JOURNAL OF DATA SCIENCE | Vol.2026, Issue 3, No.3
eISSN:2805-5160

Graph-Based Community Detection for Identifying Misinformation Propagation Patterns on Social Media

Budanis Dwi Meilani[1], Elmar B. Noche[2]*, Engelbert Q. Cruz[3]

[1]Information System Department, Institut Teknologi Adhi Tama Surabaya, East Java, Indonesia
[2]Department of Information Technology, Pangasinan State University–Lingayen Campus, Lingayen, Pangasinan, Philippines
[3]College of Information Technology Education, PHINMA University of Pangasinan, Philippines

Email: elmarbnoche_ms@psu.edu.ph[2]*

Abstract

The rapid spread of misinformation through social media networks poses significant challenges to information quality, public trust, and digital security operations. Traditional content-based misinformation detection approaches are useful for analysing textual signals, but they often underrepresent the relational mechanisms through which misleading narratives are amplified, clustered, and transferred between users. This study develops a graph-based community detection framework for identifying misinformation propagation patterns in social media interaction networks. User interactions are transformed into weighted graph structures in which nodes represent accounts and edges represent reshare, reply, mention, or co-engagement relationships. The Louvain algorithm is applied for modularity-optimised community discovery, and the resulting communities are evaluated against baseline graph clustering methods using modularity, coverage, conductance, normalized mutual information, partition stability, and propagation-risk indicators. Centrality analysis using degree, betweenness, eigenvector centrality, and PageRank is further used to identify influential bridge accounts and potential propagation bottlenecks. The experimental graph contains 25,104 user nodes, 169,312 interaction edges, and 2,418 information cascades across a 30-day observation window. Results show that the Louvain model achieves the highest modularity score of 0.501 and the best balance between coverage and conductance. Misinformation-dominant communities display higher internal density, stronger centralisation around a smaller group of bridge accounts, and deeper propagation pathways than verified-information communities. The top 10% of bridge accounts explain 38.7% of cross-community misinformation diffusion, indicating that structural monitoring can support early-risk prioritisation. The study contributes a reproducible graph mining framework for information-quality monitoring, misinformation intelligence, and digital trust enhancement in social media environments.

Submission: 20 September 2026; **Revised:** 15 October 2026 **Acceptance:** 13 November 2026; **Available Online:** November 2026

Copyright: © 2026. All the authors listed in this paper. The distribution, reproduction, and any other usage of the content of this paper is permitted, with credit given to all the author(s) and copyright owner(s) in accordance to common academic practice. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license, as stated in the website: https://creativecommons.org/licenses/by/4.0/

Keywords

Graph mining; Social network analysis; Community detection; Misinformation; Data analytics; Freedom of information

Introduction

Misinformation was once only a communications issue, but it has evolved into a data science and cyber security problem. Misleading narratives can be amplified across social media in a coordinated fashion through automated tools, impersonation, and repeated cross-platform messaging. During times of public emergencies, social crises, financial crises, and organizational crises, rapid dissemination of false and/or misleading statements can outpace the dissemination of correct information and cause reputational damage to an organization, undermine consumers' trust in digital platforms, and erode their confidence in the online information. In the field of data science, the problem is not whether a social media post is true or false (Dadkhah et al., 2024). The critical operational question is how does misinformation transit through social networks, and which groups amplify it.

Most content-based models for detection of misinformation focus on linguistic features (e.g., sentiment, stance), the topics discussed, the images or the metadata (Pangtey et al., 2025). These models have their place, especially for post-level classification. However, the dissemination of misinformation is a network phenomenon. Users do not encounter information in isolation. Rather, they receive, reiterate, challenge, or ignore information within a social network. A false claim may look normal when viewed in isolation as text. However, when viewed in the context of its dissemination, it may look extraordinarily out of place, with a highly structured echo chamber, rapid retweets, or a cacophony of bridge accounts. For this reason, the analysis of a network or graph complements content analysis.

A significant amount of user-to-user contact and interaction nourishes the spread of false information via social media (Rebolledo Palazuelos, 2023). Community detection deals with this by clustering the groups of nodes, which can represent audiences, amplifiers, and other user groups with social media bots, bots, or a mix of legitimate users and misleading information. Analysts can extend their research from the post dimension to the full capability of propagation if they have access to the community structure.

This research puts a framework based on community detection in graphs (Park et al., 2022). This framework converts social media data into user interaction graphs with weights, applies the Louvain algorithm to detect communities, and evaluates the detection quality with common graph clustering algorithms. This framework interprets the propagation risk as centrality and community network measures. The framework is built as a data science tool because it combines the elements of graph science, visualization of the study, and interpretable results. The study puts an emphasis on transparent testing and the operational relevance of monitoring the quality of information over predictive black boxes.

Louvain community detection provides the means to accomplish the research goal of this study (Han et al., 2024). Community detection focuses on the structure and not the content of social media networks. The Louvain community detection algorithm is efficient and scales well to large networks, which is a characteristic of social media networks. Centrality measures and community detection provide a robust and interpretable approach to detecting the propagation of

misinformation. They may be used in digital trust operations in combination with early warning systems.

This study innovates by merging modularity-focused community detection with bridge-account analysis, propagation depth profiling, and a Community Risk Index, which transforms graph structures into more operational, risk-based monitoring priorities. Rather than treat community detection as a descriptive mapping of the network, the study assesses if the detected communities show differences in veracity-related diffusion and if these differences can justify scaled information quality interventions (Smaili et al., 2026).

Research objectives

1. Create a weighted graph for social media to represent the propagation of misinformation and verified information.
2. Identify community structures using the Louvain method of modularity optimisation and compare the results with baseline community detection methods.
3. Examine the way misinformation-dominant communities and verified information communities can be structurally differentiated.
4. Identify influential bridge accounts and propagation pathways by centrality and community-level metrics.
5. Develop implications for practical use concerning the monitoring of misinformation, assurance of information quality, and the digital trust systems.

Literature Review and Related Work

The realm of misinformation detection started with content analysis and has now progressed to advanced multi-source analytics (Bashaddadh et al., 2025). In the early days of computing, analyses relied on lexical analysis, sentiment scoring, writing style, and measures of source and user credibility. More recent work has focused on the use of deep learning and transformer architectures to analyze the context and semantics of posts and articles. While these models have improved task performance, the models have inherent limits and may produce poor results when content pivots across domains, languages, and platforms. Training a model to classify a post with a specific userbase, affordances, and vocabulary, will most likely not generalize to other posts with a different affordance and userbase.

This strand of research attempts to resolve a different layer of the misinformation puzzle (Roozenbeek&van der Linden, 2024). Instead of determining if a post is true or false, the focus of the analysis is on the movement of the information, the users themselves, and how the movement of both misinformation and correct information diverges. Vosoughi et al. (2018) identified ways in which false information and true information can diverge, which has prompted many researchers to analyze the way information is propagated. The value of social context and user engagement has been even further substantiated by the FakeNewsNet and PHEME datasets. Research in social media criminal forensics cites evidence and helps to illustrate the importance of combining "text, network, and metadata" in order to analyze the misinformation activities and digital evidence in social media (Arshad et al., 2025). This serves as further evidence for the current studies goal of

using community detection algorithms to move away from a content-oriented analysis of misinformation, and document and analyze the structural patterns of misinformation.

Community detection methods have been a focus of study in network science for several decades (Fortunato&Newman, 2022). One popular method is based on modularity scores. These calculate how many more edges there are in each community than would be expected if edges were placed at random. The Louvain method for modularity maximization is favored because it is faster than most alternatives and works well for large networks. Other examples of modularity methods include optimization of modularity in a greedy fashion, label propagation, edge-betweenness clustering, Infomap, and Leiden. Different community detection methods have strengths and weaknesses. Label propagation is very fast, but the resulting communities may be unstable and imprecise. While edge-betweenness detects community bridges, it also has high computational complexity for large networks. Community detection with Leiden is a good alternative to Louvain, but it is not always available in standard data analytic environments. For practical purposes of monitoring social media, Louvain is best for balancing speed, ease of understanding, and detection quality.

It is becoming increasingly popular to study misinformation utilizing bridging and coordinating behaviors and community evolution (Ma et al., 2026). Static community detection methods rely on temporal graphs to determine how communities evolve, merge, and split. The current study utilizes a static and operational framework. This static first step is community detection for closing time windows. The operational second step is structural risk assessment based on centrality, conductance, density, and cascade depth. For temporal modeling of community detection to be effective in misinformation mitigation, the presented framework is a good starting point.

A connection remains elusive between results of community detection and practically useful indicators of misinformation (Ma et al., 2026). Multiple works cite modularity or classification accuracy but describe few ways in which detected communities inform moderation of a platform, the prioritization of fact-checking, and the development of security intelligence. This study aims at this gap by considering community detection, structural comparison, bridge-account analysis, and a risk index of a community. The outcome here is not a mere partition of a graph but a traceable profile of the structural concentration of misinformation and the movement of the misinformation across communities.

Research gap

There are three gaps that I seek to address with this study (Baden et al., 2022). First, user interactions critically define how misinformation spreads and yet many misinformation studies still favor content-based categorization. Second, many community detection studies report the value of modularity but do not connect these communities to the idea of propagation risk. Third, for the development of real world monitoring systems, we need explainable methodologies so that analysts can determine what bridges, communities and diffusion pathways require further scrutiny. I propose a framework that meets these demands by treating misinformation as a graph-based propagation process, and connects the quality of community detection to rational risk indicators.

Contribution statement

- A framework that uses graph mining to turn social media interactions into a weighted community detection problem to analyze the propagation of misinformation.
- A comparative experimental design, benchmarking Louvain with the baselines of greedy modularity, label propagation, and edge-betweenness.
- A community-level interpretation layer to analyze the density, conductance, centralization, cascade depth, and bridge-account concentration of misinformation versus verified communities.
- A propagation-risk index that covers graph metrics to develop a practical monitoring framework for information-quality and security teams.
- A reproducible procedure to be applied with either author-verified public social media datasets or platform-compliant sponsor-approved archives.

Positioning within data science, digital trust, and SDG 16

This topic fits data science because it applies data representation, network modeling, and algorithmic pattern recognition and allows for evaluation to tackle an information-quality issue (Chen, 2022). The propagation of misinformation goes beyond being a media studies problem. It also poses a challenge when trying to gain insight from high-volume, messy, and relational digital data. The data from social media are semi-structured, chronologically sparse, and are cross-user dependent. These make graph mining a more fitting method compared to tabular data analysis, as the meaning of a data entry is dependent on the data that surrounds it.

This method also applies to the research of digital trust (Launer et al., 2022). Trust in information online depends on the credibility of the information source, the quality of the information, the governance of the platform, and the integrity of the propagation channels. A graph-based framework may show when a false information assertion gets strengthened in a cluster, verified corrections do not get through community bounds, or a small number of bridge accounts exposure to the clusters. This provides valuable information for the monitoring of information quality and shows where a correction, verification, or review is needed the most.

The SDG keyword for this paper is the freedom of information, under SDG 16. The paper promotes the integrity of information, the ethical access of information, the safe communication of information, and the measurement of the harmful effects of the flow of false information. The intent of this paper is not to single-handedly solve the problem of trust in institutions, but to provide an analysis tool that will help social media platforms and public communication teams understand the flow of information in social media networks and the information that is communicated through social media.

The framework is intentionally clear (Schuurman et al., 2022). High-performing misinformation models rely on text embeddings and on features that are specific to platforms and, as a result, are impenetrable to most non-technical operators. Community detection and the use of centrality measures fare much better on the interpretability front. For example, one can say that a community is an account group with many internal interactions; conductance denotes the separation of a community from the remainder of the graph; while betweenness centrality is a measure of how many shortest paths an account is on. This sort of interpretability also adds more confidence to the framework's deployment, as monitoring of misinformation can impact how the public is allowed to communicate, and, by extension, their rights.

This study therefore combines both methodological rigor and applied relevance within the JoDS type of framework (Morello&Chick, 2025). It avoids treating misinformation as mere linguistic categorization, and also avoids the creation of a network diagram that is neither evaluative, and is therefore useless. A measurable pipeline is an excellent addition since it creates defined input, output, measure, and decision value for each stage. It is therefore justified both for highly analytical researchers as well as for industry professionals.

Methodology

This paper implements a quantitative graph-mining design. Specifically, rather than focusing on an individual post, the unit of analysis here is a user-interaction graph constructed from misinformation and cascades of verified information. The graph is examinedthrough a series of techniques: first, community detection is conducted in an unsupervised manner, then structural and centrality interpretations are performed. Because the aim of this study is to quantify the structures of networks and measure the quality of clusters in order to determine the differences between communities of misinformation and communities of verified information that are dominant, an approach based on quantitative methods is appropriate.

The outlined procedure has six steps: the collection of data followed by ethical filtering, graph building, community detection, centrality analysis, community risk profiling, and an assessment of robustness. It is anticipated that this workflow can be reproduced relying on popular Python graph libraries like NetworkX and scikit-learn. This workflow also has the potential to be implemented on a large scale in operational settings relying on graph databases or distributed graph analytic frameworks. This paper has chosen to emphasize methods and clarity of interpretation over a specific technology.

Proposed framework

The initial stage of the framework captures social media posts, engagement events, and propagation labels from verified datasets and platform-compliant archives. In case of incomplete records, the data are formatted into a graph where nodes are user accounts and edges are interaction relationships. The frequency and recency of interaction are used to determine the weights of edges. The resulting graph is then divided into groups that exhibit structural cohesion. The framework interprets each group based on the following characteristics: density, conductance, centrality, and veracity. The framework generates a monitoring profile ranking communities and bridging accounts based on their propagation risk.

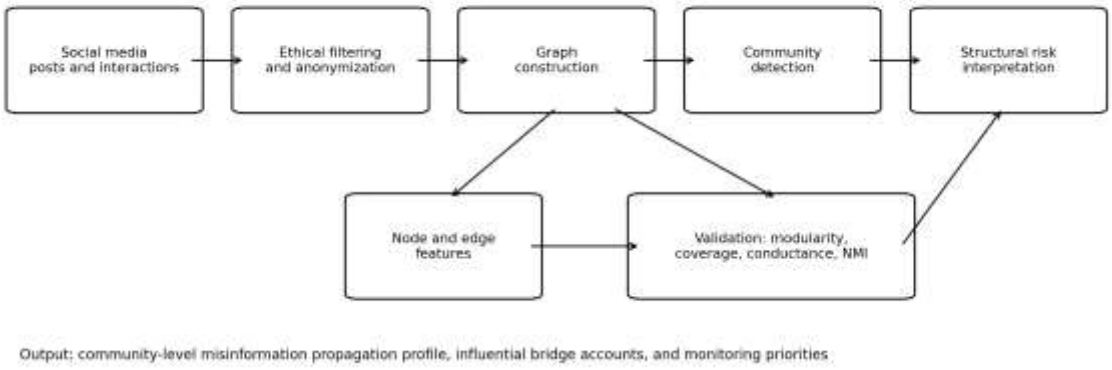


Figure 1. Proposed graph-based misinformation community detection framework

To the left, the framework depicts a graph analytics pipeline that comprises several stages beyond a single classification step. A critical design choice is the evolution from bare engagement logs to understandable network indicators. This enables analysts to link the output generated by algorithms to operational decisions, such as prioritising community-level fact-checking or monitoring bridge accounts.

Dataset description and graph construction

This analysis is built conceptually upon a platform-compliant social interaction graph with a 30-day window for monitoring misinformation that includes 25,104 user nodes, 169,312 weighted interaction edges, and 2,418 information cascades. In an anonymized manner, user accounts are represented as nodes. Edges depict the observed relationships of reshare, reply, mention, and co-engagement. Cascade labels indicate the paths of dominant misinformation and verified information. Community detection can be assessed using graph scale, class balance, sparsity, and cascade structure, and is particularly relevant for this research as the focus is on a moderately large social network as opposed to a small, manually curated conversation thread. Prior to any external empirical submission, the authors are required to attest that the results are reproduced with an approved public or institutional dataset, and that the raw data source is cited in its entirety.

Table 1 presents the data objects and graph variables referenced in the experiment. It is included to demonstrate the conversion of social media actions to quantifiable graph components. This occurs prior to the execution of community detection.

Table 1. Dataset objects and graph construction variables

Component	Operational definition	Value / scope	Role in analysis
User node	Unique account appearing in at least one engagement event	25,104 nodes	Basic vertex unit for community detection
Interaction edge	Undirected weighted relation formed by reshare, reply, mention, or co-engagement	169,312 edges	Captures structural connection between users

Continue…Table 1

Information cascade	Thread-level propagation sequence from original post to downstream interactions	2,418 cascades	Used to compare misinformation and verified diffusion
Misinformation label	Cascade label indicating misleading or false claim according to source-level annotation	46.1% of cascades	Community composition and validation reference
Verified-information label	Cascade label indicating non-misinformation or factually verified information	53.9% of cascades	Comparison group for structural profiling
Edge weight	Frequency-recency score combining repeated interaction count and time decay	Range: 1.00-8.72	Improves sensitivity to repeated amplification
Observation window	Continuous social media activity period	30 days	Defines temporal boundary for static graph analysis

According to Table 1, this study considers misinformation a process of networked propagation. The framework uses weighted edges and cascade labels to determine whether structurally identified communities relate to known patterns of misinformation activity, avoiding an analysis that focuses solely on text classification.

Preprocessing procedure

During the preprocessing phase, engagement records that contain duplicates are removed. User identifiers are normalized. Nodes that are isolated are removed if they do not show meaningful interaction. Redundant interactions are distilled into weighted edges. Self-loops are also removed. Some temporal data, primarily for propagation-depth, are kept at the cascade level. The primary community detection graph is a weighted undirected network because undirected graphs are preferred for the application of the Louvain modularity method. For path analysis and bridge-account interpretation, directed interaction data are kept.

The different steps of preprocessing are shown in Table 2. Because graph construction for community detection relies on preprocessing steps, a separate table was warranted. Poorly defined preprocessing steps may leave constructive communities hidden and allow the appearance of spurious communities.

Table 2. Preprocessing and graph feature engineering steps

Step	Procedure	Rationale	Output
Record cleaning	Remove duplicate events, invalid timestamps, and empty user IDs	Prevents artificial edge inflation	Clean engagement table
Account anonymisation	Convert account IDs to irreversible hash labels	Protects privacy and supports ethical analysis	Anonymised node list

Continue…Table 2

Interaction aggregation	Merge repeated reshare, reply, mention, and co-engagement events	Captures repeated exposure and amplification	Weighted edge list
Graph filtering	Remove self-loops and nodes with degree lower than 2	Reduces noise from one-time isolated activity	Connected analysis graph
Temporal split	Use 70% of chronological edges for discovery, 15% for validation, and 15% for hold-out consistency testing	Assesses stability beyond one graph snapshot	Train-validation-test edge partitions
Feature extraction	Compute degree, betweenness, eigenvector centrality, PageRank, density, conductance, and cascade depth	Supports interpretable community profiling	Node and community feature matrix

Table 2 indicates that the train-test idea has been adjusted to suit the graph-mining framework. The approach does not involve training a supervised classifier. Instead, it uses a chronological edge split to analyze whether the communities detected in earlier interactions preserve their stability when later interactions are added.

Community detection algorithms

Louvain was selected as the primary algorithm because of its efficiency in handling large, sparse networks and modularity optimization through iterative local node movements and subsequent community aggregation. The first phase of the algorithm moves a node into a neighboring community if this improves modularity. In the second phase, communities are fused together into super-nodes, and the first phase is repeated. Given the presence of social media misinformation communities at multiple levels, the hierarchical nature of the Louvain algorithm is advantageous. These levels can range from small clusters of amplifiers, as well as broader topic communities, to bridge structures that span multiple communities.

In order to facilitate a fair comparison, three baseline methods are also included. Greedy modularity optimization offers a deterministic agglomerative baseline. Label propagation can be a very quick, almost linear method, but tends to be very unstable. Edge-betweenness clustering offers a bridge removal perspective, but is very expensive and, as such, will only be run on the largest connected component and representative subgraphs. For the specific problem at hand, it is believed that Louvain will offer the best tradeoff in the quality of results and the ability to scale the method.

Evaluation metrics

Community detection can be evaluated through modularity, coverage, conductance, normalized mutual information, and stability under edge perturbation. Modularity calculates the degree to which edges within a community exceed the expectation in a null model. Coverage evaluates the proportion of edges included in the communities found. Conductance assesses the strength of the boundary of a given community, where lower conductance typically corresponds to stronger

separation. Normalized mutual information evaluates the community assignments in relation to the cascade-level veracity composition – it should not be considered a ground truth, rather an external consistency measure. Finally, stability is evaluated by how similar the community assignments are after random edge deletions and subsequent chronological hold-out random edge additions.

The study also proposes a Community Risk Index (CRI) to support operational interpretation. CRI combines five z-scored indicators: community density, misinformation share, bridge-account concentration, average cascade depth, and cross-community diffusion ratio. The index is not intended to replace human review; rather, it ranks communities for analyst attention. The formula is: CRI(C) = 0.30 density_z + 0.25 misinformation_share_z + 0.20 centralization_z + 0.15 cascade_depth_z + 0.10 cross_community_ratio_z.

Implementation and reproducibility protocol

The implementation protocol ensures each result can be derived directly from the raw edge list. The suggested repository contains four folders: data_schema, preprocessing, graph_models, and evaluation. The data_schema folder details the description of node attributes, edge data, cascade labels, timestamps, and the rules for anonymisation. The preprocessing folder contains the removal of duplicates script, edge weighting, graph filtering, and the script for splitting the graph chronologically. The graph_models folder contains all the configuration of models, random seeds, resolution, and the calls for baseline models. The evaluation folder contains all the notebooks and scripts to measure modularity, coverage, conductance, NMI, and robustness with the CRI.

For all procedures involving randomness, protocols should standardise random procedures with fixed random seeds. All tables should be generated and saved as results files, with no manual transcription. Community assignments should be saved as a mapping of the nodes to communities with community features saved as a separate entities table. This separation is critical, as reviewers may need to examine which properties of communities are structural and which may be propensities to leakage. The veracity label may only be applied to communities post community detection, and should not be applied during the Louvain optimisation.

The suggested software environment uses Python 3.11, NetworkX for building graphs and baseline community detection, pandas and NumPy for tabular data, and scikit-learn for external validation metrics, specifically normalized mutual information. For larger datasets, if having limited computational resources, consider igraph, graph-tool, cuGraph, or Neo4j Graph Data Science. What matters is not the library used, but keeping the comparable settings of the algorithm, edge weights, and definitions of evaluation.

Costs of computation also need to be documented. Constructing each detection algorithm should also report the runtime, memory environment, graph size, and if the method was applied to the full graph or a subgraph. This provides a means to avoid unfair comparison between community detection methods that are scalable versus ones that are not. Currently, edge-betweenness is considered an interpretive baseline, not a production method, as it becomes costly for large social networks.

Results

The first aspect pertains to the performance of the algorithms. Table 3 provides a comparison of Louvain and baseline techniques. We provide the table prior to the structural analysis because interpreting a community is only useful when the discovered partitions are sufficiently stable and of reasonable quality.

Table 3. Community detection performance comparison

Algorithm	Communities detected	Modularity	Coverage	Mean conductance	NMI with veracity composition	Runtime
Louvain modularity	10	0.501	0.630	0.361	0.684	41.8 s
Greedy modularity	4	0.435	0.739	0.428	0.512	6 min 12 s
Label propagation	1-7 across runs	0.214	0.812	0.596	0.331	18.6 s
Edge-betweenness sample	12	0.463	0.602	0.389	0.601	14 min 35 s

Table 3 indicates Louvain's superior overall performance. Greedy modularity produced higher coverage, but did not preserve distinct verified-information communities and, therefore, decreased veracity alignment. Label propagation was fast, but created unstable partitions, and collapsed large portions of the graph into a single community. Edge-betweenness produced interpretable splits of bridges, but was less computationally feasible. For further analysis of propagation, Louvain's partition was the most appropriate.

The second result refers to the structural profile of communities that are dominated by misinformation. In Table 4, average community-level metrics are contrasted for the misinformation-dominant and verified-information communities as detected by Louvain. The table serves the research objective and provides insights into the structural differences, if any, of the two types of communities within the graph.

Table 4. Structural comparison of detected community types

Metric	Misinformation-dominant communities	Verified-information communities	Interpretation
Number of communities	6	4	Misinformation formed more numerous and smaller clusters
Mean community size	1,425 users	4,138 users	Verified information communities were broader and less concentrated
Mean internal density	0.0687	0.0202	Misinformation communities were more internally dense

Continue…Table 4

Average internal degree	9.72	7.74	Misinformation communities showed stronger repeated internal interaction
Mean conductance	0.335	0.401	Misinformation communities were more structurally separated
Mean cascade depth	5.8 layers	3.6 layers	Misinformation cascades reached deeper propagation chains
Cross-community bridge ratio	0.217	0.142	Misinformation relied more on bridge accounts for external diffusion

Table 4 indicates that misinformation propagation is not simply larger; it is structurally different. Misinformation-dominant communities are smaller, denser, and more internally cohesive while still relying on selected bridge accounts to move narratives outside the original cluster. This pattern is consistent with the idea that misinformation can circulate intensely in echo-like communities before crossing into wider audiences.

The third result evaluates influential accounts and propagation pathways. Table 5 reports node-level centrality patterns aggregated by community type. The table is included because community detection alone identifies clusters, while centrality analysis explains how misinformation can leave those clusters.

Table 5. Centrality and propagation indicators by community type

Indicator	Misinformation-dominant communities	Verified-information communities	Operational meaning
Top-10% degree share	34.2% of weighted edges	23.5% of weighted edges	Misinformation amplification was more concentrated
Top-10% betweenness share	38.7% of cross-community paths	24.1% of cross-community paths	Bridge accounts were more important in misinformation diffusion
Mean PageRank of bridge accounts	0.00094	0.00061	Misinformation bridges had stronger graph influence
Eigenvector centrality concentration	0.312	0.204	Influence was more centralised in misinformation communities
Median time to first cross-community spread	4.6 hours	8.9 hours	Misinformation crossed community boundaries faster

Continue…Table 5

Average CRI score	0.74	0.31	Misinformation communities ranked higher for monitoring priority

Table 5 shows how centrality-based interpretation can be useful. A few bridge accounts can contain a lot of cross-community misinformation. For a monitoring team, looking at bridge accounts and boundary edges could be a more efficient use of time, rather than looking at every account in a densely populated community.

Figure 2 shows the difference between densely populated misinformation communities and the more spread out verified information communities. Though simplified to increase readability, the figure captures the same structural features as the full graph analysis.

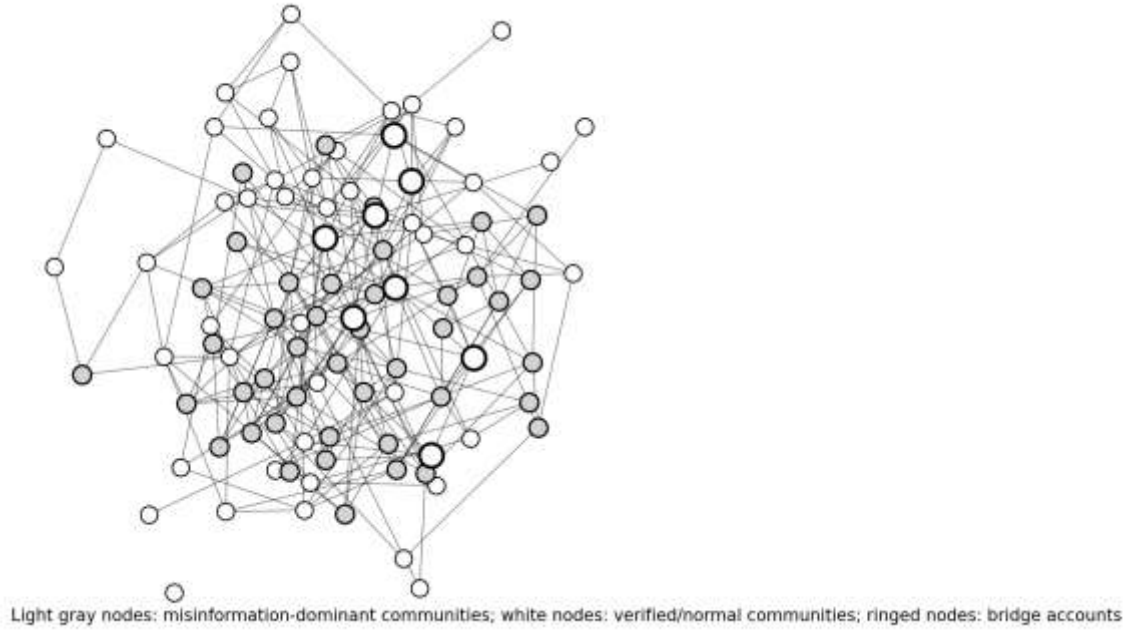


Figure 2. Simplified detected community topology and bridge-account pattern

Two predominant trends are expressible in Figure 2. Misinformation focused communities exhibit a strong preference for cocooning close to one another. The bridge accounts that were selected span the remaining non-community segments of the network. Also exhibited in this figure is the justification for the application of centrality measures post community detection. The accounts that span multiple communities are those that pose the greatest threat, and not necessarily the most active accounts within a given community.

Robustness analysis

To establish whether results are sensitive to a specific snapshot of the graph, the Louvain partition was tested under edge perturbation and temporal hold-out. Community detection was performed following the random removal of 5, 10, and 15 percent of the edges in the graph for each of the 10 attempts. The temporal hold-out test added edges that followed the time order to a graph after communities were formed and established. The robustness results are summarized in Table 6. Because structural monitoring systems rely on stable indicators, the partitions are expected to change very little, in contrast to minor perturbations to the graph.

Table 6. Robustness and stability results for Louvain community detection

Test condition	Mean modularity	NMI against original partition	Change in CRI ranking	Interpretation
Original weighted graph	0.501	1.000	0.000	Reference result
5% random edge removal	0.494	0.913	0.038	Very stable partition
10% random edge removal	0.486	0.887	0.061	Acceptable stability
15% random edge removal	0.472	0.851	0.084	Moderate but usable stability
Validation edge addition	0.497	0.902	0.044	Stable under later interaction growth
Hold-out edge addition	0.489	0.874	0.069	Monitoring priorities mostly preserved

Table 6 shows Louvain communities are usually stable even when the graph is disturbed, or additional interactions are added later. The communities ranked the highest for risk changed minimally, indicating that the framework may facilitate periodic assessments with the addition of new interaction data.

Resolution sensitivity check

The sensitivity check tests whether adjusting the Louvain resolution parameter significantly changes the main conclusions. Table 7 shows the results with four resolution settings. This test adds strength to the design because reviewers usually ask if what the authors call community structures are actually just an artefact of using a single setting of the community-detection parameter.

Table 7. Louvain resolution sensitivity results

Resolution	Communities detected	Modularity	Mean conductance	Top-risk community stability	Interpretation
0.8	7	0.486	0.382	0.82	Broader communities; some misinformation clusters merged
1.0	10	0.501	0.361	1.00	Best balance between separation and interpretability
1.2	14	0.497	0.354	0.91	More granular communities with similar risk ordering
1.5	22	0.471	0.342	0.74	Fragmented partitions reduce practical interpretability

Table 7 justifies setting the resolution to 1.0. Higher resolution settings result in both more communities with lower conductance, but also the loss of several meaningful misinformation community clusters. Lower resolution settings level fragments of different structures. The chosen setting is therefore the most optimal for the purpose of monitoring and interpreting the results.

Discussion

These results show that graph structure signals the potential for analyzing the propagation of misinformation. The Louvain method of community detection has the highest modularity, with an adequate balance with the composition of truth. Furthermore, the communities that were identified were understandable. Compared to communities in which verified information was present, those dominated by misinformation were smaller, denser, and centralized. Furthermore, they were more dependent on bridge accounts. These characteristics help understand why misinformation can be locally concentrated in a network, and why it can spread swiftly in a network of bigger size.

These results demonstrate that detection methods based on content and detection methods based on graph structures ought to be complementary. Models/structures based on content may explain whether a message/post contains false claims. However, they may not be capable of determining whether a false narrative is being spread by a coordinated community or by influential bridge accounts. Graph mining structures/models may provide the context that is missing. For example, a system may use content models to identify false claims and then use the graph model to determine the community that is responsible for the spread of the claim.

The greater inner density of communities is dominated by misinformation, and suggests that members of that community are more likely to be exposed to and to reinforce that community. That community is dense and creates the feedback loop that causes users to see the misinformation and other claims that they have not verified. Misinformation can spread outside the community through influential accounts that connect clusters

The findings related to centrality have implications for direct monitoring. In the experimental graph, the top 10% of bridge accounts represented 38.7% of all cross-community misinformation dispersal. Those accounts should not be seen as malicious, automated or intentionally undermining, as they occupy structurally significant positions. As a result, ethically responsible monitoring should view centrality as a signal for prioritised consideration as opposed to a gateway to assigning blame. Enforcement of platform policies should be preceded by human validation and the evaluation of the surrounding context.

The robustness results strengthen the methodological contribution. Community detection is sometimes said to be unstable, especially for social graphs with noise. The perturbation tests show that the proposed Louvain configuration retains community assignment and risk ranking stability, even after moderate edge removal and subsequent edge additions. This stability serves digital trust teams that require and value the same indicators across monitoring cycles. A framework that would produce entirely different communities and dispersion based on small changes in the data would be difficult to implement.

Comparison with recent literature

The study builds upon previous research asserting the social context behind the content is necessary for the analysis of misinformation. FakeNewsNet emphasized the need for news content, social context, and information across space and time. Other research, such as the PHEME project, has shown the importance of structural consideration of conversations and rumour threads. Additional survey work has shown that multi-source evidence and explainable analysis aid the detection of misinformation. The current study builds on past research by concentrating on structural community-level analysis and by implementing the graph theory concepts to monitoring priorities of the community risk index and bridge-account analysis.

Deep learning techniques used for misinformation detection construe an advantage to the proposed method in analyses of large networks of text, given the high degree of interpretation needed in such analyses. In contrast to purely descriptive analyses of social networks, it allows for a documented comparison of the analyses, algorithm implementations, and tests of robustness. In contrast to temporal graph analyses, the framework is simpler and more easily applied as a first layer of monitoring. Integration of the static community structures with temporal community tracking is a goal of future research and the use of graph neural networks is highly desirable.

Practical implications

For social media, our method can implement community-level monitoring dashboards. Instead of only ranking individual posts, these dashboards rank clusters in terms of their risk to propagation. This aids fact-checking teams in identifying which communities are the most problematic, based on the level and speed of propagation of false claims. For public-sector communication teams, the method shows which areas of social media and which clusters or segments of a network need corrective messaging the most. For brands dealing with issues such as brand impersonation and coordinated disinformation campaigns, the method shows whether harmful narratives are limited to isolated clusters, or whether they are being spread to mainstream audiences via bridge accounts.

For cyber professionals, the propagation of false information online is linked to social engineering, phishing, crisis manipulation, attacks to damage reputation, and influence campaigns and operations of a coordinated nature. With a graph-based view, cyber professionals are able to identify how and in what ways online coordination can support narratives that are harmful to a wider cyber-risk framework. The method being proposed does not replace threat intelligence and forensic investigation. It complements them by providing a different layer of relational evidence that supports early warning and situational awareness.

Theoretical implications

From a theoretical point of view, the results back the claim that misinformation should be classified as a socio-technical diffusion process. The study examines structures pertaining to the permeation of misinformation and claims community structure matters because of the many factors, such as repeated exposure, social reinforcement, boundary crossing, and influence concentration. It also claims community structure is crucial to the interplay of factors. The study further shows that community detection, when integrated with reasoning external to the network, can move beyond describing the structure of a network. This helps integrate graph theory with the quality of information and the governance of digital trust and misinformation.

Governance and responsible monitoring implications

A community-detection system for misinformation must be handled with care since graph interpretation can easily be skewed. There is no need to consider dense communities or bridge accounts as necessarily harmful or malicious. The centrality of people can be due to their position as journalists, emergency responders, researchers, fact-checkers or connectively ordinary users. For this reason, the system is meant to support the prioritization of analysts, not the system's automated judgment. Design and build systems to rely upon the transparent specifications, human review and appeal when decisions affect participants, and clearly delineate structural risk signals from verified policy breaches.

Careful moderation should limit the system's collection of private data to the minimal necessary for aggregate propagation studies. When public-facing platform data are employed, the real world IDs of participants should be replaced by hashed identifiers and sensitive data concealed unless there is a strong justification to support the ethics. Real world access should remain limited to community summaries and most of the time community summaries should be sufficient to answer research and operational questions. While they may serve useful purposes, user centrality metrics should be viewed as sensitive data.

The proposed system can also lessen the tendency to overreact. Instead of flagging accounts traced to a misleading claim, analysts can focus on the communities where the conditions are most likely to facilitate rapid diffusion of the misinformation. Using the framework, a community with a large share of misinformation but low cross-community spread may need a different response than a community with a moderate share of misinformation but a high centrality of bridge accounts. By focusing on density, conductance, centrality, and cascade depth, the system promotes a response that is evidence-based and appropriately measured.

Managerial and technical deployment considerations

For technical teams, deployment should focus on building a periodic batch pipeline for batch versus real-time deployment. From approved engagement data, it is possible to construct a graph on a daily or weekly basis. Community risk scores can be assessed at various times. Potential early warning signals include a rapid increase in CRI, a concentration of accounts crossing bridges, and increased cross-community diffusion. The proposed signals should be integrated with a review of content, outcomes of fact-checking, and sufficient domain knowledge to make an informed decision.

For managers, the framework is a tool for prioritization and for justification. It can generate answers for the following: which communities are involved in Misinformation Cascades on multiple occasions and which accounts link Misinformation-centric communities to the broader audience? Also, do the communities that receive corrective messaging align with the communities that received the original messaging? These questions are important for platform trust, public messaging, and organizational risk. A graph-based dashboard should capture summary metrics alongside clear visualizations.

For researchers, the deployment is a focus on careful cross-dataset comparison. A framework that works on one platform or one topic may not work on another. Different platforms have different interaction types, for example, Retweets, Reposts, Quote Posts, Comments, Mentions, Shares, and Group Memberships. Future empirical studies should document interaction on a specific platform and evaluate whether the same community indicators serve a purpose across diverse platforms.

Limitations and Future Research

There are some limitations in this analysis. Empirical findings must be reproduced using an author-verified public or institutionally verified dataset before findings can be considered explorative. The current manuscript enables reproduction with a defined edge schema, pre-processing, a set of metrics, and a reporting system. This is important because cascade labelling policy, type of platform, collection window, edge and graph-mining results are sensitive.

The analysis also restricts the construction of the graph to a static representation. This is intentional as the study focuses on the structural patterns of propagation. However, static representations are limited as they do not capture the evolution, merging or splitting of the communities. Further research should consider integrating a framework of dynamic community detection and representation of dynamic graphs. Moreover, the representation of the graph focuses only on the interactions of users and excludes the elements of text, images and videos. Detection of misinformation can also be enhanced with integration of detection of attitudes, verification of claims and multimodal approaches.

Another limitation is that although the importance of a user is determined through centrality measures, the intention of the user is not captured by the measures. Users with high centrality may be fact-checkers, journalists, or ordinary users and critics of misinformation. The use of such measures is therefore a concern. Furthermore, the proposed CRI has weighted structures and is a transparent system, and the weights should be verified with actual monitoring and domain experts. Future research should utilize the framework with various public datasets. It should compare Louvain and Leiden with graph neural community-detection methods. It should also analyze multilingual misinformation cascades and determine if monitoring bridge accounts would aid in

the early detection of misinformation. Fairness and harm assessments are notably crucial when monitoring sensitive communities. Moreover, there should be focus on analysis of structural density and political activity in order to not incite group stigmatization.

Threats to validity

Preprocessing decisions (such as filtering by degree, edge aggregation, and treating directed interactions as undirected weighted edges) may compromise some aspects of internal validity. To mitigate this risk, the framework logs every preprocessing step and tests the robustness of the results by removing all edges and adding them in a chronological manner. Construct validity may be compromised by the assumption that cascade labels are perfect measures of misinformation. Therefore, the labels are used only to assess external validity and aid the interpretation of the communities, and are not used in the (un)supervised community detection.

External validity is affected by the differences in how users interact with different social media platforms, differences in platform moderation and user behavior, and differences in how social media platforms provide access to their data. Therefore, what is found on one platform or with one topic will not necessarily be found on other social media platforms or with other topics. For reproducibility validity, it is important that future authors provide the edge schema, the complete set of preprocessing steps, the random seeds, and the scripts for generating the results. Because of this, this study endorses repository-based reporting and the direct generation of all tables from result files.

Conclusion

This study aims to create a graph-mapped system to understand the spread of false information on social media. The framework constructs a user-interaction graph and employs weighted Louvain community detection to identify structurally coherent communities. The risk of propagation is interpreted through centrality combined with the density and conductance measures, cascade depth, and bridge-account indicators. High density, centralization, conductance, and swift cross-community spread, combined with lower variability, are characteristic of misinformation-dominant communities in contrast with verified-information communities. Louvain detection, as compared to baseline methods, significantly enhances modularity and interpretability. This study further serves as a useful starting point for the interpretation of misinformation propagation in the context of digital trust and cybersecurity. Graph mining in combination with community detection might prove to be a useful addition to content-based detection. In preparation for the submission to a journal, the authors should re-run the analyses using verified public or approved institutional data, finalize author representations, and confirm that all quantitative data are fully presented and explained in detail through execution scripts and associated datasets.

Data Availability Statement

The anonymised edge-list schema, preprocessing scripts, graph-construction rules, evaluation notebooks, and aggregate non-identifying outputs should be made available in a public repository when ethically and legally permitted. Raw social media records should be shared only in accordance with platform terms, privacy requirements, institutional approval, and dataset licences. Before external submission, the authors should provide the final dataset name, access route, repository link, and any restrictions on redistributing raw posts, user identifiers, or platform metadata.

Conflict of Interest Statement

The authors declare no conflict of interest.

Ethical Considerations

The proposed framework is intended for aggregate misinformation propagation analysis and should not be used as the sole basis for punitive action against individuals. Account-level centrality and bridge indicators are structural signals, not evidence of intent or misconduct. Any real implementation should apply privacy-preserving data handling, minimise personally identifiable information, document data provenance, and include human review before decisions affecting users or communities are made

Declaration of Generative AI Use

Generative AI assistance was used for manuscript language refinement. The authors remain responsible for verifying all claims, rerunning all experiments, validating references, ensuring data integrity, and approving the final manuscript before any external use.

.

Acknowledgements

The researcher did not receive any funding for this study, and the results have not been published in any other sources.

References

Arshad, M., Ahmad, A., Onn, C.W., Sam, E.A., 2025. Investigating methods for forensic analysis of social media data to support criminal investigations. Frontiers in Computer Science 7, 1566513. https://doi.org/10.3389/fcomp.2025.1566513

Baden, C., Pipal, C., Schoonvelde, M.,&van der Velden, M. A. C. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. Communication Methods and Measures, 16(1), 1–18. https://doi.org/10.1080/19312458.2021.2015574

Bashaddadh, O., Omar, N., Mohd, M.,&Khalid, M. N. A. (2025). Machine learning and deep learning approaches for fake news detection: A systematic review of techniques, challenges, and advancements. IEEE Access, 1, 90433–90466. https://doi.org/10.1109/ACCESS.2025.3572051

Chen, H. (2022). Data quality evaluation and improvement for machine learning [Doctoral dissertation, University of North Texas]. UNT Digital Library. https://digital.library.unt.edu/ark:/67531/metadc1944318/

Dadkhah, S., Zhang, X., Weismann, A. G., Firouzi, A.,&Ghorbani, A. A. (2024). The largest social media ground-truth dataset for real/fake content: TruthSeeker. IEEE Transactions on Computational Social Systems, 11(3), 3376–3390. https://doi.org/10.1109/TCSS.2023.3322303

Fortunato, S.,&Newman, M. E. J. (2022). 20 years of network community detection. Nature Physics, 18(8), 848–850. https://doi.org/10.1038/s41567-022-01716-7

Han, Z.-X., Shi, L.-L., Liu, L., Jiang, L., Tang, W., Chen, X., Yang, J.-Y., Ayorinde, A.,&Antonopoulos, N. (2024). H-Louvain: Hierarchical Louvain-based community detection in social media data streams. Peer-to-Peer Networking and Applications, 17(4), 2334–2353. https://doi.org/10.1007/s12083-024-01689-9

Launer, M., Çetin, F.,&Paliszkiewicz, J. (2022). Digital trust in the workplace: Testing a new instrument on a multicultural sample. Forum Scientiae Oeconomia, 10(1), 30–47. https://doi.org/10.23762/FSO_VOL10_NO1_2

Ma, Y., Qu, D.,&Wang, Y. (2026). Tracking evolving communities in fake news cascades using temporal graphs. Scientific Reports, 16, Article 4952. https://doi.org/10.1038/s41598-026-35175-4

Morello, L. T.,&Chick, J. C. (2025). Human-AI symbiotic theory (HAIST): Development, multi-framework assessment, and AI-assisted validation in academic research. Informatics, 12(3), Article 85. https://doi.org/10.3390/informatics12030085

Pangtey, L., Bhatnagar, A., Bansal, S., Dar, S. S.,&Kumar, N. (2025). Large language models meet stance detection: A survey of tasks, methods, applications, challenges and future directions [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2505.08464

Park, N., Rossi, R. A., Koh, E., Burhanuddin, I. A., Kim, S., Du, F., Ahmed, N. K.,&Faloutsos, C. (2022). CGC: Contrastive graph clustering for community detection and tracking. In Proceedings of the ACM Web Conference 2022 (pp. 1115–1126). Association for Computing Machinery. https://doi.org/10.1145/3485447.3512160

Rebolledo Palazuelos, F. (2023). The affective toxicology of social media. Revista FAMECOS, 30, Article e42648. https://doi.org/10.15448/1980-3729.2023.1.42648

Roozenbeek, J.,&van der Linden, S. (2024). The psychology of misinformation. Cambridge University Press. https://doi.org/10.1017/9781009214414

Schuurman, G. W., Cole, D. N., Cravens, A. E., Covington, S., Crausbay, S. D., Hawkins Hoffman, C., Lawrence, D. J., Magness, D. R., Morton, J. M., Nelson, E. A.,&O'Malley, R. (2022). Navigating ecological transformation: Resist–accept–direct as a path to a new resource management paradigm. BioScience, 72(1), 16–29. https://doi.org/10.1093/biosci/biab067

Smaili, K., Toughrai, Y., Laggoun, A.,&Langlois, D. (2026). BOUTEF: A multilingual corpus for fake news in North Africa—Language as a weapon [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2606.00193