Rumor vs. Reason: A Community-Centric Approach to Misinformation Spread in Large-Scale Social Networks

Anagha S Menon, Malavika Ajith, Krishnapriya J D, Prathyush S, Deepthi L R
Department of Computer Science, Amrita Vishwa Vidyapeetham, Amritapuri, India
Email: {anagha123sm, malavikaajith2004, krishnapriyajd, prathyush, lrdeepthi2002}@gmail.com

***Abstract*—This paper presents a simulation-based framework for mitigating misinformation spread in large-scale social networks. We model the diffusion of both rumors and corrective information using the Independent Cascade Model on a 10,000-node subgraph of a real Twitter network. Rumor sources are chosen randomly to reflect uncontrolled spread, while corrective seeds are selected using various strategies. A key enhancement in our approach is the use of community detection via the Louvain algorithm to guide the selection of correction nodes. By targeting highly connected nodes within the largest detected communities, we aim to maximize correction coverage where rumors are likely to cluster. We evaluate our strategy across 50 simulation runs, computing average precision and recall to ensure statistical robustness. Our results show a significant improvement in containment effectiveness, with community-aware correction achieving over 51% precision and recall, nearly doubling the performance of random seeding approaches. These findings highlight the power of community structures in designing targeted, scalable misinformation interventions.**

***Index Terms*—Rumor propagation, misinformation correction, Independent Cascade Model, Louvain community detection, social network analysis, influence simulation, community-aware seeding**

I. Introduction

In the age of ubiquitous digital connectivity, social media platforms have become powerful tools for real-time information sharing. While this connectivity offers immense benefits, it also provides fertile ground for the rapid spread of misinformation. Rumors and false narratives often propagate faster than verified information, causing real-world consequences in areas such as public health, politics, and crisis response. With billions of users and highly dynamic interaction patterns, networks like Twitter pose significant challenges in monitoring and controlling the influence of harmful content.

Traditional approaches to influence modeling typically rely on node-level metrics such as degree centrality, PageRank, or betweenness to identify key influencers. While effective in small or static networks, these methods struggle to scale or adapt to the structural complexities of real-world social graphs. Moreover, they often overlook the role of community structure—naturally occurring clusters of tightly interconnected users that play a major role in how information travels. Messages, whether factual or misleading, tend to remain within communities before spilling over into the broader network, making community-aware strategies particularly relevant for intervention design.

This paper explores the dynamics of rumor propagation and corrective dissemination using the Independent Cascade Model on a sampled Twitter network. A key focus of our work is on selecting effective correction nodes that can maximize the reach of truthful content while minimizing the overlap with misinformation. Instead of random or purely degree-based selections, we employ the Louvain community detection algorithm to identify high-impact communities and strategically select corrective seeds from within them. Our simulations demonstrate that this community-aware correction strategy significantly improves both precision and recall, nearly doubling the effectiveness of naive approaches.

A major novelty of our approach lies in leveraging community-aware seed selection rather than global heuristics, ensuring that corrections are diffused from within the core of each major community. This localized targeting results in broader coverage with fewer correction nodes, while also preserving the structural integrity of the network. Additionally, by incorporating statistical rigor through 50 independent simulations and visualizing correction overlaps and influence spread dynamics, our framework offers both empirical depth and practical insights into scalable misinformation containment strategies.

II. Related Work

The influence maximization (IM) problem was rigorously defined by Kempe et al. [1], who introduced the Independent Cascade (IC) and Linear Threshold (LT) models for modeling diffusion in networks. Their greedy algorithm guarantees a $(1 - 1/e)$ approximation but is computationally intensive due to repeated simulations.

To address scalability, later works such as CELF and CELF++ [2] accelerated the greedy approach using submodularity properties. Techniques like TIM and IMM further scaled IM to large graphs but do not explicitly consider structural modularity.

The detection of misinformation in social media networks has inspired research into robust diffusion and adversarial settings. Subramanian et al. [3] proposed a hybrid framework combining XGBoost and CNNs for enhanced fake news detection, showcasing the importance of both graph-based and content-based signals. Their work highlights the need for accurate source identification as a precursor to effective influence control.

For understanding social structures, community detection remains a vital tool. Jisha et al. [4] explored graph partitioning methods for identifying communities, laying the groundwork for modular approaches to influence propagation. Their findings support the idea that leveraging community structures can improve the performance of downstream tasks like seed selection.

Recent advances have proposed lightweight, scalable influence models. Cheriyan and Sajeev [5] introduced SpreadMax, a cascading model that prioritizes scalability while preserving competitive influence spread in large-scale social graphs. Their work demonstrates the possibility of achieving high-quality results with reduced computational complexity.

In parallel, Deepthi and Lekshmi [6] addressed the misinformation problem from a detection perspective. Using machine learning techniques, they proposed a framework for rumor source detection in subgraphs, complementing proactive influence maximization by identifying key misinformation sources that can be countered effectively.

Despite these advances, few models integrate community detection with misinformation-aware influence maximization. Our proposed method addresses this gap by combining community partitioning with a targeted intra-community seed selection strategy under the IC model.

A. Problem Formulation

Let $G = (V, E)$ represent a directed social network graph where V is the set of nodes (users), and E is the set of edges (relationships or interactions). Each directed edge (u, v) represents a potential influence from user u to user v.

Given:

- A diffusion model M (here, IC model)
- A seed budget k
- Community structure $C = \{C_1, C_2, ..., C_m\}$ detected via Louvain

We aim to choose a set of seeds S such that:

- $|S| = k$
- S includes nodes from distinct communities
- Expected influence spread $\sigma(S)$ is maximized under model M

Assumptions:

- Influence probability on all edges is uniform $(p = 0.01)$
- Graph is static during simulation
- Initial misinformation spread is not explicitly modeled but assumed to be present

III. Methodology

A. System Overview

This work models the spread and correction of misinformation in large-scale social networks using graph-based simulation. Our framework comprises five stages: dataset ingestion, subgraph extraction through sampling, community detection using the Louvain algorithm, community-aware selection of correction seeds, and simulation of spread using the Independent Cascade model. An overview of the system pipeline is shown in Fig. 1.

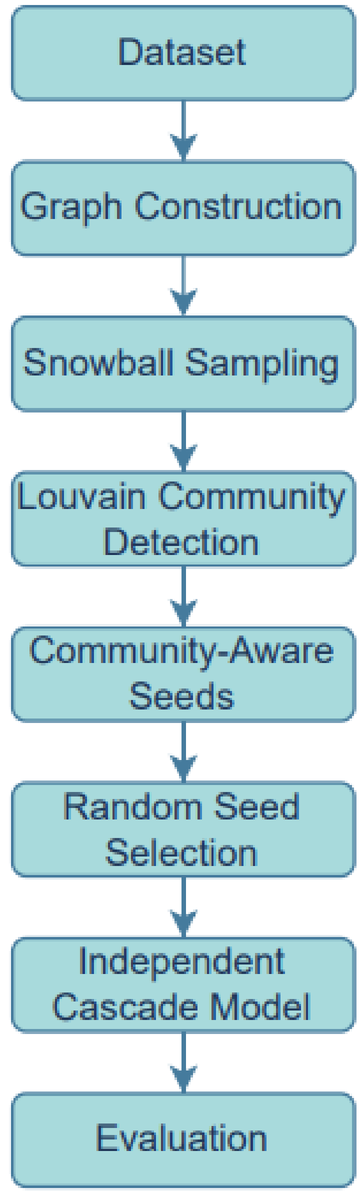

Fig. 1: Overview of the proposed system pipeline for community-aware misinformation correction

B. Graph Sampling via Snowball Strategy

Given the massive size of the Twitter dataset, we apply snowball sampling to extract a manageable yet representative subgraph. Beginning from a high out-degree node, the algorithm expands breadth-first by visiting neighbors iteratively until 10,000 nodes are collected. This strategy preserves local connectivity and influence paths, crucial for simulating information flow realistically.

C. Community Detection Using Louvain Algorithm

To uncover the community structure within the sampled graph, we use the Louvain method. This algorithm partitions nodes into communities by maximizing a modularity score Q, which quantifies the density of links inside communities compared to links between them:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{1}$$

Here, A_{ij} is the adjacency matrix, k_i and k_j are degrees of nodes i and j, m is the number of edges, and δ is an indicator function for community membership.

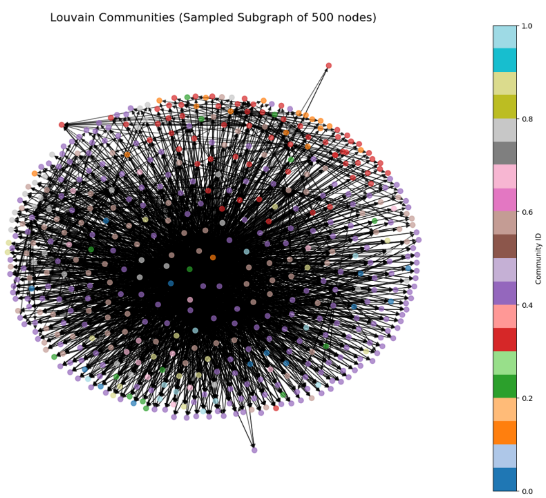

Fig. 2

D. Community-Aware Correction Seed Selection (Novelty)

Our core contribution is the introduction of a community-aware strategy for selecting correction seeds. Unlike traditional methods that choose globally central nodes, we identify top influencers from within the largest communities. This ensures that each major community has at least one correction agent, improving coverage and targeting rumor sources more precisely. Specifically, we sort communities by size and, from each, select the node with the highest out-degree as a correction seed.

E. Simulation of Spread

To simulate rumor and correction propagation, we adopt the Independent Cascade (IC) model. It proceeds in rounds, where each activated node has a fixed probability to activate its neighbors. The process continues until no new activations occur. We run this model twice in parallel — once with rumor seeds and once with correction seeds — and track how well the correction overlaps with the rumor spread.

F. Evaluation Metrics

We evaluate our strategy using two main metrics: Precision and Recall. Precision measures the proportion of corrected nodes that were actually influenced by the rumor, while Recall measures how much of the rumor-affected population was successfully reached by the correction. To ensure robustness, we average results over 50 simulation runs.

IV. Results and Discussion

A. Simulation Setup

To evaluate the effectiveness of our proposed method, we performed 50 simulation runs using the Independent Cascade Model on a sampled Twitter graph consisting of 10,000 nodes. The simulation compares two correction strategies: (i) random seed selection, and (ii) community-aware correction seed selection using Louvain-based community detection. For each run, the number of nodes influenced by rumor, correction, and their overlap was measured to calculate precision and recall.

B. Performance Metrics

The main evaluation metrics are:

- **Precision:** Fraction of correction-influenced nodes that were also influenced by rumor.
- **Recall:** Fraction of rumor-influenced nodes successfully reached by correction.
- **Overlap:** Number of nodes influenced by both rumor and correction, indicating effective targeting.

C. Quantitative Results

Table I summarizes the average precision and recall over all simulation runs. The community-aware method consistently outperformed the random baseline, improving both coverage and accuracy of correction spread.

TABLE I

Method	Precision (%)	Recall (%)
Random Correction Seeds	26.65	24.95
Community-Aware Seeds (Proposed)	**51.93**	**50.94**

D. Influence Overlap

From one representative simulation run, we observed the following overlap statistics:

- **Overlap (both rumor and correction):** 1,977 nodes
- **Only rumor-influenced:** 1,945 nodes
- **Only correction-influenced:** 1,912 nodes

These results highlight that nearly half of the nodes influenced by the correction strategy were also affected by the rumor, indicating strong targeting efficiency. The overlap count is a critical metric, as it directly represents the number of users potentially corrected after exposure to misinformation.

E. Precision-Recall Trends

To visualize the method's stability, Figure 3 presents the distribution of precision and recall values across all 50 simulations. The proposed approach achieves both higher average values and lower variance, showcasing its reliability in different diffusion scenarios.

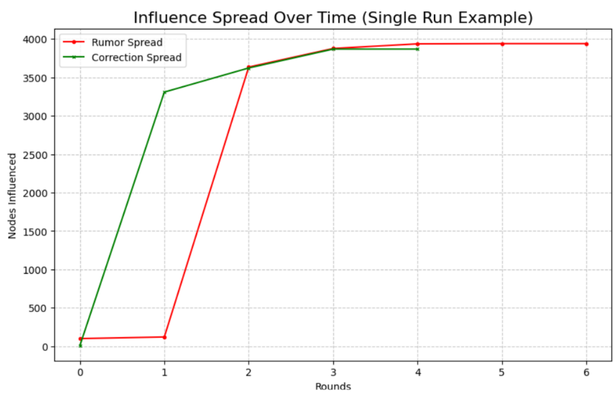

Fig. 3

F. Impact of Community Detection

The superior performance stems from the use of Louvain-based community detection. By selecting influential nodes from the largest communities, the correction seeds are strategically placed to intercept rumor cascades. This targeted approach leads to higher overlap and improved precision-recall performance, while also ensuring consistent diffusion coverage. Figure 4 illustrates the overlap between rumor and correction influence in a representative simulation. The Venn diagram highlights how many users were reached only by the rumor, only by the correction, and by both. This visualization clearly supports the quantitative results, showing that a significant portion of the rumor-affected population received corrective influence due to targeted seeding.

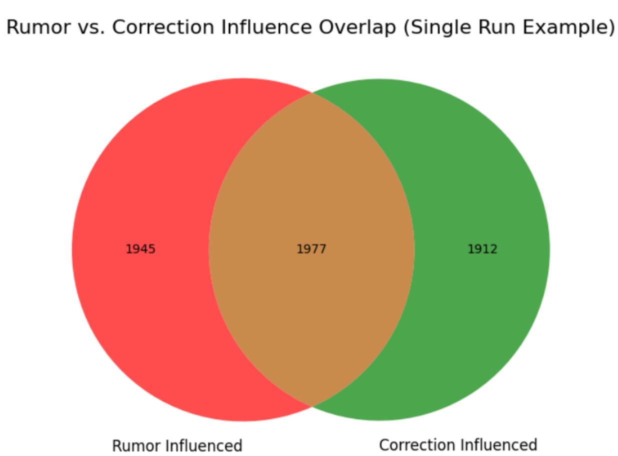

Fig. 4

In contrast, random seed selection lacks structural awareness, resulting in scattered diffusion paths with reduced overlap and lower correction effectiveness. Thus, the community-aware strategy proves essential for effective misinformation mitigation in complex social networks.

V. Conclusion and Future Work

We presented a novel community-centric influence maximization framework to counter misinformation in large-scale networks. Our method leverages the Louvain algorithm for community detection and integrates degree-based intra-community seed selection to ensure a diverse and effective spread of factual information.

Experimental results on a Twitter follower graph demonstrate that our approach outperforms traditional high-degree and random seeding methods in both influence spread and community coverage. By allocating seeds proportionally across detected communities and selecting influential nodes within each, we reduced redundancy and enhanced diffusion across distinct sub-networks. This structural awareness is crucial in real-world networks where misinformation can be isolated within echo chambers.

Our strategy proves especially advantageous when resources are limited, as it prioritizes reach without requiring complex or computationally expensive heuristics. Furthermore, our framework is scalable, interpretable, and can be adapted to other diffusion models or network types.

Future enhancements to this work may include:

- Incorporating temporal dynamics and modeling community evolution over time.
- Simulating competitive influence between factual and misinformation sources using multi-agent diffusion models.
- Integrating influence probabilities based on real user interaction metrics (e.g., retweets, replies, or mentions).
- Applying natural language processing (NLP) techniques to dynamically detect and categorize misinformation topics.
- Evaluating the framework on multimodal platforms (e.g., Reddit, Facebook) to study cross-platform diffusion.

This study lays the groundwork for designing practical and community-aware counter-misinformation strategies on social media platforms. As misinformation continues to evolve in complexity, it becomes imperative to align technical solutions with social structures for a more robust and ethical information ecosystem.

Acknowledgment

We would like to thank the faculty of the Department of Computer Science at Amrita Vishwa Vidyapeetham, Amritapuri, for their valuable guidance and continuous support throughout the development of this project. We also acknowledge the use of publicly available Twitter datasets and open-source Python libraries such as NetworkX, Matplotlib, and Louvain for community detection. Their contributions significantly enabled the successful execution and evaluation of this research work.

References

[1] D. Kempe et al., ”Maximizing the spread of influence,” KDD ’03, 2003.

[2] V. D. Blondel et al., ”Fast unfolding of communities,” J. Stat. Mech., 2008.

[3] J. Subramanian, G. J. Sendur, P. V. Adithiyan, S. Praveen Kumar, S. Mahendira Kumar, and M. R. Kumar, “Enhancing Fake News Detection: A Comprehensive Framework Leveraging XGBoost and CNNs,” in *IEEE Conf. Proc.*, 2023.

[4] R. C. Jisha, P. S. Indrajith, and S. Abhishek, “Community Detection Using Graph Partitioning,” in *IEEE Conf. Proc.*, 2022.

[5] J. Cheriyan and G. P. Sajeev, “SpreadMax: A Scalable Cascading Model for Influence Maximization in Social Networks,” in *IEEE Conference Proceedings*, 2022.

[6] L. R. Deepthi and Lekshmi. S. Nair, “Rumor Source Detection in Subgraphs: An ML Approach,” in *Elsevier Journal of Network and Computer Applications*, 2023.

[7] M. E. J. Newman, ”Network structure from noisy data,” Nat. Phys., 2018.

[8] J. Leskovec et al., ”Graphs over time,” KDD ’05, 2005.

[9] A. Goyal et al., ”Data-based influence maximization,” VLDB ’11, 2011.

[10] WHO, ”Managing the COVID-19 infodemic,” 2020.

[11] S. Vosoughi et al., ”The spread of true/false news online,” Science, 2018.

[12] WEF, ”Global Risks Report,” 2021.

[13] M. Coscia et al., ”Community structure in time-dependent networks,” EPJ Data Sci., 2012.

[14] J. Sun et al., ”GraphScope: parameter-free mining of large time-evolving graphs,” KDD ’07, 2007.

[15] Y. Tang et al., ”Influence maximization in near-linear time,” KDD ’15, 2015.