JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Influence Index Evolution and Empirical Study of Opinion Leaders in Chinese Social Networks

Cheng Zhang*

Abstract
An influence drift index (IDI) and an influence stability index (ISI) of opinion leader nodes are proposed to capture the dynamic growth trajectory of opinion leaders in the process of public opinion evolution. The multidimensional features of user nodes are extracted to construct the influence index, and the change amplitude of the influence index in the time dimension is quantified, and this forms the basis for an influence index online evolution (INFOE) method. In this method, opinion leaders' influence is regarded as the feature vector of dynamic change. The IDI for the opinion leader-audience layer is established from the three dimensions of contribution degree, recognition degree and dissemination degree of opinion leaders. Second, fuzzy formal concept analysis is used to establish the formal concept of opinion leaders and its poset to construct an ISI for the opinion leader-audience layer. Finally, the fluctuation degree of the influence index of opinion leaders in the evolution of public opinion content is quantified to realize the online evolution of this index. INFOE is verified and analyzed in this case. The empirical results show that the IDI reveals a certain correlation between the influence index of opinion leaders and the heat of public opinion in the process of event evolution, which can be used to monitor the time node of public opinion outbreaks accurately. The ISI reflects the fuzzy dependence of the influence index among opinion leader nodes, which provides a new theoretical exploration and research perspective for the multi-angle subdivision of opinion leaders.

1. Introduction

With the rapid development of streaming media technologies, generated opinions are characterized by rapid dissemination, frequent interaction and fast expression [1]. The arbitrary and flat virtual social network easily generates false and undesirable emotional feedback, which eventually evolves into an irreversible public opinion crisis. An opinion leader, as a network subject that expresses the will of the public and guides the emotional attitudes of the public in social networks, plays a guiding role in accelerating the spread of public opinion and controlling the trend of hot spots. In these cases, identifying opinion leaders in online communities and quantifying their nodal influence is crucial for improving the control and warning of the ability of online public opinion. However, hot network events, such as topic drift [2], emotional evolution [3], and public opinion evolution [4], have strong periodic change characteristics in the time dimension in complex network environments. The dynamic changes in online

public opinion lead to genetic and variation characteristics in opinion leaders' influence [5]. The accompanying high uncertainty of opinion leader identification poses new challenges to public opinion management in two ways. First, audience groups produce a word-of-mouth effect on opinion leaders' influence, which has the possibility of multilayer bursts in the opinion transmission path. Second, opinion leaders are internally connected, which slows or stops influence. Thus, the construction of relevant indices for public opinion monitoring has important theoretical significance for determining the dynamic growth trajectory of opinion leaders and mapping public opinion.

Currently, methods for identifying opinion leaders can be divided into two main categories, namely, feature recognition based on user attributes (FRUA) and feature recognition based on user relationships (FRUR). FRUA can be used to calculate the user influence index by extracting single- or multi-dimensional feature indices such as the propagation degree, activity degree and support degree of the user nodes. From the perspective of user influence and activity, Wang et al. [6] established an index weight calculation model based on an analytic hierarchy process, effectively improving the identification effect of microblog opinion leaders. Chen et al. [7] realized accurate identification of opinion leaders by integrating the multi-dimensional features of users and text emotion calculations. FRUR takes the social network of user interaction as the core and uses social network analysis, topic analysis, and sentiment computing to quantify the information transmission and sentiment guidance of opinion leaders from different perspectives. Deng et al. [8] employed social network analysis to calculate the mediating centrality degree and proximity centrality degree of opinion leaders from the perspective of two-level information transmission to construct an interactive network. On the basis of the LDA model, Zhang et al. [9] constructed a network of forwarded Weibo comments about COVID-19, identifying topic leaders and their community topic transmission paths. In summary, the current research focuses on feature identification and the influence calculation of opinion leaders to solve the problem of the transmission and quantification of the influence on the audience. However, these methods do not consider the amplitude of the evolution of the influence index under the continuous time snapshot of online public opinions and lack an in-depth analysis of the influence traction relationships within communities.

On the basis of previous studies, the author focuses on solving three problems.

1) How can the life cycle information of event evolution be used to quantify the degree of drift of opinion leaders' influence?

2) How can the correlation calculation and evolution rate of nodal influence be studied from the perspective of interest similarity among opinion leaders?

3) What is the verification effect of the constructed drift index and stability index on typical topics and their applicability in online public opinion monitoring?

In summary, the influence drift index (IDI) and influence stability index (ISI) are constructed on the basis of information theory and fuzzy formal concept analysis and can be used to clarify the significance and role of the indicators in the dynamic identification of opinion leaders. In practice, the correlation between the change in the influence index and the popularity of events is verified through case verification and analysis of real events, as well as its effectiveness in segmenting the roles of opinion leaders.

The rest of this article is structured as follows: Section 2 presents the framework of influence index evolution. Section 3 presents an illustration of the empirical analysis by providing discussions of the influence index online evolution (INFOE) in the community, and Section 4 includes conclusions followed by acknowledgment and references.

2. Framework of the Influence Index for Opinion Leaders

An INFOE method is proposed through the construction of an IDI and an ISI, which can be subdivided into four stages, namely, the identification of potential opinion leaders, the construction of a drift index and stability index, and the dynamic analysis of influence. The characteristic values of potential opinion leaders' influence are quantified mainly by calculating the node influence of community users, including topic recognition and community topology mapping, node multi-feature analysis and influence calculation. The goal of the IDI is to highlight the change in the intensity of opinion leaders' influence on community audience users at different times. The process includes the analysis of the event evolution cycle and the construction of the IDI. The ISI reflects the degree of similarity between opinion leaders in the dimension of interest and explores the interactive dependence relationship of influence from the perspective of fuzzy formal concept analysis. The process includes calculation of interest similarity, fuzzy context mapping of opinion leaders and calculation of the stability index.

2.1 Identification of Potential Opinion Leaders

2.1.1 Public opinion topic identification and community topology mapping

Owing to the sparse features of short texts in network communities, the biterm topic model (BTM) [10] is adopted to cluster the topic content of datasets. Disordered word pairs are regarded as multiple topics for the probability distribution, and the probability distribution of document topics is inferred by semantic correlation between co-occurrence words. The BTM model is based on the assumption that the topic depends on the mixed distribution of the topic in the document, and some co-occurrence two-word pairs are selected with a certain probability. The topic extraction process of the network community is as follows. First, the distribution of topic words in the corpus is generated, and the corresponding topic distribution is generated. Afterward, it is assumed that any word pair in the set selects a certain topic according to the probability and extracts co-occurrence pairs under the topic. Finally, the topic distribution and word distribution in the stable state are obtained through sampling and iteration.

By clustering node sets with similar interest marks in the network of public opinion relationships, the network community forms multiple node relationship sets centering on a certain topic. Multiple associated user nodes can be mined, and relationship extraction among users can be realized on the basis of user interaction behavior through topic recognition of community text. Finally, a network containing nodes with different attributes and their relationships can be established. The process of community topology mapping is as follows. First, the BTM topic model is adopted to complete the topic identification of community text, realize the semantic mapping of user-topic and theme-feature words, and establish the user node set U under multiple topics. Second, the directed edges of the user relationships in node set U are calculated to realize the weight labeling of edge connections and complete the topology mapping of communities.

2.1.2 Multi-dimensional feature analysis

Opinion leaders can influence the mapping and evolution of community topology because of their dominant position in information transmission. In real scenarios, opinion leaders' influence changes with their own interest drift, which is reflected in fluctuations in the average distance and the appeal of influence in the process of carrying information [11]. In this work, opinion leaders' influence is regarded as an uncertain information variable that changes under the influence of random events. In terms of

organization and orderliness, this variable manifests as an increase or decrease in characteristic values. Following Shannon's method of information entropy to address uncertain information, the influence of user nodes in the public opinion network is taken as the amount of public opinion information, and its disorderly change within the scope of the discussion domain is interpreted as the entropy of public opinion information interacting between user nodes. The greater the entropy of public opinion information is, the greater the average amount of information that influences information after eliminating redundancy tends to be. Thus, the essence of opinion leader identification is to quantify how the correlation degree of node influence changes dynamically with the evolution of public opinion in the temporal dimension to select influence characteristic values on the basis of the entropy of public opinion information. In the process of changing the entropy of public opinion information, the evolution, strength of influence and correlation degree of influence of opinion leaders all reflect different trends in the dynamic identification process of opinion leaders. The tracking mechanism of the evolution of the influence index consists of the measurement and analysis of the abovementioned characteristic values of influence.

In summary, multidimensional influence characteristic values, including the behavioral influence index, emotional influence index and network influence index, are established through the integration of the contribution degree, recognition degree and propagation degree of user nodes and with reference to P-index theory [12]. The variables contained in the above influence index are the normalization numbers after the standardization of min–max. The behavioral influence index reflects the contribution of user nodes in the communication of public opinion, including the number of forwarded and commented-upon users' views. The calculation is shown in Formula (1), where C_f is the number of forwarded users' views, C_r is the number of commented users' views, and N_f and N_r are the total numbers of forwarded and commented users' views, respectively, within a certain period.

$$I_B = ((C_f + C_r)^2 / (N_f + N_r))^{1/3}. \tag{1}$$

The sentiment influence index reflects the support degree of the views held by the user node in the community user group. The larger the sentiment influence index is, the more users support the user. The sentiment influence index is calculated as shown in Formula (2), in which C_{su} is the total number of positive sentiment words published by opinion leaders and C_{op} is the total number of negative sentiment words published by opinion leaders. N_S is the total number of sentiment words within a certain time. The behavioral influence index reflects the contribution of user nodes to the dissemination of public opinion. The numerator represents the number of times that user viewpoints are shared and commented upon, with the square indicating a weighted emphasis on this quantity. The rationale for this choice is that the squared term signifies a nonlinear relationship regarding the importance of the aforementioned variables to the behavioral influence index. An increase in quantity has a disproportionately greater impact on overall performance. Furthermore, network effects or cumulative effects may exist between user behaviors such as sharing and commenting, wherein the execution of such actions could foster additional similar behaviors, thereby significantly enhancing the behavioral influence index. The squared term effectively encapsulates this positive feedback effect. The denominator denotes the total number of all user viewpoints that were shared and commented upon within a specific timeframe, with the use of the one-third power indicating a diminishing effect as this quantity increases; in other words, the contribution of this quantity to the behavioral influence index does not increase linearly. The rationale for this is that, in many instances, opinion leaders' influence does not exhibit a monotonically increasing trend. For example, an increase in online activity may lead to saturation effects, such that after surpassing a certain

threshold, the incremental impact of each additional action on overall influence diminishes. Thus, employing one-third of the power effectively reflects this characteristic.

$$I_S = ((C_{su} + C_{op})^2/(N_S))^{1/3}. \tag{2}$$

The network influence index analyzes the influence of user nodes from the perspective of the structural characteristics of community network topology, including degree centrality and intermediary centrality. Degree centrality measures the degree of interaction between user nodes by quantifying the number of other nodes that have edge connections with the node. Intermediary centrality reflects the degree of control of user nodes over other nodes by judging whether a node user is in the shortest path between other nodes. The calculation of the network influence index is shown in Formula (3), where C_d is the degree centrality of the user node, C_b is the number of intermediary centers of the user node, and N_d and N_b are the sums of the numbers of degree centers and intermediary centers of other nodes with edge connections to the user node, respectively.

$$I_N = ((C_d + C_b)^2/(N_d + N_b))^{1/3}. \tag{3}$$

2.1.3 Influence calculation

The quantitative measurement ability of information entropy, which takes the contribution ability, recognition degree and location characteristics of users as three dimensions to measure their influence, is used to calculate the influence of a user node. The influence calculation considers the user's behavioral appeal, emotional diffusion power and environmental communication power to obtain the user node u_i influence index at time t. The calculation is shown in Formula (4), where I_B, I_S and I_N are the behavioral influence index, emotional influence index and network influence index of u_i at time t, respectively.

$$INF_t(u_i) = -\log_2 \frac{1}{I_B + I_S + I_N + 1}. \tag{4}$$

2.2 Construction of the Drift Index

The influence index of opinion leaders is accompanied by changes in people, things, time, space and other factors in the communication of public opinions, resulting in changes in different speeds, intensities and relationships. From the perspective of the life cycle of public opinion information, the liquidity and growth of the influence index need to be studied to quantify the degree of influence drift of opinion leaders and position the best intervention time after the drift. As an important indicator for monitoring the evolution of opinion leaders' influence, the IDI can not only reflect the influence drift amplitude and evolution rate but also track the best access points of different forms of the influence index in the time series to depict the dynamic growth trajectory of opinion leaders.

The specific construction process of the IDI is as follows. First, the life cycle of public opinion events is divided according to time series, and the molecular datasets are divided on the basis of different evolution stages. Second, topic discovery and community topology mapping are carried out on the datasets, and the multidimensional characteristic attributes of the nodes are analyzed to calculate the user influence indices of different time nodes. Afterward, the value of the influence index under time snapshots is used to reflect the change range of influence of opinion leaders on audience groups. Finally, the IDI is calculated as shown in Formula (5), where $INF_{t_1}(u_i)$ and $INF_{t_2}(u_i)$ represent the influence indices of user ui under time nodes t_1 and t_2, respectively.

$$IDI(u_i) = \frac{INF_{t_1}(u_i) - INF_{t_2}(u_i)}{t_1 + t_2}. \tag{5}$$

2.3 Construction of the Stability Index

The opinion community can be regarded as a network of specific topics composed of several opinion leaders and audience groups. The network interaction behavior of opinion leaders reflects not only their social authority to seek information dissemination and public opinion guidance but also their status desire to express and share their own interests and preferences. As the intrinsic inducement of the explicit behavior of opinion leaders, interest is an important factor affecting their own behavior pattern. In a real scenario, opinion leaders' interest can be stimulated by various environmental factors, such as their own needs and the development of public opinion, resulting in dynamic changes such as interest drift and long-term and short-term interest renewal. Thus, the fuzzy correlation information within the opinion leader group can be quantified by calculating the interest similarity and exploring the interactive dependence relationship. In that case, knowledge discovery and role subdivision of opinion leaders in terms of the interest dimension can be realized.

The authors introduced fuzzy formal concept analysis (FFCA) [13] into the construction of an ISI by extracting the similarity of interest between opinion leaders. The multi-granularity analysis of the influence index was realized from the perspective of fuzzy relationships on the basis of the clustering of fuzzy relationships between opinion leaders. The construction of the IDI can be divided into three stages, namely, interest similarity calculation, opinion leader fuzzy context mapping and stability index calculation. The specific process is as follows. First, the user nodes with a high influence index are selected for the set of potential opinion leaders. Afterward, the degree of interest in the annotation of opinion leaders based on the term frequency-inverse document frequency (TF-IDF) model is calculated. Second, a fuzzy formal context containing opinion leaders and their similarity relationships is established to construct a fuzzy concept lattice. Finally, the strong correlation between generalization and specialization among opinion leaders is analyzed by defining and calculating the stability index oriented to the fuzzy concept of opinion leaders.

The TF-IDF model was used to calculate the degree of interest of opinion leaders in the document topic to depict the similarity of interest among opinion leaders accurately. The degree of interest closeness is subsequently calculated by cosine similarity. The calculation process of the label interest of opinion leaders is as follows. First, the documents labeled by opinion leaders are extracted from the set of potential opinion leaders successively, and the initial document set D is obtained after preprocessing. Additionally, the BTM model is used for topic extraction to obtain the feature word set under multiple topics. Second, the TF word frequency of feature words in documents and the IDF word frequency of reverse documents were determined. Finally, the normalized TF×IDF is calculated to obtain the interest degree of opinion leader u_i in document d_n. The calculation is shown in Formula (6), where f_{t_m,d_n} denotes the number of annotations of feature word tm by u_i in the document d_n and where $\sum_{m=1}^{k} f_{t_m,d_n}$ denotes the total number of occurrences of all keywords in d_n. n denotes the total number of opinion leaders in d_n. n_{t_m} represents the number of opinion leaders with document feature word tm that have been annotated in the document library. Formula (6) was used to obtain the opinion leader's interest in d_n. Combined with cosine similarity, the interest similarity of opinion leaders u_i and u_j was calculated as shown in Formula (7).

$$S(u_i, d_n) = \times = \frac{f_{t_m, d_n}}{\sum_{m=1}^{k} f_{t_m, d_n}} \times, \tag{6}$$

$$m(u_i, u_j) = \frac{S(u_i, d_n) \cdot S(u_j, d_n)}{\|S(u_i, d_n)\| \times |S(u_j, d_n)\|} = \frac{\sum_{i=1}^{k} S(u_i, d_n) \times S(u_j, d_n)}{\sqrt{\sum_{i=1}^{k} S(u_i, d_n)^2} \times \sqrt{\sum_{j=1}^{l} S(u_j, d_n)^2}}. \tag{7}$$

2.3.1 Fuzzy context and fuzzy concept

In the public opinion network, the community, opinion leaders and the audience have a strong semantic correlation, so opinion leaders and audience have a limited dependence on the influence transmission relationship. In that case, the community is affected by various social activities of the audience groups. On the basis of fuzzy formal concept analysis theory [14], this paper maps the opinion leader set to the object-attribute set, as well as the interest similarity relation set to a fuzzy relation set, to build the fuzzy context of the opinion leader. The fuzzy concept of the opinion leader is defined as follows.

Definition 1. If $K = (U_M, U_N, R)$ is a fuzzy context, U_M and U_N are different sets of opinion leaders, and R is a set of fuzzy relations (interest similarity) on $U_M \times U_N$. There is a mapping that makes $\mu: U_M \times U_N \to [0,1]$, and $\mu(u_i, u_j) \in [0,1]$. $\mu(u_i, u_j)$ is called the membership degree of the fuzzy relation R on (u_i, u_j), and K is the fuzzy context of the opinion leader in relation to R. The fuzzy context of opinion leaders is based on the binary interest similarity between user nodes, which represents the three-way fuzzy mapping of opinion leaders in the public opinion network.

Definition 2. If the fuzzy context $K = (U_M, U_N, R)$ of the opinion leader exists for any subset $O \in U_M$, $A \in U_N$ as shown in Formulas (8) and (9), where α is the fuzzy confidence threshold, and $O^* = A$, $A^* = O$. (O, A) is the fuzzy concept of the fuzzy context K of the opinion leader, where O is the denoted leader and A is the connotation leader.

$$O^* = \{ a \in A \mid \forall o \in O, \mu(a, o) \geq \alpha \}, \tag{8}$$
$$A^* = \{ o \in O \mid \forall a \in A, \mu(a, o) \geq \alpha \}. \tag{9}$$

Definition 3. If (O_1, A_1) and (O_2, A_2) are two fuzzy concepts in the fuzzy context $K = (U_M, U_N, R)$ of the opinion leader satisfying the partial order relation $\leq$ shown in Formula (10), then (O_1, A_1) is a subconcept of (O_2, A_2), or (O_2, A_2) is a parent concept of (O_1, A_1). It is easy to prove that the lattice composed of fuzzy context K with its posets is the opinion leader fuzzy concept lattice, denoted as OLFCL $(K, \leq)$.

$$(O_1, A_1) \leq (O_2, A) \Leftrightarrow O_1 \subseteq O_2 (A_1 \subseteq A_2). \tag{10}$$

2.3.2 Calculation of the stability index

On the basis of fuzzy concept analysis, Kuznetsov [15] quantified the dependency between attributes and objects in the fuzzy concept into a set of mapping constraints to propose the concept of the stability of the fuzzy concept, as shown in Definition 4, which analyzes the relationship between the stability of the fuzzy concept and the number of objects quantitatively. Additionally, Singh et al. [16] proved the following hypothesis. If the stability of the fuzzy concept peaks ($\frac{2^{|O|}-1}{2^{|O|}}$), the attributes contained in the fuzzy concept and the object have a relatively stable membership relationship.

Definition 4. If (O, A) is a fuzzy concept defined on the fuzzy context $K = (U_M, U_N, R)$, the stability degree of (O, A) is as shown in Formula (11), where m is the maximum number of objects contained in the lower concept of (O, A), and $|O|$ represents the number of objects contained in (O, A).

$$S(O, A) = \frac{\{m \subseteq O \mid m^* = A\}}{2^{|O|}}.$$ (11)

The dependence degree of influence between leader nodes is quantified, and multi-granularity analysis of the hierarchical relationships between opinion leaders is realized on the basis of the similar relationships between opinion leaders in the dimension of interest. To meet this need, the strong correlation between generalization and specialization among opinion leaders is recognized by calculating the probability of the object contained in the fuzzy concept of opinion leaders. The stability index of the opinion leaders is calculated via Formula (12), where $|O|$ is the object number of the fuzzy concept (O, A). represents the number of objects contained in all parent concepts of the fuzzy concept (O, A).

$$ISI(O, A) = \frac{2^{|O|} - |Sup(O)|}{2^{|O|}}.$$ (12)

3. Empirical Analysis

3.1 Data Preprocessing

Crawler tools are utilized to preprocess the Bao Yuming sexual assault event as experimental data published from Sohu, Sina Weibo and the Tiaotiao community from April 1, 2020, to October 1, 2020. Public opinion events are divided into five subdata sets as shown in Fig. 3 according to the Baidu heat index (https://index.baidu.com/). The data are preprocessed as follows. First, the word segmentation software ICTCLAS is used to conduct Chinese word segmentation, word removal and stop processing on the dataset. Moreover, the topic information on different subdata sets is extracted via the BTM model. Afterward, the optimal topic number is selected according to the degree of confusion, and the user set U under multiple topics is established. Finally, Pajek is used to create a network topology to obtain social network indicators, and the directed edges of the user relationship are calculated to realize weight annotation of edges between user nodes.

3.2 Analysis of Experimental Results

3.2.1 Drift index analysis

On the basis of topological information node multi-feature analysis, the INFOE model calculates the node influence index to obtain the user-influence index matrix $R(i, j)$ in different periods, where r_{ij} represents the influence value of user u_i at time t_j. The top influential user nodes at time t_j are selected from the matrix $R(i, j)$ to calculate the IDI. Considering the large fluctuation range of influence of public opinion under different evolution stages, the top 10 users according to the influence index are selected, and the drift index is calculated. The experimental results are shown in Tables 1 and 2. Table 1 shows that each period has significant fluctuations in public opinion activity and interaction levels. Table 2 lists the influence index drift of key opinion leaders at different stages of public opinion. It details the

maximum influence index for each user during specific time frames along with the corresponding public opinion stage.

Table 1. Network topology information of the Bao Yuming sexual assault incident

No.	Time period	Retweets and comments	Number of edges	Number of user nodes	Number of emotion words	Public opinion content
1	4.1–4.8	34572	10848	2014	558	The adopted daughter reported the crime.
2	4.9–4.10	206371	35037	8430	1341	Parties involved in the heated discussion, public opinion erupted.
3	4.11–4.13	84663	12953	10231	603	Bao Yuming responded, and all parties confirmed the information. The case is anxious.
4	4.14–9.17	127492	16608	3342	775	The authority notified the result that the person was not considered guilty of sexual assault, but deported.
5	9.18–10.1	26905	7845	1547	362	Judicial system, ethics and other issues highlighted the long tail effect.

Table 2. Drift index of opinion leaders

User number	Time frame	Max influence index	Stage of public opinion	Mean migration index
U21	4.9–4.10	0.705	Burst period	0.0384
U117	4.9–4.10	0.633	Burst period	0.0411
U346	4.11–4.13	0.728	Recession period	-0.0368
U508	4.14–9.17	0.582	Diffusion period	0.0408
U1145	9.18–10.1	0.607	Recession period	-0.0526

3.2.2 Stability index analysis

In this paper, the data from the outbreak period are taken as an example to calculate the similarity of opinion leaders, and the fuzzy context of the top 10 opinion leaders is obtained, as shown in Table 3, where the sparse symmetric matrix with a similarity threshold is 0.35. On this basis, the fuzzy concept lattice of opinion leaders is constructed. The stability indices of the fuzzy concepts of opinion leaders are calculated, as shown in Table 4. The following information can be obtained from Table 4. First, the stability

Table 3. Fuzzy context of opinion leaders (top 10)

	OL_1	OL_2	OL_3	OL_4	OL_5	OL_6	OL_7	OL_8	OL_9	OL_{10}
OL_1	1.00	-	0.46	-	-	-	-	0.48	-	-
OL_2	-	1.00	0.66	-	0.38	0.55	-	-	-	-
OL_3	0.46	0.66	1.00	-	-	0.43	-	0.50	0.47	-
OL_4	-	-	-	1.00	-	-	0.67	-	-	0.62
OL_5	-	0.38	-	-	1.00	-	-	0.52	0.49	-
OL_6	-	0.55	0.43	-	-	1.00	-	-	-	-
OL_7	-	-	-	0.67	-	-	1.00	-	-	0.48
OL_8	0.48	-	0.50	-	0.52	-	-	1.00	-	-
OL_9	-	-	0.47	-	0.49	-	-	-	1.00	-
OL_{10}	-	0.42	-	0.62	-	-	0.48	-	-	1.00

Table 4. Fuzzy concepts of opinion leaders with their stability indices

User number	Extension leader	Implied leader and their degree of membership	Concept hierarchy	Stability index
C_6	OL_4, OL_7, OL_{10}	(OL_4/0.62, OL_7/0.48, OL_{10}/0.42)	L_2	0.875
C_8	OL_5, OL_9	(OL_5/0.38, OL_9/0.41)	L_3	0.25
C_{10}	OL_2, OL_5	(OL_2/0.38, OL_5/0.44)	L_3	0.25
C_{11}	OL_3, OL_9	(OL_3/0.49, OL_9/0.47)	L_3	0.25
C_{12}	OL_5, OL_8	(OL_5/0.38, OL_8/0.48)	L_3	0.25
C_{14}	OL_1, OL_3, OL_8	(OL_1/0.46, OL_3/0.49, OL_8/0.48)	L_3	0.625
C_{15}	OL_2, OL_3, OL_6	(OL_2/0.38, OL_3/0.49, OL_6/0.43)	L_3	0.625

index of fuzzy concept C_6 located at the L_2 layer is its maximum value, indicating that there is a strong interest correlation among opinion leaders OL_4, OL_7 and OL_{10}, who have certain homogeneity characteristics, and that preference relationship clustering can be carried out in the form of an opinion leader community. Second, opinion leader OL_5 has multiple fuzzy concepts at the same time, and its stability index is low, indicating that the interest types are relatively wide and that the overall interest intensity is not high. As a result, the probability of interest drift in the later period is high. Finally, opinion leader OL_3 has a high degree of membership in fuzzy concepts C_{14} and C_{15}, which reflects its development potential to evolve into a core opinion leader.

3.3 Model Verification

3.3.1 Comparison between the influence index and the Baidu heat index

To verify the effectiveness of the index in monitoring public opinion events, the influence index of user nodes is calculated on the basis of the evolution process of public opinion events, and the drift indices of the top 50 users are calculated. The average drift index of all opinion leaders is taken as its observation index to compare with the Baidu heat index (https://index.baidu.com/). The results are shown in Fig. 1. The analysis reveals the following information. On the one hand, there is a regular correlation between the general evolution trend of the heat index and the migration index, both of which show three obvious upward trends within the observation period. Specifically, when the heat index increases to the maximum peak of 779,089 for the first time, the migration index increases to 38.26%. At this time, opinion leaders rapidly increased their influence with the emergence of public opinion. After a certain period of heat decline (diffusion period), the migration index experienced a second rapid increase (22.66%) at the time node $[7.29, 8.4]$. Moreover, the heat index increases to 420,090 because of the disclosure of the nationality information of Bao Yuming. The last time when the migration index changes significantly is $[9.16, 9.22]$, and the reason for the heat rise is that the court gave no prosecution and the public security department issued a notice of deportation. On the other hand, the migration index has certain monitoring and early warning advantages compared with the heat index in the diffusion stage of the evolution of public opinion events, which improves the tracking ability in the steady evolution process of public opinion. For example, the migration index of opinion leaders has declined significantly by twofold (-10.97% and -15.67%, respectively), and new hot spots of public opinion have appeared in the subsequent time nodes. The short-term negative drift of the migration index indicates the possibility of a subsequent crisis during the evolution of public opinion. Moreover, a public opinion intervention mechanism should be launched in time to guide the emotional direction of internet users actively.
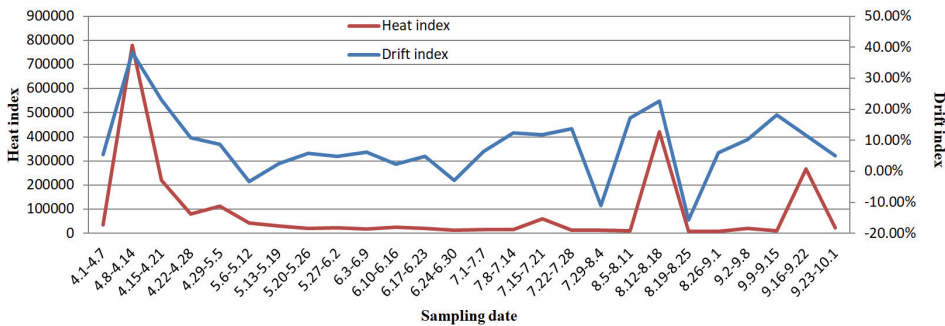
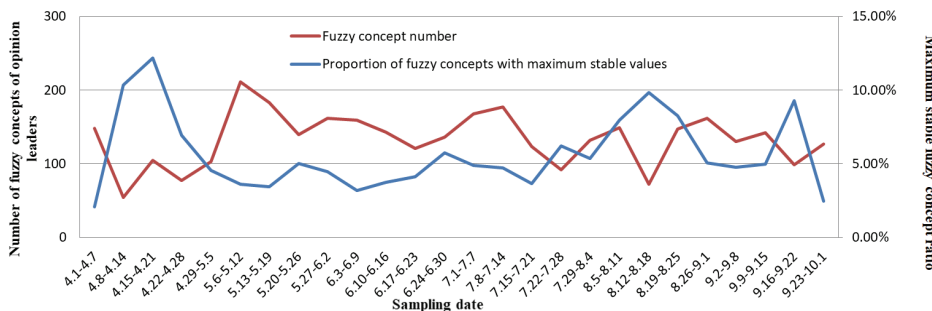**Fig. 1.** Comparison between the drift index and heat index.

Fig. 2. Trend chart of the ISI during the evolution period of public opinion.

Additionally, the fuzzy concepts of opinion leaders under different time slices are calculated to analyze the role of the ISI in opinion leaders' interest clustering. The degree of correlation between the ISI and the evolution of public opinion events is verified by extracting the number of fuzzy concepts and the proportion of fuzzy concepts that reach a maximum stable value. The empirical value of the fuzzy confidence threshold is 0.425 [17], and the results are shown in Fig. 2. The analysis provides the following information. First, the number of fuzzy concepts of opinion leaders reflects the intensity of interest clustering within the opinion leaders, which indicates that the greater the number of fuzzy concepts is, the greater the number of types of interest among opinion leaders. Additionally, in the time snapshot of the three outbreaks of public opinion, the number of fuzzy concepts shows a rapid decline, indicating that the opinion leaders' interests and preferences are concentrated and introspective to become the topic focus. Second, fuzzy concepts with great stability tend to have a more stable internal structure in the leadership community, indicating that the interest of the opinion leaders contained in them is relatively stable. When the number of fuzzy concepts reaches the maximum ISI and the number of fuzzy concepts decreases, the focus of public opinion events at this stage is concentrated, and the possibility of a public opinion crisis breaking out in the short term is high. Additionally, the above two types of digital characteristics of the ISI are basically consistent with the overall trend of the heat index, indicating that the ISI can perceive the outbreak time node of public opinion in the opinion leaders' interest dimension and enrich the measurement information of public opinion control.

3.3.2 Knowledge discovery of influence evolution

The evolution knowledge of opinion leaders on the basis of the influence index focuses on the trend

evolution of the drift index IDI and stability index ISI to track the connotational nature of public opinion development. On this basis, we further explore the characteristics of implicit interest correlation and explicit communication in the growth of opinion leaders. Combined with the empirical analysis of real incidents, the knowledge discovery of the evolution of opinion leaders can be summarized as follows.
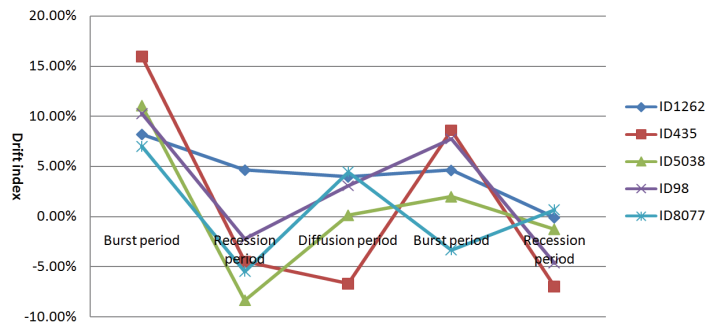

Fig. 3. Schematic diagram of tracking the drift of opinion leaders in the Bao Yuming sexual assault incident.
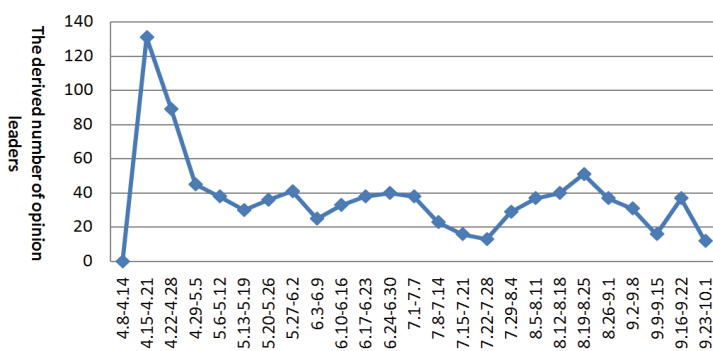

Fig. 4. The number of opinion leaders involved in the Bao Yuming sexual assault incident.

The drift index focuses on measuring the ability of opinion leaders to guide the audience's information and dominate their emotions. The time point with the highest influence value of opinion leaders in the evolution cycle of public opinion is selected, and the average IDI under the continuous time snapshot is plotted to quantify the drift degree of opinion leaders in the time dimension. Some of the results are shown in Fig. 3. The following information can be extracted from Fig. 4. First, the influence index of opinion leaders changes continuously throughout the evolutionary cycle of public opinion. For example, the IDI of the influence of *ID1262* has a small change range after the first outbreak. The analysis revealed that the registered identity of the user is internet celebrity, the drift rate of opinion leaders is slow, and the overall degree of change is relatively low. This indicates that its influence is less affected by the development of public opinion, which makes its influence index always high. Second, some opinion leaders are positively correlated with the popularity of public opinion. For example, the IDI of *ID435* shows a large numerical increase in the two outbreak periods and a large numerical decline in the period, indicating a high degree of active participation in the evolution of public opinion. In the later period, in-depth semantic analysis of such user texts can be conducted to predict the time nodes of public opinion crises accurately. Finally, opinion leaders may drop out of influence. For example, the IDI of *ID5038* has

a small change range after the first decline. After analysis, their influence index ranks low, indicating that their participation in public opinion is low in the final stage, leading to their status as ordinary users.

The stability index aims to identify the subleaders derived from the opinion leader group by using the similarity of interest among the opinion leaders. The fuzzy concept of the opinion leader is extracted with its ISI by constructing the fuzzy concept lattice of the opinion leader under adjacent time slices. Considering that the Bao Yuling sexual assault incident is a network emergency and that the formation of opinion leaders occurred mainly during the first outbreak period [4.8, 4.14], the experiment takes this time piece as the initial observation point and counts the number of derived opinion leaders meeting the above rules according to the evolution period of public opinion. A few important sets of information can be obtained. First, the number of derived opinion leaders shows a slow declining trend, and it peaks in the recession period after the initial outbreak. The reason is that the evolution of public opinion enters a period of information diffusion and emotional fluctuation, and the interest views of opinion leaders present a spreading trend, leading to the emergence of many new interest groups, and giving birth to many new opinion leaders. Second, after a long period of interest divergence and heat decline, there was a trend toward divergence of interest types and a decline in interest intensity in the community. The increase in heat fails to offset the reduction in the number of fuzzy concepts reaching the maximum value, leading to a rapid decline in the number of derivative opinion leaders. Additionally, the change rules of the external influence and internal interest degree of opinion leaders can be deeply perceived by combining the IDI with the ISI. For example, when the ISI of the fuzzy concept of opinion leaders reaches the maximum value at time t, the IDI of all the nodes of opinion leaders experience positive drift at time $t + 1$. This finding indicates that the influence of the above opinion leader group is further strengthened through the influence transmission of similar interests and that the influence of the opinion leader group is influenced by the intensity of their interest. In the later stage, the monitoring of such user groups should be improved, and close attention should be given to the degree of their influence diffusion.

3.3.3 Comparison analysis

To verify the validity and rationality of INFOE, the dataset was divided into five subsets (according to the time information in Table 1), labeled D1–D5. Additionally, the F value was used to evaluate the performance of opinion leaders identified by INFOE on five experimental datasets. The methods provided in the literature [11,16,17] were taken as comparison objects and named OLCV, FSUSE, and OFCL, respectively. Fig. 5 shows that the average performance across all datasets for the four methods is as follows: OLCV at 76.18%, FSUSE at 75.95%, OFCL at 78.64%, and INFOE at 80.52%. Among the
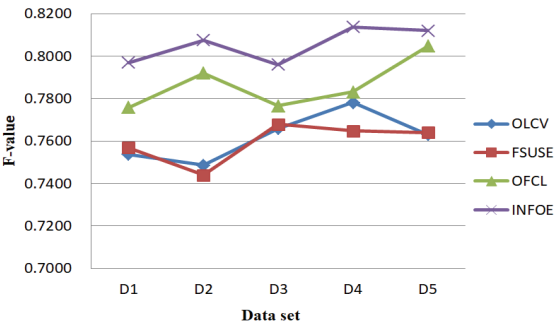

Fig. 5. Comparison of the F values of different methods.

four methods, INFOE achieves the best computational performance, as it establishes an IDI that targets the leader-audience layer and constructs an ISI that focuses on the leader layer. This approach facilitates the online evolution of influence indices for opinion leaders, thereby enhancing the accuracy of identifying opinion leaders during the evolution of public sentiment.

4. Conclusion

An INFOE is proposed, which uses the multidimensional characteristics of user nodes to calculate the influence index to construct a drift index to study the changing trend of influence between opinion leaders and audiences. Concepts such as the drift of opinion leaders and the derivation of opinion leaders are analyzed theoretically. The possibility of tracking the change in influence is discussed by establishing the drift index in the time dimension and the stability index in the interest dimension. Additionally, the fuzzy concept of opinion leaders and its stability index are calculated by calculating the similarity of interest under different time snapshots to explore the interactive dependence relationships between opinion leaders. The results show that the drift index and stability index are measures used to monitor the evolution rate of opinion leaders' influence. The degree of interest progression can provide practical guidance for public opinion control and crisis response.

We assume that user interest is the quantitative standard used to judge whether opinion leaders have homogenized characteristics without considering the reference variables of opinion leaders' dissimilation. Furthermore, the fuzzy interest portraits and emotional fluctuation degrees of opinion leaders in the later stage can be studied further. Additionally, the drift index and stability index have the characteristics of a cold start and are unable to predict the network's public opinion of emergencies, so this problem could be addressed in later stages of this project.

Conflict of Interest

The author declares that they have no competing interests.

Funding

None.

References

[1] X. Wang, G. Li, J. Mao, and G. Ye, "A review of research on identification and analysis of public opinion on emergencies," *Library and Information Science*, vol. 32, no. 1, pp. 93-102, 2021. https://doi.org/10.13366/j.dik.2021.01.093

[2] H. Zhu, L. Qian, X. Yang, and J. Wei, "Research on topic drift path of online public opinion," *Journal of Intelligence*, vol. 41, no. 6, pp. 108-113+119, 2022. https://doi.org/10.3969/j.issn.1002-1965.2022.06.016

[3] W. Feng, K. W. So, Y. Du, and B. Qi, "Research on the evolution and guidance mechanism of network public

opinion in social livelihood events," *Journal of Intelligence*, vol. 41, no. 8, pp. 112-120, 2022. https://doi.org/10.3969/j.issn.1002-1965.2022.08.016

[4] Y. Xu, W. Huang, S. Guo, and Y. Xe, "Multimedia network public opinion topic evolution tracking trend and mechanism analysis," *Intelligence Theory and Practice*, vol. 43, no. 12, pp. 156-162, 2020. https://doi.org/10.16353/j.cnki.1000-7490.2020.12.023

[5] L. Huang, "Opinion leader mining method based on IIOLR model," *Journal of Intelligence*, vol. 40, no. 6, pp. 163-170, 2021. https://doi.org/10.3969/j.issn.1002-1965.2021.06.022

[6] J. Wang, P. Wu, F. Chen, Y. Wang, and S. Ding, "An empirical study on the identification and influence of opinion leaders in emergencies," *Journal of the China Society for Scientific and Technical Information*, vol. 35, no. 2, pp. 169-176, 2016. https://doi.org/10.3772/j.issn.1000-0135.2016.002.006

[7] F. Chen, P. Chen, P. Wu, and C. Xue, "Identification of online opinion leaders by integrating user characteristics and multi-level text tendency analysis," *Intelligence Theory and Practice*, vol. 41, no. 7, pp. 143-148, 2018. https://doi.org/10.16353/j.cnki.1000-7490.2018.07.025

[8] J. Deng, C. Zhong, Y. Chang, W. Zhang, and S. Sun, "A study on the interactive behavior of opinion leaders in online health communities based on the two-step communication theory: a case study of Baidu's "Autism Forum"," *Intelligence Theory and Practice*, vol. 45, no. 1, pp. 127-133+111, 2022. https://doi.org/10.16353/j.cnki.1000-7490.2022.01.017

[9] L. Zhang, X. Wang, B. Huang, and T. Liu, "Research on topic clustering graph and topic propagation path of Weibo users related to COVID-19 epidemic based on LDA model," *Journal of the China Society for Scientific and Technical Information*, vol. 40, no. 3, pp. 234-244, 2021. https://doi.org/10.3772/j.issn.1000-0135.2021.03.002

[10] M. R. Bhat, B. Bashir, M. A. Kundroo, and N. A. Ahanger, "Understanding Twitter hashtags from latent themes using biterm topic model," *Recent Patents on Engineering*, vol. 14, no. 3, pp. 440-447, 2020. http://dx.doi.org/10.2174/1872212113666190328183517

[11] X. Cheng, M. Zhang, F. Ding, and Y. Lu, "Research on the influence of characteristic values of opinion leaders in developer communities," *Modern Intelligence*, vol. 42, no. 7, pp. 114-124, 2022. https://doi.org/10.3969/j.issn.1008-0821.2022.07.010

[12] G. Prathap, "The 100 most prolific economists using the p-index," *Scientometrics*, vol. 84, no. 1, pp. 167-172, 2010. https://doi.org/10.1007/s11192-009-0068-0

[13] Y. Wang and Y. Wang, "Discovery and comparative analysis of public opinion topics among different communicators during the dissemination stage," *Modern Intelligence*, vol. 38, no. 9, pp. 28-35, 2018. https://doi.org/10.3969/j.issn.1008-0821.2018.09.005

[14] K. Ravi, V. Ravi, and P. S. R. K. Prasad, "Fuzzy formal concept analysis based opinion mining for CRM in financial services," *Applied Soft Computing*, vol. 60, pp. 786-807, 2017. https://doi.org/10.1016/j.asoc.2017.05.028

[15] S. O. Kuznetsov, "On stability of a formal concept," *Annals of Mathematics and Artificial Intelligence*, vol. 49, no. 1, pp. 101-115, 2007. https://doi.org/10.1007/s10472-007-9053-6

[16] P. K. Singh, A. K. Cherukuri, and J. Li, "Concepts reduction in formal concept analysis with fuzzy setting using Shannon entropy," *International Journal of Machine Learning and Cybernetics*, vol. 8, pp. 179-189, 2017. https://doi.org/10.1007/s13042-014-0313-6

[17] Y. Li, X. Guo, S. Wang, and J. Liang, "Research on automobile evaluation knowledge discovery method based on connotation fuzzy concept lattice," *Journal of Chinese Information Processing*, vol. 31, no. 3, pp. 69-76, 2017.

Cheng Zhang https://orcid.org/0009-0000-7853-7017

He received master degree in library and information science from St. Paul University Philippines in 2023. He is currently a librarian in Library of Zhaoqing Medical College. His current research interests include information retrieval and knowledge modeling.