J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 1493--1510

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Modeling the Clustering Structure of Complex Networks and Calculating Propagation Thresholds for Bandwagon Behavior of Social Media Papers from the Perspective of Knowledge Reproduction and Anti-Regulatory Discipline

Cunjie Song[1], Shangwen Chen[1,✉], Xiaoyuan Tang[1]

[1] *The School of Journalism and Communication, Guangxi University, Nanning, Guangxi, 530004, China*

ABSTRACT

This paper constructs a heterogeneous network adjacency matrix containing multiple user relationships from the connotation of professional organizations and other guides to individual behaviors covered by the take-read mechanism. The GAT algorithm is used to learn the embedding of its heterogeneous network in order to obtain the embedding vectors of user nodes, which serves as the basis for the analysis of the spreading influence of group behavior. An event recognition method based on word embedding and hierarchical cohesive clustering is proposed to analyze the recognition and evolution of social media essay-carrying behavioral events (group behavioral events) for complex networks. We point out that the distribution of group behavior affects the dynamics of information dissemination, set the adoption threshold parameter of the group, and analyze the dissemination pattern of individuals' (individual information) participation in essay-reading behaviors. Analyze the emergence and evolution of thesis-reading behavior in social media, and explore the influence of individual's own attributes and the attitude of neighboring nodes on the evolution of group behavioral events in complex networks. The spreading degree analysis is conducted for different relational social media bandwagon behaviors. When $I_{i,j}$=0.6 and $Ex_{i,j}$=0.8, the individual's decision is supported by the neighbor's viewpoints, and the users who have already participated in the paper band-reading activities have a strong attraction to the individual. When the strong degree increases to a certain value, the individual decides to participate in the dissertation banding activity, at which point the individual is no longer influenced by the external environment. The degree of the initial node for the propagation of thesis banding behavior in random networks and small-world networks is linearly and negatively correlated with the percentage of the information audience.

Keywords: gat algorithm; complex networks; heterogeneous networks; small-world networks; group behavior

✉ Corresponding author.
E-mail address: m19571771987@163.com (S. Chen).

Received 30 May 2024; Revised 15 August 2024; Accepted 10 January 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-082

1. Introduction

Essay writing is not an examination of knowledge already acquired in a particular discipline or course, but rather focuses on the student's ability to apply relevant knowledge or theories to conduct an in-depth inquiry into a practical or academic problem [1]. Its purpose is to help students achieve critical reading, familiarize themselves with the writing process, and learn how to effectively use sufficient evidence, credible research resources, and the perspectives of others to improve their thinking and writing skills [2-4]. To complete a dissertation of high quality, students must not only be able to deeply understand and apply their professional knowledge, but also have the ability to think independently and creatively, as well as good academic writing skills [5-7]. However, dissertation writing is not only to solve the problems of format specification and academic diction and rhetorical skills, but more importantly, students should think, discern, criticize and then argue through certain literature reading [8-9]. The rapid development of new network media has subversively changed the reading habits of contemporary people, and the behavior of social media paper with reading can help students read the paper efficiently and master the overall structure of the paper, so that students can use fragmented time to read all kinds of literature and classic books [10-13]. Only by reading much, widely and deeply, accumulating to a certain amount with richness, can there be sparks of ideas.

Since behaviors in the network are generated by users interacting with each other with data, social media-based dissertation bandwagon behaviors cannot exist independently of a single user [14-15]. There are more data connections between multiple users that make up a certain network behavior, and this phenomenon is manifested as an association clustering phenomenon in the network topology graph [16-18]. Thus, using complex network clustering algorithm to divide the user nodes in the network into different behavioral sub-clusters can make the user nodes in the network in the same behavioral mode aggregated with each other, and then each behavioral sub-cluster can be labeled with behavior to complete the calculation of the network behavior for the purpose of [19-22]. The use of complex network clustering structure to calculate network behavior can strengthen the network administrators to master the spread of band reading behavior.

In this paper, we analyze the foundation of complex network, and build heterogeneous network based on complex network by combining the mechanism of social media paper reading behavior. It analyzes the data preparation process of heterogeneous network construction and designs the basic structure of heterogeneous network based on social media paper reading behavior. Adopting the user relations of following and being followed, mentioning and being mentioned, forwarding and being forwarded, replying and being replied as the edge types of heterogeneous network, we use the method of neighbor matrix splicing to construct the heterogeneous network based on multiple user relations of essay reading behavior. The graph embedding processing layer, instance specification layer, input layer, and GAT layer form a heterogeneous network with end-to-end features. We analyze the evolution of thesis-reading events, and use the group behavior distribution dynamics to analyze the spreading influence of thesis-reading behaviors in social media.

2. Behavioral social media communication based on complex network analysis

2.1. Complex networks

Complex networks are networks with a high degree of complexity for which there is no clear definition. But generally speaking, complex networks are characterized by the following aspects. Firstly, the structure of the network is highly complex, which is manifested in the large number of nodes, for example, the nodes in the economic network can reach tens of thousands. Secondly, the node connections in the network show a high degree of dynamic characteristics, and the structure of the network will constantly change [23-25].

2.1.1. Topology of complex networks:

1) Degree and degree distribution. Degree is a simple yet important concept in the properties of individual nodes. The degree k of node i is defined as the number of other nodes connected to that node.

The degree distribution of a node in a network can be represented by the distribution function $P(k)$. The distribution function $P(k)$ indicates that the probability that the degree of a node is equal to k is P.

2) Average path length. The heart of the distance between two nodes i and j in the network is defined as the number of edges included in the shortest path connecting these two nodes. Define the diameter of the network $D(diameter)$ as the maximum value of the distance between any two nodes in the network is, notation:

$$D = \max d_{i,j} \tag{1}$$

Define the average value of the distance between any two nodes as the average path length of the network L. The average path length of the network is also called the characteristic path length of the network i.e:

$$L = \frac{1}{0.5 * N(N+1)} \sum_{i>j} d_{i,j} \tag{2}$$

Where N is the number of network nodes.

3) Clustering coefficient. Another important network characteristic is the degree of network "grouping", the higher the degree of grouping, the higher the degree of aggregation of the network. An important index to measure the degree of network "clustering" is the clustering coefficient.

For example, in a network of one's friends, your friends' friends may be friends with each other, and this type of phenomenon in a network is known as the clustering nature of the network. Suppose a node i in the network is connected to other nodes through k_i edges, and the k_i nodes connected to it can be considered neighbors of node i. And there may be at most $\frac{k_i(k_i-1)}{2}$ edges between these k_i nodes. And the clustering coefficient C_i is defined as the ratio of the number of edges E_i that actually exist between these k_i nodes to the total number of possible edges, that is:

$$C_i = \frac{2E_i}{(k_i(k_i-1))} \tag{3}$$

By averaging the clustering coefficients C_i of all nodes i one obtains the clustering coefficient of the whole network C. It is clear that the clustering coefficients take values from 0 to 1. When the whole network does not have any nodes connected to each other $C=0$, i.e. all the nodes are isolated. When any two nodes in the network are directly connected C = 1, i.e. the network is fully coupled.

2.1.2. Small World Network. In order to characterize the real network, the small world network model, referred to as WS network, is proposed. Two statistical parameters characterize the small world network, average path length and clustering coefficient.

2.1.3. Scale-free networks. WS small-world networks reproduce the large aggregation coefficients and small average shortest paths that characterize real networks, but the degree distribution of WS networks obeys an exponential distribution. However many realistic network distributions are power-law distributed, with a few nodes occupying most of the connections. This class of networks is also known as scale-free networks.

In order to generate complex networks that conform to a power law distribution, a scale-free network construction methodology, called the BA model, has been proposed. It is considered that complex networks contain two important characteristics. The first one is the growth characteristic. The size of the network is not constant but dynamically growing, and the size of the network in reality is generally increasing. The second characteristic is the preferential connectivity property, i.e., newly generated nodes are more inclined to connect with those nodes that are more connected. That is, the stronger the stronger, this phenomenon is also known as the "Matthew effect".

2.2. Heterogeneous network construction based on social media essay banding behavior

The "Reading Mechanism" refers to the system of professional reading institutions, organizations and individuals who are enthusiastic about reading, which is based on the premise of constructing a rigorous, systematic and effective operation mode and a solid reading culture orientation, and relying on modern technological means and strategies such as video-type platforms, and spirally driving, encouraging and guiding reading-difficulty groups to participate in, get accustomed to, and love reading. The promotion service system is conducive to enhancing their reading literacy and confidence, enlightening their thinking and expanding their horizons, and cultivating good reading habits and tastes.

Social behavior between accounts can be a direct relationship, such as an account A publishing a tweet while @ another account B. It can also be an indirect relationship. For example, if an account A publishes a tweet T and another account B retweets this tweet T, then an indirect relationship is formed between account A and account B through the tweet T. Through the direct and indirect relationships between accounts, a large heterogeneous network can be constructed based on the social behavior of accounts. This heterogeneous network contains five types of nodes, namely account, tweet, event, person, hashtag, and the connecting edge relationship between them. The study of the behavioral characteristics of accounts in the heterogeneous network helps to accurately classify accounts. In order to construct this heterogeneous network, some preparations are needed, including data collection and preprocessing, event detection, and identification and extraction of related characters.

2.2.1. Data preparation for constructing heterogeneous networks:

1) Design and Implementation of Data Acquisition. The method of obtaining data through API interfaces has higher efficiency and is usually used as the preferred solution for social network data collection.

Through the application programming interfaces (API interfaces) provided by social tools, access is made according to the rules of the API interfaces to obtain the corresponding data.

API interfaces provided by social media are categorized into two main types, which are REST API and Streaming API. Where REST API mainly provides historical data collection, and Streaming API mainly provides real-time sampling results of current social media data. In this paper, the data collection method adopts Streaming API as the main method and REST API as the supplementary method. Firstly, the Streaming API is used to obtain the tweets of interest and the accounts corresponding to them. Then for the accounts of interest, the REST API is used to obtain the historical tweets of these accounts.

2) Social media based event detection methods. An event in social networks is an exact activity that occurs at some specific time interval. Events can be categorized into large and small events based on their influence and reach. It is possible that the duration interval of a big event may contain numerous small events with relatively small influence, which may have an impact on the big event.

Social media, as a mainstream social tool, allows every user of social media to discuss hot events of current interest. Therefore, tweets related to the events and the corresponding Twitter accounts can be detected in the tweets after collection and preprocessing. Usually, after each event, the text of tweets discussing the event has a certain similarity relationship, so the text similarity can be utilized to cluster the tweets to obtain a collection of tweets related to the event.

A hashtag is a hashtag that starts with a # sign, which usually refers to a certain topic, and by using hashtag in social media it is possible to clearly express the topic that the social media wants to discuss.

The captured tweets have been preprocessed to get the list of tweet words after disambiguation. Based on the list of tweet words, named entity recognition can be performed to find relevant names of people, places, and organizations. Where the names of people are the real people in the real world that are needed.

2.2.2. Basic structure of heterogeneous networks. Typically, complex networks refer to homogeneous networks, that is, networks with only one node type and connected edge type. However, the construction method of this homogeneous network usually contains only the direct relationship between accounts and does not use the indirect relationship. The constructed network is more sparse and the connecting edge relationship between nodes is not complete enough. Heterogeneous network is a special kind of complex network, compared with homogeneous network, heterogeneous network has more node types and edge types. Based on different node types and edge types, direct and indirect relationships between different nodes can be described. The social media paper with reading behavior data collected in this paper can be abstracted into multiple node types, so it is a valuable method and tool for social relationship modeling based on heterogeneous networks.

Unlike the construction of homogeneous networks, heterogeneous networks can contain multiple node types, which enables more accurate abstraction of different objects. Therefore, for the construction of heterogeneous networks, it is first necessary to determine the basic node types of heterogeneous networks. Based on the preparatory work above, our heterogeneous network contains five types of basic nodes, namely account nodes, tweet nodes, event nodes, character nodes and hashtag nodes, which can not only establish a variety of concatenated relationships with other nodes, but also contain inherent attributes describing the node's basic information, and according to the node's attributes, the node's basic characteristics can be grasped information.

2.2.3. Construction of Heterogeneous Networks. In this paper, we use sampling a fixed-size subnet from user u's r-hop network instead of directly using that user's r-hop network. Therefore, in the actual sampling, this paper uses the Randomized Wandering Algorithm with Restart (RWR). The optimization objective function is rewritten as:

$$L(\Theta) = -\sum_{i=1}^{N} \log(P_{\Theta}(s_{v_i}^{t_i+\Delta t} \mid \bar{G}_{v_i}^{r}, \bar{S}_{v_i}^{t_i})) \tag{4}$$

In general, there are four main types of user relationships in social networks that can be observed based on user behavior in social networks, which are:

1) The following and being followed relationship based on user following behavior.
2) Mention and mentioned relationship based on user mention behavior.
3) Forwarding and being forwarded relationship based on user forwarding behavior.
4) Reply and be replied relationship based on user reply behavior.

In the selection of heterogeneous network nodes, this paper adopts social network users as nodes and the above four types of user relationships as edge types of heterogeneous networks, so as to construct heterogeneous networks based on multiple user relationships. When dealing with different types of points or edges, there are usually two methods, which are based on meta-path and based on neighbor matrix splicing.

In this paper, since only one type of vertex and its corresponding multiple relationships are used, it is more appropriate to take the approach of constructing a neighbor matrix for each type of user relationship separately based on neighbor matrix splicing, so that a heterogeneous network neighbor matrix based on multiple user relationships can be obtained.

2.2.4. End-to-end model based on heterogeneous networks:

1) Graph embedding processing layer. In this paper, Deep Walk algorithm is mainly used to realize the conversion from graph to vector.

Deep Walk is based on the Word2vec method for vectorized representation of graph vertices, which is an innovative approach for learning potential representations of vertices in networks.

Deep Walk treats a sequence of nodes obtained through the random walk algorithm as a sentence in Word2vec, and obtains part of the information of the network from the truncated random walk sequence, and then learns the potential representation of the nodes after another part of the a priori information.

2) Instance normalization layer. Instance normalization is a new technique recently proposed in the field of computer vision, which mainly addresses the problem of internal covariate shift (ICS) in deep learning.

Applying the instance normalization technique to this model, for each user $u\in\bar{\Gamma}_v^r$, after obtaining the feature vector x_u of this user from the embedding layer, the normalization vector is defined as:

$$y_{sd}=\frac{x_{sd}-\mu_d}{\sqrt{\sigma_d^2+\varepsilon}} \tag{5}$$

$$\mu_d=\frac{1}{n}\sum_{u\in\bar{\Gamma}'_v}x_{ud},\sigma_d^2=\frac{1}{n}\sum_{u\in\bar{\Gamma}'_v}(x_{ud}-\mu_d)^2 \tag{6}$$

where $d=1,\dots,D$ denotes each embedding layer dimension, μ_d and σ_d denote the mean and variance, and ε is a numerically minimal number in order to make sense of the equation.

After this normalization the instance-specific mean and variance can be eliminated, allowing the downstream model to focus on the relative position of the user in the potential embedding space rather than the absolute position of the user. And instance normalization can help avoid overfitting during training.

3) Input layer. In the input layer, each user is characterized as a feature vector, in addition to the user heterogeneous network features from the upper layer that have been processed by graph embedding and instance normalization. There are also 8 user features to be added, which can be categorized into user state and indication features and vertex features according to the content of the features.

4) GAT algorithm. And Graph Attention (GAT) is a recently proposed technique to introduce the attention mechanism into GCN. The attention mechanism can be used to distinguish the importance of different nodes, i.e., to give weight values.GAT defines the matrix $A_{CAT}(G)=[a_{ij}]_{n\times m}$ through the attention mechanism.Typically, the importance of vertex i to vertex j is first computed through the attention function: $attn:\mathbb{R}^{F'}\times\mathbb{R}^{F'}\rightarrow\mathbb{R}$. Thus there is:

$$e_{ij}=\arctan(Wh_i,Wh_j) \tag{7}$$

While traditional self-attention mechanisms compute the attention coefficients between all instances, in the framework of this paper, GAT will only compute the e_{ij} that satisfy the condition $(i,j)\in E(\bar{G}_v^r)$ or $i=j$. Doing so allows GAT to better utilize and capture the structural information of the graph. In order to make the coefficients comparable between the vertices, the attention coefficients are normalized using the softmax function:

$$a_{ij} = \text{softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \Gamma_i^i} \exp(e_{ik})} \tag{8}$$

The attention function is constituted by dot product of vectors and instantiation of the leakage activation function. For an edge or a self-loop (i, j), the dot product is performed between the parameters c and a cascade of eigenvectors at the two endpoints Wh_i and Wh_j:

$$e_{ij} = \text{LeakyReLU}(c^{\mathrm{T}}[Wh_i \,\|\, Wh_j]) \tag{9}$$

where $\|$ denotes the connection operation. The normalized attention factor can be expressed as:

$$a_{ij} = \frac{\exp(Leaky\,\mathrm{Re}\,LU(c^T[Wh_i \,\|\, Wh_j]))}{\sum_{k \in \Gamma_i^3} \exp(Leaky\,\mathrm{Re}\,LU(c^T[Wh_i \,\|\, Wh_k]))} \tag{10}$$

The above completes the definition and computation of the single-head GAT layer. In addition, in order to stabilize the training process, the algorithm incorporates a multi-head attention mechanism. The multi-head attention mechanism executes K independent single-head attention process in parallel.

In summary, the model proposed in this paper first constructs a heterogeneous network of multiple user relationships based on their following, retweeting, replying and mentioning relationships. Then the embedding vectors of user nodes are obtained by embedding learning on the heterogeneous network using the GAT algorithm. Finally, the embedding vectors of users are passed through softmax layer to realize user influence prediction.

Complex Network Oriented Group Behavior Event Recognition and Evolution

In this paper, we propose complex network-oriented identification and evolution of social media paper banding behavioral events (group behavioral events). Specifically, this paper proposes an event recognition method based on word embedding and hierarchical cohesive clustering. The model does not only consider only the text itself, but also data-oriented semantic features using self-learning word embeddings for event recognition. In addition, the model does not only rely on the change of clusters to explore the recognition and evolution of group behavior events, but also uses the dynamics of clusters and the dynamics of vocabularies to describe the recognition and evolution of group behavior events.

The model for complex network-oriented group behavior event recognition and evolution research is divided into four main parts: ① Data flow module. ② Word vector embedding learning module. ③ Event recognition and time-series evolution module. ④ Keyword extraction module. Self-learning word embedding is used in the event recognition and temporal evolution and event word extraction stages.

1) Data flow module. By considering the performance of existing methods and the requirements of temporal event recognition, the event recognition and evolution study of group behavior for complex networks will use a sliding window model with a fixed time horizon for event recognition in social media data streams. The stream chunker is the module that helps to separate the data streams into time windows. Depending on the evolution of the events to be recognized, the chunker can adjust the length of the time horizon.

2) Word Vector Embedding Learning Module. In this paper, word embeddings are used to facilitate the combination of statistical features, syntactic features and semantic dynamics features in text and they are used for group behavior event recognition and evolution oriented research. Without using

pre-trained word embeddings, the model learns embeddings on the target corpus to capture its unique features such as modified or misspelled words and emoticons.

3) Event Recognition and Temporal Evolution Module. In this paper, windows larger than the set threshold (γ) are recognized as event windows. From the data sample point of view, this set threshold is mainly used to define the importance of the target event, and a lower threshold (γ) will identify less important things. The model will use normalized values to measure event changes, so the possible values for threshold (γ) are between 0 and 1.

In this study, the model examines the change of words or phrases over time by calculating the clustering change. To facilitate the calculation, the model needs to generate the corresponding clusters for each time window W_t.

In selecting the clustering method, Hierarchical Aggregate Clustering (HAC) is chosen here. Among the distance calculation methods, hierarchical agglomerative clustering uses the average distance method to help the model to be able to calculate the distance between the cluster classes in a way that involves all the elements belonging to the clusters with the following formula:

$$D(c_i, c_j) = \frac{1}{|c_i||c_i|} \sum_{w_l \in c_j} \sum dist(w_k, w_j) \tag{11}$$

$$dist(w_k, w_l) = \cos(w_k, w_l) \tag{12}$$

Where $dist(w_k, w_l)$ denotes the distance between cluster elements w_k and w_l belonging to clusters c_i and c_j respectively, which will be calculated using the cosine distance cos. After generating the clusters, the model uses a matrix-based approach to compute cluster changes to detect the occurrence and disappearance of group behavioral events.

The model decides to measure the labeling similarity based on the clustering changes using the Tree-Level (TL) similarity of the dendrogram. For a given dendrogram, then the similarity measure for a word pair w_k and w_l is the value after normalizing the shared level between these two words and the root of the dendrogram. The formula is as follows:

$$TLSimilarity_{(w_k, w_l)} = \frac{tl_{(w_k, w_l)}}{Max(tl_{r \to w}) + 1}. \tag{13}$$

The formula represents the number of shared dendrogram levels between word pairs w_k and w_1 starting from the root. The denominator represents the maximum number of tiers between the root and leaf nodes. Also, while calculating the maximum number of tiers this model considers the tier values of the leaf nodes to be considered for calculation to ensure that the similarity between the same vectors is $(TLSimilarity_{(w_k, w_k)})$.

In order to obtain comparable values over all time windows, the model calculates the vocabulary change at time point W_{t+1} compared to time point W_t by using equation (14). The formula is as follows:

$$Vochange_{(t,t+1)} = \frac{|vocab_{t+1} - vocab_t|}{|vocab_{t+1}|} \tag{14}$$

where the numerator of the formula represents the absolute value of new words in the glossary at time point W_{t+1} compared to time point W_t. The denominator represents the size of the glossary at time point W_{t+1}.

2.3. Calculating the Communication Impact of Social Media Essay Bandwagon Behavior

2.4.1. Characterizing the dynamics of group behavior distribution. In social settings, the acceptance of information is significantly higher when it originates from familiar individuals, such as family members, friends, or experts. This phenomenon reflects a unique pattern of social interaction within the group, involving varying degrees of intimate connections between friends, family members, and coworkers. This heterogeneity of intimacy within groups significantly affects the dissemination of information. In response to researchers' unsystematic and insufficient studies of group behavior, multidimensional intimate interactions on multilayer networks are further considered and a multilayer weighted network model is proposed. The study shows that group behavior distribution has an impact on the mechanism of information dissemination dynamics. There is a correlation between the adoption threshold in the group behavior distribution and the number of connections of cluster nodes as well as the parameters conforming to the truncated-tailed normal distribution. The theory of connecting edge division based on group intimacy relationship and group behavior distribution threshold is proposed to quantitatively analyze the information propagation mechanism. The results obtained after simulation are consistent with those predicted by theoretical analysis, and both reveal the intrinsic information propagation characteristics and laws of the network.

2.4.2. Adoption threshold models. Given the existence of differentially close connections between group nodes, information is more easily disseminated among nodes that are close to each other. The degree of such close connections between nodes is indicated by the weights of the edges, while weighted edges are used to reflect the connectivity between nodes. Thus, the use of edge weight distribution can reveal the diversity of connections.

For neighboring nodes i and j, their edge weights in layer X are denoted as w_{ij}^X. The edge weights follow two independent distributions $g_A(w)$ and $g_B(w)$. When an A-state node i carries out a message, the probability that its S-state neighboring node j, located in the same layer, obtains the message is:

$$\lambda_\omega^X = \lambda_X(\omega_{ij}^X) = 1-(1-\beta_X)^{\omega_{ij}^X} \tag{15}$$

where β_X is the unit information propagation probability in layer X. If $\omega_{ij}^X = 1, \lambda_\omega^X = \beta_X$, i.e., information propagation is not affected by edge weights. Further, as ω_{ij}^X keeps rising, $\lambda_x(\omega_{ij}^X)$ shows a monotonically increasing trend.

To avoid information duplication, the group node j will not receive the same information from that neighbor again. Further, to investigate the role of diversity of thresholds on information dissemination, μ and σ are used to represent the mean and standard deviation of a normal distribution. The adoption threshold parameter of the population can be calculated by the following equation:

$$f(\varphi) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\varphi-\mu)^2}{2\sigma^2}} \tag{16}$$

φ represents a parameter that obeys a truncated normal distribution restricted to the range $0 \le \varphi \le 1$. The adoption threshold of node i on layer X can be determined by its number of connections (degree) on layer X. The magnitude of the parameter is influenced by μ (mean) and σ (standard deviation), which has an impact on the adoption threshold of the message. By adjusting μ and σ, the diversity of group behavior during information dissemination can be studied. When the amount of information m_i^X received by a susceptible node i reaches or exceeds its adoption threshold T_i^X, the node switches from a susceptible state to an adopted state.

2.4.3. Characterization of the distribution of side weights and group behavior. On this basis, a communication model based on the theoretical foundation of group behavior distribution is constructed, and the (individual information) communication pattern of individuals participating in the behavior of essay banding is analyzed. Taking ω as the weight of the edges, in the network layer $X(X \in \{A,B\})$, they show the distribution characteristics of $g_X(w)$ due to their randomness. In layer

X, the edges that have not yet made a message until moment t have probability $\theta_\omega^X(t)$ and their weights are labeled as ω. Therefore, at moment t, the probability that node i has not yet acquired a message passed from its neighbors in layer X is as follows:

$$\theta_X(t) = \sum_\omega g_A(\omega)\theta_\omega^X(t) \tag{17}$$

Where in layer X, the edge weight is ω and the probability that no message propagation occurs on that edge at moment t is $\theta_\omega^X(t)$.

The cumulative probability of node i receiving m_X messages in layer X can be expressed as:

$$\phi_{m_X}^A(k_i^A, t) = C_{k_i^X}^{m_X} \theta_X(t)^{k_i^A - m_X}[1-\theta_X(t)]^{m_X} \tag{18}$$

On a multilayer network model, the degree distribution of node i can be described by $\vec{k}_i = (k_i^A, k_i^B)$, while its information adoption threshold is defined by $T_i^x = \varphi_i^x k_i^x$. Thus, the probability that node i remains in a susceptible state at time t can be expressed as:

$$s(\vec{k}, t) = (1-\rho_0)\sum_\varphi f(\varphi)[\sum_{m_A=0}^{\lceil \varphi k^A \rceil - 1} \phi_{m_A}^A(k^A, t) \times \sum_{m_B=0}^{\lceil \varphi k^B \rceil - 1} \phi_{m_B}^B(k^B, t)] \tag{19}$$

Thus, the percentage of nodes in the network that are in a susceptible state in moment t can be expressed as follows:

$$S(t) = \sum_{\vec{k}} P(\vec{k}) s(\vec{k}, t) \tag{20}$$

A node in the network must exhibit one of the three states, S-state, A-state and R-state, and $\theta_w^X(t)$ can be represented as:

$$\theta_w^X(t) = \xi_{S,w}^X(t) + \xi_{A,w}^X(t) + \xi_{R,w}^X(t) \tag{21}$$

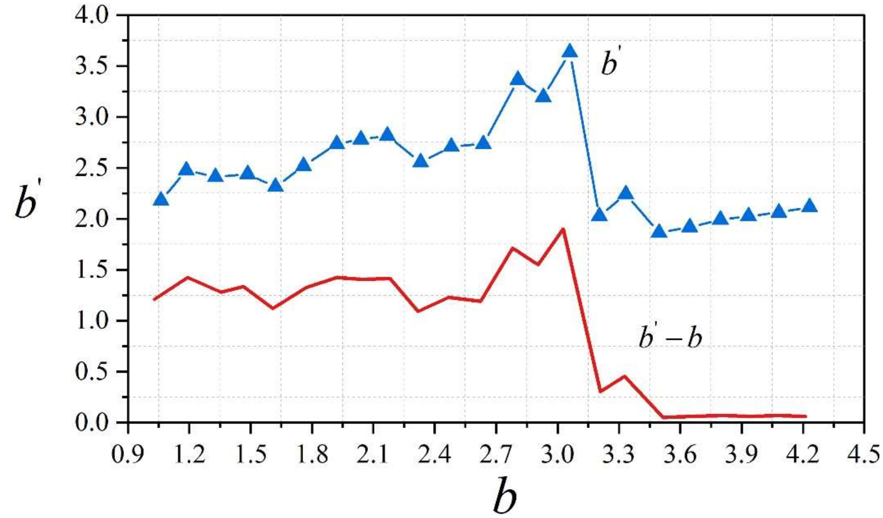
where $\xi_{S,w}^X(t), \xi_{A,w}^X(t)$ and $\xi_{R,w}^X(t)$ correspond to the probability that a node of degree k_i^X is in the susceptible, adopted, and recovered states, respectively, and has not passed information to neighboring nodes.

3. Dissemination of social media papers with reading behaviors

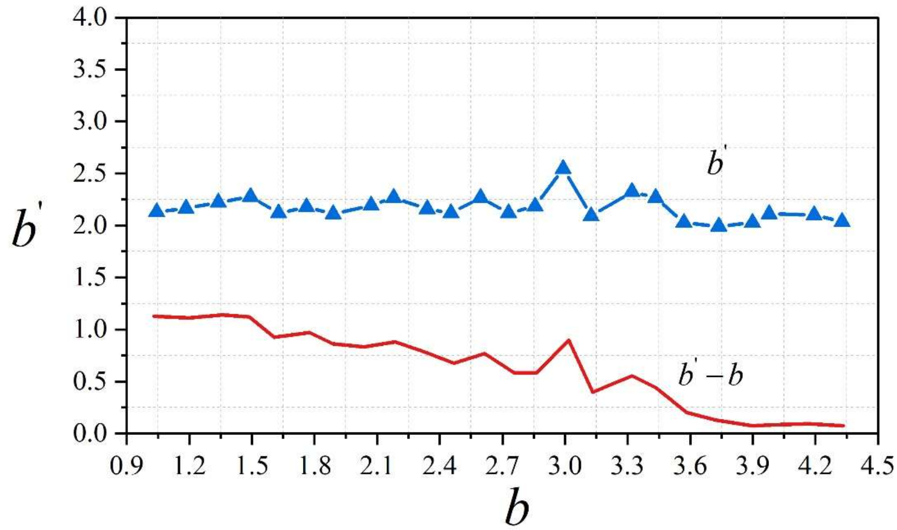
3.1. Generation and maintenance of "group cooperation" behavior

Considering the group behavior of social media essay with reading behavior as group cooperation, and the generation and maintenance of group cooperation as a continuous evolutionary process, we analyze the impact of betrayal gain b. That is, the strategy of each subnetwork node is set randomly under the initial b value, and after the subnetwork evolves and stabilizes and forms group cooperation, the value of b is further increased to analyze whether the group cooperation can survive the gradually deteriorating external environment. As the BA subnet is more conducive to cooperation, the generation and maintenance of group cooperation behavior in the BA subnet is shown in Figure 1.

The results show that BA subnetworks, once they form group cooperation, can maintain cooperative behavior in harsher external environments. For example, in the $3\times1000-4q$ -network, when the initial b = 3.2, the group cooperation formed by the BA subnetwork can effectively resist the intrusion of betrayal behaviors in the ER subnetwork and the WS subnetwork when the betrayal gain increases to a maximum of b' = 3.75. However, when the initial b-value is large and close to the maximum b-value for the formation of group cooperation, the BA subnetwork cannot resist the effect of the deterioration of the game environment. Meanwhile, the results of Fig. (a) and Fig. (b) show that the heuristic connecting edges are more conducive to resisting the intrusion of outside behaviors, which is related to the high returns of the highly nodes in the cooperative group.



(a) heuristic edge



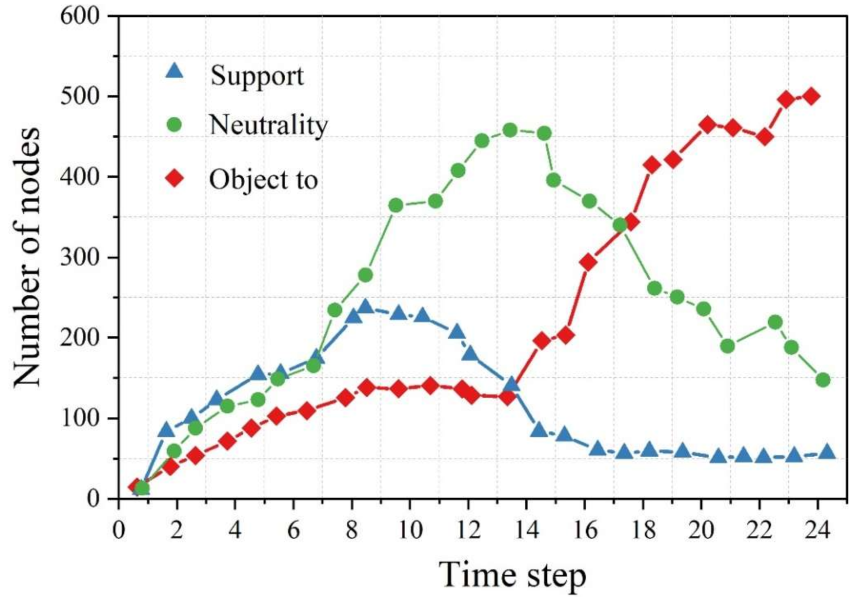
(b) random connection

Fig. 1. The cooperative behavior of bazinet group is produced and maintained

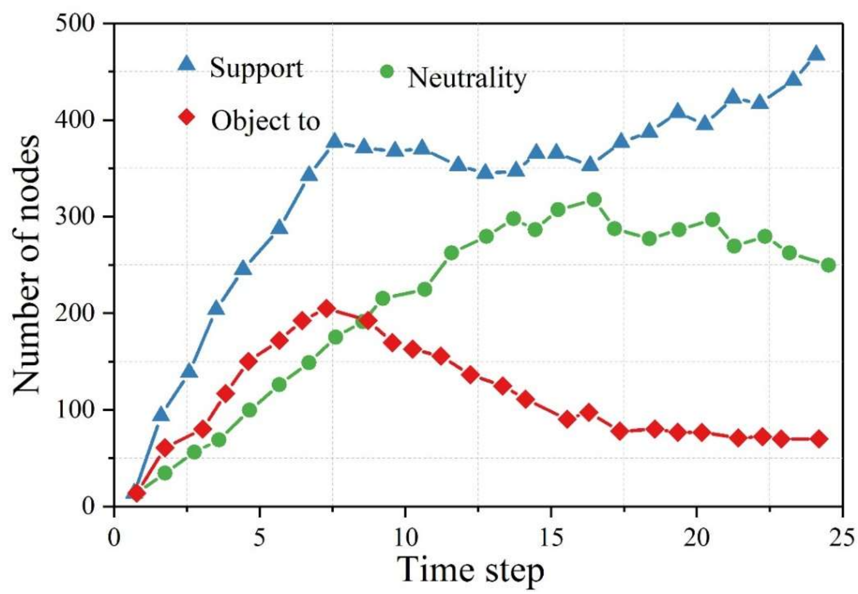
3.2. *Evolutionary Analysis of Complex Network Group Behavior Events*

3.2.1. Influence of Individuals' Own Attributes on Group Behavior. In the real network, different individuals are influenced by various aspects that lead them to have different views on the same event. Individuals with strong cognitive ability are better at analyzing information than others, so they often have their own unique views on events and do not change their views easily. Some individuals are easily influenced by other people's opinions, and thus blindly comply with them. Here, the degree of individual paranoia is selected as the main factor, and the effect of individual paranoia on group behavior is analyzed by adjusting the degree of individual paranoia in the simulation model.

The effect of individual paranoia degree on the evolution of group behavior (thesis banding behavior) is shown in Figure 2. The figure represents the evolution of group behavior when the individual node paranoia degree is 0.1 and 0.3, respectively. Individuals' attitudes towards group events are categorized into support, neutrality and opposition. In Fig. (a), when the time step is 13 time nodes, the three attitudes of individuals turn around respectively. As can be seen from the figure, with the growth of time, all three groups are in the state of growth at the initial stage of the event, and the individuals with high degree of paranoia grow faster. With the further evolution of time, the growth of the population in the state of support and neutrality tends to moderate, and the individuals in the state of opposition gradually change their views in the environment of group pressure, thus showing a downward trend. The curves are all characterized by wavy highs and lows. Thus, it seems that the paranoid individual nodes in individual nodes will have an important influence on the behavior of other individuals in the evolution of group behavior.



(a) Individual paranoia = 0.1



(b) Individual paranoia = 0.3

Fig. 2. The influence of individual paranoia on the evolution of group behavior

3.2.2. Influence of Neighbor Node Attitudes on Decision Propagation. This experiment focuses on testing the effect on the propagation decision by adjusting two parameters: the neighbor attitude $I_{i,j}$ that affects the individual's decision and whether the user who participates in the dissertation take-home activity is attractive to the individual $Ex_{i,j}$. The effects of neighbor node attitudes on decision propagation are shown in Figure 3. The experiment was set up with four groups of $I_{i,j}$ =0.6 and $Ex_{i,j}$ =0.8, $I_{i,j}$ =-0.6 and $Ex_{i,j}$ =0.8, $I_{i,j}$ =0.6 and $Ex_{i,j}$ =0.4, and $I_{i,j}$ =-0.6 and $Ex_{i,j}$ =0.4, respectively.

When $I_{i,j}$ =0.6 and $Ex_{i,j}$ =0.8, the individual's decision is supported by the neighbor's point of view. Users who have already participated in the dissertation banding activity have a strong attraction to the individual, so the individual's willingness to participate in the group behavior then becomes strong. When the degree of intensity increases to a certain value, the individual decides to participate in the thesis banding activity, at which point the individual is no longer influenced by the external environment.

When the neighbor's attitude is set to -0.6 and the attraction of users who have already participated in the essay take-home activity is unchanged, it means that the neighbor's view is opposite to the individual's own. However, since the friends who have already participated in the thesis banding activity are sufficiently attractive to them, this driver is strong enough to make the individual engage in purchasing behavior.

While changing the neighbor's attitude back to 0.6 and decreasing the buddy's attractiveness to 0.4, the individual's attitude to participate in the dissertation take-home activity is reinforced because the prevailing attitude is in line with the individual, and the individual will eventually participate in the dissertation take-home activity.

However, when the individual willingness is opposite to the neighbor's viewpoint and there are no close friends around to participate in dissertation banding group activities, then the individual willingness first decreases slowly. When it increases to a certain degree, individual willingness decreases rapidly and eventually gives up participating in the group behavior.

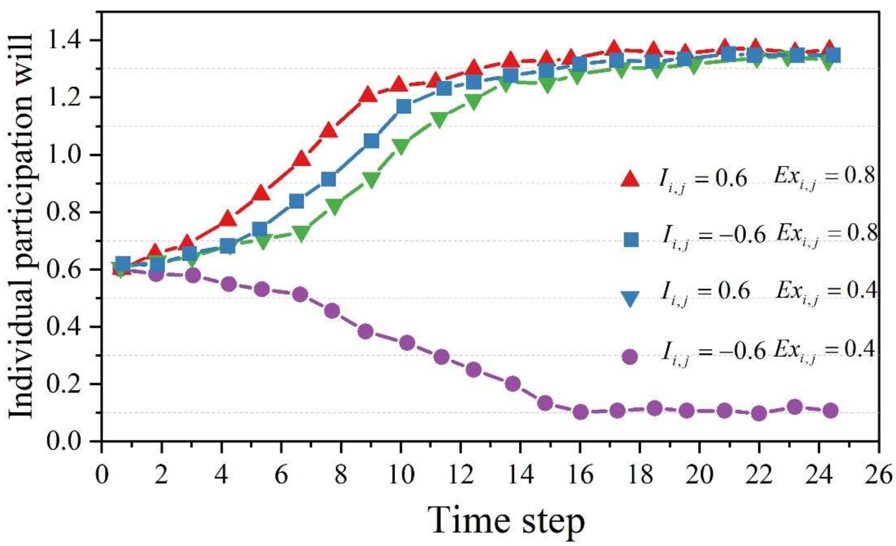


Fig. 3. The influence of the neighbor's attitude on the spread of the strategy

3.3. Analysis of the spread of social media essay bandwagon behavior

3.3.1. Weakly relational social media. Sina Weibo as one of the mainstream weak-relational social media platforms. Sina Weibo has provided data resources for research on the theory of weakly relational social media and for further improving the communication effects of weakly relational social media platforms, improving corporate communication strategies, and the role of the government in social governance.

The purpose of this study is to optimize the communication effect and evolution of information released by enterprises and other organizations on the weak-relationship social media platform. Therefore, the microblog accounts of three organizations in microblog hotspots are selected as the representatives of organizations with better communication effects on weakly relational social media platforms. The microblog information of these three microblog accounts from June 15, 2023 to March 15, 2024 including the number of retweets, the number of likes, the number of comments, as well as the microblog content retweeting structure data whose retweets exceeded the average value were obtained through the crawler program respectively.

In order to exclude the influence of social environment background on the results, the dataset is divided into two major backgrounds of general situation as well as social media official domination for analysis. Calculations were made to obtain the index data of microblog content dissemination topology pattern in the general period and social media official dominated events (thesis with reading behavior), and the microblog content dissemination topology pattern under different periods is shown in Table 1. Where C_s and C_d are the size of the cascade and the depth of the cascade, respectively. D_r represents the average shortest path length between forwarding nodes. D_N is the degree of propagation viralization.

The average cascade size C_s of the information forwarding network of microblog content is larger, indicating that in general period, the main forwarders of microblog content of large organizations are first-level disseminators. The depth of the information cascade of microblog content dissemination C_d is shallow.

The results of the cascade size and depth of the retweet network indicate that, in general, the content of large organizations on weakly relational social media is mainly disseminated through direct reading by users and direct retweeting by users, with only a small number of multilevel retweets occurring in a small number of cases.

Table 1. The topology pattern of microblogs in different times

Different periods	Variable	Mean	Median	Std
The topology pattern of micro-blog content propagation in general period	C_s	89.658%	93.291%	15.362%
	C_d	4.227	3.515	1.507
	D_r	4.034	3.167	0.413
	D_N	1.68×10^{-5}	4.701×10^{-6}	3.086×10^{-5}
The pattern of the spread of microblogs during the social media official leading event (the paper takes reading behavior)	C_s	71.253%	81.003%	26.798%
	C_d	4.114	2.072	1.893
	D_r	3.245	2.801	0.672
	D_N	10.223×10^{-6}	2.993×10^{-6}	1.615×10^{-5}

In order to further understand the role of broadcasting mode and viral mode on the influence of microblog content dissemination, the correlation between the degree of viralization of microblog content dissemination and the influence of dissemination is analyzed. A logistic regression model is set up to analyze the role of microblog content dissemination mode in dissemination influence.

The correlation between the variables in the general period is shown in Figure 4. The set of variables assessing the degree of viralization of microblog content dissemination is concentrated in the upper left corner of the matrix. The degree of dissemination viralization is positively correlated with cascade depth from the general period (Spearman correlation coefficient 0.66, $p<0.001$).

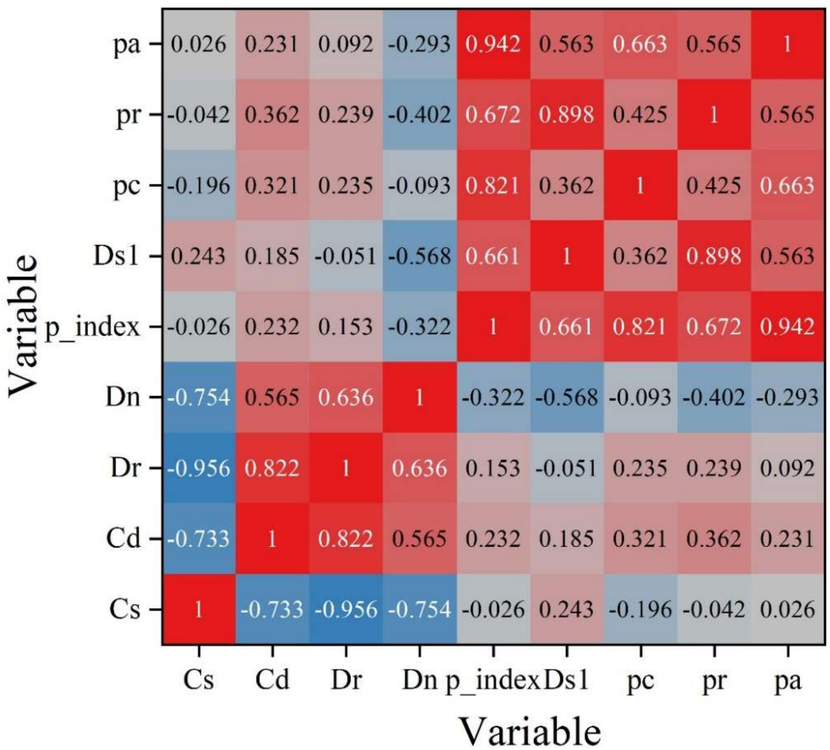


Fig. 4. The correlation between variables in general periods

The correlation of variables during the official social media dominated events (thesis bandwagon behavior) is shown in Figure 5. The other category is the set of variables evaluating the impact of microblogging content dissemination, which is concentrated in the lower right corner of the matrix. It shifts to a very weak negative correlation (Spearman correlation coefficient -0.072, $p<0.05$) during the official social media dominated event.

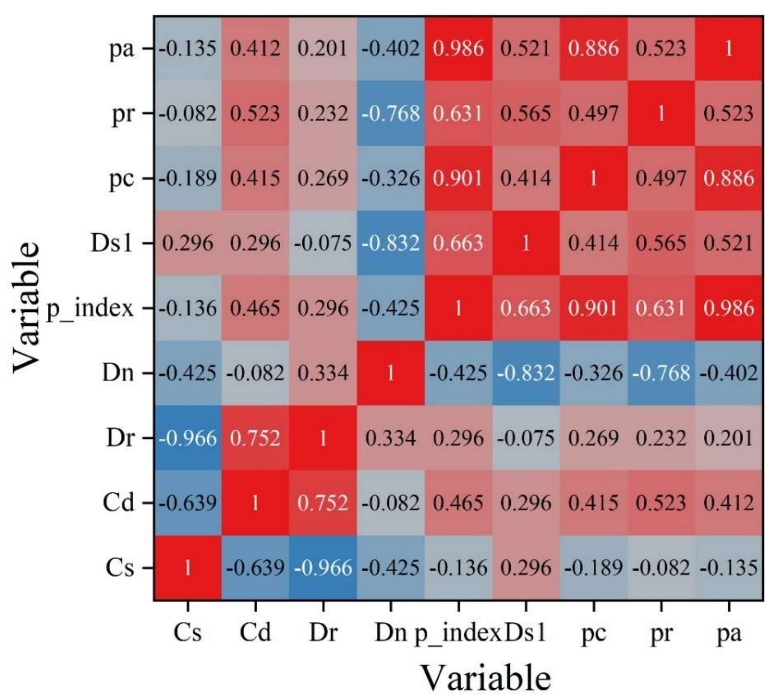


Fig. 5. The correlation between variable variables during social media official events

3.3.2. Strongly relational social media. In order to minimize the influence of the difference in the values of the variables other than the research factors in the model on the experimental results, the variables other than the research factors are constantized in the simulation experiments, and the fixed values are used as the initial state to investigate the operation law of the model of the propagation of the essay-carrying behavior in the strong relational social media platform with the constant change of the time variable t.

The initialization of the model simulation involves two aspects, the constant quantization of the variables, and the fixation of the random number sequence. The simulation experiments done in this paper are implemented under the following initialization conditions.

1) Constantization of variables. Thesis bandwagon behavior propagation onset moment $t_0 = 0$ in the value of information benefits model, $v_0 = 1$, $\lambda_0 = 1$, $\psi = 0.6$, $\sigma = 0.6$ in the pusher influence model, and $\varepsilon = 12$ in the public opinion environment model.

2) Random number sequence fixation. The random values on [0,1] are used to reflect the degree of the user's communicative personality influencing factors Q_i, the pushers' pushing strength β_i, and the intimacy between users r_i. In order to reduce the extreme values, it is assumed that the sequence of random numbers on [0,1] for all three obeys a normal distribution $N\left(\frac{1}{2},\frac{1}{6}\right)$.

The initial node degree is used as the independent variable, and the viewpoint resonance degree A_i is used as the control variable obeying a uniform distribution. The experiments simulate the evolution of thesis band-reading behavior letters sent from a particular node in different strongly relational social media network topologies. The correlations between the initial nodes with different node degrees and the percentage of spreaders, the percentage of silencers, and the sum of the two percentages under different network topologies are counted. The correlation coefficients between node degree and dissemination effect are shown in Table 2. In this case, the information audience refers to the sum of propagators and silencers.

The linear correlation between the initial node degree and the percentage of information audience for the dissemination of essay bandwagon behavior in random networks and small-world networks is relatively obvious and negative. The negative linear correlation of small-world network is larger than that of random network, i.e., the larger the initial node degree of propagation of thesis bandwagon behavior, the smaller the scope of propagation. In the scale-free network and Facebook social network, the initial node degree of the propagation of thesis banding behavior and the proportion of information audience are relatively obvious and positively correlated, i.e., the larger the initial node degree of the propagation, the larger the number of information audience.

Table 2. The correlation coefficient of the node degree and the propagation effect

	Communicator	Silent	Information audience
Random network node degree	0.0375	-0.052	-0.175
Small world network node degree	-0.412	0.000	-0.428
Unmarked network node degree	0.134	0.167	0.167
Facebook network node degree	-0.306	0.452	0.454

4. Conclusion

This paper establishes a heterogeneous network model for social media essay-reading behavior and analyzes the evolution of essay-reading behavior (group behavior) driven by social media. The distributional dynamics of group behavior is used to analyze the spreading power of social media essay-reading behavior.

1) Build an end-to-end heterogeneous network model of social media essay banding behavior and analyze the evolution of this group cooperation. When the BA subnetwork forms group cooperation, it is able to maintain cooperative behaviors in a harsher external environment. In $3\times1000-4q$ network with initial b of 3.2, the group cooperation formed by the BA subnetwork is able to effectively resist the intrusion of betrayal behaviors in the ER subnetwork and the WS subnetwork at (Betrayal Gain) $b^{'}$ = 3.75.
2) During the evolution of social media thesis banding behavior, the individual's own attributes will have an impact on the behavior of other individuals. In the early stage of the creation of thesis banding behavior events, and the individuals with high paranoia grow faster. The parameters of neighbor attitude $I_{i,j}$ and attraction $Ex_{i,j}$ of users participating in thesis banding events affect the decision of individuals to participate in thesis banding behavior to different degrees, respectively.
3) In weak-relationship social media, the main ways of spreading social media thesis-reading behavior are direct reading and direct forwarding by users, and only a few of them have a small amount of multilevel forwarding.

In the strong-relationship social media, the propagation degree of thesis-reading behavior is affected by the degree of the initial node of propagation, i.e., the larger the degree of the initial node of propagation, the larger the number of information audience and the greater the influence of the behavior.

References

[1] Pablo, J. C. I., & Lasaten, R. C. S. (2018). Writing difficulties and quality of academic essays of senior high school students. Asia Pacific Journal of Multidisciplinary Research, 6(4), 46-57.

[2] Paul, J., & Criado, A. R. (2020). The art of writing literature review: What do we know and what do we need to know?. International business review, 29(4), 101717.

[3] Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. Journal of planning education and research, 39(1), 93-112.

[4] Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies. Computers & education, 131, 33-48.

[5] Jo, C. W. (2021). Short vs. extended adolescent academic writing: A cross-genre analysis of writing skills in written definitions and persuasive essays. Journal of English for Academic Purposes, 53, 101014.

[6] Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2018). Peer feedback on academic writing: undergraduate students' peer feedback role, peer feedback perceptions and essay performance. Assessment & Evaluation in Higher Education, 43(6), 955-968.

[7] Bulqiyah, S., Mahbub, M., & Nugraheni, D. A. (2021). Investigating Writing Difficulties in Essay Writing: Tertiary Students' Perspectives. English Language Teaching Educational Journal, 4(1), 61-73.

[8] Turmudi, D. (2017). Rethinking academic essay writing: Selected genres in comparison. Premise: Journal of English Education and Applied Linguistics, 6(2), 119-138.

[9] Goctu, R. (2017). Metacognitive strategies in academic writing. Journal of Education in Black Sea Region, 2(2).

[10] Greetham, B. (2022). How to write better essays. Bloomsbury Publishing.

[11] Rahmawati, H., & Sholihah, F. A. (2019). The students' perception on academic reading log for essay writing literacy: Some influential points. JEES (Journal of English Educators Society), 4(1), 66-70.

[12] Yu, S., & Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. Assessing Writing, 48, 100525.

[13] Greenhalgh, T. M., & Dijkstra, P. (2025). How to read a paper. John Wiley & Sons.

[14] Ruijuan, Z., Jing, C., Mingchuan, Z., Junlong, Z., & Qingtao, W. (2016). User abnormal behavior analysis based on neural network clustering. The Journal of China Universities of Posts and Telecommunications, 23(3), 29-44.

[15] Garlisi, D., Martino, A., Zouwayhed, J., Pourrahim, R., & Cuomo, F. (2023). Exploratory approach for network behavior clustering in LoRaWAN. Journal of Ambient Intelligence and Humanized Computing, 14(12), 15745-15759.

[16] Sivanathan, A., Gharakheili, H. H., & Sivaraman, V. (2020). Detecting behavioral change of IoT devices using clustering-based network traffic modeling. IEEE Internet of Things Journal, 7(8), 7295-7309.

[17] Artioli, P., Maci, A., & Magrì, A. (2024). A comprehensive investigation of clustering algorithms for User and Entity Behavior Analytics. Frontiers in big Data, 7, 1375818.

[18] Wang, Y. (2020). User online behavior based on big data distributed clustering algorithm. International Journal of Advanced Robotic Systems, 17(2), 1729881420917293.

[19] Wohwe Sambo, D., Yenke, B. O., Förster, A., & Dayang, P. (2019). Optimized clustering algorithms for large wireless sensor networks: A review. Sensors, 19(2), 322.

[20] Qian, X., Yuemaier, A., Yang, W., Chen, X., Liang, L., Li, S., ... & Song, Z. (2023). A self-organizing multi-layer agent computing system for behavioral clustering recognition. Sensors, 23(12), 5435.

[21] Amutha, J., Sharma, S., & Sharma, S. K. (2021). Strategies based on various aspects of clustering in wireless sensor networks using classical, optimization and machine learning techniques: Review, taxonomy, research findings, challenges and future directions. Computer Science Review, 40, 100376.

[22] Turčanik, M. (2020, October). Web users clustering by their behaviour on the network. In 2020 New Trends in Signal Processing (NTSP) (pp. 1-5). IEEE.

[23] Ibtisam Aldawish & Rabha W. Ibrahim. (2024). A new self-organization of complex networks structure generalized by a new class of fractional differential equations generated by 3D-gamma function. Journal of King Saud University - Science(11),103512-103512.

[24] Siqin Su,Jintao Liu,Yongqi Wang & Keqiang Dong. (2024). Research on Topological Characteristics of Spatial Network Based on Complex Network Theory and Its Applications. Mathematics(23),3746-3746.

[25] Viney Kumar,Chris T Bauch & Samit Bhattacharyya. (2024). A game theoretic complex network model to estimate the epidemic threshold under individual vaccination behaviour and adaptive social connections.. Scientific reports(1),29148.