Network-wide shocking events through the lens of node representation shift

Manuel Dileo[1,*,†], Matteo Zignani[1,†]

[1]Department of Computer Science, University of Milan, Milan, Italy

Abstract

In the context of online social networks, network-wide shocking events, i.e. events widely recognized by most network users, are worthy of attention since they significantly impact users' behavior and interactions and have the potential to disrupt the distribution of temporal data. However, how users behave before, during, and after such events is still not clear. To address this gap we rely on the framework of graph representation learning, particularly focusing on Temporal Graph Neural Networks (TGNNs). In particular, we investigate the dynamics of node representations returned by TGNNs during a network-wide shocking event and examine how shifts in node representation can mirror the effects of such events on users' habits. We utilize a dataset from Steemit, a blockchain-based online social network, which experienced a network-wide shocking event, i.e. an important user migration caused by a hard fork in the supporting blockchain. Our findings highlight that node representations are influenced by the occurrence of the shocking event. We observe shifts in node representations, indicating changes in individual users' behavior during the event.

Keywords

Online Social Networks, Blockchain-data, Temporal Networks, Graph Neural Networks

1. Introduction and Related Works

In the context of online social networks (OSNs), a *network-wide shocking event* is an event that is widely recognized by the majority of network members and can significantly influence both the characteristics of individual users and their interactions [1]. Instances of network-wide shocking events encompass scenarios such as the transition of ownership of a social platform - exemplified by the acquisition of Twitter or significant modifications to the fundamental functionalities of a social network. Although the existence of these shocking events is acknowledged, there remains a lack of comprehensive understanding regarding the behavior exhibited by network members before, during, and after such events.

The framework of graph representation learning offers an interesting perspective for gaining a more profound understanding of the effects of network-wide shocking events on users' behavior and relationships thanks to its capability to learn representations of nodes that encapsulate their attributes and structure. In particular, we deal with the Temporal Graph Learning (TGL) framework, which has recently demonstrated its effectiveness in learning from evolving networks [2, 3, 4]. One of the most promising solutions within this field is Temporal Graph Neural Networks (TGNNs) [5], a family of graph neural networks (GNNs) that extends the concept of message passing to temporal graphs. One of the main advantages of TGNNs is the ability to return a node representation changing over time, which may adapt to the evolution of the network.

In this paper, we mainly focus on unfolding the dynamics of the node representation as returned by TGNNs in a scenario characterized by a network-wide shocking event, looking for shifts in the representation; and investigating a specific network-wide shocking event in an online social network. In particular, the node representations are obtained through the learning of future links within the network,

Discovery Science - Late Breaking Contributions 2024

*Corresponding author.
†These authors contributed equally.
manuel.dileo@unimi.it (M. Dileo); matteo.zignani@unimi.it (M. Zignani)
https://manuel-dileo.github.io/ (M. Dileo); https://connets.di.unimi.it/ (M. Zignani)
0000-0002-4861-455X (M. Dileo); 0000-0002-4808-4106 (M. Zignani)

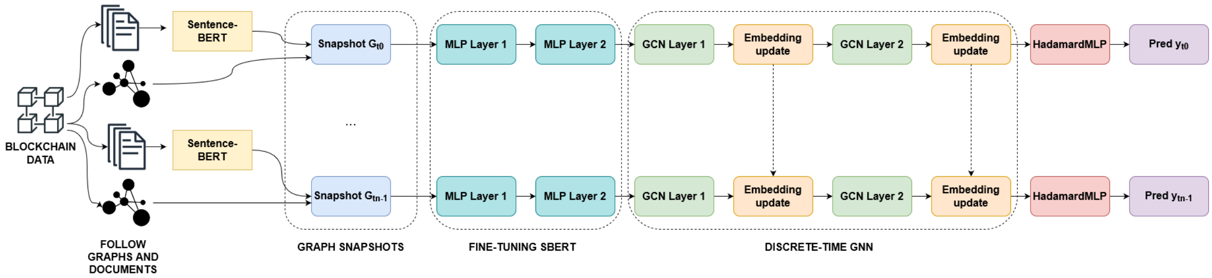


Figure 1: Pipeline of our methodology to obtain snapshot-based node embeddings on online social network data gathered from a blockchain.

without any prior knowledge of when or if a shocking event occurs during the observed periods. The methodology has been applied to a dataset collected from Steemit [6], a novel blockchain-based online social network that enables the retrieval of validated high-resolution temporal information, which experienced a user migration event caused by a hard fork in the supporting blockchain. This event was presented and studied in a few recent works [6, 7, 8]. To characterize nodes - users - we utilize both their structural features and the textual content they publish on the platform. The main findings demonstrate that the occurrence of a shocking event influences the node representation of the TGNN model. This work contributes to a deeper understanding of temporal graph learning dynamics during network-wide shocking events, highlighting their potential for change point detection and shedding light on the behavior of networked systems in the face of significant events.

2. Methodology and Dataset

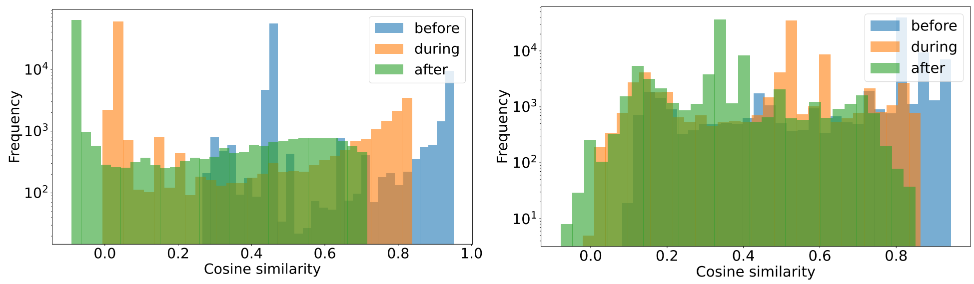
We summarize our methodology in the pipeline depicted in Figure 1. We model "follow" links and text information as a dynamic attributed graph using the snapshot-based representation [2]. To obtain the initial features for each node, we calculate the average of the document embeddings on each user's collection of posts and comments, leveraging a pre-trained SBERT LLM [9]. At the same time, we define an alternative initial node representation based only on the "follow" interaction patterns: for each node we compute some well-known structural and centrality indices such as PageRank, in-degree, out-degree, average neighbor degree, and clustering coefficient, replacing them as node features instead of textual content. We consider this alternative representation to emphasize the distinct contributions of relationships and text when users' behavior is influenced by a network-wide shocking event. Finally, we obtain temporal node representations relying on ROLAND [2], the state-of-the-art model for discrete-time dynamic graph learning, using future link prediction as a self-supervised task.

Once we obtain node representations for each snapshot, we are interested in analyzing how they change over time. Starting from the sequences of node embeddings associated with each user in the list of graph snapshots, we calculate the cosine similarity between the node representation from two months before the shocking event and all subsequent node representations obtained from the monthly snapshots following the event.

We apply our methodology to data collected from Steemit. Data have been obtained using official APIs. Of particular significance is the date March 20, 2020, which marks a notable event in the form of a hard fork in the underlying blockchain. This hard fork represents an irreversible bifurcation of the blockchain resulting from a modification in the consensus protocol and triggering a substantial debate within the Steemit community. Consequently, we denote this hard fork as the network-wide shocking event. Additionally, there were speculations regarding the acquisition of Steemit approximately one month before the hard fork. Considering this sequence of events, we focus on four specific snapshots: *i)* First month, before the event; *ii)* Second month, during the shocking event; *iii)* Third month, after the hard fork; and *iv)* Fourth month, two months after the discussion about the fork started, and the shocking event moved towards its conclusion. Overall, we processed $301,653$ new "follow" operations and $2,155,551$ comment operations.

3. Results and Conclusions

Code, data, and all the information about the model architecture and hyperparameters can be found in our GitHub repository[1]. In Figure 2 we display the distributions of the cosine similarity between the



(a) (b)

Figure 2: Distributions of cosine similarity between node representations two months before the hard fork (stable month) and i) one month before (blue), ii) during the user migration (orange), and iii) two months after (green), with textual attributes (a) and based on structural information only (b).

node states over time. We repeat the computation considering both structural and textual information. In the case of a representation combining social relationships and textual content, the cosine similarity between the representations of the same node decreases over time, especially between *before* and *during* the periods of the network-wide shocking event. In fact, we observe a shift of the distributions toward lower values of similarity as we move ahead from the first stable period. This change of representation over time suggests that *single users change their behavior during the fork*. On the other side, *taking into account structural information only, the distributions are quite comparable so the similarity among the node representations over time does not decrease so much*. This evidence aligns with the findings presented in [6], where they demonstrate that the variations of a few structural features of the social graph between the pre-fork and post-fork periods are minimal.

Conclusions. In this work, we rely on temporal node embeddings to analyze how user behavior, discussions, and interactions are influenced by a network-wide shocking event in OSNs. As a case study, we analyzed a user migration due to a hard fork event in Steemit. Our analysis reveals that the node representation reflects changes in the behavior of individual users. In future works, we will focus on accurate explanations for the detected change points.

Acknowledgments

This work has been partially funded by the Italian Ministry of University and Research (MUR) and the European Union – NextGenerationEU in the framework of the PRIN 2022 project "AWESOME: Analysis framework for WEb3 SOcial MEdia" – CUP: I53D23003680006.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] S. Huang, Y. Hitti, G. Rabusseau, R. Rabbany, Laplacian change point detection for dynamic graphs, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '20, ACM, 2020. doi:10.1145/3394486.3403077.

[1] https://github.com/manuel-dileo/nodeshift

[2] J. You, T. Du, J. Leskovec, Roland: Graph learning framework for dynamic graphs, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '22, ACM, New York, NY, USA, 2022. doi:10.1145/3534678.3539300.

[3] M. Dileo, M. Zignani, S. Gaito, Temporal graph learning for dynamic link prediction with text in online social networks, Machine Learning (2023). doi:10.1007/s10994-023-06475-x.

[4] M. Dileo, M. Zignani, Discrete-time graph neural networks for transaction prediction in web3 social platforms, Machine Learning (2024). URL: https://doi.org/10.1007/s10994-024-06579-y. doi:10.1007/s10994-024-06579-y.

[5] A. Longa, V. Lachi, G. Santin, M. Bianchini, B. Lepri, P. Lio, franco scarselli, A. Passerini, Graph neural networks for temporal graphs: State of the art, open challenges, and opportunities, Transactions on Machine Learning Research (2023). URL: https://openreview.net/forum?id=pHCdMat0gI.

[6] C. T. Ba, A. Michienzi, B. Guidi, M. Zignani, L. Ricci, S. Gaito, Fork-based user migration in blockchain online social media, in: 14th ACM Web Science Conference 2022, 2022, pp. 174–184.

[7] C. T. Ba, A. Galdeman, M. Dileo, M. Zignani, S. Gaito, User migration prediction in blockchain socioeconomic networks using graph neural networks, in: 2023 ACM Conference on Information Technology for Social Good, 2023, p. 333–341. doi:10.1145/3582515.3609552.

[8] C. T. Ba, M. Dileo, A. Galdeman, M. Zignani, S. Gaito, Analyzing user migration in blockchain online social networks through network structure and discussion topics of communities on multilayer networks, Distrib. Ledger Technol. (2024). doi:10.1145/3640020.

[9] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using Siamese BERT-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, ACLs, 2019. doi:10.18653/v1/D19-1410.