Expanding the BEND Framework: Enhancing Influence Prediction in Social Media Networks with Advanced Metrics

Rebecca Marigliano[1,2][0009-0003-8271-9990], Jacob Shaha[1,2][0000-0002-3963-1462], and Dr. Kathleen Carley[1,2][0000-0002-6356-0238]

[1] Carnegie Mellon University, Pittsburgh PA 15213, USA
[2] CMU Center for Computational Analysis of Social and Organizational Systems (CASOS), Pittsburgh PA 15213

Abstract. The influence of social media on societal narratives requires a thorough understanding of information spread and impact. Building on previous work using the BEND framework, this study incorporates new network metrics such as changes in hub centrality, authority centrality, ego-net density (bimodal and unimodal), and ego closeness to enhance the analysis of influence dynamics on social media. Using extensive Twitter data from polarized topics like the Russian-Ukraine conflict and the COVID-19 pandemic, we developed predictive models with these metrics. Despite increased complexity, the multivariate model showed promise in capturing social media influence. Our findings suggest that while new metrics provide valuable insights, further refinement, and data expansion are needed to improve accuracy. This research aids efforts to combat misinformation and support healthy online discourse.

Keywords: BEND · Network analysis · Temporal analysis

1 Introduction

The pervasive influence of social media platforms on societal narratives necessitates a robust understanding of how information spreads and impacts online communities [3] [5]. Our previous work leveraged the BEND framework to analyze how agents build influence over time through sequenced messaging within social media networks. This initial study provided critical insights into identifying optimal points for intervention to disrupt harmful narratives using network centrality and engagement metrics.

However, the complexity of influence within social networks requires a more comprehensive analysis [2]. This paper builds on our previous findings by incorporating additional metrics that offer a deeper understanding of influence dynamics [4].

Our central research question for this paper is: How can the incorporation of additional metrics enhance the identification of optimal intervention points to counteract harmful influence campaigns on social media?

To address this question, we extend our methodology to analyze temporal patterns and sequences of social media interactions, incorporating the new metrics to quantify the cumulative impact of influence campaigns. We aim to identify key moments where interventions can effectively mitigate the spread of harmful narratives without suppressing healthy discourse.

We collected extensive Twitter data focused on highly polarized topics, including the Russian-Ukraine conflict and the COVID-19 pandemic. By applying the BEND framework alongside the new metrics, we aim to develop predictive models that provide actionable insights for real-time interventions.

The ultimate goal of this research is to enhance our ability to counteract misinformation and malicious influence on social media platforms. By understanding the detailed dynamics of influence through a broader set of metrics, we aim to provide valuable insights for policymakers, social media platforms, and researchers working to preserve the integrity of online discourse.

2 Background

Researchers have demonstrated the value of quantifying an entity's reach and influence by examining the network's structure and the agent's position within it [7]. Metrics derived from graph theory offer reliable insights into how an agent's location within a network can affect their potential impact [7]. A well-constructed network model allows for a precise measurement of an individual's influence within that community. Our previous work utilized this approach by employing the BEND framework to analyze and predict how sequenced messaging campaigns can build influence on social media platforms.

In our initial study, we used network analysis and the BEND framework to investigate the temporal dynamics of influence-building on Twitter [4]. We focused on identifying key intervention points that could disrupt harmful narratives by examining centrality and engagement metrics. This analysis highlighted the critical role of network structure and agent positioning in the spread of influence.

However, existing research often overlooks multiple social media actions' cumulative and interconnected effects over time. Researchers like Wei and Carley have made significant strides in analyzing dynamic social networks and tracking individual behaviors to detect meaningful deviations [8]. These methods transition static measures of influence into dynamic, time-aggregated assessments. Our original study aligned with this approach but indicated the necessity for a more comprehensive set of metrics to capture the complexity of influence dynamics fully.

The BEND framework, developed by Carley and Beskow, simplifies the analysis of social media communication by categorizing messages into discrete influence maneuvers [1]. This framework reduces the complexity of social media interactions, making it possible to conduct categorical analysis and predict the general patterns of influence campaigns.

While our initial application of the BEND framework yielded promising results, the multifaceted nature of influence within social networks calls for a broader array of metrics. This paper expands on our previous research by looking at new metrics such as hub centrality, authority centrality, and closeness centrality. These metrics provide a more detailed and nuanced understanding of influence dynamics.

3 Methodology

To understand and counteract influence campaigns on social media, we analyze agents' actions through a series of coordinated interactions to alter the network or narrative. By employing the BEND framework, we reinterpret these interactions as time-sequenced maneuvers, allowing us to evaluate both the immediate and cumulative impacts. Our goal is to pinpoint optimal intervention moments, thereby preempting harmful narratives' significant influence without stifling genuine discourse.

3.1 Data Collection and Pre-processing

Our research utilized Twitter data collected using search terms relevant to highly polarized discussions. We selected influencers based on their demonstrated engagement and reach within these topics, focusing on those with a significant but not overwhelming influence. This ensured our analysis included agents actively shaping narratives.

We used the same two main datasets for this study:

Russian-Ukraine Conflict Dataset: Collected over 4.5 million Tweets from February 11 to August 30, 2022, using conflict-related search terms. Influencers were identified through their use of relevant hashtags and their demonstrated influence.

COVID-19 Pandemic Dataset: Included Tweets from 2019 to 2020, focusing on US debates about COVID-19 lockdowns and reopening businesses, filtered to capture the "re-open" debate in key states.

To prepare the data for analysis, we implemented a preprocessing pipeline that included tokenizing the tweets to break down the text into analyzable components, standardizing text formats for uniformity, and removing irrelevant terms and stop words. This ensured that our datasets were consistent and focused on substantive content, enhancing the accuracy of our subsequent natural language processing tasks.

3.2 Describing an Influence Campaign

Using the same key influencers, with the same assumptions that all messages sent by these users were part of building influence for their influence campaign, we extracted semantic features. Using Netanomics' Netmapper software and applying the BEND framework using Netaomics' ORA network analysis tool to classify each message.

Each tweet was assigned specific BEND maneuvers which allowed us to conceptualize each communication as a series of influence actions. We also considered the temporal clustering of tweets; if a significant amount of time elapsed between two tweets from the same user, we treated the latter as the start of a new campaign. This approach ensured our analysis reflected coherent influence sequences rather than isolated activities.

3.3 Measuring an Influence Campaign

To measure the success of influence campaigns, we built upon our previous research by incorporating additional network metrics. We again focused on identifying the cumulative impact of these messages throughout a campaign. We tracked changes in network metrics surrounding each tweet to capture this cumulative effect, providing a more accurate representation of an agent's influence over time.

Our network model is a bimodal-directed graph that maps connections between users and tweets based on message propagation. In our initial study, we employed measures such as in-degree centrality, out-degree centrality, and ego network density to understand the influence dynamics. Building on this foundation, we looked at different metrics to provide a more nuanced analysis of influence campaigns.

To derive additional unimodal measures, we first folded the bimodal user-tweet network into a unimodal user-user network. This process involved transforming the original network, which captures interactions between users and tweets, into a simplified network that focuses on direct interactions between users. By folding the network, we could analyze how users are directly connected, rather than through intermediary tweets.

Specifically, we studied the following metrics:

1. **Change in In-Degree Centrality (bimodal):** Measures the incoming connections of a given agent, reflecting the number of messages directed at them or mentioning them. Changes in this metric provide insight into the importance of the agent and the desire of other agents to solicit their attention or feedback. Our previous work was able to predict this metric with a high degree of accuracy.
2. **Change in Out-Degree Centrality (bimodal):** Measures the outgoing connections of a given agent, reflecting their level of messaging activity within the network. A change in this metric may be indicative of a more receptive or permissive network. Our previous work was able to predict this metric with a high degree of accuracy.
3. **Change in Betweenness Centrality (bimodal):** Measures the number of shortest paths within the network that include the agent in question. A change in this metric measures the extent to which an agent is becoming a key information broker within the network.
4. **Change in Hub Centrality:** Evaluates the influence of an agent based on their connections to highly connected nodes. This metric helps us understand

the agent's prominence within the network. By measuring the change in hub centrality, we can determine how an agent's strategic positioning among key nodes fluctuates over time and influences the broader network.

5. **Change in Authority Centrality:** Measures an agent's impact based on connections from authoritative nodes within the network. Tracking changes in authority centrality allows us to assess how an agent's credibility and influence among recognized authorities evolve, providing insights into their ability to sway important influencers.
6. **Change in Ego-Density (Bimodal):** Assesses the local network density around an agent in a bimodal context, providing an understanding of their immediate influence within the network. Changes in the ego-net density highlighted how closely connected an agent's immediate network becomes, indicating the effectiveness of their efforts to build a cohesive and influential sub-network.
7. **Change in Total Degree Centrality (unimodal):** Measures the number of agents connected to a given agent. Changes in this measurement reflect an agent's increasing connectedness or, conversely, isolation.
8. **Change in Betweenness Centrality (unimodal):** As previously described. Measured in the unimodal network, this provides a different context for identifying key information brokers.
9. **Change in Eigenvector Centrality (unimodal):** Indicates the extent to which an agent's connections are, themselves, connected. Agents with rising eigenvector centrality can be seen as increasingly "in the know" across the network. Our previous work was able to predict this metric with a high degree of accuracy.
10. **Change in Ego Closeness (unimodal):** Shows how efficiently an agent can disseminate information throughout the network. This metric indicates the agent's ability to reach other nodes quickly. By examining changes in ego closeness, we can understand how an agent's efficiency in spreading messages evolves, which is crucial for evaluating their real-time impact on information dissemination.
11. **Change in Ego-Net Density (Unimodal):** Examines the local network density around an agent in an unimodal context, showing an agent's immediate influence in a simplified network structure. This metric helps us understand the concentration of influence in the agent's immediate environment, and how this concentration changes in response to their actions.

For modeling and prediction, we used a random forest regression model to handle non-linear effects among the 16 binary features. Each user's tweets were excluded in turn during leave-one-user-out cross-validation, and the model was trained on the remaining data. The trained model was tested on the excluded user's tweets, and this process was repeated for each user. All metric values were normalized to the range [0,1] to ensure comparable Root Mean Squared Error (RMSE) for each variable. Each model's predictive accuracy was measured with the R^2 coefficient of determination, and the final performance of the model was scored using the root mean squared error (RMSE). We selected RMSE as our measurement because the measured values all fell below 1.0, making the mean squared error (MSE) an unreliable indicator of error magnitude.

4 Analysis and Follow-on

Our analysis of the predictive models involved examining the performance metrics for each of the influence metrics we studied. We summarized the model results in Table 1, showing the RMSE and R^2 values achieved by the best model for each variable [6].

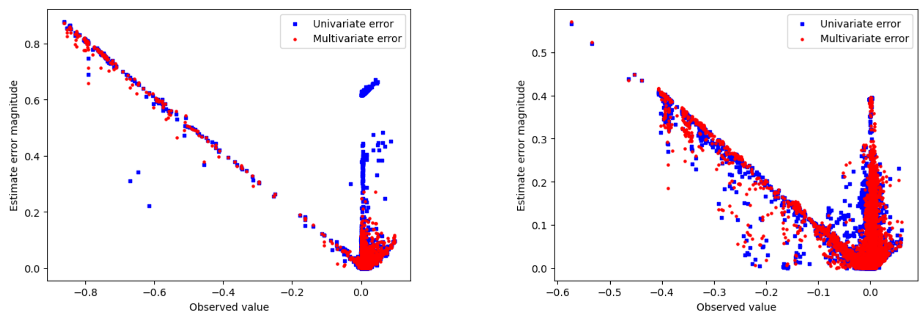
Summary of Predictive Models			
Dependent Variable	*Average R^2 (Training)*	*Best R^2 (Training)*	*Average RMSE (Validation)*
Multivariate	-1.107	-0.00714	0.0255
Δ Bi In-Deg Cnt	-3.318	0.0014	0.00418
Δ Bi Out-Deg Cnt	-0.679	0.00496	0.00541
Δ Bi Betw Cnt	-0.344	-0.00549	0.027
Δ Bi Hub Cnt	-1.703	0.05	9.98E-05
Δ Bi Authority Cnt	-4.129	-0.000607	0.00178
Δ Bi Ego-net Dns	-0.507	0.082	0.0386
Δ Uni Deg Cnt	-1.038	0.0138	0.00523
Δ Uni Betw Cnt	-0.407	-0.00665	0.0256
Δ Uni Eigen Cnt	-1.051	0.00103	0.0123
Δ Uni Closeness Cnt	-1.051	0.00228	0.142
Δ Uni Ego-net Dns	-0.493	0.0754	0.0353

Table 1: Performance Metrics of Predictive Models

Model accuracy was significantly lower than in our previous work, where our best models scored over 0.95 R^2. Alarmingly, our new model underperformed even for the outcome metrics (in- and out-degree centrality) successfully predicted by a similar approach in that paper. This discrepancy could be attributed to the new metrics being weaker indicators of influence, suggesting they are less directly tied to BEND maneuvers. Additionally, the inherent complexity and variability introduced by these additional metrics might have diluted the predictive power observed with the original set.

However, the most likely explanation is the inclusion of new data, indicating that the BEND tendencies within these two separate populations were divergent enough to weaken the predictor. This suggests that the patterns of influence and behavior in the new data set differ significantly from those in the previous data set, complicating the model's ability to generalize effectively. Notably, the strongest model was the multivariate model, which was not the case previously. This indicates that there is value in the additional metrics, as they may capture different measurements of influence not accounted for by the original metrics.

Additionally, we examined the residuals to identify interesting behaviours. Below are two plots for Change in Ego Closeness and Change in Ego-Net Density (Unimodal), showing atypical behaviour in the below figures.



(a) Residuals for: Change in Ego Closeness

(b) Residuals for: Change in Ego-Net Density (Unimodal)

Fig. 1: Residual analysis for selected metrics

In both plots, the multivariate model (red) shows tighter clustering around the observed values compared to the univariate model (blue), indicating better overall performance. However, there are notable outliers where the error magnitude is significantly higher. This suggests that certain data points are not well-predicted by either model, highlighting areas where the model's assumptions or the data itself may need further refinement.

Most alarming is the clear linear shape to the residuals as the measured value moves away from zero. This strongly indicates that our models have converged toward an inappropriately linear solution, likely due to the preponderance of data clustered around 0; while the "shape" of the model may be accurate at predicting small change, it clearly breaks down the further the values move away from the origin.

Given the findings, we suggest several steps for further research to enhance the robustness and accuracy of our models. First, adding more data would provide more comprehensive insights and potentially improve predictive accuracy by capturing a broader range of behaviours and patterns. Second, including temporal features, such as t-n lag maneuvers for tweet t, can capture more complex temporal dynamics. Third, dropping less relevant features, such as narrative maneuvers (categories E and D in the BEND framework) could streamline the model and reduce the overfitting tendency seen in some of our results. Lastly, conducting a thorough analysis of outliers and considering their removal may improve model performance, as indicated by the residual chart. In fact, the residuals indicate a data transform may be necessary, implying that our selected impact metrics are either insufficiently independent of the BEND maneuvers, or that our model as presented insufficiently deals with an inherently autocorrelated phenomenon.

Despite these poor results, we remain optimistic that a predictive methodology can be developed. By addressing these areas and refining our approach, we plan to build on our current findings and develop a more robust framework for predicting influence dynamics in social media networks. This ongoing research is critical for effectively countering harmful narratives and supporting healthy discourse on online platforms.

Acknowledgments. We are indebted to the many researchers who have built a robust and thorough foundation in social network research. We especially recognize the contributions of Carnegie Mellon's CASOS and IDeaS centers. These contributions include the BEND framework which, regardless of its shortcomings, provides a useful and standardized methodology to quantitatively consider issues like this one. Our ability to understand and explore our network data was vastly enhanced by Netanomics ORA and NetMapper software.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Carley, K.: Social cybersecurity: an emerging science. Comput Math Organ Theory **26**, 365–381 (2020)
2. Carley, K.: Disinformation: A Multi-National, Whole of Society Perspective, chap. A Political Disinfodemic. Springer, Cham (2022), https://doi.org/10.1007/978-3-030-94825-2_1
3. Holmstrom, M.: The narrative and social media. Defence Strategic Communications **1**, 118–132 (2015)
4. Marigliano, R., Shaha, J., Carley, K.: Logical momentum: Leveraging the bend framework to determine optimal patterns and sequences for network influence. In: Proceedings of the International Conference on Behavioral Representations in Modeling and Simulation (BRiMS). Carnegie Mellon University, Pittsburgh, PA, USA (2023)
5. Muhammed T, S., Mathew, S.K.: The disaster of misinformation: a review of research in social media. International journal of data science and analytics **13**(4), 271–285 (2022)
6. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
7. Wasserman, S., Faust, K.: Social network analysis. Cambridge University Press (1994)
8. Wei, W., Carley, K.M.: Measuring temporal patterns in dynamic social networks. ACM Trans. Knowl. Discov. Data **10**(1) (jul 2015). https://doi.org/10.1145/2749465, https://doi.org/10.1145/2749465