Human Factors in Artificial Intelligence and Social Computing, Vol. 203, 2026, 68–79
https://doi.org/10.54941/ahfe1007319

Identification of Influential Nodes and Discourse Features Within Synthetical Hierarchical Communities in Online Social Networks

Yufan Wu, Zhouhai Chen, and Peihan Wen

School of Management Science and Real Estate, Chongqing University, Chongqing 400044, P. R. China

ABSTRACT

Social fragmentation and information gap are leading to a growing number of communication barriers and social issues. To analyse root causes of the above phenomena from the view of interactive influence, we develop two models to identify influential nodes and discourse features in online social networks to reveal their influence on information dissemination in synthetical hierarchical communities. Firstly, a Node Influence Calculation Model is constructed based on network topology by integrating multi-dimensional indicators such as degree centrality, closeness centrality, and betweenness centrality, to evaluate the influence of nodes with the holistic information in online social network. Secondly, a Discourse Features Model is built based on semantic information by incorporating topics and sentimental features with fine-grained semantic analysis to decode the strategic adjustments and sentimental polarization mechanisms of influential nodes. Finally, with empirical research in real networks, the above two models effectively calculate the influence of nodes and reveal their roles in reinforcing communication effects, cross-community connections, and fostering community integration through themes and emotions. The findings can provide theoretical basis and guiding strategies to promote balanced information dissemination for online social network management, public opinion guidance and societal cohesion.

Keywords: Influential nodes, Discourse features, Online social network, Synthetical hierarchical communities

INTRODUCTION

With the popularization of social media and the Internet, the dissemination of information has entered the age of information explosion. However, despite the huge amount of information, the dissemination of information is not equal, there is a significant gap in access to information between different communities. This is known as the information gap, which directly affects the interaction and connection between social communities, and thus triggers the phenomenon of social fragmentation. The existence of the problems of information gap and social fragmentation forces us to rethink how to promote the effective dissemination of information and bridge the gap between communities in online social network.

Received November 16, 2025; Revised April 22, 2026; Accepted May 9, 2026; Available online July 20, 2026

In this paper, Influential Nodes (INs) in social networks have become an important entry point. These nodes in social networks not only help to promote the rapid dissemination of information, but also can influence the thoughts, behaviours, and emotions within the Synthetical Hierarchical Communities (SHCs).

How to identify INs and further understand their influence mechanisms within SHCs in online social network constitutes the central question of this study. This paper focuses on how high-influence nodes facilitate or hinder interaction and integration between different SHCs in terms of discourse features. A centrality measure (including degree centrality, closeness centrality and betweenness centrality) based on social network topology is adopted to identify INs in social networks. In addition, this study incorporates discourse analysis methods to delve into the role of INs in information dissemination and analyse how they affect interaction and information flow within SHCs.

Currently, mainstream literature focuses on topological indicators to construct influence models. The local and global influence model (LGI) was proposed, which constructs node influence metrics by improving the degree centrality and k-shell methods (Qiu et al., 2021). The Local Entropy Weighting Algorithm (LENC) was proposed (B. Wang et al., 2022) to calculates the influence of a node by using the multistep neighbourhood edge weight distribution. The selection of topological indexes in this kind of methods is generally limited to one-dimensional or multi-dimensional features, and the moderating effect of user attributes on information dissemination is neglected. Research on individual networks mainly focuses on the attributes of individuals and the explanation of the influence of networks (Li et al., 2021), and thus mostly focuses on identifying community features, user identification, user sensitivity prediction etc. Some researchers fuse user social attribute features with social attribute trust to obtain overlapping community structure of multi-layer social networks (Jia et al., 2021). Others fuse user profiles (gender/age) and behavioural data (posting frequency/keywords) to achieve cross-platform user identification (Xing et al., 2021).

By identifying and analysing the attributes of influential nodes, this study not only reveals how these nodes influence the formation and development of SHCs, but also further explores the importance of influential nodes in the process of information dissemination. Identifying influential nodes and understanding their speech characteristics in depth has important theoretical significance and practical value for social network analysis and guidance of public opinion. The contributions of this paper are specified as follows:

(1) The Node Influence Calculation Model and Discourse Features Model are proposed. By integrating network topology and attribute features, we utilize users' multidimensional information to calculate their influence. Meanwhile, the influence of high-influence nodes on SHCs is analysed from the discourse aspect by combining content analysis and sentiment analysis.

(2) In this study, we use multicentricity evaluation criterion in the Node Influence Calculation Model, and consider the information of the whole network. The network centrality value in the Node Influence Calculation Model is constructed by degree centrality, closeness centrality and betweenness centrality. It not only covers the network local information between nodes and their immediate neighbours, but also the network-wide information of distance and shortest path between nodes. The evaluation from local to global network makes the Node Influence Calculation Model more reasonable and accurate.

(3) This paper proposes a Discourse Features Model that integrates topic feature and sentiment feature to analyse the discourse features of high-influence nodes by considering content and sentiment at the same time. In the topic dimension, the LDA model is used to extract features. In the sentiment dimension, the intensity polarity and score are constructed, and the fine-grained annotations and logical connective constraint rules of several dictionaries, such as Dalian University of Technology Emotion Ontology, Emoji Dictionary and One-Word Dictionary are integrated, so as to analyse the influence mechanism under different semantics. The research framework is shown in Figure 1.

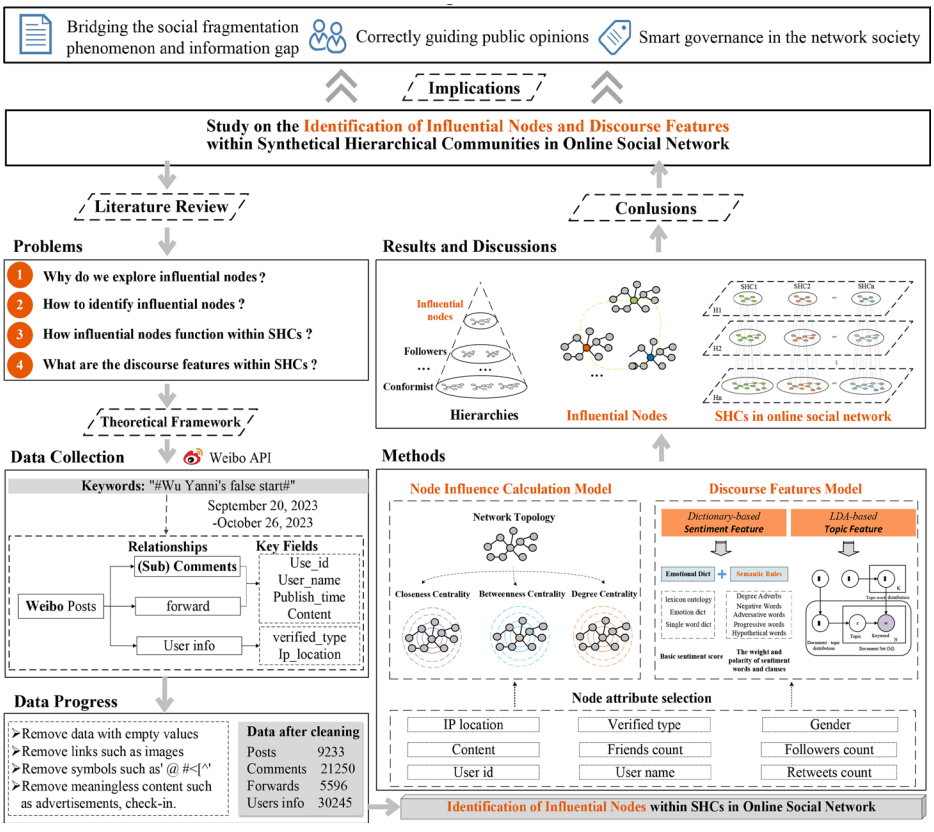

Figure 1: Research framework.

NODE INFLUENCE CALCULATION MODEL

The Node Influence Calculation Model incorporates the network topology into the calculation of node influence. Network topology is a consideration of the network structure characteristics of INs in the social network space, which is used to reflect the importance of nodes on the network structure. Meanwhile, social networks are constructed with users as nodes and interactive relationships such as retweets and comments as edges, and node influence is evaluated using three parameters: degree centrality, betweenness centrality and closeness centrality, which not only consider local information but also the network-wide information into account. In addition, the interaction between INs and netizens can also be screened, which can indirectly reflect the influence of comment published by influential nodes.

A social network is a collection of nodes and links. Sina users and their interrelationships (comments, retweets, and likes) lead to the formation of a large social network among users as they explore evens. This paper take the users who participated in the discussion of "Wu Yanni's false start" as nodes, and the interactions between users, including likes, comments and retweets, as the edges between the nodes.

Network topology is used to measure a user's network position and influence. It consists of three attributes, degree centrality, betweenness centrality, and closeness centrality. Degree centrality is a measure of a node's centrality using the node's "degree", which is the number of other nodes that are directly connected to it. The formula for calculating the degree centrality of a node is as follows, where k_i denotes the number of existing edges connected to node i, and N–1 denotes the number of edges of node i that are connected to all other nodes.

$$DC_i = \frac{k_i}{N-1}$$

Betweenness centrality is the number of times a node acts as the shortest path between two other nodes and can reflect whether the node plays a mediating role or not. The formula for the betweenness centrality of a node is as follows, where n_{st}^i denotes the number of paths that pass-through node i and are shortest paths; and g_{st} denotes the number of shortest paths connecting s and t.

$$BC_i = \sum_{s \neq i \neq t} \frac{n_{st}^i}{g_{st}}$$

Proximity centrality is the reciprocal of the distance from a node to all other nodes, the higher the closeness centrality, the shorter the distance between the node and other nodes, the faster it can communicate and disseminate information to other nodes in the network, and it is an important hub in the network. The proximity centrality CC_i of a node is calculated as follows, where d_i denotes the average distance from node i to the rest of the points, and the reciprocal of the average distance is the proximity centrality.

$$d_i = \frac{1}{N-1}\sum_{j=1}^{N} d_{ij} \quad CC_i = \frac{1}{d_i}$$

Based on degree centrality, betweenness centrality and closeness centrality, this paper constructs the network topology index of influential nodes, which not only considers the local information but also the whole network information.

DISCOURSE FEATURES MODEL

According to Nudge Theory (Thaler & Sunstein, 2008) and Dual Path Theory (Petty & Cacioppo, 1986), all kinds of interaction behaviours in social media will invariably shape users' attitudes and behaviours. A small number of people will think deeply about external information and generate opinions, and the majority of people will follow the existing tendency of public opinion to strengthen their personal attitudes and behaviours, which is a process that requires low cognitive ability and relies on the source of information, the amount of information and other factors. Therefore, those INs will have a strong influence on the discourse and behaviour of other users. If they are guided incorrectly, they can easily induce a split within the communities, which in turn generates a public opinion vortex or even social multi-polarization. This paper analyses the discourse characteristics of high-influence users from the perspective of topics and sentiment.

Topic Features

In online social networks the topics of content posted by high-influence nodes affect users' perception of information, and identifying topics and semantic information of users' posts in online social networks helps to understand their positional tendencies, so there is a need to study the content posted by high-influence users by comparing the similarity of their thematic content with that of other users, and to generalize their discourse thematic features. Thus, how they influence the SHCs in online social networks can be analysed. In this paper, LDA model is used for topic mining of user's posts and the highest probability distribution values of user's posts in different topics are used as user's topic feature attributes.

$$p(w_n \mid M_m) = \sum_{k \epsilon K} p(w_n \mid K_k)\, p(K_k \mid M_m)$$

The probability formula represents the probability of the word w_n appearing in document M_m. In other words, it is the product of the probability of the word wn appearing in topic K_k and the probability of topic K_k appearing in document M.

Sentiment Features

This paper utilizes sentiment knowledge to construct a sentiment lexicon, which identifies fine-grained sentiments by matching them with words in user comment texts and then calculating the sentiment features of the text. The choosing process relies on the Chinese sentiment lexicon ontology released by the Information Retrieval Research Lab at Dalian University of Technology (DUT) (Xu et al., 2008). The lexicon covers three sentiment features: sentiment categories (Happy, Good, Sadness, Disgust, Fear, Anger, and Surprise), sentiment polarity (positive or negative), and sentiment intensity. The polarity and intensity of sentiment words can be used to quantify the overall sentiments of users' posts to be input features.

The first step is to construct a sentiment analysis lexicon specialized for user posts.

In the second step, the effects of different combination patterns of emotion words, degree adverbs and negation words on the sentiment of users' posts are considered. We utilized 219 adverbs of degree provided by HowNet, which are divided into six hierarchies. We also used a dictionary of negatives provided by the CSDN website. Table 1 lists the different combination patterns of sentiment words, degree adverbs and negation words in the microblog text and the corresponding sentiment weight values.

Table 1: Different combinations and weights of sentiment word.

Combination	Weight
Only Sentiment Words	1
Degree Adverbs and Sentiment Words	ad
Negation Words and Sentiment Words	-1
Degree Adverbs and Negation Words and Sentiment Words	$-1*ad$
Negation Words and Degree Adverbs and Sentiment Words	$0.5*ad$
Negation Words and Negative Words and Sentiment Words	1

Then, calculate the sentiment value of a single word in users' posts. The sentiment value v_i of the i-th sentiment word in the text:

$$v_i = \omega * p * (-1)^m * ad$$

where ω denotes the strength of the sentiment word, taking values of 1, 3, 5, 7, and 9; p denotes the polarity of the sentiment word, taking values of 1 for positive, 0 for neutral, and –1 for negative; m denotes the number of negative words; and ad denotes the weight of the degree adverb appearing before the sentiment word.

At last, calculate the sentiment value of users' posts by adding up the sentiment values of all words. S_j denotes the sentiment value of the j-th text.

$$S_j = \sum_{1}^{n} v_i$$

Where n is the number of sentiment words in users' posts. If $S_j > 0$ means users' post has positive sentiment; if $S_j < 0$ means the users' post has negative sentiment; if $S_j = 0$ means the users' post has neutral sentiment. The calculated sentiment value is used as one of the discourse features of high-influence users, which will help to explain the influence of high-influence users on SHCs in online social networks from the perspective of topic feature.

RESULTS AND DISCUSSION

High-Influence Users Identification

Based on MSGCN (Wen et al., 2025), we delineate the communities in Weibo online social network under the topic of "Wu Yanni's false start", as shown in Figure 2. The users with the high influence scores are then calculated using three parameters: degree centrality, betweenness centrality and closeness centrality.

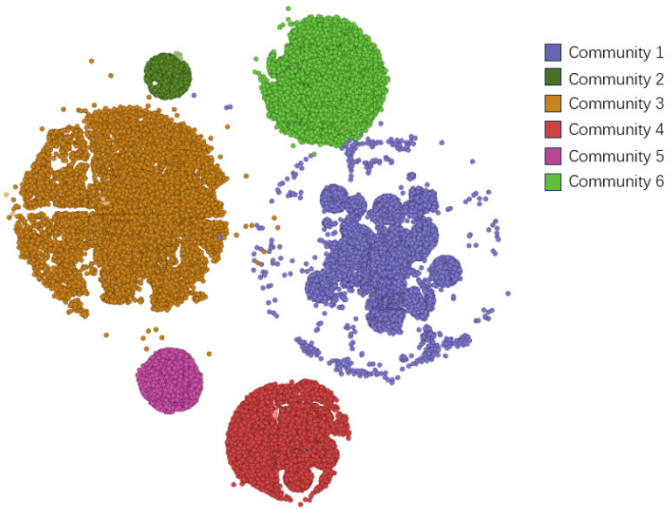

Figure 2: Communities delineated using the MSGCN.

The entire relationship network consists of 30,245 user nodes and 24,119 edges, with an average degree of 0.797. The network diameter is 5, the average path length is 1.255, and the average clustering coefficient is 0.006. We obtained the entire relational network consisting of 30,245 user nodes and 24,119 edges, with an average degree of the graph of 0.797, a network diameter of 5, an average path length of 1.255, and an average clustering coefficient of 0.006. We found that there is a significant decrease in the network topology values starting from the 4th ranked user, and that the influence scores are not differentiated much from those of the 13th ranked user and maintained at a low level. Therefore, we take the top-12 users as high-influence users, among which there are 2 media users, 8 celebrity, 1 enterprise user and 1 general user.

Table 2: TOP-12 high-influence users.

User Name	Identity Type	Number of Fans	Influence Rank	Community
Public Discourse Analysis	celebrity	14,475	1	3
FLYCO	enterprise	124,041	2	6
HE Xiaolong	general	324,445	3	1
Entertainment Lord	celebrity	1,210,298	4	1
Weibo Surfer	celebrity	297,226	5	1
Beijing Youth Sports	media	2,754,029	6	3
YUAN Qicong	celebrity	2,081,449	7	1
Study Reminder	celebrity	24,435	8	1
ZHAO Zhao (Commentator)	celebrity	453,752	9	1
caozenghui	celebrity	729,350	10	6
Sifang Cuisine & Sifang Yoga	celebrity	698,522	11	1
Sichuan Observation News Lens	media	397,506	12	3

We find that the high influence nodes are mainly concentrated in community 1, community 3 and community 6, where community 1 has 7 users with high influence value, community 3 has 3 users with high influence value and community 6 has 2 users with high influence value. These three communities have the largest number of users in terms of size, which means that these three types of users have the most comments, likes and retweets on the topic of "Wu Yanni's false start", and the highest centrality of interaction. The size of the community is directly proportional to the number of high-influence users. At the same time, we found that the correlation between the influence rank of users and their existing number of followers is not high. Users with high influence do not have the highest number of followers; users with high number of followers do not necessarily have the highest influence ranking. However, it is undeniable that users with a certain degree of influence already have a certain number of followers.

Sentiment Features

In the whole user network, these three emotions accounted for 24.5%, 58.65%, and 16.85% respectively. Among the statements with positive emotions, the majority of users expressed encouragement for the contestants themselves, hoping that they would not be discouraged and would continue to work hard next time. Some users also expressed their praise for the contestants, saying that there is nothing wrong with having a flamboyant personality, and some expressed their liking and expectation for the contestants. Among the statements with neutral emotions, most netizens expressed the attitude that it

doesn't matter whether they win or lose, saying that the contestant's behavior is understandable, and more netizens chose to look at this matter rationally. In the speeches with negative emotions, users expressed accusations of the players' behavior, held a nasty attitude towards the players themselves, and even felt angry at the social impact caused by the players.

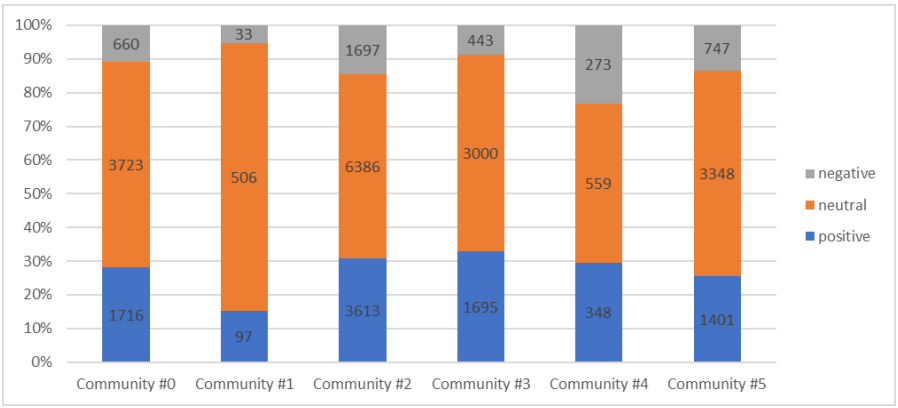

Figure 3: The proportion of sentiments in every community.

In order to better reveal the influence of high-influence users in SHCs, we plotted the distribution of sentiment types in each community as well as the sentiment tendencies and intensities of the top 12 high-influence nodes, as shown in the Figure 3.

Topic Feature

To in order to better reveal the influence of high-influence users on SHCs, we plotted the share of topics in the six communities (see Figure 4) as well as the distribution of topics with the highest probability for the top 12 high-influence users.

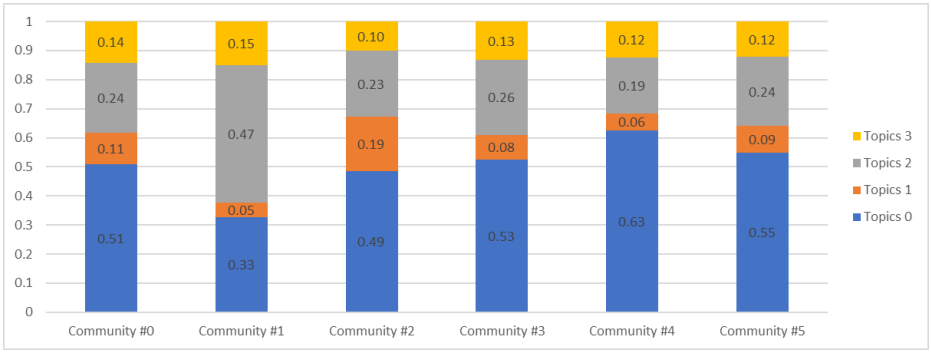

Figure 4: The share of topics in every community.

The distribution of topics within community 1 is correlated with the formation of these sub-communities. Because high-influence users express

their views on events from different perspectives, the interaction and discussion within the communities are numerous and balanced. The high-influence nodes in community 3 are all commenting around topic 1, and the other users in community 3 are influenced to post comments with the same topic, which indicates that the high-influence users are influencing the formation of the SHCs in terms of content. High-influence users in community 6 are mainly commenting on topic 2, which shows that celebrity and corporate nodes respect the facts. Although the content is neutral, they can still stimulate discussion and interaction among the fans of the parties involved.

CONCLUSION

Based on "Wu Yanni's false start" event, this paper reveals the features of high-influence users and their role in the SHCs. After the analysis, we preliminarily draw the following conclusions.

First, there is a synergistic effect between high-influence users and community size. The number of high-influence users is positively correlated with community size. This phenomenon suggests that the intensity of user interaction (comment, like and retweet) directly affects the expansion of communities, while high-influence users become the core driver of community activity through multi-dimensional participation (e.g., initiating topics, sentimental expression, or content production). It is worth noting that user influence shows a weak correlation with fan base, suggesting that the actual communication efficacy of users in breaking public opinion events depends more on the quality of their content interactions than on the static fan size.

Second, the role of high-influence nodes on SHCs in terms of sentiment was found. Community 1 formed a nested structure of multiple sub-communities through the multiple sentimental expressions of high-influence users, which promoted the spread of the topic. Community 3 formed a positive public opinion field with a high degree of emotional unity due to the centralized guidance of the media nodes, which verified the "Anchoring Effect" of the INs in the public opinion setting; and community 6 showed a path of public opinion diffusion of "the event subject-fan community-peripheral public" through the interaction of the nodes within the SHCs.

Finally, it is summarized that high-influence users dominate the direction of community content evolution through topic focusing ability. The high concentration of topic 1 in community 3 (accounting for 85.7%) shows that INs can effectively lead the group to converge in cognition; the neutral content of topic 2 in community 6 has not triggered controversy directly, but has still formed a local hotspot of discussion through the secondary interpretation of the fans. This two-layer communication pattern of "core topic driven-edge topic derived" reveals that there is a mainstream narrative framework for online public opinion, but it also maintains the complexity of dynamic evolution.

Regulators should pay more attention to the dynamics of high-influence users in large-scale communities, and identify their dual roles of sentiment guidance and topic focusing. Meanwhile, it is necessary to pay attention to the

"double-edged sword" effect of media nodes in the setting of public opinion, which can accelerate the clarification of information, but also strengthen the polarization of the communities. Future research can further quantify the influence of life cycle of public opinion on the SHCs, and provide theoretical support for cross-platform communication modelling.

ACKNOWLEDGMENT

The authors would like to acknowledge National Social Science Foundation of China (grant No. 23BSH029) and the Fundamental Research Funds for the Central Universities, China (grant No. 2024CDJSKJJ20) for the financial support.

REFERENCES

Ding, D., Yi, J., Xie, J., & Chen, Z. (2024). Meta-path aware dynamic graph learning for friend recommendation with user mobility. Information Sciences, 666, 120448. https://doi.org/10.1016/j.ins.2024.120448

Hou, L., Pan, X., Liu, K., Yang, Z., Liu, J., & Zhou, T. (2023). Information cocoons in online navigation. iScience, 26(1), 105893. https://doi.org/10.1016/j.isci.2022.105893

Jia, J., Liu, P., Du, X., & Zhang, Y. (2021). Multilayer Social Network Overlapping Community Detection Algorithm Based on Trust Relationship. WIRELESS COMMUNICATIONS & MOBILE COMPUTING, 2021, 9268039. https://doi.org/10.1155/2021/9268039

Li, C., Bai, J., Zhang, L., Tang, H., & Luo, Y. (2019). Opinion community detection and opinion leader detection based on text information and network topology in cloud environment. Information Sciences, 504, 61–83. https://doi.org/10.1016/j.ins.2019.06.060

Li, N., Huang, Q., Ge, X., He, M., Cui, S., Huang, P., Li, S., & Fung, S.-F. (2021). A Review of the Research Progress of Social Network Structure. COMPLEXITY, 2021, 6692210. https://doi.org/10.1155/2021/6692210

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In R. E. Petty & J. T. Cacioppo (Eds.), Communication and persuasion: Central and peripheral routes to attitude change* (pp. 1–24). Springer.

Qiu, L., Zhang, J., & Tian, X. (2021). Ranking influential nodes in complex networks based on local and global structures. APPLIED INTELLIGENCE, 51(7), 4394–4407. https://doi.org/10.1007/s10489-020-02132-1

Song, T., & Ho, Y.-C. (Chad). (2024). Relational Resonance and Content Creation. Management Science, mnsc.2023.00830. https://doi.org/10.1287/mnsc.2023.00830

Teng, F., Liu, X., Dong, X., & Liu, P. (2024). A novel consensus model considering individual and social behaviors under the social trust network. Information Sciences, 670, 120587. https://doi.org/10.1016/j.ins.2024.120587

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Wang, B., Zhang, J., Dai, J., & Sheng, J. (2022). Influential nodes identification using network local structural properties. SCIENTIFIC REPORTS, 12(1), 1833. https://doi.org/10.1038/s41598-022-05564-6

Wang, L., Mou, J., Dai, B., Tan, S., Cai, M., Chen, H., Jin, Z., Sun, G., & Lu, X. (2024). Influential nodes identification based on hierarchical structure. CHAOS SOLITONS & FRACTALS, 186, 115227. https://doi.org/10.1016/j.chaos.2024.115227

Wang, Z., Huang, R., Yang, D., Peng, Y., Zhou, B., & Chen, Z. (2024). Identifying influential nodes based on the disassortativity and community structure of complex network. SCIENTIFIC REPORTS, 14(1), 8453. https://doi.org/10.1038/s41598-024-59071-x

Wen, P., Wu, J., Wu, Y., & Fu, Y. (2025). A novel synthetical hierarchical community paradigm for social network division from the perspective of information ecosystem. Technology in Society. https://doi.org/10.1016/j.techsoc.2024.102784

Xing, L., Deng, K., Wu, H., Xie, P., Zhang, M., & Wu, Q. (2021). Exploiting Two-Level Information Entropy across Social Networks for User Identification. WIRELESS COMMUNICATIONS & MOBILE COMPUTING, 2021, 1082391. https://doi.org/10.1155/2021/1082391

Xu, L., Lin, H., Pan, Y., Ren, H., & Chen, J. (2008). Constructing the affective lexicon ontology. Journal of the China Society for Scientific and Technical Information. https://doi.org/10.3969/j.issn.10000135.2008.02.004.