dynamic graph for misinformation detection. (iii) This study is the first to introduce forecasting of user engagements to misinformation analysis.

2   Related Work

To figure out the differences between true and false statements, most researchers conduct studies from three approaches: textual content, multimedia features and social context. Misinformation often contains opinionated language [2], which motivates textual content-based detection [1]. Sentiment features like positive words (e.g., love, sweet) and negating words (e.g., not, never) are reported to help detect rumours [6]. Misinformation also relies on sensational images to provoke an emotional response in consumers. As an example, Deepfakes [3] employed deep learning to generate fake images and videos to convey misleading information.

In social media, every piece of news is correlated to other posts and users. User engagements (e.g., commenting) provide rich reference evidence in two ways: by aggregation with relevant posts for a specific affair, and by temporal evolution. The first way relies on the "wisdom of crowds" to locate potential misinformation [1], while the second way captures temporal propagation patterns. For example, Hawkes processes are used to analyze how user stance changes temporally in [11]. However, these methods neglect geometric propagation features.

Graph neural networks can extract geometric propagation patterns. Graph Convolutional Networks (GCN) are used in [14] to encapsulate the propagation structure of heterogeneous data. Graph-Aware Co-Attention Network (GCAN) is proposed in [4,8] to utilise the co-attention mechanism in graph modeling. Each of these works use static graphs and researchers neglect temporal information.

3   Problem Formulation

This section gives definitions and describes notation. A source claim takes the form of $c = (x, t)$, where x is a concatenation of the posting user account features and the claim's text features, i.e. $x = [u \parallel M]$. Here, u is the user account representation and M is the text message representation. t is initially

zero, as ensuing dissemination events are timestamped with respect to the source claim.

Suppose the claim c is accompanied by a sequence of social engagements $S = \{v_1, v_2, \ldots, v_j, \ldots, v_N\}$, where $v_j = (x_j, t_j)$. Similarly, x_j is the feature of an engaging node and t_j is the engagement time with respect to claim post time. Social engagements include all forms of interactions that users conduct with claims on social media platforms, such as reposting, commenting and tagging.

Our temporal, dynamic graph is represented as a sequence of time-stamped snapshots $G = \{\mathcal{G}(t_0), \mathcal{G}(t_1), \cdots, \mathcal{G}(t_j), \cdots, \mathcal{G}(t_N)\}$, where the first snapshot simply represents the source claim node and further snapshots are added with each representing the state of the dissemination network when a new node is connected. Let $\mathcal{G}(t) = < \mathcal{V}(t), \mathcal{E}(t) >$ denote the state of the temporal graph $\mathcal{G}$ at time t, where $\mathcal{V}(t) = \{c, v_1, v_2, \ldots, v_j, \ldots, v_{N(t)}\}$, with $N(t)$ being the number of nodes to have directly or indirectly interacted with the claim c as of time t. A new graph snapshot $\mathcal{G}(t_{j+1})$ is generated when a node v_{j+1} is added to the sequence of social engagements. The graph structure of an exemplary false claim's dissemination tree is demonstrated in Fig. 1.
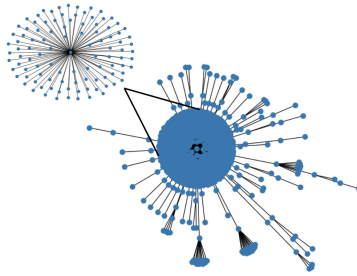
Fig. 1. Graph representation of source claim dissemination tree, where nodes represent interaction events such as comments and retweets.

4   Model Description

With the temporal evolution of the propagation graph $\mathcal{G}(t)$, new engagement nodes will establish edges with existing nodes and thus update the graph. To capture both geometric and temporal propagation features, we view the addition of new engagement nodes as the chronological events and develop a temporal point process that generates node embeddings of the dynamic graph $\mathcal{G}(t)$.

4.1   Propagation by Temporal Point Processes

A temporal point process (TPP) is a stochastic process that is realised as a list of discrete events in the continuous time domain $t \in \mathbb{R}^+$. TPPs usually rely on

an intensity function, which is defined as the probability of the occurrence of an event in an infinitesimal time interval [22], to describe the temporal dynamics. They have been used to model dynamic graphs in [10,17,25].

In our propagation graph use-case, the timestamped event sequence comprises static graph snapshots. This static propagation graph represents the final state of the misinformation dissemination tree. Symbolically, $\mathcal{S} = \{(\boldsymbol{x}_j, t_j)\}_{j=1}^{N}$, where $\boldsymbol{x}_j$ are the event features (previously node features) and t_j is the timestamp of the jth event in the sequence $\mathcal{S}$. Intuitively, the added edge $\boldsymbol{e}_{i,j}$ between the source node $\boldsymbol{v}_i$ and the new node $\boldsymbol{v}_j$ are influenced by not only $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$, but also the history nodes of $\boldsymbol{v}_i$. With this assumption, we define the intensity function associated with adding the new edge $\boldsymbol{e}_{i,j}$ as,

$$\lambda_{i,j}(t) = g(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{i' \in \mathcal{H}^i} \alpha_{i'j}(t) f(\boldsymbol{x}_{i'}, \boldsymbol{x}_j) \kappa(t - t_{i'}). \tag{1}$$

where $\mathcal{H}^i$ contains history events of the node i. The function $g(\cdot)$ calculates the affinity between two nodes, which is implemented as a bilinear interaction with the trainable parameter $\mathbf{W}_1$, i.e., $f(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i * \mathbf{W}_1 * \boldsymbol{x}_j$. A non-linear activation ReLU is used to define the base intensity $g(\cdot) = ReLU(f(\cdot))$.

The influence from history nodes are measured via the self-attention mechanism as proposed in [21,22]. For history nodes before time t, we calculate attention weight for each node,

$$\alpha_{i'j} = \frac{\exp(f(\boldsymbol{x}_{i'}, \boldsymbol{x}_j))}{\sum_{k \in \mathcal{H}^i} \exp(f(\boldsymbol{x}_k, \boldsymbol{x}_j))}. \tag{2}$$

With the intensity function, we derive the probability of having a new node $\boldsymbol{v}_j$ following an existing node $\boldsymbol{v}_i$ at the timestamp t,

$$p\left(\boldsymbol{v}_i, \boldsymbol{v}_j \mid \mathcal{H}^i(t)\right) = \frac{\lambda_{i,j}(t)}{\sum_{i' \in \mathcal{H}^i(t)} \lambda_{i',j}(t)}. \tag{3}$$

The objective function to minimize is the negative log-likelihood of all the events in the sequence, $\mathcal{L}_{TPP} = -\sum_{t \in \mathcal{T}} \sum_{(\boldsymbol{v}_i, \boldsymbol{v}_j, t) \in \mathcal{E}} \log p\left(\boldsymbol{v}_i, \boldsymbol{v}_j \mid \mathcal{H}^i(t)\right)$. Negative sampling is used to generate non-existing edges in the objective function as done in [9], so that the learnt node embeddings are able to distinguish which two nodes are connected and which two are not, i.e., the geometric structure. Maximizing the intensity at occurrence timestamps while minimizing the intensity otherwise will enforce the node embeddings to capture temporal dynamics.

4.2   Predictive Task

Macro-dynamics describe the evolution pattern of the network scale. We assume the network scale can be described with a certain dynamics equation. Given a dynamic graph $\mathcal{G}$, we have the cumulative number of nodes $N(t)$ by timestamp t. We empirically find that $N(t)$ increases in a power law, which is presented

in Sect. 5. To approximate the power law, we define the following predictive equation

$$\hat{N}(t) = N_{max} * (1 - \alpha * \exp(-\beta * t)), \tag{4}$$

where N_{max}, α and β are learnable parameters. N_{max} is the maximum number of nodes that this graph will contain while α and β control how fast the graph scale will increase. Predictive loss is measured by $\mathcal{L}_{Pred} = (N(t) - \hat{N}(t))^2$.

4.3 Veracity Classification

We have designed a temporal point process to capture the geometric structure and temporal evolution of the propagation graph. With node embeddings, we obtain the graph embedding by concatenating the mean pooling and the maximum pooling of all nodes as well as the source claim being verified, $\boldsymbol{x}_G = [MeanPool(\mathcal{S})||MaxPool(\mathcal{S})||\boldsymbol{c}]$. The graph embedding is then concatenated by parameters in predictive tasks, i.e., $\boldsymbol{x} = [\boldsymbol{x}_G||N_{max}||\alpha||\beta]$. The veracity prediction is conducted by a Multi-Layer Perceptron (MLP) $\hat{\mathbf{y}} = \text{softmax}(\text{ReLU}(\mathbf{W}_2\boldsymbol{x} + \mathbf{b}))$, where $\mathbf{W}_2$ and $\mathbf{b}$ are trainable parameters. And the classification loss is calculated by cross-entropy: $\mathcal{L}_{MLP} = -y\log(\hat{y}_1) - (1-y)\log(1-\hat{y}_0)$. We take the weighted sum of the TPP loss, predictive loss and the MLP loss as the final loss function $\mathcal{L} = \mathcal{L}_{TPP} + \omega_1 * \mathcal{L}_{Pred} + \omega_2 * \mathcal{L}_{MLP}$.

5 Experiments

We use two Twitter datasets [12], i.e., Twitter15 and Twitter16, in the experiments. Each dataset has a collection of stories with a source tweet being verified and a sequence of its retweets. We pick "True" and "False" source tweets to make the experimental datasets, and split the dataset into training, validation and test

Table 1. Statistics of the used datasets.

	Twitter 15	Twitter 16
# Source Tweets	742	412
# True	372	205
# False	370	207
# Users	190,868	115,036
Avg. retweet per story	292.19	208.70

sets with 70%, 10% and 20% respectively. We train the model with the training set, tune hyperparameters with the validation set and report performance on the test set. We crawl user information according to their user IDs via Twitter API (Table 1).

As we set out to tackle the misinformation detection task, we compare our model with state-of-the-art baselines. RFC [5] is a random forest model with features from the source tweets and engaged user profiles. CRNN [7] combines convolutional neural networks and recurrent neural networks to extract features from engaged users and retweet texts. CSI [15] incorporates relevant articles and analyses the group behaviour of engaged users. dEFEND [16] uses a co-attention mechanism to study the source claims and user features. The graph-based baseline GCAN has been explained in Related Works.

6   Results and Analysis

To demonstrate the dissemination trends of true and false claims, we plotted the mean number of nodes within temporal graphs associated with each veracity classification at 5 min time intervals for the first 200 min following a source Tweet's posting time. In Fig. 2, we make three interesting observations. (1) Both claim veracity types exhibit a similar power-law trend of plateauing gradient. (2) Contrary to much of the misinformation literature, which suggests that fake news spreads faster than true news [18], within our datasets, true news stories spread faster and reach more users on average. (3) There is a far greater disparity between the mean spreading plots in the Twitter16 dataset than there is in the Twitter15 dataset. This would indicate that it is easier to extract temporal features that are consistent within a given veracity classification in Twitter16.

We show the misinformation detection performance of our model against state-of-the-art baselines on test subsets. From Table 2, we can tell that we are able to achieve comparable performance with GCAN. Specifically, we beat GCAN on the Twitter16 dataset. This can be explained by the fact that Twitter16 displays greater disparity between the mean spreading of true and false claims, and our model captures such patterns to reach higher performance.
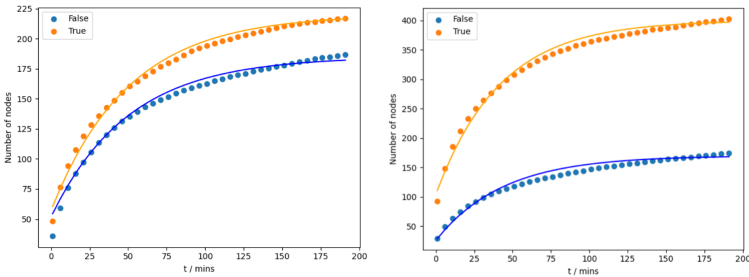
Fig. 2. Plots of average number of nodes comprising a dissemination tree with respect to time from the moment of source claim publication. The left is Twitter15 while the right is Twitter16. The solid curves follow the power law approximation.

Table 2. Test results on the two experimental datasets.

Model	Twitter15				Twitter16			
	F1	Recall	Precision	Accuracy	F1	Recall	Precision	Accuracy
RFC	0.4642	0.5302	0.5718	0.5385	0.6275	0.6587	0.7315	0.6620
CRNN	0.5249	0.5305	0.5296	0.5919	0.6367	0.6433	0.6419	0.7576
CSI	0.7174	0.6867	0.6991	0.6987	0.6304	0.6309	0.6321	0.6612
dFEND	0.6541	0.6611	0.6584	0.7383	0.6311	0.6384	0.6365	0.7016
GCAN	**0.8250**	**0.8295**	**0.8257**	**0.8767**	0.7593	0.7632	0.7594	0.8084
Ours	0.7698	0.7643	0.7754	0.7988	**0.7774**	**0.7741**	**0.808**	**0.8453**

7   Conclusion

This study sets out to detect and forecast misinformation. We model the misinformation propagation as a continuous-time dynamic graph, and employ Temporal Point Processes to capture geometric and temporal patterns of the graph. We also develop a power law equation to forecast the growth of the graph scale. Experiments show the effectiveness of our model to achieve state-of-the-art performance in misinformation detection tasks. Future works will investigate more comprehensive methods to combine temporal and geometric features for propagation-based misinformation detection.

References

1. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on Twitter. In: Proceedings of the 20th International Conference on World Wide Web, pp. 675–684. ACM (2011)
2. Chen, Y., Conroy, N.J., Rubin, V.L.: Misleading online content: recognizing clickbait as false news. In: Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, pp. 15–19. ACM (2015)
3. Floridi, L.: Artificial intelligence, deepfakes and a future of ectypes. Philos. Technol. **31**(3), 317–321 (2018)
4. Huang, Q., Yu, J., Wu, J., Wang, B.: Heterogeneous graph attention networks for early detection of rumors on twitter. arXiv preprint arXiv:2006.05866 (2020)
5. Kwon, S., Cha, M., Jung, K.: Rumor detection over varying time windows. PLoS ONE **12**(1), e0168344 (2017)
6. Kwon, S., Cha, M., Jung, K., Chen, W., et al.: Prominent features of rumor propagation in online social media. In: International Conference on Data Mining. IEEE (2013)
7. Liu, Y., Wu, Y.F.B.: Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
8. Lu, Y.J., Li, C.T.: GCAN: graph-aware co-attention networks for explainable fake news detection on social media. arXiv preprint arXiv:2004.11648 (2020)
9. Lu, Y., Wang, X., Shi, C., Yu, P.S., Ye, Y.: Temporal network embedding with micro- and macro-dynamics. arXiv preprint arXiv:1909.04246 (2019)
10. Lu, Y., Wang, X., Shi, C., Yu, P.S., Ye, Y.: Temporal network embedding with micro-and macro-dynamics. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 469–478 (2019)
11. Lukasik, M., Srijith, P., Vu, D., Bontcheva, K., Zubiaga, A., Cohn, T.: 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), pp. 393–398 (2016)

12. Ma, J., Gao, W., Wei, Z., Lu, Y., Wong, K.F.: Detect rumors using time series of social context information on microblogging websites. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 1751–1754. ACM (2015)
13. Ma, J., Gao, W., Wong, K.F.: Rumor detection on twitter with tree-structured recursive neural networks. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1980–1989. Association for Computational Linguistics (2018)
14. Monti, F., Frasca, F., Eynard, D., Mannion, D., Bronstein, M.M.: Fake news detection on social media using geometric deep learning. arXiv preprint arXiv:1902.06673 (2019)
15. Ruchansky, N., Seo, S., Liu, Y.: CSI: a hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 797–806. ACM (2017)
16. Shu, K., Cui, L., Wang, S., Lee, D., Liu, H.: Defend: explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 395–405 (2019)
17. Trivedi, R., Farajtabar, M., Biswal, P., Zha, H.: Dyrep: learning representations over dynamic graphs. In: International Conference on Learning Representations (2019)
18. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**, 1146–1151 (2018). https://doi.org/10.1126/science.aap9559
19. Weiss, K., Khoshgoftaar, T.M., Wang, D.D.: A survey of transfer learning. J. Big Data **3**(1), 1–40 (2016). https://doi.org/10.1186/s40537-016-0043-6
20. Zhang, Q., Liang, S., Lipani, A., Ren, Z., Yilmaz, E.: From stances' imbalance to their hierarchical representation and detection. In: The World Wide Web Conference, pp. 2323–2332 (2019)
21. Zhang, Q., Liang, S., Yilmaz, E.: Variational self-attention model for sentence representation. arXiv preprint arXiv:1812.11559 (2018)
22. Zhang, Q., Lipani, A., Kirnap, O., Yilmaz, E.: Self-attentive hawkes processes. arXiv preprint arXiv:1907.07561 (2019)
23. Zhang, Q., Lipani, A., Liang, S., Yilmaz, E.: Reply-aided detection of misinformation via Bayesian deep learning. In: The World Wide Web Conference, pp. 2333–2343. ACM (2019)
24. Zhang, Q., Yilmaz, E., Liang, S.: Ranking-based method for news stance detection. In: Companion Proceedings of the The Web Conference 2018. ACM Press (2018)
25. Zuo, Y., Liu, G., Lin, H., Guo, J., Hu, X., Wu, J.: Embedding temporal network via neighborhood formation. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2857–2866 (2018)