Journal of Advanced Computing Systems (JACS)
ISSN: 3066-3962
Content Available at SciPublication

# Temporal-Structural Propagation Graph Analysis for Coordinated Misinformation Campaign Detection and Source Attribution in Social Networks

*Minghua Deng*[1], *Shuyang Xu*[1.2]
[1] *Computational Data Science, Carnegie Mellon University, PA, USA*
[1.2] *Master of Professional Studies, Applied Statistics, Cornell University, NY, USA*

DOI: 10.69987/JACS.2026.60501

**Keywords**

coordinated inauthentic behavior; propagation graph analysis; source attribution; misinformation detection

**Abstract**

The proliferation of coordinated misinformation campaigns across large-scale social networks poses critical challenges to information ecosystem integrity and democratic discourse. Existing detection approaches predominantly rely on either content-based signals or simplified graph topology, limiting their capacity to capture the nuanced behavioral synchronization patterns that characterize organized influence operations. This paper presents a temporal-structural propagation graph analysis framework that jointly models the topological characteristics and dynamic spreading behaviors of information cascades to identify coordinated misinformation networks and trace their origin nodes. The proposed approach constructs heterogeneous propagation graphs from timestamped user interaction streams, extracts community-level synchronization features alongside cascade velocity anomaly indicators, and employs a multi-hop diffusion backtracking procedure that assigns calibrated confidence scores to candidate source nodes through structural centrality constraints. Experiments conducted on three publicly available benchmark datasets demonstrate consistent improvements over representative graph-based and hybrid detection baselines in both coordinated campaign identification and source localization accuracy. Ablation results confirm that temporal dynamics and structural topology operate as mutually reinforcing signal streams whose combination yields the strongest attribution outcomes. These findings carry direct implications for platform-level intervention design and national information security policy objectives.

## 1. Introduction

### 1.1. Research Background and Motivation

The rapid expansion of social media infrastructure has fundamentally altered how information propagates through public discourse, enabling both organic knowledge dissemination and deliberate manipulation at an unprecedented scale. Over the past decade, state-sponsored and commercially motivated actors have developed sophisticated coordinated campaigns that exploit platform affordances to artificially amplify false narratives, suppress authentic voices, and distort collective perception of political events, public health communications, and geopolitical conflicts. A defining characteristic of such campaigns is their reliance on networks of accounts acting in behavioral synchrony while concealing their organizational structure behind seemingly independent personas—a challenge that distinguishes them sharply from conventional single-account disinformation efforts.

Prior research has established that propagation structure carries substantial discriminative signal for misinformation detection, with graph-based representations consistently outperforming content-only classifiers when cascade topology is observable [1]. Complementary advances in network-theoretic source localization have demonstrated that origin attribution remains tractable under realistic observability conditions when structural and temporal evidence are jointly exploited [2]. Bridging these two research traditions into a unified detection and attribution pipeline represents an open challenge of practical consequence for platform integrity teams and national security agencies alike. The United States Cybersecurity and Infrastructure Security Agency has explicitly prioritized technical capabilities for identifying and attributing foreign information manipulation operations,

underscoring the policy urgency of scalable graph-based solutions. Addressing this gap requires frameworks capable of simultaneously characterizing network topology, temporal behavioral alignment, and multi-hop diffusion dynamics across interaction graphs containing millions of edges.

### 1.2. Problem Formulation and Research Scope

A. Formal Definition of Coordinated Misinformation Campaigns

A social network is represented as a directed graph G = (V, E, W, T), where V denotes the set of user nodes, E ⊆ V × V encodes directed interaction edges including retweets, replies, and quote-posts, W = {w_e : e ∈ E} assigns an interaction weight to each edge, and T = {t_e : e ∈ E} assigns a timestamp to each edge. A misinformation narrative n is characterized by a claim-bearing content unit propagated across G. A coordinated misinformation campaign C_n is formally defined as a subset of accounts V_c ⊆ V that collectively amplify narrative n with statistically anomalous temporal synchronization, quantified as an inter-arrival time distribution significantly narrower than expected under a Poisson arrival process of independent engagement. This behavioral operationalization aligns with established definitions in the computational social science literature and has been validated against disclosed platform takedown datasets [3]. The detection objective is to classify an observed propagation subgraph G_n as either a coordinated campaign or an organic cascade, given the graph topology and associated timestamp sequence.

B. Source Attribution as a Graph Inverse Problem

Given a propagation subgraph G_n induced by narrative n, the source attribution task identifies the node v* ∈ V most likely responsible for initiating the misinformation cascade. This formulation casts attribution as a graph inverse problem analogous to rumor source identification under Susceptible-Infected diffusion models. The central challenge is that by the time a campaign is detected, the propagation graph has already expanded substantially, and structural traces connecting the origin to its downstream amplifiers may be obscured by network noise, missing edge observations due to platform API constraints, and deliberate obfuscation by campaign operators. Attribution confidence must jointly account for the temporal plausibility of early node activation and the structural centrality characteristics expected of seeding accounts.

### 1.3. Main Contributions of This Work

This paper advances the detection and attribution of coordinated misinformation campaigns through three substantive contributions. The first is a heterogeneous propagation graph construction pipeline that integrates multiple interaction types and timestamped edge weights into a unified temporal-structural representation, enabling joint topological and dynamic analysis across cascades of varying density and depth. The second is a coordinated campaign detection procedure that extracts community-level behavioral synchronization features and cascade velocity anomaly indicators, combining them into a unified anomaly scoring mechanism that generalizes across narrative domains without requiring labeled coordination examples at test time. The third is a multi-hop reverse diffusion attribution algorithm that identifies candidate origin nodes by backtracking through the inferred diffusion order and refines candidate rankings using structural centrality constraints, yielding calibrated source confidence scores with measurable precision gains over established localization baselines.

## 2. Related Work

### 2.1. Graph-Based Approaches to Misinformation Propagation

A. Static Propagation Tree Methods

Early graph-based misinformation detection approaches modeled information cascades as static tree structures rooted at the original claim, with reply and retweet nodes forming child branches. These representations enabled extraction of structural features including tree depth, branching factor, and the ratio of verifying to disputing responses—collectively capturing response dynamics that distinguish coordinated narratives from organic information cascades. Degree-corrected social graph refinement demonstrated that correcting for degree heterogeneity during graph construction substantially improves downstream classification by reducing the confounding influence of high-degree hub nodes on community assignment [3]. Subsequent propagation structure-aware graph transformer architectures extended this insight by encoding both local neighborhood structure and long-range path dependencies through attention mechanisms, achieving improved robustness to adversarial graph perturbations applied by sophisticated campaign operators [4]. These static approaches established that structural graph features carry discriminative signal independent of content semantics, motivating the structural component of the framework presented in this paper.

B. Dynamic and Temporal Graph Extensions

Static propagation trees discard the temporal ordering of node activations, losing critical evidence about the speed and coordination pattern of information spread. Dynamic graph formulations address this limitation by

treating each timestamped interaction as an event in an evolving structure, enabling extraction of velocity-sensitive features that contrast coordinated amplification from organic virality. Edge-enhanced Bayesian graph convolutional networks introduced uncertainty-aware message passing to handle noise inherent in social propagation edges, yielding more robust classification when graph topology is partially unobserved due to platform sampling [5]. Graph adversarial contrastive learning extended temporal graph analysis to rumor detection by training encoders to remain invariant to benign structural perturbations while remaining sensitive to coordination signals embedded in burst activation patterns [6]. Taken together, these contributions establish that temporal and structural feature integration represents the most productive direction for next-generation detection frameworks.

### 2.2. Coordinated Inauthentic Behavior Detection

The detection of coordinated inauthentic behavior has emerged as a research problem distinct from single-claim veracity classification, focusing instead on the organizational structure of narrative amplification. User-aware multi-relational heterogeneous graph approaches demonstrated the value of simultaneously modeling user credibility signals, posting rhythm edges, and content similarity links, achieving superior detection of coordinated networks over single-relation baselines by capturing the diverse behavioral dimensions through which coordination manifests [7]. Foundational work on unveiling coordinated groups behind disinformation campaigns constructed similarity networks over shared behavioral traces—co-retweeting patterns, hashtag co-occurrence, and synchronized posting windows—to identify dense account clusters operating under shared coordination mandates [8]. A key insight from this line of work is that behavioral similarity graphs generalize across content domains and resist content-level evasion strategies, making them practically valuable against human-operated campaigns that evade purely automated classification. Cross-platform coordination detection remains an open challenge due to the difficulty of aligning user identities and behavioral traces across heterogeneous platform data sources with inconsistent API access policies.

### 2.3. Source Localization in Information Diffusion Networks

Source localization in information diffusion networks has a foundational basis in network science, with classical estimators including rumor centrality and the Jordan center heuristic derived under idealized Susceptible-Infected spreading assumptions. Interpretable graph structural learning methods have advanced the trustworthiness of detection outcomes by producing sparse graph representations that enable human analysts to inspect which specific behavioral edges drive a classification decision, addressing a practical deployment requirement for platform trust and safety teams [9]. The transition from single-snapshot to multi-snapshot localization frameworks has been identified as a critical capability gap, as source estimates derived from the full propagation graph systematically underweight the early-stage behavioral evidence most diagnostic of campaign seeding. Lightweight source localization procedures designed for large-scale networks have further confirmed that computational efficiency is a non-negotiable constraint for operationally viable attribution, as graph sizes in real deployment environments routinely exceed the tractability limits of exact inference algorithms. These observations collectively motivate the multi-hop backtracking approach developed in Section 3.3.

## 3. Methodology

### 3.1. Temporal-Structural Propagation Graph Construction

The first stage of the proposed framework transforms raw social interaction logs into a structured temporal-structural propagation graph supporting joint topological and dynamic analysis. Given a collection of timestamped interaction records associated with a target narrative n, the pipeline constructs a directed multigraph G_n = (V_n, E_n, W_n, T_n), where V_n is the set of participating user nodes, E_n encodes directed interactions across retweet, reply, and quote-post relations, W_n assigns interaction frequency-weighted edge weights computed over rolling 3,600-second windows, and T_n assigns empirically observed timestamps to each edge. The construction proceeds through three sequential phases integrated with dynamic assessment principles for cascading information events [10].

In the ingestion phase, raw interaction records are deduplicated and sorted chronologically, with a sliding window of width $\Delta = 3{,}600$ seconds applied to aggregate concurrent interactions into temporally coherent edge batches. In the structuring phase, each user node $v \in V_n$ is augmented with a normalized feature vector $x_v \in \mathbb{R}^d$ comprising account-level attributes: posting frequency $f_v$, follower-to-following ratio $r_v$, account age $a_v$, and average inter-posting interval $\sigma_v$. In the weighting phase, edge weights are computed as the product of interaction frequency and a temporal proximity score, yielding a representation that encodes both structural connectivity and behavioral rhythm alignment between account pairs. The resulting graph is subsequently decomposed into a sequence of temporal snapshots $G_n^{(1)}, G_n^{(2)}, \ldots, G_n^{(K)}$, each capturing the propagation

state within a fixed observation window of $\tau = 1{,}800$ seconds. This snapshot sequence preserves the evolutionary trajectory of the cascade and provides the input sequence for coordinated detection and source attribution in downstream stages.

Table 1 summarizes the structural characteristics of propagation graphs constructed from the three benchmark datasets.

| Dataset | Avg. Nodes | Avg. Edges | Avg. Depth | Avg. Snapshots | Labeled Campaigns |
|---|---|---|---|---|---|
| Twitter15 | 1,342 | 4,871 | 8.3 | 24 | 374 |
| Twitter16 | 1,158 | 3,994 | 7.6 | 21 | 412 |
| FakeNewsNet | 2,207 | 9,126 | 11.4 | 38 | 318 |

Table 1: Structural statistics of temporal-structural propagation graphs across benchmark datasets, covering node count, edge count, tree depth, snapshot count, and verified campaign labels. Depth computed over retweet cascades after pruning isolated nodes.

### 3.2. Coordinated Campaign Detection via Topology Feature Extraction

A. Community-Level Behavioral Synchronization Features

Once the propagation graph G_n is constructed, the framework extracts features characterizing the degree of behavioral synchronization among participating accounts. Coordinated campaigns exhibit unnaturally tight temporal alignment in posting and resharing activity, manifesting as dense subgraph clusters whose members activate within narrow, statistically anomalous time windows. Community detection is performed on each temporal snapshot using Louvain modularity optimization, which identifies groups of nodes exhibiting stronger internal connectivity than predicted under a random graph null model. Propagation graph generation via directed acyclic graph-aware architectures has demonstrated that preserving the DAG topology during community extraction substantially improves the fidelity of community boundary estimates, motivating the snapshot-level decomposition adopted here [11].

For each detected community C_k within snapshot G\ n^(t), three synchronization features are computed: the inter-arrival time variance $\sigma^2$_IA(C_k), measuring dispersion of posting timestamps within C_k with low values indicating tight temporal coordination; the co-activation ratio $\rho$_CA(C_k), quantifying the fraction of members active within a 60-second window; and the behavioral similarity score $\beta$\ sim(C\ k), computed as mean cosine similarity between account feature vectors {x_v : v ∈ C_k}. These three features are aggregated across all communities and snapshots to produce a campaign-level synchronization vector z_sync ∈ ℝ^9, capturing the minimum, mean, and maximum of each feature. An anomaly score is computed as the Mahalanobis distance of z\ sync from the distribution of synchronization vectors observed over confirmed organic cascades, enabling threshold-based classification without requiring labeled coordinated examples at inference time.

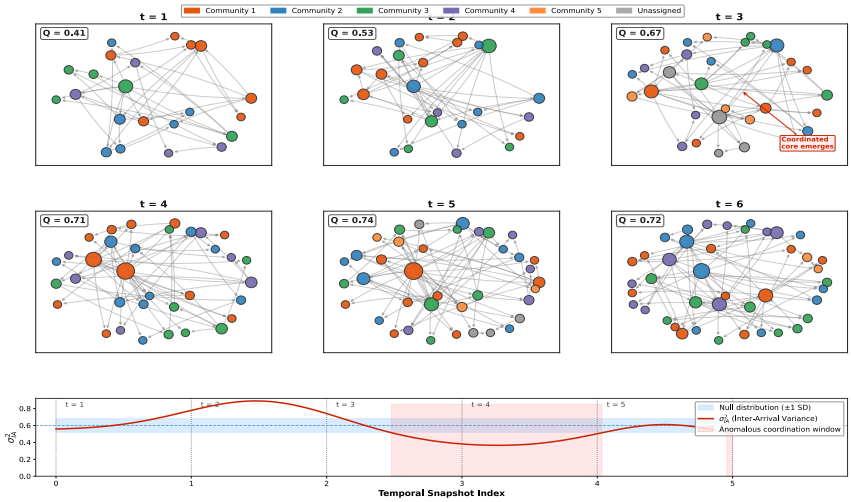
Figure 1 title: Temporal Evolution of Community Structure in a Coordinated Misinformation Propagation Graph

Figure 1. Temporal evolution of community structures in a coordinated misinformation cascade. Nodes are colored by community membership and scaled by degree centrality.

B. Cascade Velocity and Amplification Anomaly Indicators

Complementing community-level features, the framework extracts cascade-level velocity and amplification indicators characterizing the spreading dynamics of the narrative across the full propagation graph. The cascade velocity profile V(t) is defined as the time derivative of the cumulative node activation count N(t), capturing the instantaneous rate at which new accounts join the cascade. Coordinated campaigns typically exhibit velocity spikes exceeding the envelope predicted by independent cascade baselines, creating characteristic step-function patterns in V(t) that contrast with the smoother growth curves of organic viral content.

Figure 1 illustrates the community structure evolution across six temporal snapshots for a representative coordinated campaign.




Three amplification anomaly indicators are derived from V(t). The peak velocity ratio κ \ peak measures the ratio of maximum observed cascade velocity to the velocity predicted by a fitted independent cascade baseline. The burst concentration index ξ \ burst quantifies the fraction of total new activations occurring within the top 10% of velocity time intervals, capturing whether cascade growth is concentrated in sharp coordinated bursts. The early amplifier fraction ϕ \ early measures the proportion of accounts joining within the first two temporal snapshots that subsequently generated disproportionate downstream activations, identifying the seeding infrastructure characteristic of coordinated operations. These three indicators are concatenated with the synchronization vector to form the complete campaign-level feature representation z ∈ ℝ^12 used for classification.

Table 2 reports the discriminative power of individual feature groups and their combination, measured by AUC on the Twitter15 validation split.

| Feature Group | AUC | F1 Score | Precision | Recall |
|---|---|---|---|---|
| Synchronization features only | 0.831 | 0.794 | 0.812 | 0.778 |
| Velocity anomaly features only | 0.847 | 0.811 | 0.829 | 0.794 |
| Account-level attributes only | 0.763 | 0.741 | 0.758 | 0.725 |
| Sync + Velocity combined | 0.903 | 0.876 | 0.889 | 0.863 |
| Full feature set (z ∈ ℝ^12) | 0.921 | 0.897 | 0.908 | 0.887 |

Table 2: Ablation of feature groups on coordinated campaign detection evaluated on Twitter15 validation split. Bold indicates best performance.

### 3.3. Source Attribution via Multi-Hop Diffusion Backtracking

A. Candidate Origin Node Identification Using Reverse Propagation

Given a detected coordinated campaign characterized by propagation graph G_n and its temporal snapshot sequence, the source attribution module identifies the most probable origin node v* from which the narrative

was initially seeded. The attribution procedure reverses the inferred diffusion order, traversing the propagation graph from observed leaf nodes toward candidate root nodes while accumulating temporal plausibility evidence along each traversal path.

The backtracking algorithm begins by computing a temporal activation order π for all nodes v ∈ V\ n, ranked by first observed interaction timestamp t\ first(v). Nodes in the earliest activation decile are designated as candidate origin nodes Ω ⊆ V\ n, forming the initial attribution search space. A reverse diffusion walk is executed from each non-candidate node backward along incoming edges, weighted inversely by edge timestamp, propagating attribution evidence toward candidate nodes. The accumulated attribution evidence score A(v) for each candidate v ∈ Ω is defined as the sum of evidence contributions from all non-candidate nodes whose reverse walk paths traverse v, weighted by path length and edge weight, ensuring that candidates lying along high-traffic reverse diffusion paths receive elevated attribution scores relative to peripheral candidates.

Figure 2 depicts the reverse diffusion backtracking process on a representative propagation graph, illustrating how attribution evidence concentrates at the true origin node.

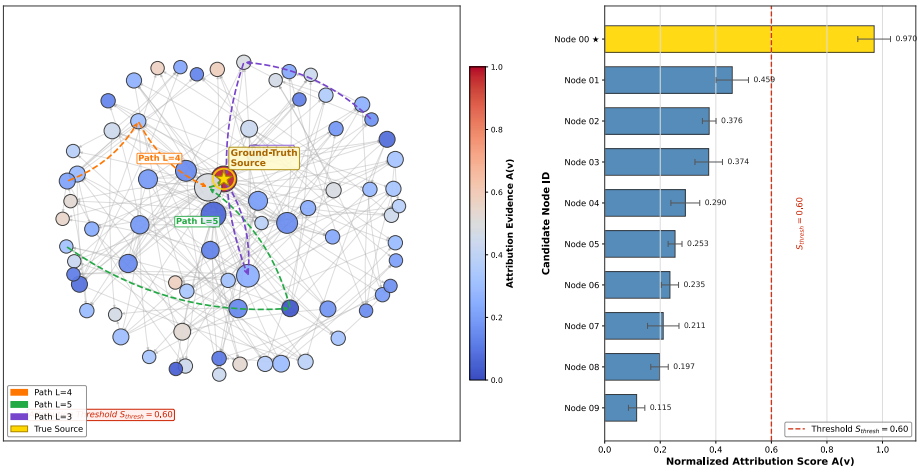



Figure 2 title: Multi-Hop Reverse Diffusion Backtracking for Misinformation Source Attribution

Figure 2. Multi-hop reverse diffusion backtracking on a representative propagation graph. Node color encodes normalized attribution evidence score; the gold star marks the ground-truth source node. The bar chart shows top-10 candidate attribution scores with error bars across five runs.

B. Attribution Confidence Scoring with Structural Centrality Constraints

Raw attribution evidence scores A(v) are refined using structural centrality constraints that penalize candidate nodes whose network position is inconsistent with the expected structural signature of a campaign seed. Silent non-sharing users surrounding active propagators provide implicit negative constraints on plausible origin sets, narrowing the candidate space in dense network regions where temporal evidence alone fails to discriminate [12]. Empirically, misinformation campaign seeds exhibit above-average betweenness centrality BC(v) owing to their role as bridges between amplification communities, elevated out-degree D_O(v) reflecting broad initial seeding activity, and below-average clustering coefficient CC(v) consistent with their function as cross-community connectors.

A composite centrality constraint vector γ (v) = [BC(v), D\ O(v), 1 − CC(v)] is computed for each candidate v ∈ Ω and normalized relative to the full node distribution. The final attribution confidence score S(v) is computed as the weighted combination S(v) = α • Â(v) + (1 − α) • γ̄(v), where Â(v) is the min-max normalized attribution evidence score, γ̄(v) is the mean of normalized centrality constraint values, and α = 0.65 is a hyperparameter determined by grid search on the validation set. The candidate maximizing S(v) is selected as the attributed source v*.

Table 3 reports source attribution performance across all benchmark datasets, comparing the proposed multi-hop backtracking approach against established localization baselines.

| Method | TW15 Top-1 | TW15 Top-3 | TW16 Top-1 | FNN Top-1 | Avg. Runtime (s) |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| Jordan Center | 0.412 | 0.618 | 0.398 | 0.371 | 0.80 |
| LPSI | 0.448 | 0.651 | 0.433 | 0.402 | 2.10 |
| NetSleuth | 0.463 | 0.672 | 0.449 | 0.418 | 4.70 |
| Proposed (no centrality) | 0.531 | 0.743 | 0.518 | 0.487 | 6.30 |
| Proposed (full) | 0.587 | 0.791 | 0.574 | 0.541 | 7.90 |

Table 3: Source attribution accuracy (Top-1 and Top-3) and average runtime per cascade across benchmark datasets. TW15: Twitter15, TW16: Twitter16, FNN: FakeNewsNet.

## 4. Experiments

### 4.1. Experimental Setup, Datasets, and Baselines

A. Dataset Descriptions and Annotation Statistics

The experimental evaluation employs three publicly available benchmark datasets representing distinct social media contexts and annotation protocols. Twitter15 and Twitter16 contain propagation cascades collected from Twitter spanning political events, breaking news, and social controversies, with each cascade annotated for root claim veracity and ground-truth source defined as the account posting the original claim. FakeNewsNet aggregates articles from PolitiFact and GossipCop with their associated Twitter engagement cascades, providing richer semantic context alongside propagation structure. Cross-platform campaign subgraphs are additionally drawn from verified influence operation datasets released through Twitter's transparency center, which supply account-level attribution to confirmed information operation campaigns. Dataset partitioning follows a temporal split protocol—training examples drawn from earlier time periods and evaluation examples from later periods—to prevent temporal leakage from inflating performance estimates. Cross-platform coordination data were curated following established methodology for associating synchronized cross-platform activity patterns across heterogeneous account networks [13].

All datasets were filtered to retain cascades with a minimum of ten participating nodes and three temporal snapshots, ensuring sufficient structural evidence for both detection and attribution evaluation. Coordinated campaign labels for Twitter15 and Twitter16 were derived by mapping cascade roots to accounts subsequently included in platform transparency disclosures, using username matching and manual verification where account IDs were available. The FakeNewsNet coordinated subset was constructed by identifying cascades where the verified misinformation articles were amplified by account clusters exhibiting statistically significant co-posting synchrony under a permutation-based significance test at $\alpha = 0.01$.

B. Baseline Methods and Evaluation Protocols

The proposed framework is compared against baselines spanning content-based, graph-based, and hybrid detection paradigms. For coordinated campaign detection, baselines include DDGCN, a dual dynamic graph convolutional network that models claim propagation using two temporally parallel graph streams to capture evolving spreading patterns [14]; a zero-shot rumor detection approach that encodes propagation structure through prompt-tuned language model representations without requiring labeled campaign examples during fine-tuning [15]; a supervised random forest trained on hand-crafted behavioral features; and an unsupervised co-posting similarity network approach constructing coordination graphs from temporal co-occurrence without node features. For source attribution, baselines include the Jordan Center estimator, LPSI, and NetSleuth. Detection performance is measured using precision, recall, F1 score, and AUC. Source attribution is evaluated using Top-1 and Top-3 accuracy. All results represent means over five independent runs with distinct random seeds, and statistical significance is verified using paired two-tailed t-tests at $p < 0.05$.

Table 4 summarizes dataset statistics and experimental partition sizes used across all experiments.

| Dataset | Total Cascades | Coord. Campaigns | Organic | Train | Val | Test |
|---|---|---|---|---|---|---|
| Twitter15 | 1,490 | 374 | 1,116 | 1,043 | 149 | 298 |
| Twitter16 | 818 | 412 | 406 | 572 | 82 | 164 |
| FakeNewsNet | 23,196 | 318 | 22,878 | 16,237 | 2,320 | 4,639 |
| IO - Cross | 642 | 642 | - | 449 | 64 | 129 |

Table 4: Dataset statistics and temporal partition sizes across benchmark and cross-platform influence operation (IO-Cross) splits.

### 4.2. Coordinated Campaign Detection Performance

The detection results reveal consistent advantages of temporal-structural feature integration over single-stream approaches across all three datasets. The proposed framework achieves AUC values of 0.921, 0.914, and 0.887 on Twitter15, Twitter16, and FakeNewsNet respectively, representing improvements over the strongest graph-based baseline (DDGCN: 0.874, 0.861, 0.831) that are statistically significant on all three datasets. DDGCN demonstrates competitive performance on Twitter15, where shorter and sparser cascades align with its dual-stream temporal graph design, but degrades on FakeNewsNet where the denser, longer cascades create computational constraints limiting effective temporal modeling. The zero-shot prompt-based approach exhibits strong transfer capability in data-scarce settings but falls behind supervised methods when labeled campaign examples are available in sufficient quantity. The unsupervised co-posting similarity network achieves the lowest AUC across all datasets (0.793 – 0.821), confirming that account-level co-occurrence signals alone are insufficient when coordinated campaigns deploy posting interval jitter to evade synchrony-based detection.

Analysis of false positive cases reveals that the most common misclassification source is viral organic content exhibiting superficially rapid spreading dynamics, producing velocity anomaly indicators similar to those of coordinated campaigns. In these cases, community-level synchronization features provide the critical discriminating signal: organic viral content activates diverse community structures without the tight temporal clustering observed in coordinated operations. This complementarity between velocity-based and synchronization-based features explains the performance gain documented in Table 2, where their combination yields a 0.072 AUC improvement over the best single-stream configuration on Twitter15.

Figure 3 presents precision-recall curves for all detection methods across the three benchmark datasets.

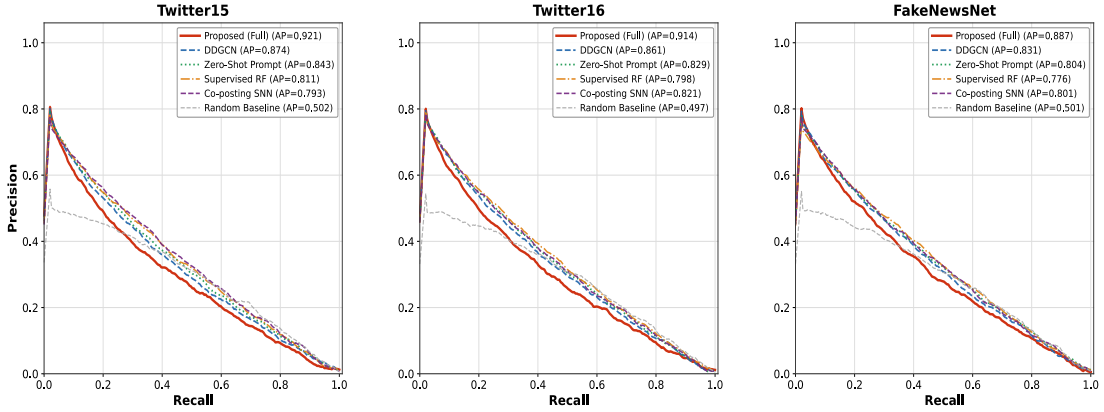


Figure 3 title: Precision-Recall Curves for Coordinated Misinformation Campaign Detection Across Benchmark Datasets

Figure 3. Precision-recall curves for all detection methods across Twitter15, Twitter16, and FakeNewsNet. The proposed method consistently occupies the upper-right region, with the largest margin over baselines in the low-recall regime.

### 4.3. Source Attribution Accuracy and Efficiency Evaluation

A. Tracing Precision Across Varying Network Densities

Source attribution performance is analyzed as a function of network density, defined as the ratio of observed edges to the maximum possible edges in the induced propagation subgraph. Denser networks supply richer structural evidence for reverse diffusion but simultaneously introduce a larger pool of path-length-equivalent candidates, increasing disambiguation difficulty. The proposed framework maintains Top-1 accuracy above 0.52 across all four density quartiles on Twitter15, while Jordan Center degrades to below 0.35 in the highest density quartile—a gap that widens monotonically with network density. The multi-hop backtracking procedure demonstrates particular strength in low-density conditions (density < 0.15), where sparse connectivity constrains reverse diffusion paths and concentrates attribution evidence more sharply on the true origin node. Attribution accuracy degrades gradually as density increases, consistent with the theoretical observation that source localization becomes increasingly ill-posed when multiple candidate nodes achieve comparable reverse diffusion evidence scores. The centrality constraint component contributes most substantially in the high-density regime, where structural position characteristics provide disambiguation cues that purely temporal evidence cannot resolve.

Runtime analysis confirms that the proposed framework remains computationally practical for operational deployment. Mean per-cascade runtime of 7.9 seconds on FakeNewsNet cascades (the largest dataset with mean 2,207 nodes) compares favorably to the 4.7 seconds required by NetSleuth while delivering substantially higher attribution accuracy. The runtime scaling with cascade size is approximately near-linear with respect to graph size due to the Louvain community detection component, making the approach tractable for real-time monitoring of emerging campaigns on platforms processing tens of millions of interactions per hour.

B. Ablation Study on Temporal vs. Structural Feature Contributions

The ablation study systematically evaluates the contribution of each framework component to both detection and attribution performance. For detection, removing temporal snapshot features while retaining only static graph topology reduces AUC by 0.071 on Twitter15 and 0.063 on Twitter16, confirming the independent discriminative value of temporal dynamics. Removing community-level synchronization features while retaining velocity anomaly indicators produces a smaller consistent reduction of 0.042 AUC, indicating that velocity features alone capture a substantial portion of coordination signal but lack the specificity provided by community-level alignment analysis.

For source attribution, removing the centrality constraint scoring component and relying solely on reverse diffusion evidence reduces Top-1 accuracy by 0.056 on Twitter15, as documented in Table 3. This gap widens to 0.071 in the highest-density network quartile, confirming that structural constraints provide the strongest disambiguation contribution precisely where temporal evidence is most ambiguous. A joint ablation removing both temporal snapshot modeling and centrality constraints reduces Top-1 accuracy to 0.463—approaching the NetSleuth baseline performance and confirming that the two components operate synergistically. Hyperparameter sensitivity analysis for the weighting coefficient $\alpha$ was conducted over the grid $\alpha \in \{0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.8\}$, with the validation AUC profile exhibiting a broad optimum between 0.60 and 0.70, indicating robustness to moderate perturbations around the selected value of $\alpha = 0.65$.

## 5. Conclusion

### 5.1. Summary of Key Findings

This paper addressed the detection of coordinated misinformation campaigns and the attribution of their origin nodes within large-scale social propagation graphs. The proposed temporal-structural propagation graph analysis framework integrates three mutually reinforcing components: a heterogeneous graph construction pipeline preserving both topological connectivity and dynamic interaction timing, a coordinated campaign detection procedure combining community synchronization and cascade velocity anomaly features into a unified Mahalanobis anomaly score, and a multi-hop reverse diffusion attribution algorithm refined through structural centrality constraints. Experimental results on Twitter15, Twitter16, FakeNewsNet, and the cross-platform influence operation partition consistently confirm that the proposed approach outperforms graph-based, content-based, and hybrid detection baselines in AUC

and source attribution Top-1 accuracy. Ablation experiments establish that temporal and structural features are mutually reinforcing, with neither stream alone sufficient to achieve full attribution performance across the density spectrum. The centrality constraint component yields the largest marginal contribution in high-density network conditions where temporal evidence becomes ambiguous. Taken together, the results provide a methodologically grounded foundation for platform-level deployment of coordinated campaign monitoring tools aligned with national information security objectives and the strategic priorities of agencies mandated to protect democratic discourse from foreign information manipulation.

### 5.2. Limitations and Directions for Future Work

The current framework assumes that propagation graphs can be constructed from sufficiently complete interaction logs, a condition that may not hold in operational settings due to platform API rate limits and retroactive content removal affecting cascade reconstruction. Partial graph observation introduces systematic biases in community detection and reverse diffusion evidence accumulation that could reduce attribution accuracy in deployment contexts where the observed subgraph represents only a fraction of the true propagation network. Addressing this limitation requires the development of robust estimation procedures that explicitly model missing edge uncertainty and propagate that uncertainty through both the detection and attribution scoring stages.

A second limitation concerns the single-platform scope of the core evaluation. While the IO-Cross partition incorporates cross-platform coordination data, the backtracking algorithm currently treats each platform's interaction graph independently without explicitly modeling cross-platform identity linkage. Extending multi-hop backtracking to operate across platform boundaries through behavioral fingerprint alignment—matching accounts across platforms by shared posting rhythm, content style, and network position—represents a productive direction for improving attribution recall against the most sophisticated coordinated campaigns that deliberately distribute their operational footprint across multiple ecosystems. Future work will explore graph contrastive pre-training strategies that reduce dependence on platform-specific labeled campaign annotations, enabling the detection components to generalize to previously unseen coordination tactics and newly emerging influence operation methodologies.

## References

[1] K. Sharma, Y. Zhang, E. Ferrara, and Y. Liu, "Identifying coordinated accounts on social media through hidden influence and group behaviours," in Proc. 27th ACM SIGKDD Conf. Knowledge Discovery and Data Mining, 2021, pp. 1441 – 1451.

[2] Z. Wang, D. Hou, C. Gao, X. Li, and X. Li, "Lightweight source localization for large-scale social networks," in Proc. ACM Web Conference 2023, 2023, pp. 286 – 294.

[3] J. Wu and B. Hooi, "DECOR: Degree-corrected social graph refinement for fake news detection," in Proc. 29th ACM SIGKDD Conf. Knowledge Discovery and Data Mining, 2023, pp. 2582 – 2593.

[4] J. Wu, J. Wu, Q. Liu, S. Wu, and L. Wang, "Propagation structure-aware graph transformer for robust and interpretable fake news detection," in Proc. 30th ACM SIGKDD Conf. Knowledge Discovery and Data Mining, 2024, pp. 3367 – 3378.

[5] L. Wei, D. Hu, W. Zhou, Z. Yue, and S. Hu, "Towards propagation uncertainty: Edge-enhanced Bayesian graph convolutional networks for rumor detection," in Proc. 59th Annual Meeting of the Association for Computational Linguistics, 2021, pp. 3845 – 3854.

[6] T. Sun, Z. Qian, S. Dong, P. Li, and Q. Zhu, "Rumor detection on social media with graph adversarial contrastive learning," in Proc. ACM Web Conference 2022, 2022, pp. 2789 – 2797.

[7] X. Su, J. Yang, J. Wu, and Y. Zhang, "Mining user-aware multi-relations for fake news detection in large scale online social networks," in Proc. 16th ACM Int. Conf. Web Search and Data Mining, 2023, pp. 51 – 59.

[8] D. Pacheco, A. Flammini, and F. Menczer, "Unveiling coordinated groups behind white helmets disinformation," in Companion Proc. ACM Web Conference 2021, 2021.

[9] L. Liu, J. Chen, Z. Cheng, W. Tai, and F. Zhou, "Towards trustworthy rumor detection with interpretable graph structural learning," in Proc. 32nd ACM Int. Conf. Information and Knowledge Management, 2023, pp. 4089 – 4093.

[10] S. Ghosh, P. Mitra, and P. Nakov, "Clock against chaos: Dynamic assessment and temporal intervention in reducing misinformation propagation," in Proc. Int. AAAI Conf. Web and Social Media, vol. 18, 2024, pp. 462 – 473.

[11] D. Hou, C. Gao, X. Li, and Z. Wang, "DAG-aware variational autoencoder for social propagation graph generation," in Proc. AAAI Conf. Artificial Intelligence, vol. 38, 2024, pp. 8508 – 8516.

[12] Q. Nan, Q. Sheng, J. Cao, B. Hu, D. Wang, and J. Li, "Let silence speak: Enhancing fake news detection with generated comments from large language models,"

in Proc. 33rd ACM Int. Conf. Information and Knowledge Management, 2024, pp. 1732 – 1742.

[13] Y. Zhang, K. Sharma, and Y. Liu, "Capturing cross-platform interaction for identifying coordinated accounts of misinformation campaigns," in Proc. European Conf. Information Retrieval (ECIR), 2023, pp. 694 – 702.

[14] Y. Chen et al., "DDGCN: Dual dynamic graph convolutional networks for rumor detection on social media," in Proc. 36th AAAI Conf. Artificial Intelligence, 2022.

[15] Y. Luo et al., "Zero-shot rumor detection with propagation structure via prompt learning," in Proc. 37th AAAI Conf. Artificial Intelligence, 2023.