Computers and Artificial Intelligence; Vol.3, No.5; 2026
E-ISSN: 3006-3728 P-ISSN: 3006-371X
DOI: https://doi.org/10.70267/cai.v3n5.171177
Published by: Zeus Press

Key Node Identification in Dynamic Heterogeneous Hypergraphs for Sudden-Onset Disaster Public Opinion

Shiying Liu*

School of Economics and Management, Nanjing University of Science and Technology, Nanjing, 210094, China

**Corresponding author: Shiying Liu.*

Abstract

In sudden disaster events, the dissemination of social media information has distinct stages. Identifying the key dissemination nodes at each stage is of great significance for tracing the source of the information and implementing emergency interventions. This paper divides the event evolution process into 72-hour time windows and constructs three types of nodes (users, content, and topics) and two types of hyperedges (content resonance circles and topic participation circles) in a dynamic heterogeneous hypergraph. It proposes a three-component influence score IS composed of content participation degree, topic participation intensity, and structural hole bridging degree. The weights of each component are adaptively determined based on the correlation between each component and the communication ability. Taking the Weibo data of 2023 Zhuozhou Rainstorm as an example, the propagation simulation verification is carried out using the independent cascading model. The results show that IS is superior to degree centrality, k-core, HITS and other structural baselines in all 7 snapshots, leading by 12.8 points relative to the strongest baseline and statistically significant; its identification gain is concentrated in medium and low-level users, indicating that this method can discover efficient disseminators that are easily overlooked by the structural baselines.

Keywords

disaster-related public opinion, dynamic heterogeneous hypergraph, structural hole, key node identification

1. Introduction

Sudden disaster events often trigger concentrated surges in online public opinion within a short period of time, with social platforms serving as the primary arenas for information dissemination and emotional aggregation; identifying key nodes is therefore a prerequisite for conducting communication traceability analysis and public opinion guidance [1]. However, the dissemination of disaster-related public opinion exhibits distinct temporal characteristics and heterogeneity: empirical studies have demonstrated that disaster-related information behavior follows a life cycle of formation. And the topic distribution and emotional characteristics at different stages show significant differences [2]. Consequently, relying solely on global statistical analyses or single-structure analytical methods proves insufficient to capture these evolutionary patterns. Existing research primarily follows two approaches. One is to model public opinion based on hypernetwork theory, by depicting high-order correlations among multiple elements through multiple layers of subnets [3-4]. However, most models are static and do not distinguish the participation behaviors at different levels. The other is to identify key nodes based on centrality indicators, but most of these indicators originate

© 2026 The Author(s). Authors retain copyright in their articles.

from ordinary graphs and are difficult to directly adapt to the high-order connection characteristics of hypergraphs. To address these challenges, this study proposes a dynamic heterogeneous hypergraph-based method for key node identification in disaster-related public opinion: (1) Representing the event evolution as a dynamic heterogeneous hypergraph, partitioning the timeline into snapshots via time windows, and employing two types of hyper-edges to capture user participation along both content and topic dimensions; (2) Introducing a three-component influence scoring mechanism (IS) that integrates multi-dimensional features and adaptively determines optimal weights; (3) Utilizing Weibo data from the Zhuozhou rainstorm incident to validate the method's efficacy through simulation experiments, while analyzing the sources of propagation gains and its applicable boundaries.

2. Related Work

2.1 Hypergraph

Hypergraphs were first proposed by Berge [5]. Unlike ordinary graphs where edges only connect two nodes, hyperedges in hypergraphs can connect any number of nodes, naturally adapting to the high-order associations in public opinion scenario where a piece of content simultaneously relates to the publisher, content clusters, and topics. There is no need to break them down into pairwise relationships. Current research on hypergraph-based studies mostly adopts static modeling, which limit their ability to reflect the phased evolutionary characteristics of public opinion during disasters.

2.2 Key Node Identification

In the dissemination of public opinion, identifying key nodes serves as a crucial foundation for assessing the trajectory of an event. In conventional graphs, metrics such as degree centrality, k-shell, PageRank, and betweenness centrality are commonly employed for this purpose [6-9]; these metrics offer clear definitions and efficient computation. However, in hypergraph scenarios, projecting a hypergraph onto a conventional graph and then computing centrality metrics can lead to the loss of holistic information about hyperedges, and existing methods struggle to fully leverage the higher-order structural characteristics of hypergraphs.

2.3 Structural Hole Theory

The structural hole theory was first proposed by Burt [10]. Its core tenet is that nodes occupying structural holes can acquire information brokerage advantages by bridging different groups, and this theory has already been applied in contexts such as social networks and cooperation networks [11-13]. However, existing studies predominantly define structural holes on ordinary graphs, and their computations still rely on transforming hypergraphs into line graphs [14]. Furthermore, these works have not yet extended the framework to heterogeneous hypergraphs containing multiple node types, nor have they examined its applicability in identifying the propagation dynamics of public opinion.

3. Methods

3.1 Definition and Construction of Hypergraphs

Given a social media dataset D of disaster events, the observation period is discretized into M time windows of equal duration. For t time window, a dynamic heterogeneous hypergraph $H_t=(V_t,E_t,\phi)$, is constructed, where V_t denotes the set of nodes within the window, E_t denotes the set of hyperedges, and $\phi:V_t\rightarrow\{U,C,T\}$ is the node type mapping function.

Node V_t comprises three types $V_t = V_t^U \cup V_t^C \cup V_t^T$, where V_t^U is user nodes, V_t^C is content nodes. V_t^T is topic nodes. (1) Content resonance circles: composed of a given content node and all its participating users, representing the resonance among multiple users under the same content, where $e = \{c\} \cup \{u | u \in V_t^U, (u,c) \in R_t\}, c \in V_t^C$. (2) Topic participation circles: composed of a given topic node and all its participating users, representing the clustering of users around the same topic, where $e = \{q\} \cup \{u | u \in V_t^U, (u,q) \in R_t\}, q \in V_t^T$.

© 2026 The Author(s). Authors retain copyright in their articles.

3.2 Influence Score

Building upon the aforementioned hypergraph structure, this paper characterizes user influence from three complementary dimensions. (1) Content Participation Intensity: $R(u) = \sum_{e \in \mathcal{E}_t^C(u)} \log|e|$, where $\mathcal{E}_t^C(u) = \{e \in E_t^C | u \in e\}$ represents the entire set of content resonance circles that user u has participated in. (2) Topic Participation Intensity: $P(u) = \sum_{e \in \mathcal{E}_t^T(u)} \log|e|$, where $\mathcal{E}_t^T(u) = \{e \in E_t^T | u \in e\}$ measures the level of user activity at the topic level. Both R and P only capture the scale of user participation and fail to distinguish between a participator and a connector. In this work, we extend the structural hole theory to heterogeneous hypergraph scenarios by defining (3) Structural Hole Bridging Degree: $B(u) = \frac{1}{|\mathcal{E}_t^T(u)|} \sum_{e \in \mathcal{E}_t^T(u)} \frac{|N(e) \setminus N_C(u)|}{|N(e)|}$, where $N(e)$ denotes the set of member nodes of hyperedge e, and $N_C(u) = \bigcup_{e' \in \mathcal{E}_t^C(u)} N(e')$ denotes the set of nodes already covered by user u through their content resonance communities. When a user participates in no topic communities, $B(u) = 0$.

3.3 Normalization and Weight Fusion

Given the significant variation in the number of active users across different time windows, min-max normalization is applied to each component within each snapshot $IS(u) = \omega_R \hat{R}(u) + \omega_P \hat{P}(u) + \omega_B \hat{B}(u)$ to ensure cross-window comparability. The final influence score IS is defined as a weighted sum of the three components, where the weights are adaptively determined based on the Spearman rank correlation coefficient between each component and the true propagation capability: components whose correlation falls below a threshold ($\tau = 0.05$), are pruned, while the remaining components have their weights normalized according to their correlation values; if all three components fall below this threshold, the weighting scheme defaults to equal weights. To prevent components from tying in the rank correlation calculation, a random perturbation of magnitude 10^{-6} is applied to each component to break ties.

3.4 Propagation Simulation Verification

We employ an independent cascade model to conduct propagation simulations in hypergraph scenarios, thereby quantifying the actual propagation capability of nodes: if a node within a hyperedge is activated, it attempts to activate other inactive nodes within that hyperedge with a fixed probability $AvgSeedSpread@K = \frac{1}{K} \Sigma_{u \in S_K} Spread(u)$. For a given user u, the propagation scale is defined as the average number of activated nodes across NMC Monte Carlo simulations. For the Top-K seed set provided by the scoring method, the mean number of activated nodes under independent propagation from a single seed is used as the evaluation metric.

4. Experiments and Results

4.1 Dataset and Preprocessing

Using 2023 Zhuozhou rainstorm event as the research subject, we conducted keyword searches and targeted data scraping on the Sina Weibo platform, collecting 30823raw data entries containing fields such as the Weibo post content, publication time, and user ID, thereby covering the entire event lifecycle. Following BGE encoding, text clustering, and topic modeling, and after filtering out irrelevant information and anomalous data, we retained 10583 data entries and divide them into seven snapshots using a 72-hour window. Finally, the dynamic heterogeneous hypergraph comprises 6360 user nodes, 562 content nodes, and 167 topic nodes, totaling 7089 nodes and 1255 hyperedges. A propagation simulation label pool of 4066 individuals was constructed. All experiments employed a fixed random seed 42.

4.2 Experimental Setup

Using AvgSeedSpread@K as the evaluation metric, the main experiment employed a propagation probability $p = 0.1$ and $N_{MC} = 100$. Set Degree, k-core, HITS, Random methods as method baseline. The primary scheme adopted global correlation weighting (B). Additionally, four control schemes were

© 2026 The Author(s). Authors retain copyright in their articles.

implemented to assess the impact of the weighting design, which contains equal weighting (A), snapshot-by-snapshot weighting (C), absolute value weighting (D), and leave-one-out weighting (E).

4.3 Results

As shown in Figure 1, core users exhibit a distinct dual participation characteristic across both content and topic communities. In Figure 2, the degree distribution of the three types of nodes follows a long-tail distribution, with the vast majority of nodes concentrated in the low-degree range, while only a very small number of nodes possess high degrees.

Figure 1: Top-20 Core User Connection Subgraphs Across All Time Periods

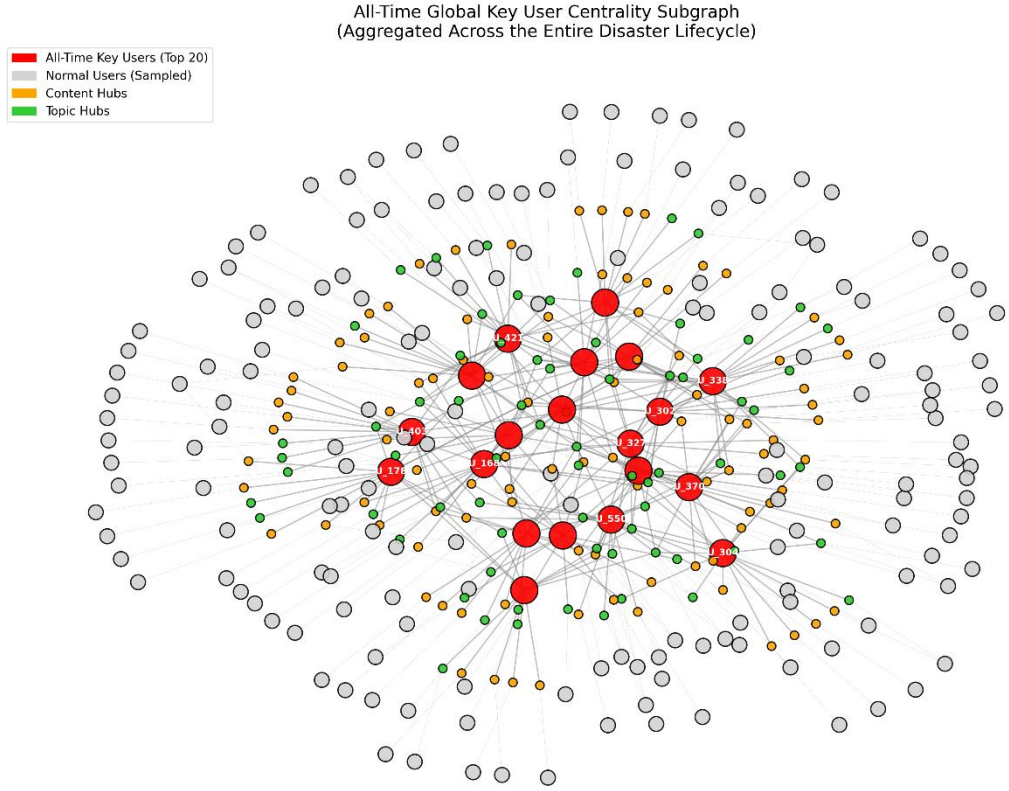


Figure 2: Degree Distribution

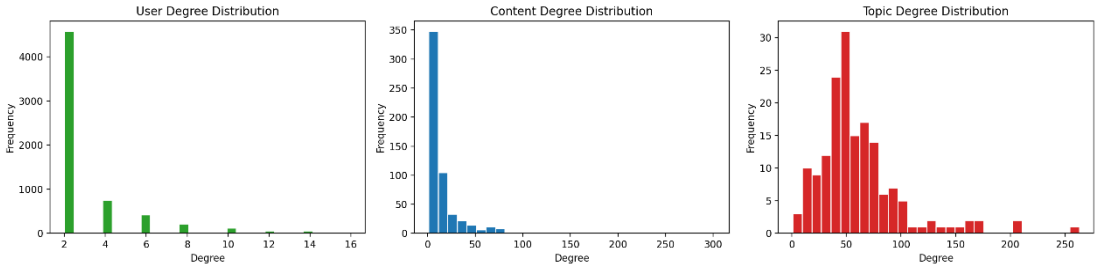


Table 1 presents the AvgSeedSpread@20 results of the five methods across various snapshots. The IS method achieves a cross-window average score of 124.6, outperforming the strongest baseline, Degree, by 12.8 points, and consistently surpasses Degree across all seven snapshots. In contrast, the k-core/HITS method exhibits high volatility and instability, whereas IS maintains consistent performance across different phases of the opinion dynamics structure. Paired Bootstrap tests demonstrate that IS's superiority over Degree and Random is statistically significant; even after excluding the window with the greatest volatility 08-03, IS retains its significant advantage, indicating that the method's performance is not contingent upon a single time window.

© 2026 The Author(s). Authors retain copyright in their articles.

Table 1: Comparison of AvgSeedSpread@20 Across Methods

Snapshot	Degree	HITS	k-core	Random	IS-ours
07-28	101.8	100.7	17.8	61.4	**102.0**
07-31	154.7	186.7	107.3	99.9	**186.9**
08-03	57.8	5.5	**148.9**	85.3	90.3
08-06	84.3	82.9	83.0	81.6	**87.4**
08-09	130.5	149.2	**150.0**	76.4	138.8
08-12	205.2	66.0	65.9	118.2	**209.4**
08-15	48.2	56.3	56.7	33.6	57.4
Mean	111.8	92.5	89.9	79.5	**124.6**

To validate the independent contributions of each component, this study conducted ablation experiments to compare the recognition performance of the three individual components and their combinations. The results show that the Full model, incorporating all three components, achieved a score of 124.6; when the bridging degree B was added to the participation component, the R+B combination scored 125.1 and the P+B combination scored 122.7, respectively, yielding positive gains over the standalone R model 118.5 and standalone P model 120.1. Notably, the Full model scored 4.8 points higher than the R+P combination 119.8, thereby confirming the supplementary value of the bridging component. B scored 101.8 which is lower than that of the participation component, but higher than the random baseline indicating that structural position alone is insufficient to fully capture transmission potential, but can still provide orthogonal structural perspective information. Further collinearity analysis revealed that the Spearman rank correlations between B and R or P were only -0.31 and +0.18, VIF=1.52. It is significantly lower than the R~P pair of 0.736 (VIF=4.12). This low collinearity constitutes the core mechanism by which the bridging component delivers performance gains.

4.4 Mechanism Analysis

4.4.1 Stratified Assessment

Divide the active users into three groups based on their degree centrality, and compare the communication scale of the top 20 users selected by Degree, P and IS within each group.

Table 2: Comparison of AvgSeedSpread@20 by Degree Stratification

Degree Stratification	Degree	P	IS
Low	89.3	118.2	114.6
Mid	67.1	116.3	108.2
High	113.6	120.0	123.3

As shown in Table 3, the advantage of IS relative to Degree is concentrated in the low-to-medium degree stratification; this is because Degree-based ranking exhibits very limited discriminative power among users with small to medium degrees, making it difficult to identify high-efficiency spreaders within this group. In contrast, IS leverages clique structural information to pinpoint such high-efficiency spreaders. A score close to the highest degree level indicates that IS has not compromised the precision of identifying top-tier users in exchange for gains in the lower-tier user segments.

4.4.2 Weighting Scheme and Correlation

Comparative results among the five proposed schemes demonstrate that the equal-weighting scheme achieved the highest score 126.5, while both the global *ρ-weighting* and the absolute-value enhancement approaches scored 124.6; the sliding-window weighting and the leave-one-out weighting schemes scored 121.3 and 122.9, respectively. The maximum score difference among these five schemes was merely 5.2 points and all outperformed the traditional structural baseline method. This indicates that the performance gain of our approach primarily stems from the structural fusion of the three components themselves, and is insensitive to the specific choice of weights. Based on the global correlation analysis, the Spearman rank correlations between the three components and propagation capacity ranking from highest to lowest were P 0.457 > R 0.278 > B 0.174, all exceeding the threshold $\tau = 0.05$, and the corresponding weights are 0.503, 0.306, and 0.191 respectively. The final IS score showed a Pearson correlation coefficient $r = 0.418$, Spearman $\rho = 0.411$ ($p < 0.001$) with respect to single-seed true propagation capacity, representing a robust performance among unsupervised structural evaluation metrics.

© 2026 The Author(s). Authors retain copyright in their articles.

4.5 Robustness Validation

The comparative experiments were repeated for transmission probabilities $p \in \{0.05, 0.1, 0.2\}$. The lead of IS relative degree remained consistent across all three transmission intensities, indicating that the conclusions are independent of specific transmission probability assumptions. Even after excluding anomalous time windows, the results remained robust; the overlap among the Top20 users in adjacent time windows was low, reflecting the characteristic of rapid user turnover driven by disaster events rather than methodological instability.

5. Conclusion and Discussion

This paper proposes a dynamic heterogeneous hypergraph-based key node identification method for disaster-related public opinion analysis. It characterizes users' multidimensional participation through content resonance circles and topic participation circles, integrates participation intensity with structural hole bridging degree to compute influence scores, and adaptively determines corresponding weights. Empirical evidence based on Weibo data from the Zhuozhou rainstorm event demonstrates that this method outperforms traditional structural baselines across multiple temporal snapshots, with performance gains primarily observed among users of medium-to-low participation level, and can effectively identify efficient disseminators that are often overlooked in centrality. The structural hole bridging degree and participation degree components are complementary, enabling the identification of information mediators spanning multiple topic circles. Such users play an important role in the initial core issue consolidation stage of disaster-related public opinion. The high degree of overlap between R and P indicates that the framework can subsequently be simplified to a two-component form comprising participation intensity and bridging degree. This method is a purely structural unsupervised approach; its evaluation relies on single-seed independent propagation capabilities and cluster-derived circle structures within a single incident case study, meaning its generalizability requires validation across additional scenarios. Future research directions include integrating structural hole bridging degree as a prior knowledge into hypergraph neural networks, thereby further combining hyperedge semantics with weight information.

References

[1] Walter, S., & Brüggemann, M. (2020). Opportunity makes opinion leaders: Analyzing the role of first-hand information in opinion leadership in social media networks. *Information, Communication & Society*, *23*(2), 267-287.

[2] Sun, Y., Wang, Q., Zhu, Y., Xu, J., Liu, L., Xiang, C., & Qian, C. (2026). Extracting public opinion on typhoon disasters in China: a sina weibo case study of landfalling typhoon Muifa (2022). *Scientific Reports*, *16*(1), 10227.

[3] Shen, W., Sun, K., Li, H., et al. (2026). Research on user role identification of social media public opinion integrating heterogeneous hypernetwork representation learning from the perspective of symbolic interaction theory [Article in Chinese]. *Journal of Information*, *45*(3), 417-432.

[4] Zhu, W., Li, M., Fan, J., et al. (2024). Identification of key public opinion elements in Weibo hypernetwork based on hypergraph [Article in Chinese]. *Operations Research and Management Science*, *33*(8), 155-161.

[5] Berge, C. (1973). *Graphs and hypergraphs*. North-Holland.

[6] Wang, C., Yan, J., Cai, W., Li, X., Cai, Y., Zhang, X., & Li, Z. (2026). Exponent-based hypergraph centrality for identifying vital nodes in higher-order brain networks. *Journal of Neuroscience Methods*, 110914.

[7] Yang, Q., Wang, Y., Yu, S., & Wang, W. (2024). Identifying influential nodes through an improved k-shell iteration factor model. *Expert Systems with Applications*, *238*, 122077.

[8] Shi, F. H., Wang, X., & Pan, W. L. (2024). Identification of key nodes in hypernetwork based on improved PageRank algorithm [Article in Chinese]. *Information Technology*, 38(3), 22-27.

© 2026 The Author(s). Authors retain copyright in their articles.

[9] He, W., Gao, C., & Jia, T. (2025). Deciphering the overtaking phenomenon initiated by low-centrality nodes in spreading processes. *Communications Physics*, *8*(1), 266.

[10] Burt, R. S. (2022). Structural holes and good ideas. In *Handbook of Sociological Science* (pp. 372-422). Edward Elgar Publishing.

[11] Lin, Z., Zhang, Y., Gong, Q., Chen, Y., Oksanen, A., & Ding, A. Y. (2021). Structural hole theory in social network analysis: A review. *IEEE Transactions on Computational Social Systems*, *9*(3), 724-739.

[12] Huang, C., Yang, C., & Su, J. (2021). Identifying core policy instruments based on structural holes: A case study of China's nuclear energy policy. *Journal of Informetrics*, *15*(2), 101145.

[13] Mu, L., Wang, N., Liu, Z., Cui, X., & Li, J. (2025). Structural hole-based heterogeneous hypergraph for group recommendation. *Expert Systems with Applications*, *286*, 127996.

[14] Liu, W., Guo, L., & Hu, F. (2025). Method for identifying key public opinion elements in social public opinion hypernetwork based on structural holes [Article in Chinese]. *Journal of Qinghai Normal University (Natural Science Edition)*, *41*(3), 68-80.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights

Copyright for this article is retained by the author (s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

© 2026 The Author(s). Authors retain copyright in their articles.