ARTICLE

Check for updates

A classification and recognition algorithm of key figures in public opinion integrating multidimensional similarity and K-shell based on supernetwork

Guanghui Wang[1,5], Yushan Wang[2,5✉], Kaidi Liu[3] & Shu Sun[2,4]

In online public opinion events, key figures are crucial to the formation and diffusion of public opinion, to the evolution and dissemination of topics, and to the guidance and transformation of the direction of public opinion. Based on the four-dimensional public opinion communication supernetwork (social-psychology-opinion-convergent), this study proposes a classification and recognition algorithm of key figures in online public opinion that integrates multidimensional similarity and K-shell to identify the key figures with differentiation in online public opinion events. The research finds that the evolutionary process of public opinion events is the joint action of key figures with different roles. The opinion leader is the key figure in the global communication of public opinion. The focus figure is the core figure that promotes the dissemination of public opinion on local subnetworks. The communication figure is the "bridge" node in the cross-regional communication of public opinion. Through the algorithm verification of the case "China Eastern Airlines Passenger Plane Crash Event", we find that the algorithm proposed in this paper has advantages in feasibility, sensitivity, and effectiveness, compared with traditional algorithms such as CI, forwarding volume, degree centrality, K-shell, and multidimensional similarity. The classification and recognition algorithm proposed in this study can not only identify multirole key figures simultaneously but also improve the recognition granularity and eliminate the interference of core-like nodes.

[1] Institutes of Science and Development, Chinese Academy of Sciences, Beijing, China. [2] The Institute for Sustainable Development, Macau University of Science and Technology, Macau, China. [3] Qingdao Institute of Humanities and Social Sciences, Shandong University, Qingdao, China. [4] School of Financial Mathematics and Statistics, Guangdong University of Finance, Guangzhou, China. [5] These authors contributed equally: Guanghui Wang, Yushan Wang. ✉email: ysw0327@163.com

Introduction

With the widespread popularity of the internet and social media, online media has created a pseudo environment of the opinion market in the era of "everyone-has-a-microphone". With the explosive growth of online data, online public opinion events occur frequently. Public opinion events, through the fermentation of social media, are prone to evolve into various emotional, provocative, destructive, and misleading issues, resulting in the "breaking circle" phenomenon of online public opinion and the "ripple effect" of public opinion diffusion. In the era of online media, the concealment of the network environment, the virtuality of network identity, and the decentralisation of network communication have made managing online public opinion extremely difficult. Key figures play an important role in promoting the formation and dissemination of public opinion. In online media, key figures with different roles and attributes play different roles in developing public opinion. For example, opinion leaders mainly guide the accelerated dissemination of public opinion within the community, and structural hole (SH) spanners are the core force of cross-community communication of public opinion. If the key figures of all types cannot be identified in time, their attributes and characteristics can be revealed, and the public opinion propagation law can be found. The negative externality of online public opinion will spread, the good atmosphere of cyberspace will be destroyed (Barnes, 2020), and the social morality of consumption will impact the social public order and good customs. Therefore, evaluating the mechanism of key figures in online public opinion events and classifying and identifying key figures are urgent issues in public opinion emergency management and comprehensive cyberspace governance.

Currently, Sina Weibo has become one of China's most popular social platforms and an important venue for the outbreak and dissemination of online public opinion (Shi et al., 2022). Interactive social platforms such as Sina Weibo have the characteristics of independent and public information release, rapid dissemination, and broad influence. This makes it possible for every user to become a key figure, which poses difficulties in the timely and effective identification of key figures. Social network theory is an important method for studying information diffusion patterns and characterising the spread of sudden public opinion. However, traditional single-layer social network analysis methods have shortcomings, such as shallow composition methods and lagging information acquisition through static analysis. The emergence of the "supernetwork" has become a new method for studying complex systems and problems. The supernetwork method provides a tool for studying the interactions and impacts between multidimensional complex networks, and it has recently been widely used to solve practical problems. A supernetwork is defined as a complex network that exists on top of existing networks and surpasses existing networks (Nagurney and Dong, 2002; Nagurney and Wakolbinger, 2005). It is a multilevel, multidimensional, multiattribute network with varying degrees of congestion and coordination (Ma and Liu, 2014). As a new method for studying complex systems and complexity problems, the supernetwork provides tools for studying the interactions and impacts between multidimensional complex networks, which is widely used to solve practical problems. There are various complex self-linking phenomena among Sina Weibo users, such as following, commenting, liking, and forwarding. When they are related to the same topic, they also form "topic" and "emotional" connections (An et al., 2021). These connection methods enable users to form various stable or dynamic network topologies, which provides an opportunity for using supernetwork methods to study online public opinion issues on social platforms such as Sina Weibo.

Research focus

Based on previous research and the practical needs of social media public opinion dissemination management, this article introduces the theory of the supernetwork to design algorithms that can simultaneously identify key individuals playing different roles in online public opinion dissemination. The motivation is to deeply explore the mining of public opinion risk points and the structure of public opinion diffusion. The research focus of this article is as follows:

Firstly, the construction of a public opinion supernetwork. Based on the theory of supernetwork, a four-dimensional network of "social-psychology-opinion-convergence" is constructed to reveal the social network structural characteristics of online public opinion dissemination from the dual perspectives of explicit and implicit relationships. A multidimensional attribute database of Weibo users was established based on the supernetwork and Weibo text data. These data can be used to reveal the role evolution, forwarding relationships, topic evolution, and emotional attribute changes in public opinion subjects. Furthermore, it can be used to describe the complex relationships between public opinion subjects.

Secondly, research has been conducted on key figure classification and attribute characteristics. Based on the characteristics of the public opinion network structure and the network influence of nodes, this study divides key figures in public opinion into three categories: opinion leaders, focus figures and communication figures. This study can also reveal the attribute characteristics of key figures in different roles and their role in public opinion dissemination.

Thirdly, classification and recognition algorithms are constructed for key figures in public opinion. To comprehensively and meticulously mine key figures in online public opinion events, based on the theory of supernetworks, this study constructs a key figure classification and recognition algorithm that integrates multidimensional similarity and K-shell, starting from the composite dimensions of network location and network attributes. This article uses AUC analysis, network destructive experiments, and fine-grained recognition tests to verify the accuracy and effectiveness of the algorithm's recognition.

Contributions

Drawing lessons from previous research using the supernetwork method to identify key nodes, this study starts from the compound dimensions of network location and network attributes, and then a public opinion key figure classification and identification algorithm that combines multidimensional similarity and K-shell is constructed. This study is expected to provide valuable research for the risk identification and control of network public opinion. The possible marginal contributions are as follows.

Firstly, existing research focuses on the explicit relationships between public opinion subjects, while more attention needs to be on the implicit relationships, such as emotions and viewpoints revealed by user-published text information. The four-dimensional network of "social-psychology-opinion-convergence" constructed in this article combines the network structure of qualitative public opinion subjects with quantitative text analysis methods to achieve structured and unstructured attributes of public opinion subjects, such as explicit relationships (forwarding) and implicit relationships (dimensions such as emotions, viewpoints, and roles). This provides a reference model for social network role recognition and dynamic evolution.

Secondly, traditional research often equates key figures in public opinion to opinion leaders, which neglects the role identification and functional analysis of public opinion subjects. This article creatively divides key figures into three categories: opinion leader, focus figure and communication figure. In addition, this article systematically describes the attribute characteristics and

action patterns of multiple types of key figures. This can reveal the complex relationship between the subject of public opinion and the development trend of public opinion events. At the same time, this can also provide scientific and effective decision-making support for the emergency management of public opinion in crisis events.

Thirdly, traditional research on opinion leaders focuses on Influence maximisation (IM) and identifying a single role. The classification and recognition method for key figures in public opinion proposed in this article, which combines multidimensional similarity and K-shell, helps to improve key figure recognition accuracy. This not only improves the fine-grained recognition of key nodes but also effectively filters out low-impact kernel-like nodes while discovering important local "bridge" nodes.

Paper organisation

To reflect the research logic of the article, the research framework is presented as follows. The "Introduction" section is the introduction. The "Related works" section discusses related work. The "Construction of the public opinion supernetwork model" section introduces the basic ideas and methods of social networks and supernetwork while constructing a "social-psychology-opinion-convergent" four-dimensional public opinion supernetwork. The "Classification of key figures in public opinion based on social network theory" section explores the role division and functional attributes of key figures based on social network theory. The "Classification and recognition algorithm for key figures based on a public opinion supernetwork" section is based on the public opinion supernetwork model to construct a key figure classification and recognition algorithm that integrates multidimensional similarity and K-shell. The "Empirical research" is based on the classification and recognition algorithm to classify and identify key figures in the "China Eastern Airlines plane crash" incident and verify the algorithm's effectiveness. The "Conclusion and discussion section analyses the research results and prospects for future research.

Related works

The influencing factors of network public opinion dissemination. Scholars believe that the dissemination of public opinion on the Internet will be affected by factors such as time distance, spatial distance, and social distance between public opinion subjects (Wang and Street, 2018; Shi et al., 2019; Li et al., 2020; Sosa and Buitrago, 2021; Boot et al., 2021). The relationship between the attributes and characteristics of public opinion subjects, such as emotions (Ramanathan and Meyyappan, 2019), opinions (Michaels, 2002), and communication motivation (Biran et al., 2012), also affects the group polarisation effect of public opinion and the dissemination of misinformation (Allcott and Gentzkow, 2017). In addition, government information release strategies (Zhang et al., 2020) and social media (Kushwaha et al., 2021) both play key roles in the public dissemination of public opinion.

The application of the social network analysis (SNA) method in identifying key figures in public opinion. Scholars use social network analysis (SNA) methods to identify key figures in public opinion. The speed of dissemination of crisis information on social media depends on the influence of the communicator (Yoo et al., 2016; Veirman et al., 2016), and the network influence of individuals located in different positions in the network varies (Wang et al., 2021), thereby determining key nodes in the network (Maji et al., 2021). Kistak et al. (2010) identified important nodes in complex networks using the K-shell algorithm based on the idea of global network location. Zeng and Zhang (2013) and Yu et al. (2020) further proposed mixed-degree decomposition

(MDD) and lowest-degree decomposition (LDD) methods. Morone and Makse (2015) proposed a low complexity collective impact (CI) centrality index based on percolation theory in Nature. The percolation theory in complex networks focuses on how to transmit and influence information through nodes and connections in the network, as well as how to predict and control the speed and scope of information dissemination. Some scholars have expanded public opinion research from single-layer networks to multilayer networks (Mucha et al., 2010) and multidimensional networks (Wang et al., 2021; Corradini et al., 2021). Ma and Liu (2014) used the SuperedgeRank algorithm to identify opinion leaders in online public opinion based on the supernetwork theory. Some scholars use "opinion dynamic models" to identify opinion leaders in social networks (Deffuant et al., 2004; Zhao et al., 2018; Bamakan et al., 2019). In addition, some scholars have considered SH spanners on the internet. Scholars commonly use network topology attributes such as the node shortest path (Zhu et al., 2018), network average distance (Rezvani et al., 2015), and weighted centrality (Bhowmik et al., 2015) to identify SH spanners.

In recent years, machine learning algorithms based on social networks have emerged and been widely applied in online public opinion research. Scholars use machine learning algorithms to study the influence of social network users, summarise user behaviour patterns through extensive training and testing of user information behaviour data, analyse user behaviour, attributes, and relationship characteristics to study the decision-making characteristics of network users and calculate their influence in the network. Phan et al. (2016) proposed an ontology-based deep learning model (SRBM+) to predict human behaviour on undirected graphs and node attribute graphs. Qiu et al. (2018) and Leung et al. (2019) are based on the deep learning social impact prediction algorithm (DeepInf), which comprehensively considers potential network structure and potential user functions to predict social impacts. Abu Salih et al. (2020) designed a system that combines semantic analysis and machine learning modules to identify social influencers.

The application of the law of attraction in public opinion dissemination. The study of multidimensional similarity originated from the law of attraction, which states that similar objects attract each other due to their similar attributes. With the development of big data technology, machine learning methods such as decision trees, support vector machines, neural networks, Bayesian networks, and BERT have been widely applied to the recognition and analysis of public opinion user behaviour, public opinion topics, and text emotions (Pang et al., 2002; Hu et al., 2013). This provides technical support for calculating the multidimensional similarity among public opinion subjects. In public opinion research, multidimensional similarity can be used to identify opinion leaders, such as text mining (Bliss et al., 2014), network topology analysis (Liben-Nowell and Kleinberg, 2003) and the random walk algorithm (Lichtenwalter et al., 2010), which are the main methods for mining opinion leaders. Scholars evaluate user influence by combining text information published by public opinion subjects (Tian et al., 2012; Nolasco and Oliveira, 2020). For example, Lee et al. (2019) constructed a comprehensive similarity index from three aspects: content similarity, time similarity, and user topology to identify opinion leaders. Additionally, multidimensional similarity can be used to study the laws of public opinion dissemination (Singh and Ho, 2000). Xiong et al. (2011) analysed the impact of public opinion content and network structure attributes on the formation of public opinion events. On this basis, Wang et al. (2021) proposed a network link prediction algorithm based on multidimensional superedge similarity.

Reviewing relevant research shows that there is currently a wealth of research on key figures and their influence on online public opinion. However, the following limitations remain. First, previous studies have identified opinion leaders based on single methods, such as social network structure characteristics or content interaction among public opinion subjects. However, the role recognition of public opinion events is a complex problem that cannot be solved by a single method. Second, existing machine learning algorithms based on social networks still need to be expanded to one-dimensional structured data. However, the explanatory ability is relatively limited for multidimensional complex structured and unstructured public opinion user interaction data. Third, the identification results of existing research in identifying key figures in public opinion are relatively coarse-grained, and they mostly focus on identifying public opinion leaders. Additionally, there is a lack of role division about key figures, especially attention to the "bridge" nodes of small characters.

As an interdisciplinary research discipline that includes big data, machine learning, graph theory, and other disciplines, network science provides a new perspective and method for the study of complex systems in nature and society. The classification and recognition algorithm proposed in this study is a key figure analysis tool that starts from network science methods, integrates multiple methods such as supernetwork and machine learning, and combines the structural features of public opinion networks with the interaction of public opinion information content. This algorithm expands the application of machine learning in network science and contributes to the further development of network science. Designing a scientific method to identify multiple key figures in public opinion is of great practical significance for strengthening the emergency management of public opinion and optimising comprehensive cyberspace governance.

Construction of the public opinion supernetwork model

Usually, the mechanism by which public opinion subjects disseminate public opinion information on Weibo is to express their opinions and wishes on a topic of interest by posting different types of Weibo texts or forwarding their views and opinions on a topic of interest to their followers (other users) on Weibo. It can be seen that Weibo networks have multidimensional complexity involving attributes such as users, texts, topics, and intentions. These characteristics all form networks; that is, Weibo networks are nested with other networks. Traditional social network analysis often starts from the explicit relationship (forwarding or following) of public opinion subjects to construct a single-layer social network. This ignores the impact of multidimensional implicit attributes such as emotions, opinions, and communication motivation on the communication power of public opinion subjects. Therefore, it is difficult to express complete network relationships and features. Meanwhile, complex network analysis based on ordinary graphs often focuses on constructing a single identification indicator starting from the network topology structure, which leads to the inability to identify influential nodes comprehensively and accurately.

Supernetworks have natural advantages in solving such problems. The multilevel and multiattribute characteristics of supernetworks can maximise the restoration of real semantic social networks. This helps to better depict the complex relationships and global features of real networks while revealing the deep information hidden within the network structure. The multilayer networks of a supernetwork can achieve a layered layout of complex networks while characterising the propagation mechanism between multilayer networks. To comprehensively analyse the correlation between public opinion subjects and quantitatively depict the topological relationships of various dimensions of public opinion elements, based on the theory of supernetwork, this study structured the complex and unstructured node influence problem. We construct a "social-psychology-opinion-convergent" four-dimensional public opinion supernetwork model from four perspectives: text forwarding relationship, emotional evolution relationship, viewpoint evolution relationship and role change relationship.

The components of the public opinion supernetwork model include public opinion subject, role-playing, emotional inclination, and opinion. The relationship between these four components is defined when the public opinion subject, who plays a certain role, expresses a certain opinion under the influence of a certain emotional tendency. The four-layer networks of the public opinion supernetwork model can be constructed from the above four components (Fig. 1): the "social network", "psychological network", "opinion network", and "convergent network". The following section explains the different networks in more detail:

Supernetwork of public opinion (SNP). An SNP is a multidimensional network composed of four-layer networks X and interlayer superedges SE, denoted as SNP = $\{N, SE\}$. Among them, $N = \{N_i; i \in \{S, K, P, C\}\}$ and S, K, P, C represent four layers of networks: the social network, opinion network, psychological network, and convergent network, respectively. $N_i = (E_i, V_i), i \in \{S, K, P, C\}$ represents the set of points and edges of networks. $SE = \{E_{ijmn} \in V_i \times V_j \times V_m \times V_n; i, j, m, n \in \{S, K, P, C\}, i \neq j \neq m \neq n\}$, SE represents the set of superedges of a supernetwork. This study processed "N" pieces of public opinion information from "i" public opinion subjects and obtained "j" viewpoints. Dividing the emotional scores of public opinion information can yield "m" psychological types. Taking into account the different influences of public opinion subjects in public opinion events, we can obtain "n" types of roles.

Social network. The social network represents the forwarding relationship between the subjects of public opinion events. A node is an individual S_i that publishes public opinion information. Directed and empowered edges are constructed based on the forwarding relationship of public opinion information. The social network is marked as $G_S = (S, SL)$, where $S = (S_1, S_2, ..., S_n)$ is the propagation subject set.

$$SL = \left\{ \left(S_i, S_j \right) \right\} = \begin{bmatrix} SL_{11} & \cdots & SL_{1j} & \cdots & SL_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ SL_{i1} & \cdots & SL_{ij} & \cdots & SL_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ SL_{n1} & \cdots & SL_{nj} & \cdots & SL_{nn} \end{bmatrix},$$

where $i, j = 1, 2, \cdots, n$, is the forwarding relationship set. If s_j forwarded s_i, SL is marked as 1; otherwise, it is marked as 0.

Opinion network. The opinion network describes the affiliation among opinions (composed of multiple keywords) in the information published by the public opinion subject. The line between opinion nodes K_i indicates that the two opinions belong to the same piece of public opinion information. The opinion network is described as $G_K = (K, KL)$, where $K = (K_1, K_2, ..., K_m)$ is the opinion set.

$$KL = \left\{ \left(K_i, K_j \right) \right\} = \begin{bmatrix} KL_{1m} & \cdots & KL_{1j} & \cdots & KL_{1m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ KL_{i1} & \cdots & KL_{ij} & \cdots & KL_{im} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ KL_{m1} & \cdots & KL_{mj} & \cdots & KL_{mm} \end{bmatrix},$$

$i, j = 1, 2, \cdots, m$ is the set of opinion attribution relationships. If $k_i I$ and k_i belong to the same public opinion information, KL is marked as 1; otherwise, it is marked as 0.
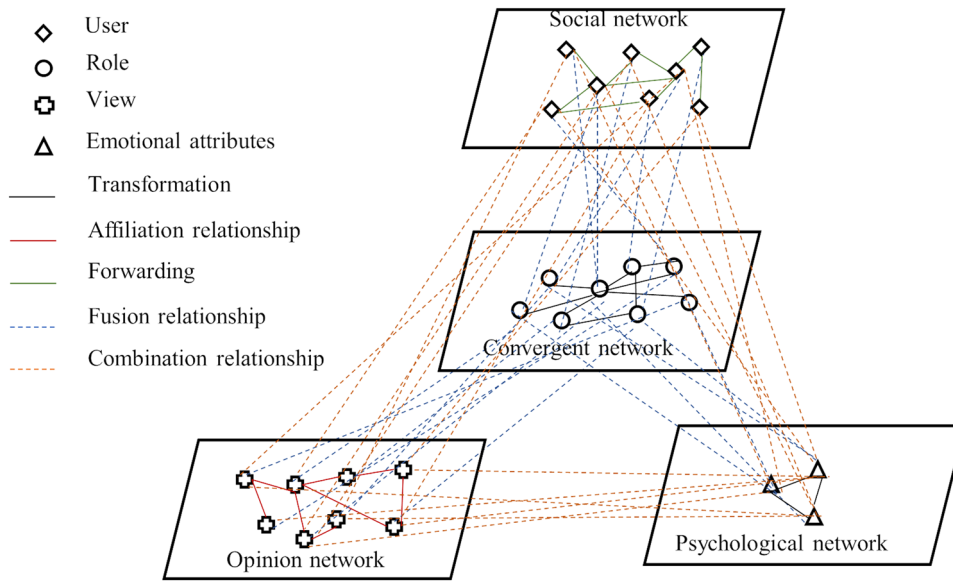
Fig. 1 Schematic diagram of the public opinion supernetwork.

Psychological network. The psychological network represents the psychological type (emotional attribute) that the subject of public opinion has when making a speech. There is a transformation relationship between different psychological types P_m. This study contains three types of affective attributes—namely, positive, neutral, and negative. The psychological network is described as $G_p = (P, PL)$, where $P = (P_1, P_2, \ldots, P_t)$ is the set of psychological types. The edge set of the psychological network is denoted as

$$PL = \left\{ \left(P_i, P_j \right) \right\} = \begin{bmatrix} PL_{11} & \cdots & PL_{1j} & \cdots & PL_{1t} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ PL_{i1} & \cdots & PL_{ij} & \cdots & PL_{it} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ PL_{t1} & \cdots & PL_{tj} & \cdots & PL_{tt} \end{bmatrix},$$

$i, j = 1, 2, \cdots, t$, and $\left(P_i, P_j \right) = \begin{cases} 1 & P_i \text{ and } P_j \text{ can be transformed into each other} \\ 0 & \text{There is no direct correlation in } P_i \text{ and } P_j \end{cases}$.

Convergent network. The convergent network represents the role c_n played by public opinion subjects—namely, opinion leaders, focus figures, communication figures, and ordinary figures—in public opinion events. The roles played by public opinion subjects in different periods can be transformed into each other. The convergent network is described as $G_C = (C, CL)$. Among them, $C = (C_1, C_2, \ldots, C_q)$ is the character type set.

$$CL = \left\{ \left(C_i, C_j \right) \right\} = \begin{bmatrix} CL_{11} & \cdots & CL_{1j} & \cdots & CL_{1q} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ CL_{i1} & \cdots & CL_{ij} & \cdots & CL_{iq} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ CL_{q1} & \cdots & CL_{qj} & \cdots & CL_{qq} \end{bmatrix},$$

$i, j = 1, 2, \cdots, q$ is the set of role transformation relationships. If C_i and C_j can be converted to each other, CL is denoted as 1; otherwise, it is denoted as 0.

SuperEdge. SuperEdge in the public opinion supernetwork model and the superedge connection SE between the four-layer networks. SE is represented by numerical symbols as $SE = \{c_n, s_i, p_m, k_j \mid \theta(s_i, k_j) = 1, \theta(s_i, p_m) = 1, \theta(s_i, c_n) = 1\}$, which indicates that the public opinion subject s_i playing role
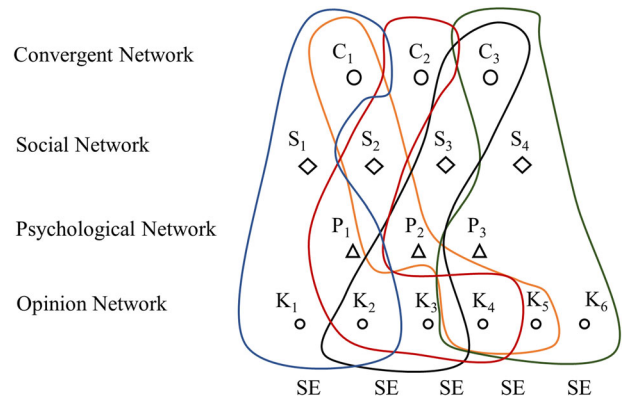

Fig. 2 Superedge schematic diagram.

Table 1 Example of the public opinion supernetwork model.

Superedge	Role	User	Emotion	Opinion
SE_1	C_1	S_1	P_1	K_1, K_2
SE_2	C_3	S_3	P_2	K_2, K_3
SE_3	C_2	S_2	P_1	K_2, K_3, K_4
SE_4	C_1	S_2	P_2	K_4, K_5
SE_5	C_3	S_4	P_3	K_4, K_5, K_6

c_n expresses opinion k_j under the action of psychological p_m. A superedge contains one node in each layer of the network. The actual network environment may be more complex. To simplify the processing, this study assumes that each superedge only contains one role c, one subject s, and one emotional tendency p but can contain one to multiple viewpoints k (Fig. 2 and Table 1).

The nodes connecting the social network, psychological network, opinion network and convergent network can establish the association matrix between different networks, namely, SP, SK, SC, PK, PC and KC, representing social-psychological, social-opinion, social-convergent, psychological-opinion, psychological-convergent and opinion-convergent boundaries, respectively.
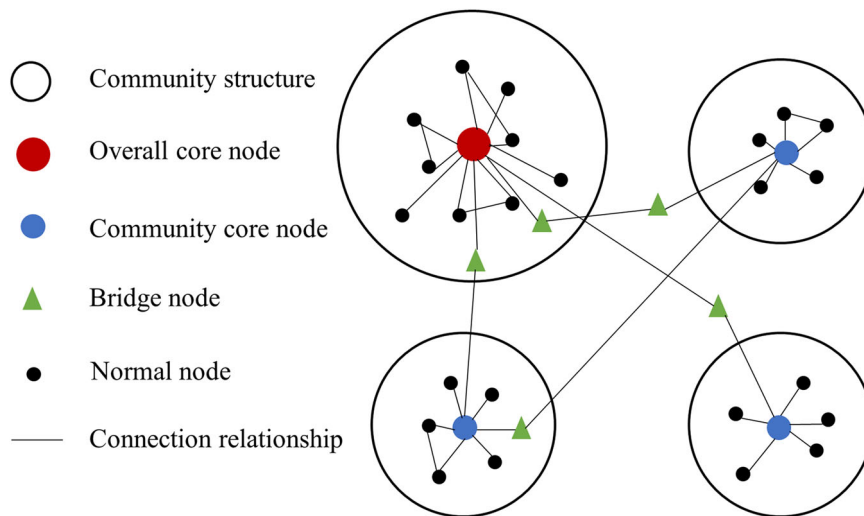
Fig. 3 Social network context map.

These six association matrices are denoted as:

$$SP = \begin{bmatrix} SP_{11} & \cdots & SP_{1j} & \cdots & SP_{1t} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ SP_{i1} & \cdots & SP_{ij} & \cdots & SP_{it} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ SP_{n1} & \cdots & SP_{nj} & \cdots & SP_{nt} \end{bmatrix} SK = \begin{bmatrix} SK_{11} & \cdots & SK_{1j} & \cdots & SK_{1m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ SK_{i1} & \cdots & SK_{ij} & \cdots & SK_{im} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ SK_{n1} & \cdots & SK_{nj} & \cdots & SK_{nm} \end{bmatrix}$$

$$SC = \begin{bmatrix} SC_{11} & \cdots & SC_{1j} & \cdots & SC_{1q} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ SC_{i1} & \cdots & SC_{ij} & \cdots & SC_{iq} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ SC_{n1} & \cdots & SC_{nj} & \cdots & SC_{nq} \end{bmatrix} PK = \begin{bmatrix} PK_{11} & \cdots & PK_{1j} & \cdots & PK_{1m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ PK_{i1} & \cdots & PK_{ij} & \cdots & PK_{im} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ PK_{t1} & \cdots & PK_{tj} & \cdots & PK_{tm} \end{bmatrix}$$

$$PC = \begin{bmatrix} PC_{11} & \cdots & PC_{1j} & \cdots & PC_{1q} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ PC_{i1} & \cdots & PC_{ij} & \cdots & PC_{iq} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ PC_{t1} & \cdots & PC_{tj} & \cdots & PC_{tq} \end{bmatrix} KC = \begin{bmatrix} KC_{11} & \cdots & KC_{1j} & \cdots & KC_{1q} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ KC_{i1} & \cdots & KC_{ij} & \cdots & KC_{iq} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ KC_{m1} & \cdots & KC_{mj} & \cdots & KC_{mq} \end{bmatrix}$$

Classification of key figures in public opinion based on social network theory

In previous studies, scholars have tended to explore the subject of public opinion by exploring opinion leaders. An opinion leader refers to a user who has a representative viewpoint and recognition by the majority of users. Based on social platforms such as Sina Weibo and Twitter as research subjects, the identified opinion leaders are usually Big V with a high number of reposts and followers. With the rapid development of the mobile internet and social media, traditional social opinion has shifted towards online public opinion. Additionally, the information acquisition and dissemination mechanism of netizens has undergone profound changes. In the era of online public opinion, the dissemination and evolution of public opinion exhibit more complex and multidimensional characteristics. The formation and dissemination of public opinion on Weibo exist not only on social networks but also on observation and emotional networks. Therefore, fan count and forwarding count indicators cannot fully identify key nodes. The key figures in public opinion should be excavated from multiple dimensions, referring not only to the big V with distinct viewpoints but also to several types of users who are more active, widely spread, and influential among them.

A social network is a complex network that reflects the interdependence between nodes. As a typical social network, the

complex interaction structure of public opinion dissemination networks plays an important role in disseminating and diffusing public opinion information. Key figures in public opinion are nodes that play a special role in the public opinion network. Based on social network theory (Fig. 3), according to the characteristics of community structure and node locations, nodes play different roles in the network, namely, overall core, community core, "bridge" node (also known as SH spanner), and ordinary node (Yang et al., 2015). The overall core is the global core node in the social network. The community core is the local core node in the community. Bridge nodes are intermediary nodes located at the edge of a community but play a connecting role between core nodes. The nodes in the network, except for the core and bridge nodes, are collectively referred to as ordinary nodes. Our goal is to systematically analyse the main characteristics and role patterns in the evolution of public opinion dissemination. Based on the division of node roles in traditional social network theory and based on the node positions and subject roles of subject networks, we creatively divide public opinion subjects into four categories: opinion leaders, focus figures, communication figures, and ordinary figures. Opinion leaders focus figures and communication figures are key figures that play an important role in the evolution of public opinion events. The classification mechanism and effects of the key figures are shown in Fig. 4. This study is expected to provide support for classifying key figures in public opinion from the theoretical level.

Opinion leader. Opinion leaders are public opinion subjects that guide and control online public opinion events' overall trend and dissemination direction. From the perspective of social network topology, opinion leaders are nodes that occupy the core position and control core network resources. Opinion leaders directly influence many key figures and influencers, and their own influence and neighbouring influence are extremely high, making them the core force for the global dissemination of public opinion. They have both a priming effect and a global effect on the evolution of public opinion dissemination.

Focus figure. In public opinion events, the public opinion subject that promotes the spread of public opinion within a certain time and space range is called the focus figure. From the social network structure perspective, the focus figures have a certain range of self-core networks. They form a local community structure and serve as the core nodes of the social network. However, focus
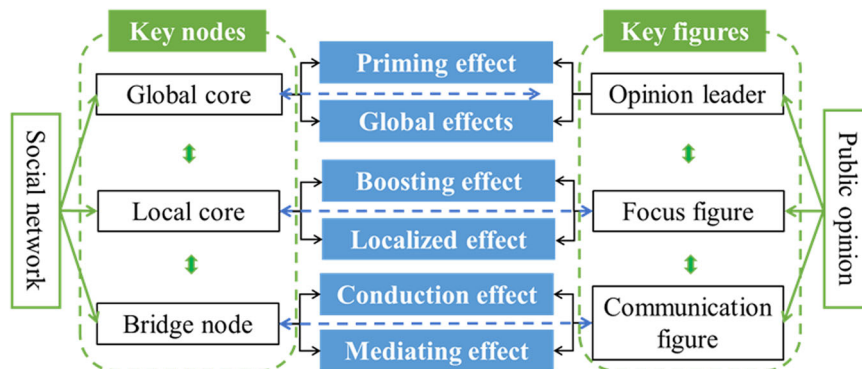
Fig. 4 Classification mechanism of key figures in public opinion.

figures are usually only connected to a limited number of influencers, which limits their overall communication influence. Therefore, focus figures are the backbones of the local dissemination of public opinion, which has a boosting effect and a local effect on the evolution of public opinion dissemination.

Communication figure. A communication figure is a subject that guides public opinion information to transcend time and space constraints and transmit between multiple subjects, themes, and emotions. From the social network structure perspective, the communication figures are SH spanners that connect multiple community subnetworks. They usually serve as bridges between public opinion leaders and focal points or between focal points, serving as intermediaries connecting various local networks. They generally do not form a large self-core network on their own. Communication figures can be seen as "bridge" nodes in the evolution of public opinion dissemination. They have both a conduction effect and a mediating effect.

Classification and recognition algorithm for key figures based on a public opinion supernetwork

Algorithm construction mechanism. Both the K-shell algorithm and multidimensional similarity index are of great significance when identifying key figures in online public opinion. Identifying the key nodes in complex networks has attracted widespread attention from an increasing number of scholars. Identifying key nodes is crucial for studying networks' various functional characteristics and practical applications (Jiang et al., 2010). In the rumour-spreading network, controlling the superspreader can suppress the development of the situation (Lu et al., 2016). Depending on the entire network location, the K-shell algorithm can accurately identify essential nodes in complex networks. The study of the "similarity effect" in social psychology derives from the law of attraction. The formation of social networks is also governed by the law of attraction, which means that network nodes with similar characteristics attract each other, resulting in connections. The higher the similarity between two nodes, the more significant the impact between them. In social networks, similarity awareness is beneficial for establishing connections between users. Users are more susceptible to the influence of similar individuals, thereby changing their attitudes. Similarity promotes implicit behaviour. Especially in social media, the multidimensional similarity of users' professional environment, thematic interests, social psychology, and other factors can bring attraction, promote interaction, stimulate a sense of connection, enhance relationship perception, and promote online relationship connections.

However, due to the different construction mechanisms of the multidimensional similarity index and the K-shell algorithm, each has advantages and disadvantages in measuring node influence.

From a functional perspective, the multidimensional similarity index considers the multidimensional similarity between nodes and domain nodes and typically assumes that nodes have a significant effect on nodes with high similarity to themselves. However, in some situations, similarity and attraction are not related. According to the self-expansion model, dissimilarity brings appeal in certain conditions. This will lower the influence of some core nodes in the network and lead to some low-influence but high-similarity core-like nodes (pseudocore nodes). The K-shell algorithm examines the important positions occupied by nodes in the overall network from the perspective of network structure. Usually, the closer the node is to the centre, the greater its influence. However, the virtual nodes obtained by the K-shell decomposition method are too coarse-grained, meaning that multiple nodes have the same k_s. Additionally, the K-shell algorithm ignores edge nodes (SH spanners) that have a relatively significant influence in areas such as emotions and viewpoints. The classification and recognition algorithm that integrates K-shell and multidimensional similarity indicators not only reflects the network node positions in social networks but also considers the influence of nodes on emotions, viewpoints, and other fields of the leader, thereby improving the fine-grained recognition of key nodes. It effectively filters out low-impact kernel nodes while discovering essential local SH spanners.

Algorithm construction. This study uses the public opinion supernetwork as the basis to model and identify key figures: opinion leaders, focus figures, and communication figures. Figure 5 shows in detail the classification and recognition algorithm of key figures.

Multidimensional similarity index. This paper selects two nodes S_i and S_j in the social network. S_{Sij} is the set of neighbour nodes in the social network connected to nodes S_i and S_j, while P_{Sij} and K_{Sij} are the set of neighbour nodes between layers of the psychological network and opinion network connected to nodes S_i and S_j, respectively. By connecting S_i and S_j with the neighbour nodes within and between layers, in turn, a polygonal network structure can be constructed to judge multidimensional similarity. As shown in Fig. 6, the number of nodes in the node sets S_{Sij}, P_{Sij} and K_{Sij} are 2, 1, and 1, respectively. They are all connected to S_i and S_j from the social network. Six nodes form a polygonal network structure. Each neighbour node is connected to S_i and S_j to form a corner of the network structure. In this case, the neighbour node can be regarded as the network corner-point $w_\angle$. For example, K_{Sij} is connected to both S_i and S_j, which indicates that a corner-point $w_{\angle K(Si, Sj)}$ is formed between superedges SK_i and SK_j. All neighbour nodes form a polygonal structure with S_i and S_j, which is the corresponding star-structured network for building multidimensional similarity.
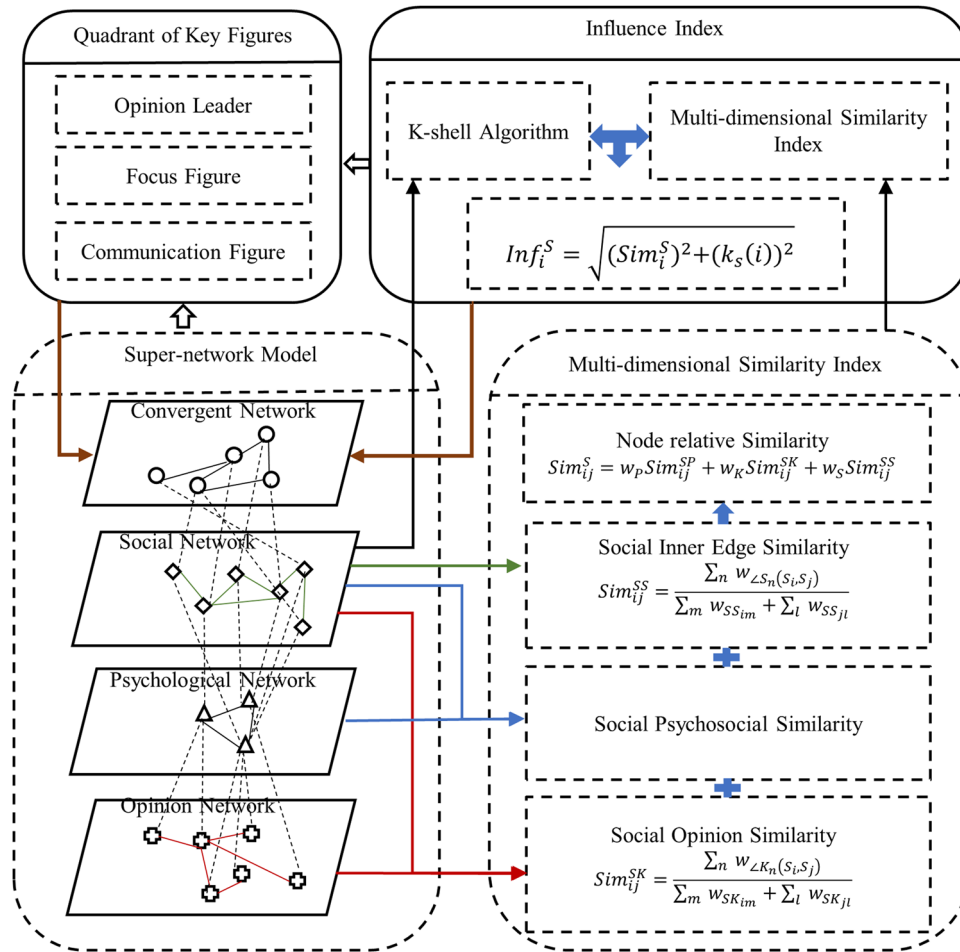
Fig. 5 Classification and recognition model of key figures.

Based on the corner-point distribution of the polygonal structure of the interlayer network and the intralayer network, the superedge similarity between nodes can be calculated. Specifically, the psychosocial superedge similarity Sim_{ij}^{SP} mainly focuses on the similarity relationship of emotional attributes between the psychological network nodes associated with the social network nodes S_i and S_j. The social opinion superedge similarity Sim_{ij}^{SK} mainly focuses on the overlapping degree of the opinion nodes in the opinion network associated with the social network nodes S_i and S_j. The inner-edge similarity of the social network Sim_{ij}^{SS} mainly focuses on S_i and S_j and the topology of its neighbour nodes within the social network.

Social opinion similarity.

$$Sim_{ij}^{SK} = \frac{\sum_n w_{\angle K_n(s_i, s_j)}}{\sum_m w_{SK_{im}} + \sum_l w_{SK_{jl}}} \quad (1)$$

In the above formula, Sim_{ij}^{SK} is the social opinion superedge similarity between nodes S_i and S_j. S_i and S_j are nodes in the social network. K_n, K_m and K_l are nodes in the opinion network. $w_{\angle K_n(s_i, s_j)}$ represents the corner-point of superedges SK_m and SK_{jn}. The value is 1 when the corner-point exists and 0 when the corner-point does not exist. This corner-point exists when K_n is connected to both S_i and S_j. The corner-point does not exist when K_n is connected to one of S_i and S_j, or not to either of them. $w_{SK_{im}}$ and wS_{Kjl} denote the superedges SK_{im} and SK_{jl}, respectively. If the superedge exists, the value is 1. If it does not exist, the value is 0.
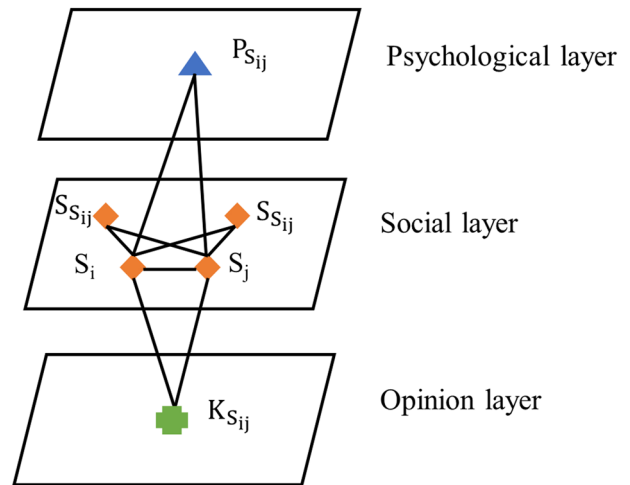

Fig. 6 Star network model.

Social-psychological similarity.

$$Sim_{ij}^{SP} = 1 - \left| \sum_x w_{SP_{ix}} \cdot w_{P_x} - \sum_y w_{SP_{jy}} \cdot w_{P_y} \right| \quad (2)$$

In the above formula, Sim_{ij}^{SP} is the social-psychological superedge similarity between nodes S_i and S_j. P_x and P_y are the nodes in the psychological network. Furthermore, w_{SPix} and w_{SPjy}

represent the superedges SP_{ix} and SP_{jy}, respectively. If the superedge exists, the value is 1. If it does not exist, the value is 0. Finally, w_{Px} and w_{py} represent the emotional intensity of node P under different emotional attributes.

Social inner-edge similarity.

$$Sim_{ij}^{SS} = \frac{\sum_n w_{\angle S_n(S_i, S_j)}}{\sum_m w_{SS_{im}} + \sum_l w_{SS_{jl}}} \quad (3)$$

In the above formula, Sim_{ij}^{SS} is the inner-edge similarity between nodes S_i and S_j in the social network. S_n, S_m and S_l are nodes in social networks. It may also be noted that represents the corner-point of sides SS_{in} and SS_{jn}. The value is 1 when the corner-point exists and 0 when the corner-point does not exist. Then, $w_{SS_{im}}$ and $w_{SS_{jl}}$ represent the interlayer connection edges SS_{im} and SS_{jl}, respectively. If the inner-edge exists, the value is 1. If it does not exist, the value is 0.

Relative similarity between nodes.

$$Sim_{ij}^S = w_P Sim_{ij}^{SP} + w_K Sim_{ij}^{SK} + w_S Sim_{ij}^{SS} \quad (4)$$

This study comprehensively considers Sim_{ij}^{SP}, Sim_{ij}^{SK}, and Sim_{ij}^{SS} and selects different combinations of superedge similarity and inner-edge similarity by adjusting the weights w_p, w_K and w_S to calculate the multidimensional weighted similarity Sim_{ij}^S between S_i and S_j. At present, the prior weight model commonly used in multidimensional similarity is the equal weight model (Li et al., 2019; Wang et al., 2021). Based on this, this study does not discuss the specific importance and contribution of each similarity index. In this study, an equal weight calculation is carried out; that is, $w_P = w_K = w_S = 1/3$.

Average similarity of nodes.

$$Sim_i^S = \sum_{j \neq i} Sim_{ij}^S / (N - 1) \quad (5)$$

In the above formula, Sim_i^S represents the average similarity between node S_i and the rest of the nodes in the social network. The larger the value of Sim_i^S is, the higher the similarity between node S_i and the remaining nodes in the network. Furthermore, N represents the number of nodes in the social network.

K-shell algorithm. K-shell decomposition is an algorithm that makes a global measure of node importance based on the idea of the location of the entire network. K-shell decomposition divides the network into different layers, from the core to the edge. This gives nodes at different levels a K-shell index (k_s, an integer greater than zero). The difference in k_s values can be used to determine the importance of different nodes. The process of K-shell decomposition is similar to peeling an onion. The specific decomposition process is as follows. First, we remove degree 1 from graph G. If there are any remaining nodes of degree 1 in the remaining subgraphs, remove them as well. This process continues until the subgraph generated by the remaining nodes (denoted as G1) no longer contains nodes of degree 1. The subgraph composed of all removed nodes is 1-shell. The K-shell index of this part of the nodes is 1, that is, $k_s = 1$ (the black nodes in Fig. 7(b1)). Second, we continue to remove the nodes with intermediate degrees of 2 in subgraph G1 until all degrees are greater than 2 in the subgraph produced by the remaining nodes (denoted as G2). The notes removed in this step generate a subgraph of 2-shell. The k_s of these nodes is equal to 2 (the blue nodes in Fig. 7(b2)). This process continues to cycle. Finally, we gradually increase the value of k until all the notes are partitioned into a k-shell layer. Each node is

assigned the corresponding K-shell index $k_s = k(k = 1, 2, ..., n)$. For example, the red nodes in Fig. 7(b3) have a k_s of 3.

In graph theory, K-core and K-shell are often used to identify the importance of network nodes. We use coreness to represent the K-core Index of the node. A maximal subgraph whose coreness is at least k is called k-core. The K-core decomposition process is similar to the K-shell decomposition process. To understand this in terms of the set theory, a subset of nodes generated by removing the (K+1)-core set from the K-core set is defined as K-shell. Set k-core is the union of all k-shells with $k_s \geq k$. The k-shell is the complement of the set (k+1)-core in set k-core. Obviously, the set (k+1)-core is a subset of set k-core, while set (k+1)-shell and set k-shell belong to different sets. Figure 7 shows the difference between the K-shell decomposition and the K-core decomposition. 1-core is the entire network (Fig. 7(c1)), which contains all the nodes in the diagram. 1-shell is a set of points between 1-core and 2-core, containing only nodes with $k_s = 1$ (Fig. 7(b1)). 2-core contains nodes which $k_s = 2$ and $k_s = 3$ (Fig. 7(c2)). 2-shell is the point set between 2-core and 3-core, containing only nodes with $k_s = 2$ (Fig. 7(b2)). Both 3-core and 3-shell contain only nodes with $k_s = 2$ (Fig. 7(c3), 7(b3)). Obviously, the coreness of the node after the final decomposition is the same as its k_s. However, it is worth noting that the coreness of some nodes will change during the decomposition process, but the k_s is unique. At the same time, by referring to related studies (Bickle, 2013; D'Arcangelis et al., 2021), we find that K-shell and K-core usually reach the same conclusion when identifying the importance of the network. Both the highest k_s and coreness mean the most important notes. In order to more clearly divide nodes into different levels to show their different importance, we prefer to use K-shell decomposition and k_s in the following discussion.

In graph theory research, Kitsak et al. (2010) pointed out that the nodes with the highest value of k_s in the different levels derived from K-shell decomposition are usually the most effective spreaders in epidemic models. Therefore, in this study, the highest value of k_s is named k_s^{max}. At the same time, we will record the k-shell layer representing k_s^{max}. Based on this, we take k_s^{max} as the classification threshold, and consider that the nodes in k_{max}-shell with the K-shell index of k_s^{max} play a more important role in the network. Based on Fig. 7, we can see that the k_s^{max} of this example network is 3. The 3-shell layer is the k_{max}-shell. We can identify the nodes with $k_s = 3$ as core nodes, that is, nodes with greater relative importance.

Influence Index. Since the multidimensional similarity index and the K-shell algorithm have specific dimensions, this study selects the linear transformation method (Formula 6) for the normalisation calculation. The values of the two indices are assigned to the interval value between 0 and 1. Thus, two sets of vectors are obtained, namely, $Sim_i^S(0, 1)$ and $k_s(0, 1)$.

$$x = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (6)$$

Then, the Euclidean distance method is chosen to couple the two types of indicators:

$$Inf_i^S = \sqrt{(Sim_i^S)^2 + (k_s(i))^2} \quad (7)$$

In the above formula, Inf_i^S represents the influence index of node S_i in the social network. The larger the value is, the greater the comprehensive influence of node S_i in the supernetwork.

Classification and identification algorithm of key figures. This study uses the multidimensional similarity index and K-shell algorithm to construct a quadrant (Fig. 8). The key figures are classified and identified according to the relative performance of the values of *Sim* and k_s of the nodes. After calculation, nodes
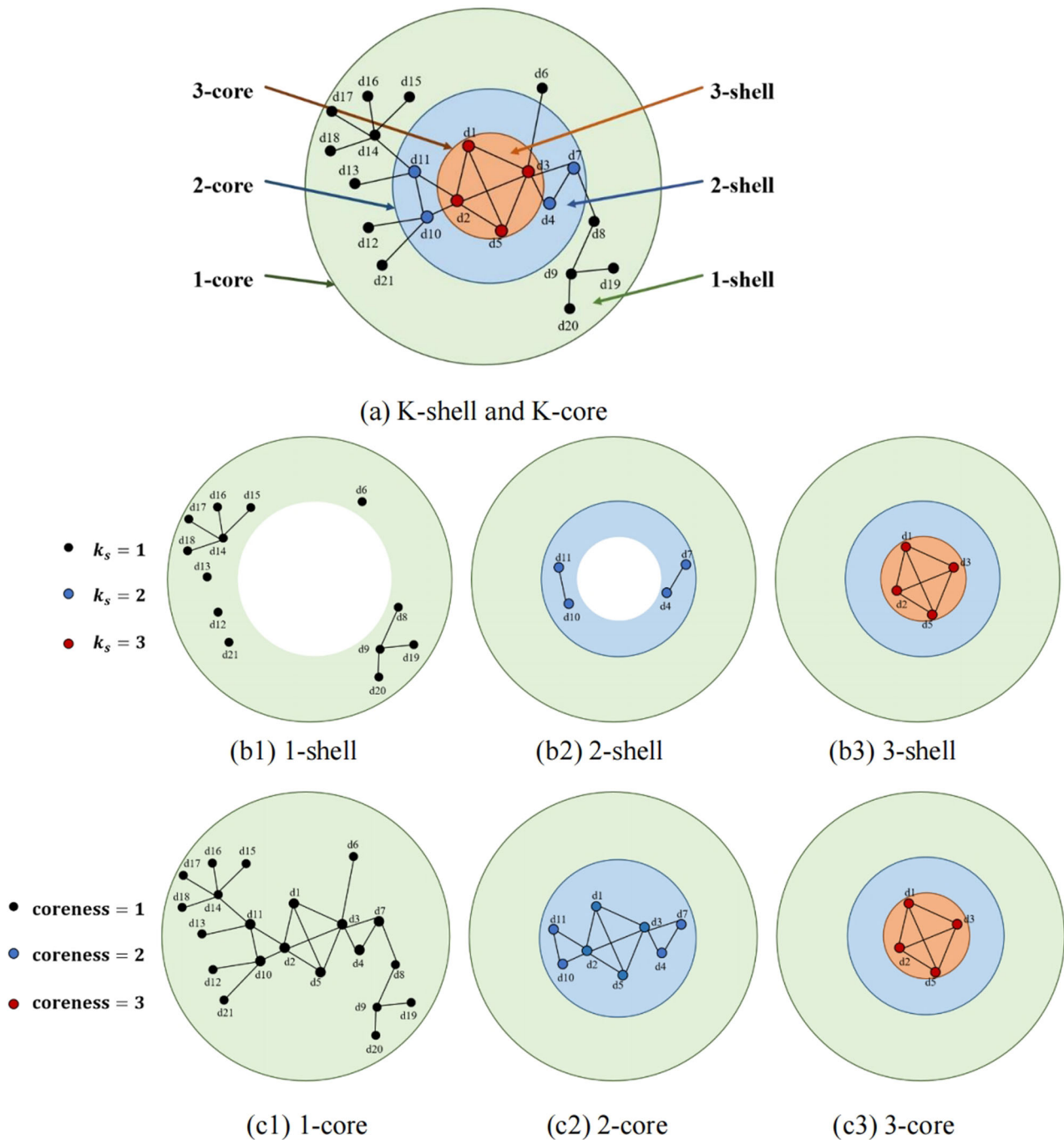
Fig. 7 K-shell decomposition and K-core decomposition of a simple network. a Shows K-core and K-shell models. **b** Shows the k-shell decomposition process. **b1**–**b3** Show the 1-shell, 2-shell, and 3-shell models respectively. **c** Shows the k-core decomposition process. **c1**–**c3** Show the 1-core, 2-core, and 3-core models respectively.

located in the HH quadrant (with both high values of *Sim* and k_s) are highly influential nodes in the network, and they are identified as opinion leaders. The nodes located in the LH quadrant (the *Sim* is relatively low and the k_s is relatively high) are focus figures, occupying a relatively important position in the network and maybe the core nodes of the local area network, with a strong influence on the local area network. However, due to the relatively low k_s, its cross-community transmission power is relatively weak. Nodes located in the HL quadrant (where *Sim* is relatively high and k_s is relatively low)

are communication figures, which typically have strong cross-regional communication capabilities and serve as bridge nodes for public opinion dissemination. The nodes located in the LL quadrant (with both low values of *Sim* and k_s) are ordinary figures.

The classification and identification algorithm of key figures requires that the high and low thresholds of *Sim* and k_s be determined to identify their roles. However, setting the threshold is subjective, and there is no clear standard. Different threshold settings have a great impact on the recognition results. Therefore,
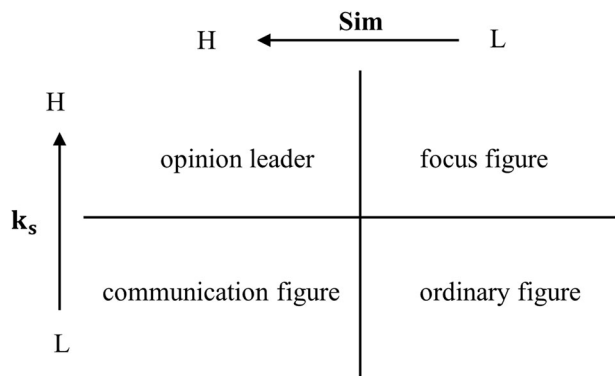
Fig. 8 Quadrant diagram of key figures.

setting a scientific threshold is very important to the performance of the classification and identification algorithm of key figures. In this study, multiple nodes in social networks can have the same k_s. Related studies usually use the highest value of k_s (k_s^{max}) as the dividing line to divide the subjects with different k_s values into two categories (Kitsak et al., 2010; Lu et al., 2016). The nodes with k_s^{max} are identified as important nodes that play an important role in the network, while the other nodes are identified as normal nodes. Therefore, the value of k_s^{max} is used as the threshold value that splits K-shell index into H and L. The process of determining the threshold is as follows: first, the k_s value of each node is obtained according to the K-shell decomposition rule. Second, the k_s values of each node are sorted from high to low. Finally, we select the k_s^{max} as the demarcation to divide the nodes into two dimensions: H and L. For *Sim*, referring to Li et al. (2019) and Wang et al. (2019), the high and low thresholds of *Sim* are set to 30%.

Classification and identification mechanism of key figures based on a public opinion supernetwork. The process of classification and recognition of key figures based on public opinion supernetwork is decomposed as follows: First, a public opinion supernetwork model is constructed, which includes a four-dimensional subnet of " social-psychology-opinion-convergent". Second, we identify the attribute characteristics of the connecting edges between the public opinion subjects in the social subnet, as well as the attribute characteristics of the superedge formed between the opinion subnet, the psychological subnet and the social subnet, based on the research methods of complex network and supernetwork. Third, we calculate the K-shell index (k_s) and the multidimensional similarity index (*Sim*) according to the network topology attributes and then count the influence index (Inf^s) of public opinion subjects participating in public opinion dissemination. Finally, the classification and recognition rules of key figures proposed in this study are used to identify three different types of key figures in public opinion events, namely, opinion leader, focus figure and communication figure. The classification and recognition framework of key figures based on the public opinion supernetwork is shown in Fig. 9.

Empirical research
Data collection and preprocessing
Data collection. This study chooses the "China Eastern Airlines passenger plane crash" incident as the case and derives data from the Sina Weibo platform. The reasons for choosing this case are twofold. ① Sina Weibo is a representative self-media communication platform. It has relatively mature channels for public opinion dissemination, attracting many netizens to participate in ongoing discussions. Additionally, it maintains a relatively

complete network of public opinion dissemination. This facilitates the extraction and sorting of public opinion information. ② The incident of the "China Eastern Airlines passenger plane crash" is a significant emergency public event, and incidents of this nature have a very low probability of occurrence. After the incident, it quickly gained the attention of online social platforms, official media, and the public. This event became a hot topic of online public opinion in 2022 and formed a complex public opinion ecology.

This study starts from the "China Eastern Airlines passenger plane crash" incident and finds that the discussion heat of the incident on Weibo platforms has almost subsided from March 21, 2022, to March 31. Therefore, by crawling all relevant Weibo posts and their corresponding comment information (direct comments and forwarded comments) during this period, a total of 49,728 pieces of data were obtained. The data collection includes the following: ① basic information, such as the username and gender of Weibo users when registering, geographic location when posting information, and account category; ② interaction information, such as likes, comments, and forwarding; ③ social information, such as following and followers' lists; and ④ original information, created by the user in Weibo and usually reflecting their actual psychological state and preferences. After removing links in the text, other users' information, stop words, and so on, 49,728 pieces of data were obtained. Based on the discussion heat of the event in Weibo space, the event was divided into its gestational period (6 h), explosion period (6–28 h), duration period (28–72 h), and recovery period (4–10 days). Figure 10 shows the overall development trend of the popularity of public opinion events.

Data preprocessing. Since Weibo data have no fixed format and are filled with network vocabulary, data preprocessing is needed. First, the non-textual information as links, emoticons, and symbols are removed from Weibo text. Second, existing word segmentation tools such as the Chinese Academy of Sciences, Chinese lexical package, and padding word segmentation are used to segment microblog text. Finally, three methods of document frequency (DF), word frequency (TF), and entropy are comprehensively used to remove meaningless stop words. This article obtained a total of 10,191 sets of valid keywords.

Division of opinion. Weibo text data do not have a fixed format and are full of network vocabulary. This article uses machine segmentation tools in R language to extract viewpoints from Weibo text. First, based on the R language environment, the Jieba programme package is called to remove keywords from Weibo content published by netizens, and syntax analysis is performed on keywords with a higher frequency of occurrence, replacing the main viewpoint words of certain Weibo content. Second, the latent Dirichlet allocation (LDA) topic model can give the topic of each document in the document set in the form of a probability distribution, and the recognition result is objective and scientific. In this study, the subject is modelled. The opinion information is clustered by calling the LDA package and assisting the subjective and objective method of manual marking. Then, text mining of the example data is carried out to realise opinion word extraction. Finally, opinion segmentation is extracted layer by layer to make the opinion information more abstract and generalised to avoid the disadvantage of poor interpretation of machine language classification.

Through the extraction and cluster analysis of the keywords of public opinion information, the opinion extraction results of the opinion network were obtained. In the China Eastern Airlines incident, 82 core opinions were formed. The topic orientation and dissemination of power covered by different opinions are varied.
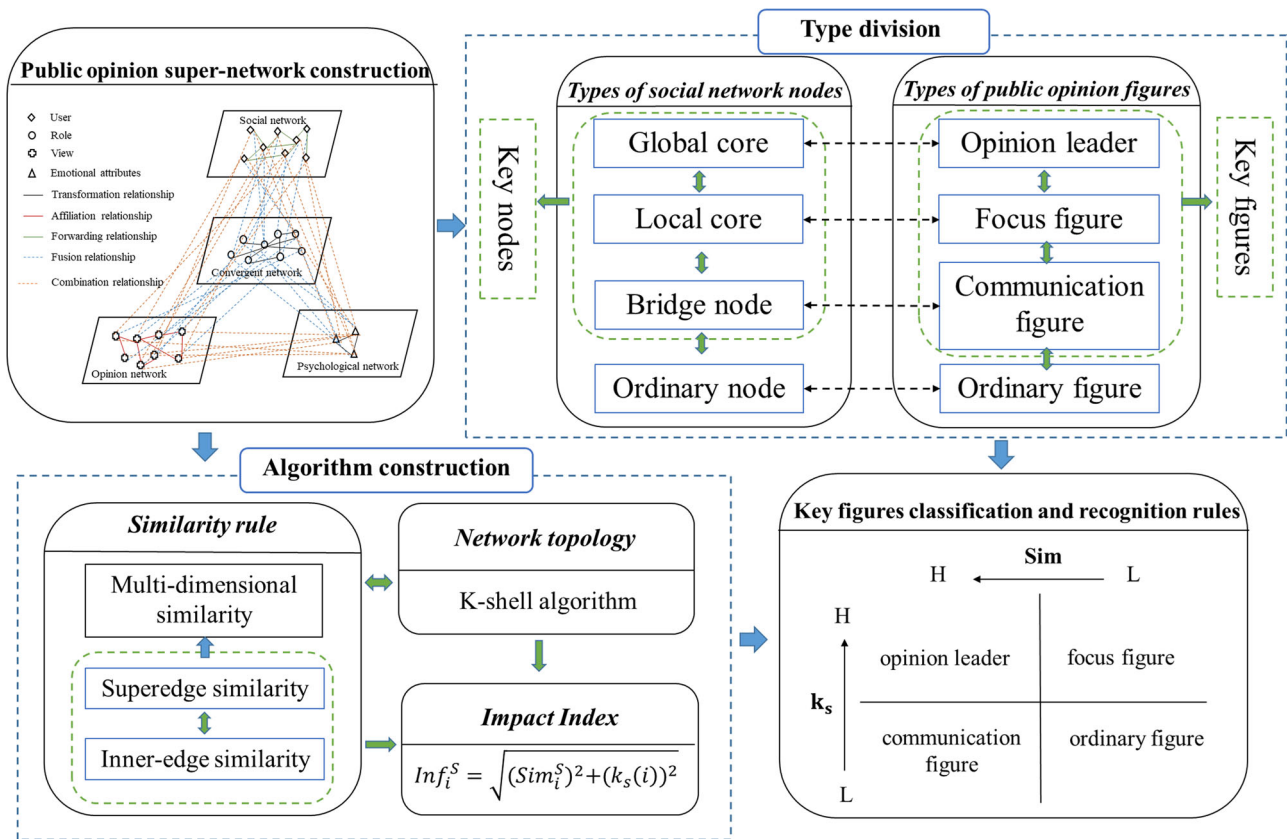
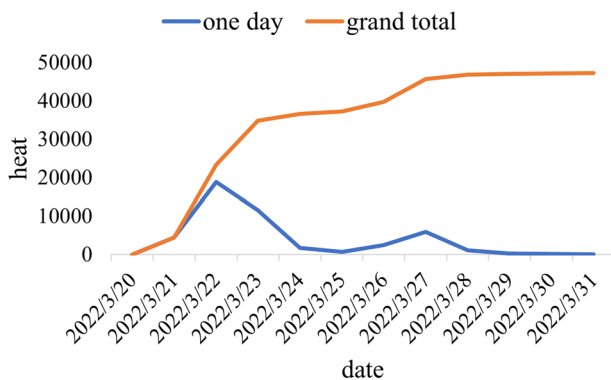**Fig. 9** Classification and recognition mechanism of key figures based on the public opinion supernetwork.

Fig. 10 Event development life cycle.

technology, we classify Weibo subject emotions into negative, neutral and positive. After that, the emotional intensity of each emotion is scored. Table 3 shows the results of the emotional analysis of some Weibo texts. The range of positive emotion intensity is (0, 1]. The value of neutral emotion intensity is 0. The range of positive emotion intensity is $[-1, 0)$.

Analysis of the topology structure of social network. Figure 12 shows the social network structure of each period. Its basic network topology characteristics are shown in Table 4, which are the number of nodes (n), the number of network edges (m), the average degree ($\langle d \rangle$), the maximum degree ($d_{\max}$), the average clustering coefficient (C) of the nodes in the network, the average shortest path between nodes ($\langle l \rangle$), the kernel value (k_s) and the level of modularisation (λ).

To illustrate the orientations, Table 2 shows some representative opinions and keywords selected from different opinion types. Figure 11 shows the relationship between core ideas and core keywords.

Sentiment analysis. Sentiment analysis is the procedure of analysing, processing, summarising, and reasoning with the use of subjective texts with emotional colour using natural language processing (NLP) methods. Intelligent classification algorithms such as machine learning can judge emotional polarity through corpus collection and part-of-speech tagging (Loureiro and Maria, 2020). BERT is a machine learning method. It is based on the bidirectional encoder representation technology of the converter and can be pretrained for plain text corpora. The original Weibo text is used as the training corpus in this study. Based on the BERT pretraining model, combined with semisupervised

Key figure recognition results and evolution analysis. Table 5 shows the phased recognition results based on the classification and recognition algorithm for key figures proposed in this article. Due to the article's length limitations, this study only showcases the top three key figures of each type. Taking the public opinion network during the explosion period as an example, the opinion leaders are users with high values of *Inf*, such as "China News Agency" ($Inf = 1.15$), "Mo Chen Mo Chen" ($Inf = 1.13$), and "China Daily" ($Inf = 1.10$). They are the global core nodes of social networks. The focus figures are users with higher k_s ($k_s = 3$) but lower *Sim*, such as "Modern Express", "Li Sweet Sauce", and "China News Network". They are the local core nodes of social networks. Communication figures are users with high *Sim* but low k_s ($k_s < 3$), such as "mind recorder", "big Qiqi is a big Qipa", and "Yabo flavour Pop Rocks". They are the "bridge" nodes to realise the cross-regional spread of public opinion in social networks.

Table 2 Partial results of opinion recognition.

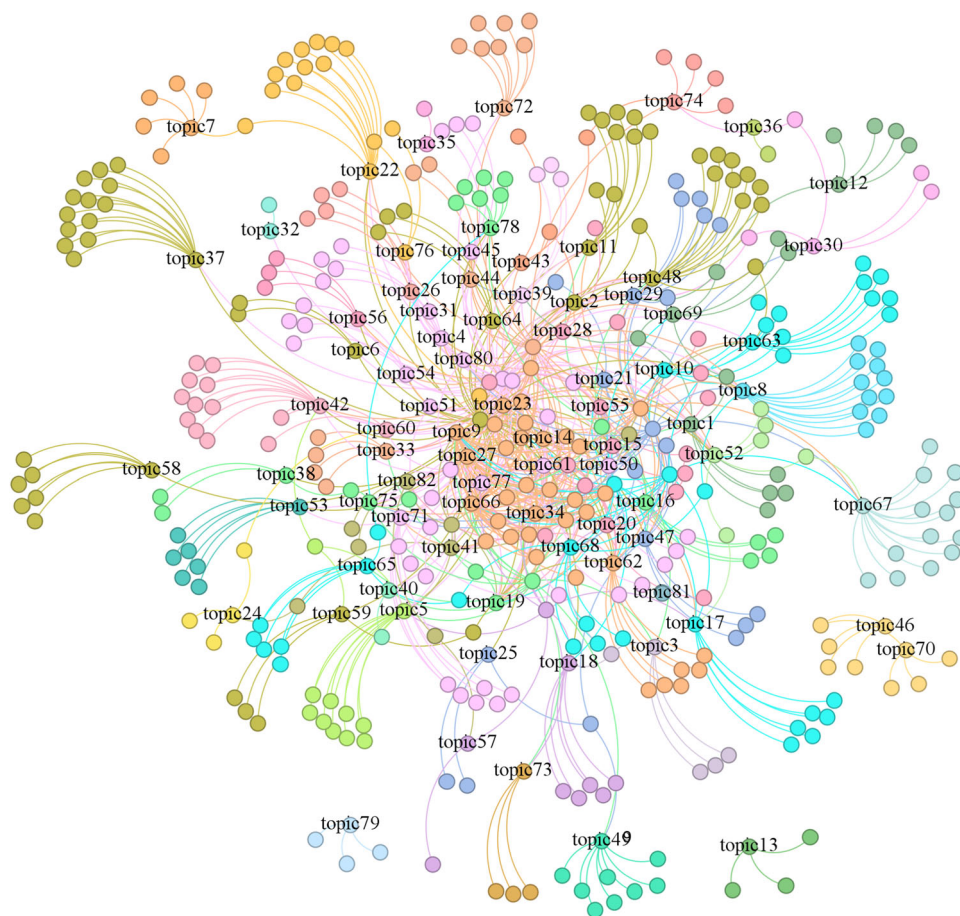| Keywords	Opinion	Opinion type
black box, China Eastern, crash, video, search, search, rescue, full force	Efforts to search for the black box of the crashed plane of China Eastern Airlines	Event discussion
life, awe, freshness, disaster, hold, heart	In the face of disaster, there should be a sense of awe	Positive view
family, media, news reports, mourning, distress, psychological, comfort, rescue	Psychological comfort for family members is another rescue	Neutral view
bottom line, human nature, tragedy, hold, malicious, business, traffic, disaster, consumption, marketing, eye-catching, heart-wrenching, heinous	Disaster marketing behaviour is outrageous	Negative view

Fig. 11 Keyword-opinion correlation diagram.

From the perspective of subject roles and node types, opinion leaders and focus figures generally have a larger self-core network, while communication figures are usually bridging nodes connecting various local area networks. As Fig. 12 shows, the blue nodes in the picture are opinion leaders, the orange nodes are focus figures, and the green nodes are communication figures. Many pink nodes are gathered around the key nodes, which are ordinary nodes that form a forwarding relationship with the key nodes. According to the network topology, opinion leaders gather around the largest number of ordinary nodes and are directly connected with a large number of focus and communication figures. This shows that opinion leaders and their neighbours' influence is extremely high. Opinion leaders play a core role in guiding, controlling and influencing global network development trends and information transmission. They are the core nodes of the global network. Focus figures can form a self-core network within a certain range. However, the network is usually only connected to a limited number of communication figures, which will limit its overall communication influence. Therefore, the focus figure is the core node of the local area network, which only controls and dominates the public opinion dissemination of a certain local network. As the SH spanners between opinion leader and the focus figures, the communication figures act as intermediaries to connect the local subnetworks. They guide the spread of information across regions and generally do not form a large self-core network. Therefore, the communication figure is the bridge node in the social network.

Model checking. Evaluating the question of whether the key node identification results are accurate has a direct impact on the detection and early warning of public opinion risks and the guidance of public opinion trends. Therefore, it is necessary to test the accuracy of the evaluation results. In this paper, AUC analysis, network destructive experiments, and fine-grained identification inspection are used to verify the model sensitivity

Table 3 Partial results of emotion score of Weibo text.

Weibo text	Affective type	Emotional intensity
This year's epidemic, war, and air crashes all remind us that the most precious thing is safety and health. Pray for a miracle and peace in the China Eastern Airlines crash.	Positive	0.9506
This time the peacock flies southeast and will never fly back.	Neutral	0
Shocking conspiracy: China Eastern Airlines flight weird vertical fall! Do you still dare to take a Boeing plane that can be controlled remotely backstage?	Negative	−0.8471

(a) Gestational period (b) Explosion period

(c) Duration period (d) Recovery period

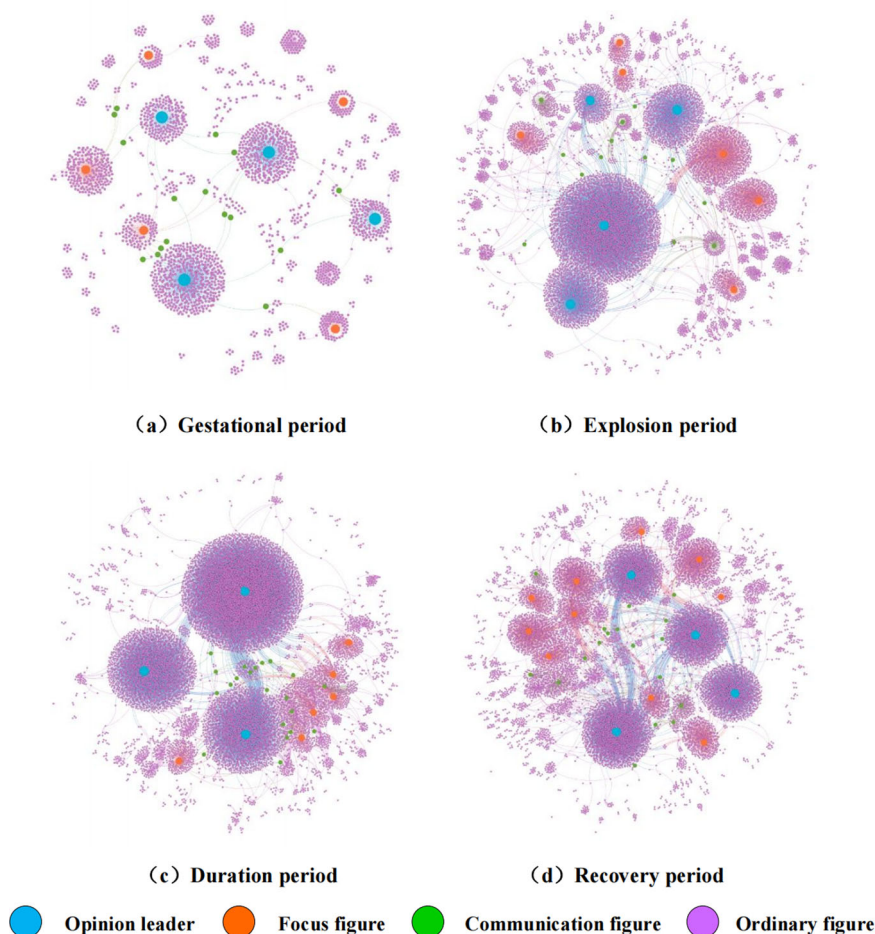● Opinion leader ● Focus figure ● Communication figure ● Ordinary figure

Fig. 12 Social network structure of different periods. a Social network structure of gestational period (6 h). **b** Social network structure of explosion period (6–28 h). **c** Social network structure of duration period (28–72 h). **d** Social network structure of recovery period (4–10 days).

and validity of the recognition results in the section "Key figure REcognition Results and Evolution Analysis".

AUC analysis. The calculation rule of the area under the curve (AUC) is simple, and the result is intuitive. It is a commonly used evaluation method to measure the classification algorithm's ability to distinguish categories and to test the accuracy of the model. It is widely used in public opinion analysis and information security (Wang et al., 2021; Goel and Sharma, 2021; Karoui et al., 2022). Generally, the higher the AUC is, the better the performance of the algorithm. In this study, AUC is applied to evaluate the rationality of the classification and identification algorithm for key figures. For our research goal, AUC can be regarded as the probability that the score of randomly selected key figures is higher than that of randomly selected non-key people in the test set. In this study, the CI, forwarding volume, degree centrality, multidimensional similarity index and K-shell algorithm are selected as the baseline methods to identify the

key figures of public opinion in the gestation period, outbreak period, duration period and recovery period. As shown in Fig. 13, the AUC values of each algorithm are different in different periods. However, the AUC value of the influence index is significantly higher than that of the baseline method. For example, the AUC value of classification and recognition (0.7229) in the explosion period is significantly higher than that of CI (0.7152), forwarding volume (0.4390), degree centrality (0.6711), K-shell (0.6416) and multidimensional similarity (0.5536). To some extent, this verifies the sensitivity of the influence index in key figure identification.

Network destructive experiment. Node influence can be reflected by simulating the degree of network damage caused by node failure. We use the "selective attack" strategy to conduct network destructive experiments. The "attack" referred to in this study is a simulated state. It is a virtual scenario where a node is unable to participate in the dissemination of public opinion information

Table 4 Basic topology information of social networks in different periods.

Period	n	m	$\langle d \rangle$	d_{max}	C	$\langle l \rangle$	k_s	λ
Gestational period	2110	2024	1.903	433	0.019	5.226	2	0.893
Explosion period	10246	10429	1.984	4134	0.002	4.064	3	0.792
Duration period	14913	15063	1.998	7557	0.007	3.480	3	0.680
Recovery period	13738	14175	2.009	2610	0.033	4.419	3	0.882

Table 5 Identification of key figures by period.

Period: Gestational period	Sim	k_s	Inf	Period: Explosion period	Sim	k_s	Inf	Role	Node type
Modern Express	0.99	2	1.41	China News Agency	0.56	3	1.15	Opinion leader	Global Core
Gibberish	0.69	2	1.22	Mo Chen Mo Chen	0.52	3	1.13	Opinion leader	Global Core
Sai Lei, three minutes	0.53	2	1.13	China Daily	0.46	3	1.10	Opinion leader	Global Core
Weibus	0.27	2	1.03	Modern Express	0.28	3	1.04	Focus figure	Local core
China News Network	0.26	2	1.03	Li Sweet Sauce	0.27	3	1.04	Focus figure	Local core
Blue Whale Financial Journalist Work Platform	0.25	2	1.03	China News Network	0.26	3	1.03	Focus figure	Local core
Deer singing string song	0.47	1	0.47	Ideological recording body	0.67	2	0.83	Communication figure	Bridge
Dismantling CT	0.39	1	0.39	Daqiqi is Daqipa	0.48	2	0.69	Communication figure	Bridge
Internet Junming Theory	0.33	1	0.33	Yeah, spinach flavour Pop Rocks	0.67	1	0.67	Communication figure	Bridge

Period: Duration period	Sim	k_s	Inf	Period: Recovery period	Sim	k_s	Inf	Role	Node type
People's Daily	0.80	3	1.28	CCTV news	0.87	3	1.33	Opinion leader	Global Core
Bian Menglong's Long-Legged Liu Ergou	0.67	3	1.20	Langfang Toutiao	0.63	3	1.18	Opinion leader	Global Core
The Great White of Brain Card	0.67	3	1.20	It's not someone's pretty girl	0.63	3	1.18	Opinion leader	Global Core
Dismantling CT	0.44	3	1.09	Sina News	0.34	3	1.06	Focus figure	Local core
New Beijing News Shell Finance	0.33	3	1.05	Diligent Brother	0.32	3	1.05	Focus figure	Local core
China Chang'an Network	0.32	3	1.05	People's Daily	0.28	3	1.04	Focus figure	Local core
The Connotation of Wind_46992	0.67	2	0.83	Nibelungen's Dream	0.67	2	0.83	Communication figure	Bridge
Litterstar33	0.67	2	0.83	Penguin 9156	0.67	2	0.83	Communication figure	Bridge
21st Century Economic Report	0.67	1	0.67	Rule of Law in Sichuan	0.67	1	0.67	Communication figure	Bridge

Note: *Sim* is the multidimensional similarity index, k_s is the K-shell index, and *Inf* is the influence index.
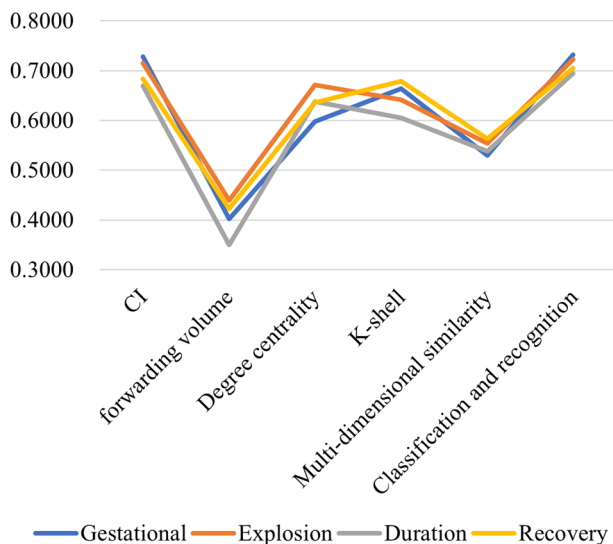

Fig. 13 AUC values of various algorithms in different periods.

due to force majeure (such as prohibition and post-deletion). In network destructive experiments, the faster the network is destroyed, the higher the node failure rate, indicating that the attacked node is more important to the network. Assuming that the node failure rule in a social communication network is: when a node is attacked, it immediately fails, and the edges connected to that node will also fail in the network.

We select Weibo text messages in the explosion period (6–28 h) to experiment on social networks according to six algorithms: CI, forwarding volume, degree centrality, K-shell, multidimensional similarity, classification and recognition. We use network destructive experiments to attack some of the top nodes identified by the six algorithms in turn. The importance of nodes is reflected by observing the degree of network damage in the failure scenario of this part of the nodes. The degree of network disruption is calculated by the ratio of network connectivity after disruption to initial network connectivity.

Figure 14 shows the degree of network damage of six algorithms over four periods in different network destructive experiments (attacking key nodes in the top 1‰, 2‰, and 5‰ of the network). Under the three attack ratios, the degree of network damage of the classification and recognition algorithm in each period is greater than that of the four algorithms: forwarding
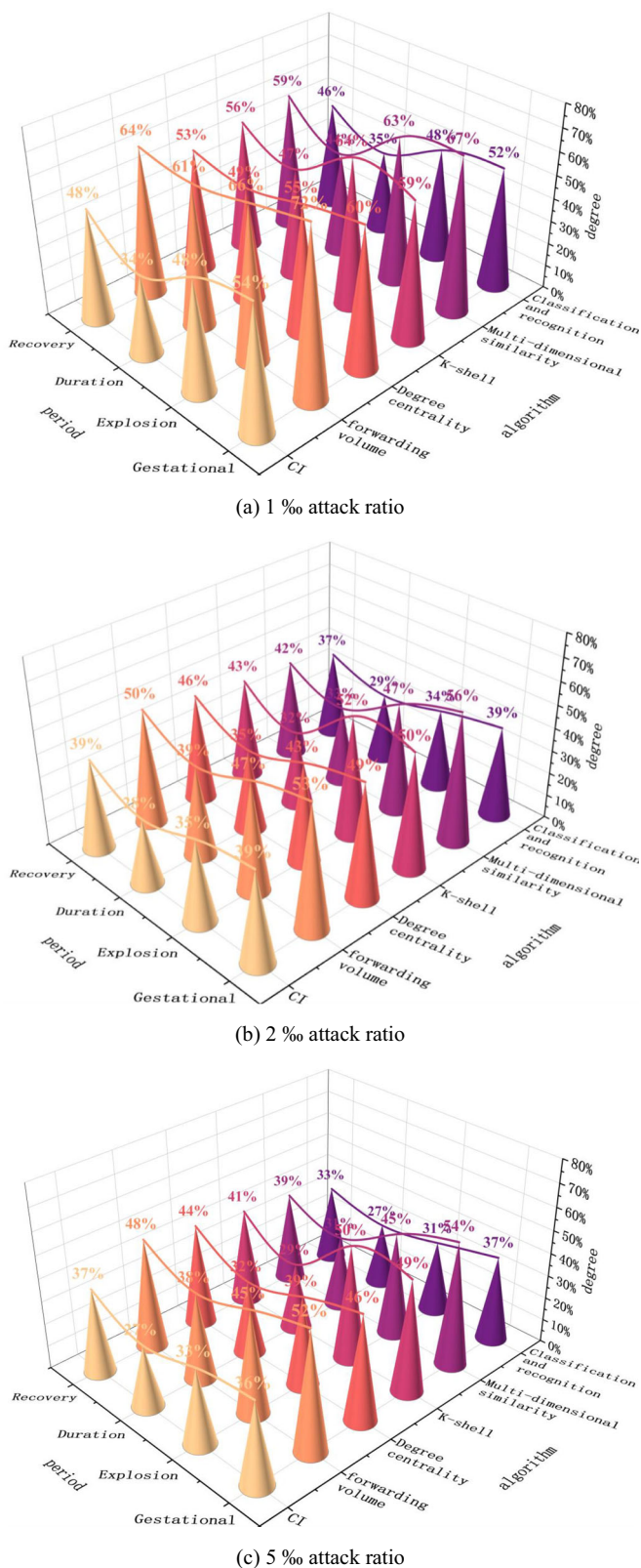
(a) 1 ‰ attack ratio

(b) 2 ‰ attack ratio

(c) 5 ‰ attack ratio

Fig. 14 Performance of different recognition methods in network destructive experiments. a Network destructive experiment at 1‰ attack ratio. **b** Network destructive experiment at 2‰ attack ratio. **c** Network destructive experiment at 5‰ attack ratio. All the network destructive experiments show the experimental results of six algorithms in four periods. The six algorithms are CI, forwarding volume, degree centrality, K-shell, multidimensional similarity, classification and recognition. The four periods are gestational period, explosion period, duration period and recovery period.

volume, degree centrality, K-shell, and multidimensional similarity. During the duration of the 1‰ attack ratio, 2‰ attack ratio, and 5‰ incubation period, the degree of CI destruction exceeds that of the classification and recognition algorithms. During the outbreak period, under a 1 ‰ attack ratio, the network connectivity obtained from the network destructive experiments of six algorithms: CI, forwarding volume, degree centrality, K-shell, multidimensional similarity, classification, and recognition were 48%, 66%, 55%, 64%, 63%, and 48% of the initial network connectivity, respectively. This indicates that the public opinion key figure classification and recognition algorithm proposed in this study has certain advantages in identifying key figures, demonstrating higher effectiveness and better performance.

Identifying fine-grained validation. This article selects Weibo text information from the explosive period (6–28 h). It compares the key figure recognition results under the CI, forwarding volume, centrality, K-shell, multidimensional similarity, and the classification and recognition algorithm mentioned in this article. From Table 6, it can be seen that CI, forwarding volume, centrality, K-shell, and multidimensional similarity can only identify public opinion leaders without considering the division of key figures. The CI method, an algorithm that comprehensively considers the influence of nodes and neighbouring nodes, performs well in identifying key nodes. However, it ignores the influence of implicit attributes such as emotions and viewpoints between node neighbourhoods and the differences in the degree of connectivity between neighbourhoods, resulting in certain limitations in its recognition of "bridge" nodes. The forwarding volume method is simple in calculation but has low recognition accuracy. However, there is a nonlinear correlation between the number of node connections and the dynamic process. The recognition effect of purely linear indicators such as forwarding volume is relatively poor, and it is easy to miss key nodes with low forwarding volume but significant influence from neighbouring nodes, such as "Mo Chen Mo Chen". As a local metric, the degree centrality method can easily confuse the entire network core with the local core, mistakenly identifying local cores such as the "Li Sweet Sauce" as the entire network core. Additionally, it cannot effectively identify key figures with lower centrality, such as "Mo Chen Mo Chen". The K-shell algorithm has a large number of nodes with the same sort, making it impossible to perform functional classification on key nodes, and it is prone to missing important "bridge" nodes. The multidimensional similarity index cannot effectively filter out class core nodes with small influence, making it easy to identify "bridge" nodes as opinion leaders, such as "thought recorders," which will also lead to omitting global core nodes in social networks, such as "China Daily". The classification and recognition algorithm proposed in this article reflects both the network position of nodes in social networks and their influence on the emotions, opinions, and other fields of the leader, effectively improving the fine-grained recognition of key nodes. It can not only identify traditional core nodes but also effectively filter core nodes, identifying functional core nodes as important SH spanners connecting local subnetworks, which denotes "bridge" nodes with a small network centrality that plays an important role in information transmission.

Conclusion and discussion
This study applies supernetwork analysis to identify key figures in public opinion. A classification and recognition algorithm based on a multidimensional similarity algorithm and K-shell index is proposed. A specific case is used to further verify the reliability of the algorithm. The results of the study are as follows:

(1)   The dissemination of public opinion events is influenced by key figures from various roles. Opinion leaders play a core role in

Table 6 Key node identification results of different algorithms.

Algorithm	CI	Forwarding volume	Degree centrality	K-shell	Multidimensional similarity	Classification and recognition
Opinion leaders	China News Agency Mo Chen Mo Chen China Daily China Chang'an Network Modern Express	China News Agency China Daily China Chang'an Network China News Network Communist Youth League Central Committee	China News Agency China Daily Communist Youth League Central Committee Li Sweet Sauce	China News Agency Mo Chen Mo Chen China Daily Modern Express Li Sweet Sauce	Ideological recording body Yeah, spinach flavour Pop Rocks China News Agency Mo Chen Mo Chen Modern Express	China News Agency Mo Chen Mo Chen China Daily China Chang'an Network
Focus figures						Modern Express Li Sweet Sauce China News Network
Communication figures						Ideological recording body Daqiqi is Daqipa Yeah, spinach flavour Pop Rocks

guiding, controlling, and influencing the development direction and information transmission of the global network. They are the core global network nodes. The focus figures can be the core node of a local network, controlling and leading the dissemination of public opinion in a certain local network. The communication figures act as intermediaries connecting various local subnetworks and guide information to spread across regions, which can be seen as SH spanners in the network.

(2) In this study, a key figure classification and recognition algorithm based on the topology and node attribute characteristics of a public opinion supernetwork model, which combines a multidimensional similarity algorithm and the K-shell index, is proposed. It can not only simultaneously identify multiple types of key figures, such as opinion leaders, focus figures, and communication figures, but also outperforms baseline algorithms in terms of sensitivity and effectiveness through validation analysis.

Future work can be carried out with the following aspects. (1) We can try to further subdivide the key figures in public opinion. For example, we can continue to explore nodes that do not play a leading role at this stage but have a certain probability of growing into key figures in the future, such as potential communication figures and potential opinion leaders. The effectiveness of public opinion governance can be improved by preempting the opportunity to intervene in advance. (2) The application of the classification and recognition algorithm in identifying key topics can be continuously verified. In addition, the applicability of the model can be improved. A single public opinion event can involve multiple viewpoints and topics. Timely and accurate access to the topic information of key figures is crucial for effective public opinion risk management and control. (3) We plan to explore machine learning methods based on supernetwork by combining multidimensional complex network structures, interactive information, and unsupervised learning processes. In the future, it can automatically identify and analyse user behaviour, public opinion topics, public opinion analysis, and evolution trends.

Data availability
The datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

References
Abu-Salih B, Chan KY, Al-Kadi O, Al-Tawil M, Wongthongtham P, Issa T, Saadeh H, Al-Hassan M, Bremie B, Albahlal A (2020) Time-aware domain-based social influence prediction. J Big Data 7(1):1–37. https://doi.org/10.1186/s40537-020-0283-3

Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. J Econ Perspect 31(2):211–235. https://doi.org/10.1257/jep.31.2.211

An L, Zhou W, Ou M, Li G, Wang X (2021) Measuring and profiling the topical influence and sentiment contagion of public event stakeholders. Int J Inf Manag 58(7):102327. https://doi.org/10.1016/j.ijinfomgt.2021.102327

Bamakan S, Nurgaliev I, Qu Q (2019) Opinion leader detection: a methodological review. Expert Syst Appl 115:200–222. https://doi.org/10.1016/j.eswa.2018.07.069

Barnes SJ (2020) Information management research and practice in the post-COVID-19 world. Int J Inf Manag 55(2020):102175. https://doi.org/10.1016/j.ijinfomgt.2020.102175

Bhowmik T, Niu N, Singhania P, Wang W (2015) On the role of structural holes in requirements identification: an exploratory study on open-source software development. ACM Trans Manag Inf Syst (TMIS) 6(3):1–30. https://doi.org/10.1145/2795235

Bickle A (2013) Cores and shells of graphs. Math Bohem 138(1):43–59. https://doi.org/10.21136/MB.2013.143229

Biran O, Rosenthal S, Andreas J, Mckeown K, Rambow O (2012) Detecting influencers in written online conversations. In Proceedings of the Second Workshop on Language in Social Media (LSM2012). Association for Computational Linguistics. pp. 37–45

Bliss CA, Frank MR, Danforth CM, Dobbs PS (2014) An evolutionary algorithm approach to link prediction in dynamic social networks. J Comput Sci 5(5):750–764. https://doi.org/10.1016/j.jocs.2014.01.003

Boot AB, Dijkstra K, Zwaan RA (2021) The processing and evaluation of news content on social media is influenced by peer-user commentary. Humani Soc Sci Commun 8(1):209. https://doi.org/10.1057/s41599-021-00889-5

Corradini E, Nocera A, Ursino D, Virgili L (2021) Investigating negative reviews and detecting negative influencers in yelp through a multi-dimensional social network based model. Int J Inf Manag 60(3):102377. https://doi.org/10.1016/j.ijinfomgt.2021.102377

Deffuant G, Amblard F, Weisbuch G (2004) Modelling group opinion shift to extreme: the smooth bounded confidence model. Physics. 1–12. https://doi.org/10.48550/arXiv.cond-mat/0410199

D'Arcangelis AM, Levantesi S, Rotundo G (2021) A complex networks approach to pension funds. J Bus Res 129:687–702. https://doi.org/10.1016/j.jbusres.2019.10.071

Goel R, Sharma R (2021) Studying leaders & their concerns using online social media during the times of crisis—a COVID case study. Soc Netw Anal Min 11(46). https://doi.org/10.1007/s13278-021-00756-w

Hu X, Tang J, Gao H, Liu H (2013) Unsupervised sentiment analysis with emotional signals. Proceedings of the 22nd International Conference on World Wide Web. 607–618. https://doi.org/10.1145/2488388.2488442

Jiang L, Hoegg J, Darren WD, Chattopadhyay A (2010) The persuasive role of incidental similarity on attitudes and purchase intentions in a sales context. J Consum Res 36(5):778–791

Karoui WF, Nesrin HF, Lotfi BR (2022) Machine learning-based method to predict influential nodes in dynamic social networks. Soc Netw Anal Min 12(1):1–18. https://doi.org/10.1007/s13278-022-00942-412.1

Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA (2010) Identification of influential spreaders in complex networks. Nat Phys 6(11):888–893. https://doi.org/10.1038/nphys1746

Kushwaha AK, Kar AK, Dwivedi YK (2021) Applications of big data in emerging management disciplines: A literature review using text mining. Int J Inf Manag Data Insights 2021(2):100017. https://doi.org/10.1016/J.JJIMEI.2021.100017

Leung CK, Cuzzocrea A, Mai JJ, Deng D, Jiang F (2019) Personalized DeepInf: enhanced social influence prediction with deep learning and transfer learning. In: 2019 IEEE International Conference on Big Data (big data). 2871–2880. https://doi.org/10.1109/BigData47090.2019.9005969

Li CL, Bai JP, Zhang L, Tang HL, Luo YL (2019) Opinion community detection and opinion leader detection based on text information and network topology in cloud environment. Inf Sci 504:61–83. https://doi.org/10.1016/j.ins.2019.06.060

Li W, Fan Y, Mo J, Liu W, Wang C, Xin M, Jin Q (2020) Three-hop velocity attenuation propagation model for influence maximization in social networks. World Wide Web 23(2):1261–1273. https://doi.org/10.1007/s11280-019-00750-5

Liben-Nowell D, Kleinberg J (2003) The link prediction problem for social networks. Association for Computing Machinery International Conference on Information and Knowledge Management

Lichtenwalter RN, Lussier JT, Chawla NV (2010) New perspectives and methods in link prediction. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

Loureiro ML, Maria Alló (2020) Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the U.K. and Spain. Energy Policy 143:111490. https://doi.org/10.1016/j.enpol.2020.111490

Lu LY, Chen DB, Ren XL, Zhang QM, Zhang YC, Zhou T (2016) Vital nodes identification in complex networks. Phys Rep. 650:1–63. https://doi.org/10.1016/j.physrep.2016.06.007

Ma N, Liu Y (2014) SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork. Expert Syst Appl 41(4):1357–1368. https://doi.org/10.1016/j.eswa.2013.08.033

Maji G, Dutta A, Malta MC, Sen S (2021) Identifying and ranking super spreaders in real world complex networks without influence overlap. Expert Syst Appl 179:115061. https://doi.org/10.1016/j.eswa.2021.115061

Michaels CL (2002) Circle communication: an old form of communication useful for 21st century leadership. Nurs Adm Q 26(5):1–10. https://doi.org/10.1097/00006216-200210000-00004

Morone F, Makse H (2015) Influence maximization in complex networks through optimal percolation. Nature 524:65–68. https://doi.org/10.1038/nature14604

Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP (2010) Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. Science 328(5980):876–878. https://doi.org/10.1126/science.1184819

Nagurney A, Wakolbinger T (2005) Supernetworks: an introduction to the concept and its applications with a specific focus on knowledge supernetworks. Int J

Knowl Cult Change Manag Ann Rev 2005(4):1–16. https://doi.org/10.18848/1447-9524/CGP/v04/50227

Nagurney A, Dong J (2002) Supernetworks: decision-making for the information age. Edward Elgar Publishers

Nolasco D, Oliveira J (2020) Mining social influence in science and vice-versa: a topic correlation approach. Int J Inf Manag 51(4):102017. https://doi.org/10.1016/j.ijinfomgt.2019.10.002

Pang B, Lee L, Vaithyanathan S (2002) Thumbs up?: sentiment classification using machine learning techniques. Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Stroudsburg. pp. 79–86

Phan N, Dou D, Piniewski B, Kil D (2016) A deep learning approach for human behavior prediction with explanations in health social networks: social restricted Boltzmann machine (SRBM+). Soc Netw Anal Min 6(1):79. https://doi.org/10.1007/s13278-016-0379-0

Qiu J, Tang J, Ma H, Dong Y, Wang K, Tang J (2018) Deepinf: social influence prediction with deep learning. In The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD2018). Association for Computing Machinery. pp. 2110–2119

Ramanathan V, Meyyappan T (2019) Prediction of individual's character in social media using contextual semantic sentiment analysis. Mob Netw Appl 24(6):1763–1777. https://doi.org/10.1007/s11036-019-01388-3

Rezvani M, Liang W, Xu W, Liu C (2015) Identifying top-k structural hole spanners in large-scale social networks. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (ACM2015). Association for Computing Machinery. pp. 263-272

Shi Q, Wang C, Chen J, Feng Y, Chen C (2019) Location driven influence maximization: Online spread via offline deployment. Knowl Based Syst 166(15):30–41. https://doi.org/10.1016/j.knosys.2018.12.003

Shi WZ, Zeng F, Zhang A, Tong C, Shen X, Liu Z (2022) Online public opinion during the first epidemic wave of COVID-19 in China based on Weibo data. Humanit Soc Sci Commun 9:159. https://doi.org/10.1057/s41599-022-01181-w

Singh R, Ho SY (2000) Attitudes and attraction: a new test of the attraction, repulsion and similarity-dissimilarity asymmetry hypotheses. Br J Soc Clin Psychol 39(2):197–211. https://doi.org/10.1348/014466600164426

Sosa J, Buitrago L (2021) A review of latent space models for social networks. Rev Colombiana de Estad 44(1):171–200. https://doi.org/10.15446/rce.v44n1.89369

Tian Z, Bai W, Wu B, Zhu C (2012) Topic correlation and individual influence analysis in online forums. Expert Syst Appl 39(4):4222–4232. https://doi.org/10.1016/j.eswa.2011.09.112

Veirman MD, Cauberghe V, Hudders L (2016) Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude. Int J Advert 36(1):1–31. https://doi.org/10.1080/02650487.2017.1348035

Wang GH, Wang YF, Li JM, Liu KD (2021) A multidimensional network link prediction algorithm and its application for predicting social relationships. J Comput Sci 53:1877–7503. https://doi.org/10.1016/j.jocs.2021.101358

Wang W, Street WN (2018) Modeling and maximizing influence diffusion in social networks for viral marketing. Appl Netw Sci 3(6). https://doi.org/10.1007/s41109-018-0062-7

Wang Y, Li H, Guan J, Liu N (2019) Similarities between stock price correlation networks and co-main product networks: Threshold scenarios. Phys A: Stat Mech Appl 516:66–77. https://doi.org/10.1016/j.physa.2018.09.154

Xiong F, Liu Y, Zhu J, Zhang YC, Zhang Y (2011) A dissipative network model with neighboring activation. Eur Phys J B 84(1):115–120. https://doi.org/10.1140/epjb/e2011-20286-7

Yang Y, Tang J, Leung C, Sun Y, Chen Q, Li J, Yang Q (2015) RAIN: Social role-aware information diffusion. Twenty-Ninth AAAI Conference on Artificial Intelligence. pp. 367–373

Yoo E, Rand W, Eftekhar M, Rabinovich E (2016) Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises. J Oper Manag 45(5):123–133. https://doi.org/10.1016/j.jom.2016.05.007

Yu Y, Jing M, Zhao N, Zhou T (2020) Lowest degree decomposition of complex networks. https://doi.org/10.48550/arXiv.2002.05358

Zeng A, Zhang CJ (2013) Ranking spreaders by decomposing complex networks. Phys Lett A 377(14):1031–1035. https://doi.org/10.1016/j.physleta.2013.02.039

Zhang W, Wang M, Zhu Y (2020) Does government information release really matter in regulating contagion-evolution of negative emotion during public emergencies? from the perspective of cognitive big data analytics. Int J Inf Manag 50:498–514. https://doi.org/10.1016/j.ijinfomgt.2019.04.001

Zhao YY, Kou G, Peng Y, Chen Y (2018) Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective. Inf Sci 426:131–147. https://doi.org/10.1016/j.ins.2017.10.031

Zhu J, Zhu L, Bao CM, Zhou LH, Wang CY, Kong B (2018) Top-k structure holes detection algorithm in social network. In 2018 9th International Conference

on Information Technology in Medicine and Education (ITME2018). IEEE. pp. 1064–1071

Author contributions

GW and YW: conceptualisation, methodology, resources, writing, validation, revisions, formal analysis and data compilation. KL and SS: review, editing, and supervision.

Competing interests

The authors declare no competing interests.

Ethical approval

Ethical approval was not required as the study did not involve human participants.

Informed consent

Informed consent was not required as the study did not involve human-related content.

Additional information

Correspondence and requests for materials should be addressed to Yushan Wang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.