International Journal of Economic Behavior and Organization
2023; 11(3): 125-134
http://www.sciencepublishinggroup.com/j/ijebo
doi: 10.11648/j.ijebo.20231103.12
ISSN: 2328-7608 (Print); ISSN: 2328-7616 (Online)

Identification and Ranking of High-Risk Nodes in Complex Financial Networks

Yishuo Wu[1, *], **Zhongjun Li**[2]

[1]International Department Sino-American Class, Jinling High School, Nanjing, China
[2]School of Management Science and Engineering, Southwestern University of Finance and Economics, Chengdu, China

Email address:
wuyishuo0302@126.com (Yishuo Wu), lizj_t@swufe.edu.cn (Zhongjun Li)
[*]Corresponding author

To cite this article:
Yishuo Wu, Zhongjun Li. Identification and Ranking of High-Risk Nodes in Complex Financial Networks. *International Journal of Economic Behavior and Organization*. Vol. 11, No. 3, 2023, pp. 125-134. doi: 10.11648/j.ijebo.20231103.12

Received: June 18, 2023; **Accepted**: July 12, 2023; **Published**: July 17, 2023

Abstract: Financial markets are ever closer interconnected, the superimposed impact and scope of multiple systemic financial risks such as epidemics, wars, and supply chains are expanding day by day, and the risk of contagion and spread of financial crises has become a problem that cannot be ignored. With the deepening of the theoretical research of complex networks, the combination of systemic financial risks and complex networks has become more closely connected. Financial networks are characterized by the accumulation of multiple risks, and their overall stability depends on the stability of specific nodes in the network. Therefore, accurately identifying and ranking high-risk nodes has become a difficult point restricting the improvement of resource utilization efficiency, and it has become extremely critical. Based on the complex network theory, firstly, a directed graph from the massive financial risk warning information is constructed in this paper, and obtains the subject community of financial risk incidents. Then the superimposed impact of multiple systemic financial risks is measured from the three dimensions of complex network topology, financial risk behavior and risk propagation probability. Finally, the influence distribution and dissemination rules of nodes are analyzed in the subject community, and a high-risk nodes identification algorithm CIRA (community-based identifying and ranking algorithm) is proposed. Experiments show that the algorithm can effectively mine potential high-risk nodes and obtain higher risk density.

Keywords: Systemic Financial Risk, Complex Network, Network Topology Structure, Risk Behavior, Risk Spreading

1. Introduction

In recent years, financial risk events such as the Russia-Ukraine war, the epidemic, and the European debt crisis have occurred continuously. These sudden and catastrophic events have brought serious humanitarian, economic and social challenges to countries around the world, and the global political and economic situation has undergone profound changes. At the same time, it has also spawned a series of new forms of international cooperation and competition, exacerbating the instability of the global financial system and the volatility of financial markets. In this context, social and economic subjects have become a community of economic interests, and the systemic characteristics of the financial crisis are also particularly prominent, and risk factors are shared. Because the interwinded feature allows risks to spread throughout the financial markets, the financial stability can be threatened through cascading in financial networks [1]. The contagion characteristic of financial risk is one of the important causes of financial crisis.

The financial crisis is an eternal phenomenon, which is always caused by the risk accumulation and finally leads to the crisis. Looking at the previous financial crises in the world, they have experienced the evolution process from risk accumulation, explosive events, transmission to crisis deepening [2]. Looking back at the process of the typical COVID-19 epidemic continuing to trigger financial crisis risks: From the perspective of the transmission mechanism, the financial crisis is transmitted to the banking crisis and economic crisis through the balance sheet recession, and the domestic crisis is transmitted to the world through the chains such as trade, external demand, financial markets and other.

From the perspective of response, the timeliness and effectiveness of policy responses will affect the degree of damage to the crisis by alleviating the liquidity crisis, restoring solvency, and blocking the spread of financial network risks in a timely manner [3].

Under the background of financial liberalization and economic globalization, financial risks present a high degree of complexity and variability [4]. Financial risk has multiple dimensions from the length of the cycle to different coverage. The cumulative contagion effect of these multi-dimensional financial risks should attract special attention, because it is generally not easy to find, and once it occurs, the impact will be very huge. Corresponding to the type and quantity of financial risks, massive systemic financial risks information data is generated. These massive, sporadic and messy financial risks data are often not equal to true and effective risk information, nor can they constitute effective crisis incidents alone, nor can even form the guiding basis for risk prevention and control measures.

On the one hand, the deluge of redundant, insignificant warning data creates enormous stress and illusion for socioeconomic subjects or regulatory departments, and even the serious warnings are disregarded. How to efficiently handle the massive amounts of financial risks warning information has become a challenging task. Due to the lack of timely assessment of financial risks, risk prevention and control measures are seriously delayed. Only by realizing intelligent early warning based on the perception of harmful behavior, intelligent filtering and correlation of warning information, it is possible to fully realize the efficient response to financial risks information [5].

On the other hand, the risk warning information is scattered and messy, and potential financial crisis incidents are not easy to find. Generally, the medium-high-risk warning information can be processed in time, while the large number of medium-low-risk warning information is easily ignored. Obviously, some scattered warning information is correlated to each other. Behind the huge information set, there are systematically consistent and real and effective subject of financial crisis incidents. Through correlation analysis and merging, it is possible to dig out the conduction mechanism hidden in the massive warning information. Regrettably, these interrelated and widely distributed low-risk warning information are always ignored and omitted due to disorder and dispersion, until the cumulative effect of risks led to serious consequences [6].

Issues of financial stability have become increasingly sensitive and important, it is necessary to proactively prevent and effectively resolve various hidden risks in the economic and financial fields, and put the prevention and control of financial risks in a more important position [7]. For social and economic entities and regulatory authorities, it is of great significance for preventing and defusing financial risks to analyse the cumulative contagion effect of financial risks and identify systemically important financial institutions in financial risk contagion. Therefore, it is necessary to establish a complete risk analysis framework for the cumulative contagion effect of multi-dimensional financial risks.

The CIRA (community-based identifying and ranking algorithm) financial incident and risk analysis system described in this paper mainly includes the following works:

First, a complex network processing model for massive scattered risk warning information is proposed. Based on the theory of complex networks, a semantic data warehousing of financial stability warning information is generated.

Second, the discovery of subject community of financial risk incidents. Data extraction and data cleaning are carried out on the raw data of financial stability warning information, and then normalization, classification, and aggregation are used to identify financial risk subject incidents, then form a subject community.

Third, the identification and ranking of potential high-risk nodes. It not only analyses the structural features and behavioral features of complex network entity nodes, but also analyses the risk propagation probability, the risk density characteristics of network entity nodes are obtained.

2. Related Work

It is found that the normal operation of functions in a complex network greatly depends on a number of important nodes. In recent years, research on key nodes in complex networks has attracted widespread attention [8]. A large number of studies have shown that large-scale complex networks abstracted from various real world have three basic statistical characteristics: small world [9], scale-free and clustering [10]. The small world phenomenon shows that the key nodes in the complex network need to be responded to with priority, and the propagation speed of financial risks will be reduced more rapidly. The scale-free characteristic and the clustering characteristic multiply the effectiveness of the countermeasures against financial risks. Under the response mechanism of the targeted strategy, the scale-free network appears extremely fragile. It is precisely because of the above three characteristics that the research on nodes importance is of great significance for the financial stability. The heterogeneity of entity nodes in financial risk complex network determines the non-equivalence of the node status.

Previous work has studied the systemic financial risk contagion of financial networks from different perspectives.

A proper assessment of firm-specific risk must take into account potential risk spill over effects from other firms [11]. The potential impact of interconnected financial institutions on the overall financial system has been a financial stability concern for central banks and regulators. The need for an economic basis for systemic risk measures is not merely an academic issue, as it involves regulators, regulators, and policymakers [12, 13]. Cai's results emphasize that reducing risk at the institutional-level through diversification ignores the negative externalities of an interconnected financial system [14]. The above research has shown the importance of cross-sectional dependencies of assets, credit exposures and volatility, which can threaten the financial stability through cascades in financial networks.

And most studies have shown that the network structure is the key to affecting systemic financial risk contagion. Some entity nodes have a greater influence on other entity nodes, and can play a key guiding role in the development of the financial risk situation, which called as "influential high-risk nodes" [15]. Shahzard et al. applies a bivariate cross-quantilogram method to examine the spill over network structure of stock markets in 58 countries. And the directionality of risk contagion and key nodes in risk contagion is analysed, so as to identify the strongest interdependencies, the directionality of the spill over risk effects, and to detect those equity markets with the potential to cause global systemic risk [16]. Based on the view of risk spill over, Liu Chao et al. apply the GARCH model and the generalized forecast error variance, and use the variance contribution calculated by the generalized forecast error variance decomposition as the adjacency matrix to construct a financial risk spill over network. The direction and intensity of risk contagion are analysed from dynamic and static perspectives, and the risk centres and their evolution are identified within crisis applying approaches of spill over index and complex network [17]. One of the important research issues on the complex networks is the quantitative analysis of the importance of all nodes, so as to discover and mine the core nodes. The research on the influence model of node in complex network has great theoretical value and practical application value, which has the applications in many fields due to its universality [18, 19]. In a complex network abstracted by financial risk information, the local features, the global features, and the random walk algorithms are comprehensively considered, which is of great significance for the study of high-risk node feature models and the design of high-risk node mining algorithms.

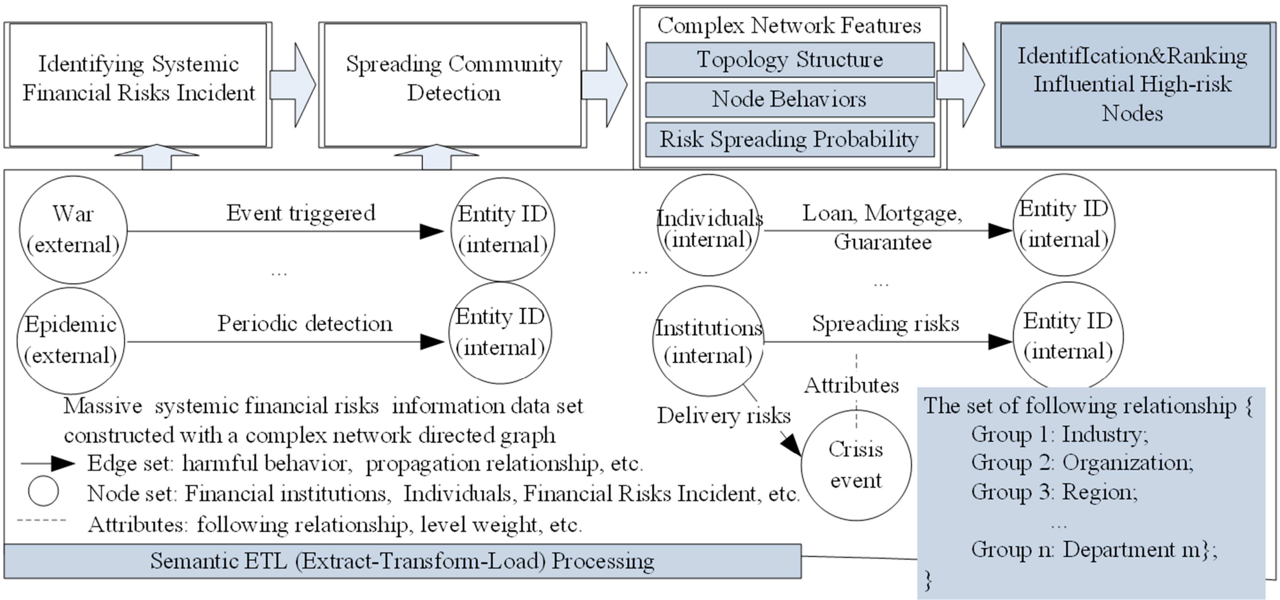

Identifying Systemic Financial Risks Incident
Spreading Community Detection
Complex Network Features
Topology Structure
Node Behaviors
Risk Spreading Probability
IdentifIcation&Ranking Influential High-risk Nodes
War (external)
Event triggered
Entity ID (internal)
Individuals (internal)
Loan, Mortgage, Guarantee
Entity ID (internal)
...
...
...
Epidemic (external)
Periodic detection
Entity ID (internal)
Institutions (internal)
Spreading risks
Entity ID (internal)
Attributes
Delivery risks
Crisis event
Massive systemic financial risks information data set constructed with a complex network directed graph
Edge set: harmful behavior, propagation relationship, etc.
Node set: Financial institutions, Individuals, Financial Risks Incident, etc.
Attributes: following relationship, level weight, etc.
Semantic ETL (Extract-Transform-Load) Processing
The set of following relationship {
Group 1: Industry;
Group 2: Organization;
Group 3: Region;
...
Group n: Department m};
}

Figure 1. *The overall framework for identification and ranking influential high-risk nodes.*

3. Constructing a Directed Graph from the Massive Financial Risk Warning Information

The massive financial risk warning information is of great analysis and mining value for data-driven financial risk incident prediction and disposal. The data assets that have undergone data pre-processing are called as risk warning metadata.

In the data pre-processing stage, the following operations are performed, grouping by time interval, valid field extraction and data cleaning. Grouping by time interval will set a time window *T* as a period, and group warning metadata containing the same behavior according to *T*, then perform subsequent aggregation processing of host node ID. Generally, the controlled nodes infected with the same risks show a certain degree of consistency in their behavior patterns. Therefore, their warning metadata has certain periodic characteristics. The setting of a reasonable time interval can make the detection of certain harmful behaviors more accurate. The effective field extraction will extract the key fields from warning metadata, such as timestamp, node ID, access domain name, etc. Data cleaning completes redundant data filtering and filtering by white list. At the stage, the processes are finished, such as data record linking, semantic integrating, and data object labeling, thereby ensuring data quality and credibility. A unified format of financial risk incident set $V= \{v_1, v_2, ..., v_n\}$ is formed, and the output structure of its elements is a sequence of <*timestamp, node ID, risk incident*>.

The Figure 1 shows the overall framework for identification and ranking influential high-risk nodes. It mainly includes four aspects of work:

1) Generate semantic data warehousing of financial risk

warning information.

2) Construct subject community of risk incident. Through the data analysis of semantic data warehousing, the clustering algorithm is used to identify risk incidents, and the node mapping of risk incidents is performed after the risk incident set is obtained, then form a subject community.
3) The features analysis of an influential high-risk nodes. From the warning data resources, the three main features of complex network nodes: the structure features, the behavior features, and the risk diffusion probability are extracted to describe high-risk nodes.
4) The identifying and ranking algorithm of influential high-risk nodes. On the basis of analyzing the features of complex network nodes, a identifying and ranking algorithm of high-risk nodes is devised.

3.1. Semantic Data Warehousing of Financial Risk Warning Information

As shown in Figure 2, the semantic ETL (extraction, transfor-mation, and loading) process for financial risk warning information is devised to enable complex network models and warning data to be dynamically integrated into the data warehouse. In which, RSD (Raw Structured Data) is a massive raw financial risk warning information. UML, the de facto standard for object-oriented visual modeling, is used as a visual modeling for warning data flow. OML (Ontology Model Language) is an engineering modeling method that describes the logic relationship between concepts and any data assets from various detection nodes of financial risk. It provides class hierarchies, profiles, properties, and equivalencies for risk warning. It also provides a means for multiple ontologies to coexist and for mappings to be defined between them such as harmful behavior, victim node. The data extractor handles CQL (Continuous Query Language) queries of RSD metadatas and the relationships between those RSD metadatas. Then, the ETL Process delivers the RSD data to a transformer service instance that provides an RSD to OML transformation. The resulting OML instance data is then loaded in the risk warning data warehouse. An automatic transformation process from warning metadata (annotated UML information models) into OML ontologies is implemented by the data ontology generator. The data warehouse of risk warning is a semantic data warehousing, allows users to define which data sources they are interested in and automates the ETL process through semantic ETL (SETL) [20] across entire classes of data sources. It also is dynamic data stores, which can model and integrate new data sources from diverse detection services of financial risk on the fly.

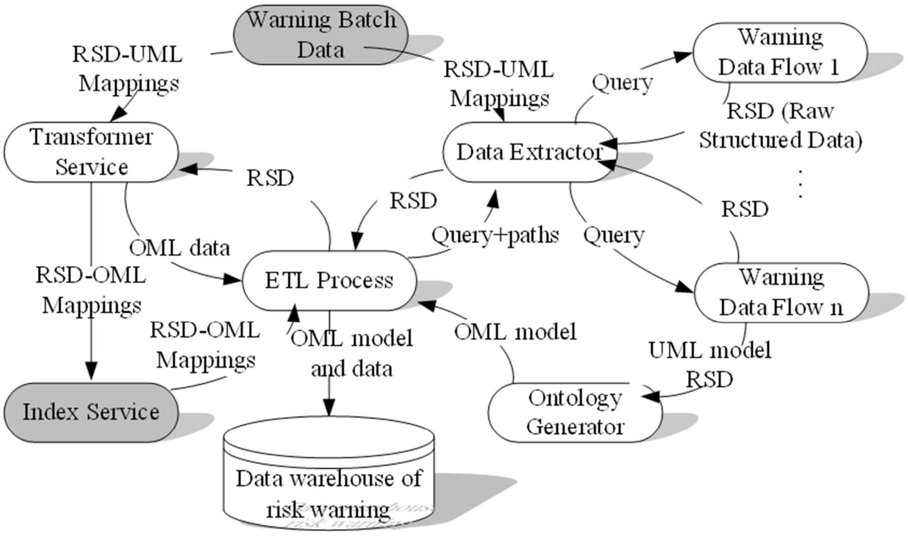


Figure 2. *The semantic ETL process for financial risk information.*

3.2. Transformer: The Generation Framework of Directed Graph

As shown in Figure 3, the generation framework of directed graph implements an automatic transformation process from warning metadata, that is RSD, into OML ontologies. The following relation is defined as a farm-out relationship or subordination relationship formed for a certain service imple-mentation, there are frequent service interactions among relationship subjects. Such as, the relationship between the financial institution and the loan enterprise, or among the guarantors in a guarantee network structure, or the relationship among the “peer nodes” in a P2P lending network. A guarantee network is used as an example to illustrate the generation mechanism of directed graph. According to the New Basel Capital Accord, the loan status can be divided into five statuses: normal, concern, secondary, and suspect, and loss. For the convenience of research, this paper only considers the two statuses of the enterprise, and the loss status can be regarded as being infected by risk status, and considers normal, concern, secondary, and suspect as vulnerable to risk status.

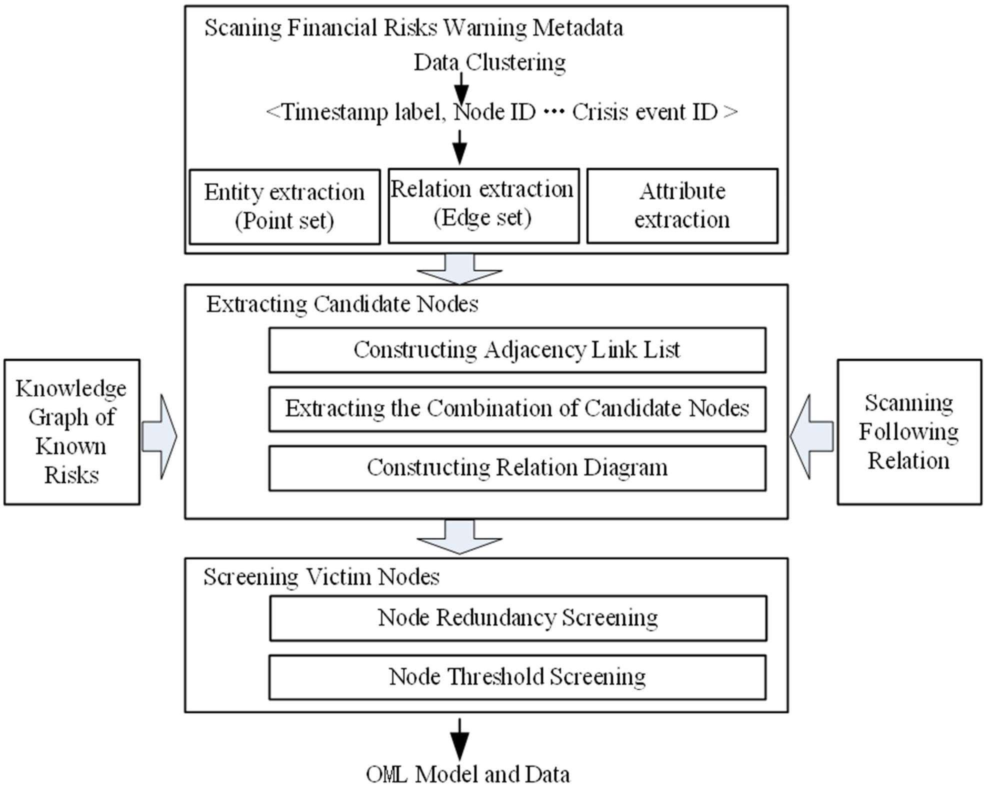

Scaning Financial Risks Warning Metadata
Data Clustering
<Timestamp label, Node ID ··· Crisis event ID >
Entity extraction (Point set)
Relation extraction (Edge set)
Attribute extraction
Extracting Candidate Nodes
Constructing Adjacency Link List
Extracting the Combination of Candidate Nodes
Constructing Relation Diagram
Knowledge Graph of Known Risks
Scanning Following Relation
Screening Victim Nodes
Node Redundancy Screening
Node Threshold Screening
OML Model and Data

Figure 3. *The generation framework of directed graph from financial risk metadata.*

Data clustering is to perform aggregation operations on risk warning metadata in a time interval. A certain type of loan is used as the key value, the guarantee network nodes are aggregated, and the clustering result is <*timestamp, loan type, node ID set*>, where the timestamp is used to mark the occurrence of these loan operations time interval. The bipartite graph method is used to construct the access relationship graph. There are two types of nodes in the bipartite graph of the access relationship. The node set *{node ID$_1$, node ID$_2$... node ID$_n$}* represents the node ID set with guarantee or loan behavior, and the node set *D = {timestamp#loan type$_1$, timestamp#loan type$_2$... timestamp# loan type$_n$}* represents the loan type set in the time interval *T*. The guarantee relationship set $A = <hi, di >, (hi \in H, di \in D)$ is defined to represent the adjacency relationship between the node ID h_i and its corresponding loan type d_i. Based on the guarantee relationship set, each loan type node and its corresponding adjacent node are constructed as a guarantee relationship linked list. In the linked list, the element in the *D* set is the head node of the linked list, and the subsequent node is the node ID node in the *H* set. The order of the node ID is in descending order according to the degree of the node in the entire relationship graph. Further, the different linked lists are arranged in order, and the rule is the arrangement in descending order according to the degree of the loan type node in the entire relationship graph.

Each path starting from the root node in the above-mentioned access relationship linked list is a combination of candidate nodes. The nodes set on the path represents the set of host IDs, and the list of <*timestamp#loan type*> of the root node in a linked list represents the collection of loan type in common by these node IDs. Then a candidate guarantee node combination C_b*=Structure<hi set, di set>*, $(hi \in H, di \in D)$ is formed. In the combined structures of all candidate nodes, if there is an inclusion relationship, it is necessary to remove all included data records through a redundant screening.

At the same time, in order to improve the accuracy of candidate risk status node combinations, a threshold screening is used to select the results with higher infection probability. For example, a threshold T_h for the number of node ID and a threshold T_d for the number of loan type are set. For all candidate node combinations, the data records that meets the following conditions are retained: $C_{bi}=\{<h_i,d_i>\}$, $sizeof\ (h_i)>T_h$, $sizeof\ (d_i)>T_d$. The nodes that show the same loan or guarantee behavior many times are considered the high-risk nodes infected by botnets. Therefore, for a data record in the screening results, if the greater the number of elements in H set and the greater the number of elements in *D* set, then the higher the probability that these nodes will be infected.

The warning metadata set $V_i = \{v_1, v_2,..., v_n\}$ is divided into two types of OML ontologies: node set and edge set. These two types of ontologies are imported during the RSD metadata to OML transformation process, and are the same for any RSD-derived OML ontology. The loan source, guarantor, and associated object information (files) involved in a financial risk incident are stored as nodes, and the node set $N_i = \{n_1, n_2,..., n_n\}$ is used to store the set of network entities involved in harmful behavior. The system assigns the

KEY value to uniquely identify the ontologies data. Edge sets are used to store the behavior relationships between network entities, and are divided into two categories: the harmful behavior relationship set $E_i= \{e_1, e_2,...., e_n\}$, and the non-harmful behavior relationship set $!E_i =\{!e_1,!e_2,....,!e_n\}$. A specific behavior relationship in the edge set is called an edge, and the system assigns the KEY value of the edge set to (source point, destination point), which is used to uniquely identify a specific edge in the directed graph. Furthermore, the corresponding relationship between the node set and the edge set is constructed. The corresponding relationship must be directed, and the same relationship may correspond to multiple source points and end points. The directed graph from financial risk warning metadata $G_i=\{g_1, g_2,...., g_n\}$ is formed. Upon receiving the request for computing attack chain and spreading chain from the system, the calculation of the global weakly connected graph is executed, then the attack chain, infection chain, or spreading area are automatically obtained.

4. Identification and Ranking of High-risk Nodes

4.1. Topology Construction of Subject Community of Financial Risk Incidents

In the process of the spread of financial risks, the behavioral relationships among complex network entity nodes are all related to specific financial risk incidents. The entity nodes in a complex network with similar risk behaviors usually form a virtual community with "financial risk incidents" as the core, which is called subject community on financial risk incidents in this paper. The features of financial risk incident in the subject community provide rich semantic support for the feature extraction of entity nodes in a complex network. It is helpful to mine potential high-risk nodes in the infection chain, and help the design of identification and ranking algorithms.

Based on the knowledge graph of financial risk incident, the corresponding interactive network topology can be constructed according to the directed graph set G_i related to the specific financial risk incident t_i. The construction process can be vividly described in Figure 4. First, the carriers of all financial risk incidents are found out in G_i of level 2 to form a entity node set U in complex network. First, the carriers of all financial risk incidents are found out in G_i set of level 2 to form an entity node set U in complex network. Then, the interaction relationships among entity nodes in the raw data set are extracted and add them to the set U, the subject community topology is obtained, as shown in level 1 in Figure 4. Algorithm 1 shows the construction process of the community topology of the financial risk incident subject.

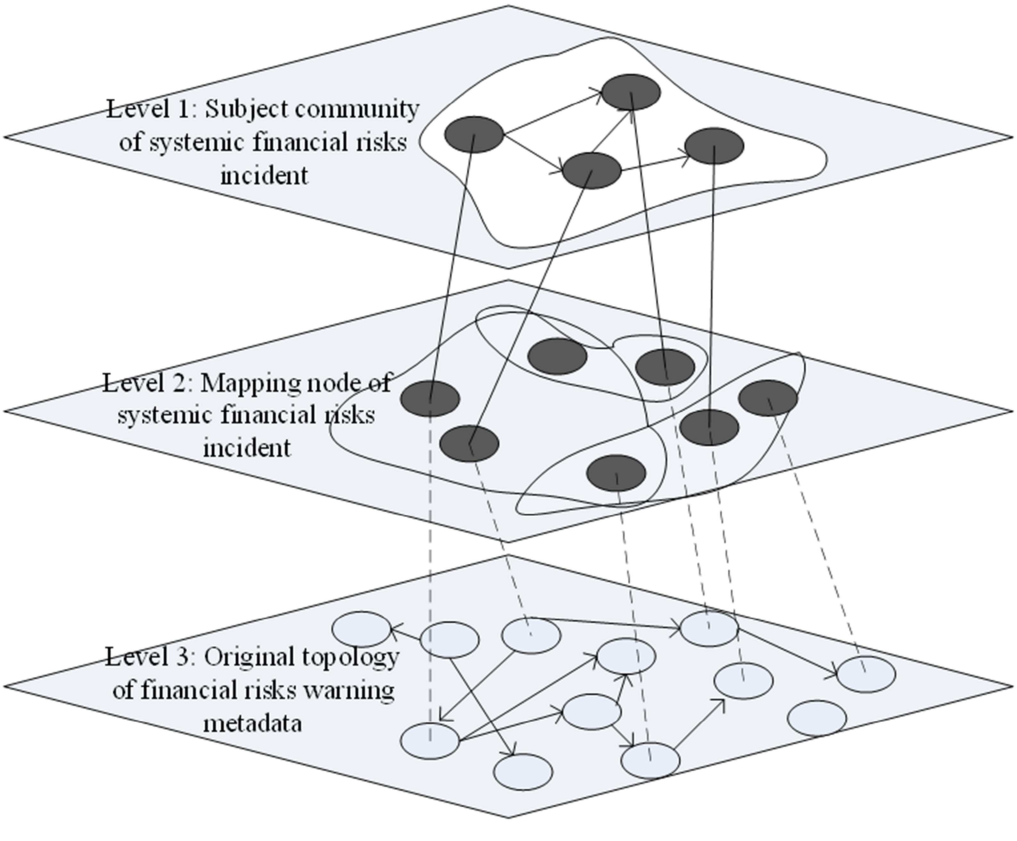
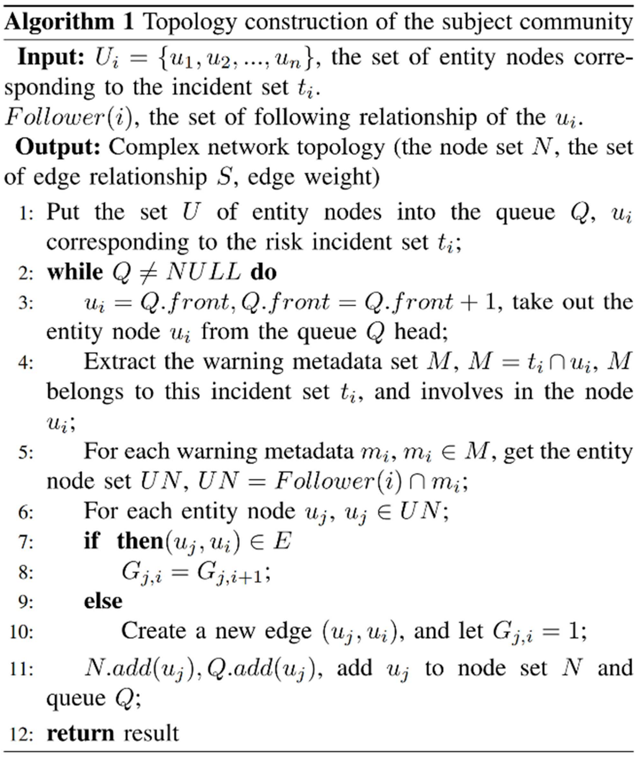

Level 1: Subject community of systemic financial risks incident
Level 2: Mapping node of systemic financial risks incident
Level 3: Original topology of financial risks warning metadata

Figure 4. *The construction of subject community of financial incident.*



Algorithm 1 Topology construction of the subject community

```
Input: $U_i = \{u_1, u_2, ..., u_n\}$, the set of entity nodes corresponding to the incident set $t_i$.
$Follower(i)$, the set of following relationship of the $u_i$.
Output: Complex network topology (the node set $N$, the set of edge relationship $S$, edge weight)
Put the set $U$ of entity nodes into the queue $Q$, $u_i$ corresponding to the risk incident set $t_i$;
while $Q \neq NULL$ do
    $u_i = Q.front, Q.front = Q.front + 1$, take out the entity node $u_i$ from the queue $Q$ head;
    Extract the warning metadata set $M$, $M = t_i \cap u_i$, $M$ belongs to this incident set $t_i$, and involves in the node $u_i$;
    For each warning metadata $m_i$, $m_i \in M$, get the entity node set $UN$, $UN = Follower(i) \cap m_i$;
    For each entity node $u_j$, $u_j \in UN$;
    if then$(u_j, u_i) \in E$
        $G_{j,i} = G_{j,i+1}$;
    else
        Create a new edge $(u_j, u_i)$, and let $G_{j,i} = 1$;
    $N.add(u_j), Q.add(u_j)$, add $u_j$ to node set $N$ and queue $Q$;
return result
```

4.2. Features Analysis of High-Risk Nodes

After the construction of the financial risk incident subject community, mining the influence features of entity nodes has become a key factor for identification and ranking high-risk nodes. The influence of entity nodes is the result of joint action of multiple complex factors. The structure, behavior and risk spreading probability of entity nodes are selected as the influence features in this paper.

4.2.1. Structural Features

The structural features reflect the structural influence of the complex network topology and the influence of the entity node itself, such as the betweenness centrality of the entity node, the nodes number of accessing and the nodes number of being accessed [21]. From the topology model of the directed graph of the financial risk, the feature value can be obtained and normalized. The maximum and minimum normalization method is adopted in this paper. Assuming that a certain feature value is quantified as f, the maximum value is f_{max}, and the minimum value is f_{min}, the normalized value f_n is:

$$f_n = \frac{f - f_{min}}{f_{max} - f_{min}} \tag{1}$$

The average value of the normalized value is used as the structural feature value of an entity node:

$$S(u) = (u_{betweenness} + u_{access} + u_{accessed})/3 \tag{2}$$

Among them, $u_{betweenness}$ is the normalized value of betweenness centrality, u_{access} is the normalized value of the nodes number of accessing, and $u_{accessed}$ is the normalized value of the nodes number of being accessed.

4.2.2. Behavioral Features

Behavioral features are summarized in the following two points:

1) Activeness: the number of effective malicious behaviors initiated, forwarded, and responded to by an entity node within a unit time, denoted as u_{active};

2) Spreading power: the effective number of malicious actions of an entity node that are forwarded and responded to by other entity node within a unit time, denoted as $u_{spread.}$

In the topological structure, these two characteristics of activeness and spreading power are quantified and normalized, and then the average value is taken to obtain the behavioral feature value of an entity node.

$$B(u) = (u_{active} + u_{spread})/2 \tag{3}$$

4.2.3. Risk Spreading Probability (Rsp) Features

For any entity node i, define its risk status as $R_{i,t}$ at time t, $R_{i,t}=1$ indicates that a financial risk has occurred, and $R_{i,t}=0$ indicates that the entity node is not yet infected by a risk.

Each entity node in the directed graph is a relatively independent node, and each node has a certain difference in its capability against financial risk due to its different susceptibility status. Network entities in reality have different possibilities of being affected by risks because of their different networking environments, management levels, and operating services. However, for different types of viruses or threats, the probability of different network entities being infected or affected is close. Just like the probability of being infected by COVID-19 viruses is usually higher than the probability of being infected by AIDS. Assuming that there is a following relationship between the entity node u_i and the entity node u_j with service interaction, when the entity node u_j has a cyber risk, the probability that the entity node u_i is infected by the risk k is β_k, which corresponds to the warning level outputted by the intrusion detection node, $\beta_k \in (high, medium, low)$. Obviously, if an entity node has a following relationship with two other nodes with service interaction, and the other two nodes have risk k, the probability that the entity node u_i is infected by risk k is $1-(1-\beta_k)^2$. Therefore, the risk spreading probability feature of a single node can be expressed as the probability that node u_i is infected by risk k at time $t+1$:

$$P(R_{i,t} = 1) = 1 - (1 - \beta_k)^{\alpha_{i,t}}, i = 1, 2, \dots, N. \tag{4}$$

in which, $k \in Risk(i)$, $Risk(i)$ are the full set of risks of entity node u_i, $\alpha_{i,t}$ is the sum of risks of all nodes that point to i, when there is an edge from node u_j to node u_i, $s_{j,i}=1$, otherwise, $s_{j,i}=0$.

$$\alpha_{i,t} = \sum_{j=1}^{N} s_{j,i} * R_{j,t}, i = 1, 2, \dots N. \tag{5}$$

4.3. Algorithm for Mining High-Risk Nodes

Considering the structural features, behavior features and risk spreading probability features of entity nodes, a high-risk node mining algorithm CIRA is proposed. The expected result of the mining algorithm is that the entity nodes with the following characteristics should have a higher priority for processing.

1) Normalize the risk quantity and warning level of entity nodes itself, and the result of the entity node is at the forefront.
2) The weighted path length between the entity node and all following nodes is at the forefront.

The algorithm uses formula 6 to calculate the financial risk influence value of an entity node.

$$CIRA(Structure, Behavior, Rsp) = INF(u_i) = (1-d)\big(S(u_i) + B(u_i)\big) + d * \sum_{k \in Risk(i)}^{|Risk(i)|} \left[P(R_{i,t}=1) + \frac{P(R_{i,t}=1)}{\sum_{l \in Follower(i)}^{Follower(i)} P(R_{l,t}=1)} * INF(u_j) \right] \tag{6}$$

Where the *INF(u$_i$)* is the influence value of the entity node u_i, the *S(u$_i$)* is the normalized structural feature of the u_i, the *B(u$_i$)* is the normalized behavior feature of the u_i, the *P(R$_{i, t+1}$=1)* is the feature weight of the risk spreading probability of the u_i, the *Followers(i)* is the full set of nodes that have a following relationship with node u_i, *d* is a damping factor and its value is set 0.8 here. The CIRA algorithm draws on the idea of PageRank, a webpage importance ranking algorithm, and believes that the financial risk influence of an entity node is not only closely related to its structural and behavioral features, but also depends on the risk spreading probability of its following nodes. If the following node of the u_i has a greater risk spreading probability to the entity node u_i, the greater the corresponding influence weight is also, and the greater the contribution to the u_i influence. On the other hand, the financial risk influence of u_i is also related to the influence of its following nodes. If the financial risk influence of its following nodes is generally high, it will greatly contribute to the financial risk influence of the u_i. Therefore, the *CIRA (Structure, Behavior, Rsp)* algorithm not only has the advantages of PageRank, but also combines the features of risk spreading probability to discover in-depth influencing factors. The detailed description is as follows algorithm 2.

Assuming that the total number of entity nodes in the topology model is *n* and the number of iterations is *K*, then the time complexity of the above CIRA algorithm is *O(Kn^2)*.

Algorithm 2 CIRA mining algorithm for high-risk node by multi-feature

```
Input: Complex network topology relationship (the set $N$ of node, the set $S$ of edge, and risk propagation probability weight $P$)
Output: Node influence ranking (Top-n)
Initialize $\varepsilon$, let $\varepsilon > \varepsilon_{threshold}$;
while $\varepsilon > \varepsilon_{threshold}$ do
    $\varepsilon = 0$;
    Calculate the contribution of following nodes $v_j$ to $v_i$: $INF(v_i) = INF_{old}(v_i) + d * \sum_{k \in Risk(i)}^{|Risk(i)|} P(R_{i,t+1} = 1) * INF_{old}(v_j)$;
    Calculate the sum of the difference between the two iterations before and after: $\varepsilon += |INF(v_i) - INF_{old}(v_i)|$;
return Top-n nodes with greater influence.
```

5. Experiment Analysis

5.1. Data Description

The main parameters that measure the constructed complex network are shown in Table 1. Among them, the "maximal connected subgraph" is the number of nodes contained in the subgraph, the "average out-degree" and the "average in-degree" refer to the measured value of the node with a removal degree of 0. The "average clustering coefficient" and the "average path length" are the measurement values that the maximal connected subgraph is transformed into an undirected graph. Although the average path length is deviated from the optimal value of 6 in small world network theory, and the average clustering coefficient is larger, it is in the acceptable range. These two indicators reflect that the financial risk information network conforms to the characteristics of the small world network. The out-degree and frequency of the node are measured, and the double logarithmic form is used to draw as shown in Figure 5. It can be seen that the out-degree distribution of the nodes conforms to the power law distribution well. It is a scale-free network with power law coefficients of -3.0267. The in-degree distribution is similar and will not be repeated here. Its small-world characteristics and scale-free characteris-tics show that the directed graph of financial risk information is a complex network.

Table 1. *Basic measurements of complex network about financial risk.*

Parameter name	Parameter value
Number of nodes	36878
Number of edges	31496
Maximal connected subgraph	8815
Average out-degree	1.5211
Average in-degree	1.6152
Maximal out-degree	16
Maximal in-degree	144
Average clustering coefficient	0.1214
Average path length	13.1524

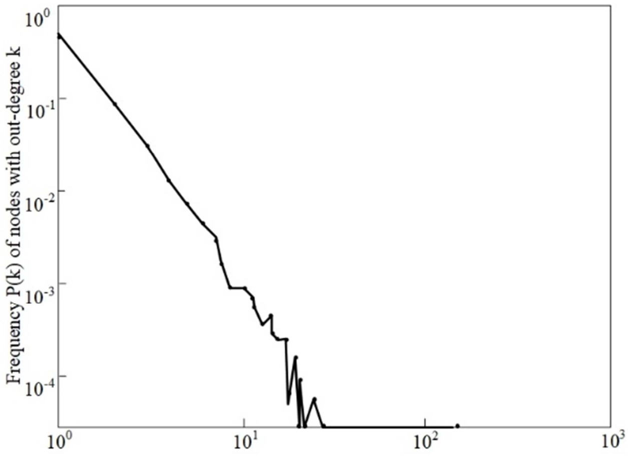


Figure 5. *Double logarithm diagram of the out-degree distribution.*

5.2. Comparison Algorithms

In order to verify the effectiveness of the CIRA algorithm, two comparison algorithms [22, 23] and comparison experiments are designed in this paper, are described as follows.

RspRank algorithm: A ranking algorithm based on the features of risk spreading probability. In the subject community of financial risk incident, the ratio of the Rsp weighted value of an entity node to the weighted value of the total Rsp is defined as the influence of the entity node.

RspHITS algorithm: On the basis of the original structure of HITS, take the Rsp of the entity node as the weight of centrality and authority, and calculate the weighted centrality and authority according to formula 7:

$$Authority(v_i) = \sum_{V_i,V_j \in E} W_{ji} * Hub(v_j)$$

$$Hub(v_i) = \sum_{V_i,V_j \in E} W_{ji} * Authority(v_j) \qquad (7)$$

5.3. Experimental Results and Analysis

The evaluation indicator Risk Density (RD) is used. Define the risk density at the time *t* as: $\rho_t(v_i)=1/N*|v_j|$, that is, the number of harmful behaviors (including the weight of risk level, which has been normalized) infected or spread by the node for the proportion of all harmful behaviors. *N* represents the total number of harmful behaviors in the community, and $|v_j|$ represents the number of harmful behaviors initiated and received between node v_i and all following nodes v_j. For directed networks, the lower $\rho_t \rightarrow \infty$ means the stronger the anti-risk capability of the node. Conversely, the nodes with higher ρ_t at a certain moment need more attention.

Figure 6 shows the comparison results of the three algorithms. It can be seen that the CIRA algorithm and the RspHITS algorithm by multi-features are obviously better than the RspRank algorithm for direct ranking. When the selected Top-K is smaller, the CIRA algorithm can obtain a higher risk density. Although the RspRank algorithm considers the features of risk spreading probability, it is only an evaluation of the influence of individual entity node, while the CIRA algorithm not only considers individual factors, but also comprehensively considers the potential impact relationship among following nodes. This also confirms the spreading law of the influence interaction among entity nodes in complex networks. Similarly, the RspHITS algorithm uses authority and centrality to reflect the mutual influence between network nodes, which has greater advantages than the RspRank method.

6. Conclusion and Recommendations

As more and more raw data sets are available from financial quantitative analysis nodes, the effective data analysis methods are needed to improve the efficiency of financial risk prevention and control. A processing model established for the massive amount of scattered financial information is proposed in this paper, the risk information metadata is normalized, classified and aggregated. They are used to generate complex networks that contain logical relationships among harmful behaviors, entity nodes, and propagation paths. On this basis, starting from the topological structure and risk spreading probability features in the complex network, the issue of mining high-risk nodes is studied based on the multi-feature analysis method. Various comparative experiments show that the CIRA algorithm proposed in this paper can more effectively find out the high-risk nodes in domain-specific, and the obtained high-risk nodes have a higher risk density.

The dynamics of financial risk complex networks is an issue that needs attention in the future. From the perspective of the dynamic development of the financial risk situation, new crisis incidents will continue to emerge and new harmful behaviors will emerge between nodes. The real-time changes of the complex network itself determine that the entire complex network is an evolving dynamic graph. For the identifying and ranking of high-risk nodes, such development and changes need to be considered. At the same time, the influence of overlapping communities and structural holes between communities needs to be carefully considered in the methodology to further improve the accuracy of the identifying and ranking of influential high-risk nodes.

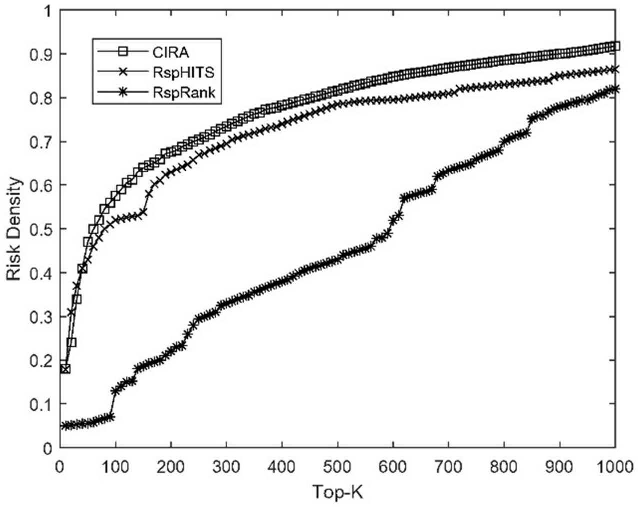


Figure 6. Comparison results of three algorithms considering Rsp features: CIRA, RspRank and RspHITS.

As the core of modern economy, financial stability is a major prerequisite for economic stability. In order to achieve the goal of financial stability and economic stability, it is necessary to proactively prevent and effectively resolve various risks and hidden dangers in the economic and financial field, and put the prevention and control of financial risks in a more significant position. The method in this paper has the following important policy implications: First, in order to achieve the goal of healthy operation of the financial market, the risk management and the supervision for financial institutions should be strengthened; the healthy operation of the financial market should be promoted; the various processes and orders should be standardized; and the cumulative contagion effect of financial risks should be analyzed on a regular basis. Regulatory authorities should strengthen the identification and supervision of financial institutions with the feature of systemically important in financial risk contagion, prevent the accumulation and spread of systemic risks, and cut off the domino effect of financial risk contagion. Secondly, regulatory authorities need to improve the evaluation indicators of systemic importance; establish and improve the index system of scale, relevance and complexity, etc.; fundamentally explore the risk contagion path of financial institutions; and improve the accuracy and operational efficiency of comprehensive response measures to global financial turmoil.

References

[1] Meryem Duygun, Daniel Ladley, et al., "Challenges to global financial stability: Interconnections, credit risk, business cycle and the role of market participants", Journal of Banking & Finance, vol. 112, 105735, 2020.

[2] Yang Zihui, Chen Yutian, et al., "A Literature Review of Systemic Risk: Status, Development and Prospect", Journal of Financial Research, no. 1, pp. 185-206, 2022.

[3] Ikram, Muhammad, Y. Sayagh, "The Consequences of COVID-19 Disruption on Sustainable Economy in the Top 30 High-Tech Innovative Countries," Global Journal of Flexible Systems Management, vol. 24, no. 2, pp. 247-269, 2023.

[4] Tafakori, L., Pourkhanali, et al. "Measuring systemic risk and contagion in the European financial network", Empirical Economics, vol. 63, pp. 345–389, 2022.

[5] Shi, Yong, Zheng, Yuanchun, Guo, Kun, Jin, Zhenni, Huang, Zili, "The Evolution Characteristics of Systemic Risk in China's Stock Market Based on a Dynamic Complex Network", ENTROPY, vol. 22, no. 6, doi: 10.3390/e2206061-4, 2020.

[6] De Santis R A. "Unobservable systematic risk, economic activity and stock market", Journal of Banking & Finance, vol. 97, 51-69, 2018.

[7] OuYang Zisheng, Yang Xite, et al., "Research on Systemic Risk Contagion Effect of Chinese Financial Institutions Considering Network Public Opinion Index", Chinese Journal of Management Science, vol. 30, no. 4, pp. 1-12, 2022.

[8] X. Zhou, F. M. Zhang, et al., "Finding vital node by node importance evaluation", Acta Phys. Sin, vol. 61, no. 5, pp. 0502011-0502017, 2012.

[9] D. J. Watts, S. H. Strogatz, "Collective dynamics of 'small-world' networks", Nature, vol. 393, no. 6684, pp. 440-442, 1998.

[10] S. H. Strogatz, "Exploring complex networks", Nature, vol. 410, no. 6825, pp. 268-276, 2001.

[11] Nikolaus Hautsch and others, Financial Network Systemic Risk Contributions, Review of Finance, vol. 19, no. 2, pp. 685–738, March 2015.

[12] Viral V. Acharya, Anjan V. Thakor, "The dark side of liquidity creation: Leverage and systemic risk," Journal of Financial Intermediation, vol. 28, pp. 4-21, 2016.

[13] Adams, Z., Füss, R., et al., "Spillover Effects among Financial Institutions: A State-Dependent Sensitivity Value-at-Risk Approach", Journal of Financial and Quantitative Analysis, vol. 49, no. 3, pp. 575-598, 2014.

[14] Jian Cai, Frederik Eidam, et al., "Syndication, interconnected-ness, and systemic risk", Journal of Financial Stability, vol. 34, pp. 105-120, 2018.

[15] Syed Jawad Hussain Shahzad, Jose Areola Hernandez, et al., "A global network topology of stock markets: Transmitters and receivers of spillover effects," Physical A: Statistical Mechanics and its Applications, vol. 492, 2136-2153, 2018.

[16] Shahzad, S. J. H., et al., "A global network topology of stock markets: Transmitters and receivers of spillover effects", Physica A: Statistical Mechanics and its Applica-tions, vol. 492, 2136-2153, 2018.

[17] Liu Chao, Xu Junhui, Zhou Wenwen, "Study on risk spillover effect of financial markets in China based on methods of spillover index and complex network", Systems Engineering-Theory & Practice, no. 4, pp. 831-842, 2017.

[18] Li M, Zhang R, Hu R, et al., "Identifying and ranking influential spreaders in complex networks by combining a local-degree sum and the clustering coefficient", International Journal of Modern Physics B, vol. 32, no. 06, 1850118, 2018.

[19] Wang P, Tian C, Lu J A, "Identifying influential spreaders in artificial complex networks", Journal of Systems Science & Complexity, vol. 27, no. 8, pp. 650-665, 2014.

[20] Nath R P D, Hose K, et al. "SETL: A Programmable Semantic Extract-Transform-Load Framework for Semantic Data Warehouses", Information Systems, vol. 68, no. 8, pp. 17-43, 2017.

[21] Rosvall M, Bergstrom C T, "An information-theoretic framework for resolving community structure in complex networks", Proceedings of the National Academy of Sciences, vol. 104, no. 18, pp. 7327-7331, 2007.

[22] J. G. Liu, Z. M. Ren, et al., "Node importance ranking of complex networks", Acta Phys. Sin, vol. 62, no. 17, pp. 1789011- 1789019, 2013.

[23] Berkhin, Pavel, "A Survey on PageRank Computing", Internet Mathematics, vol. 2, no. 1, pp. 73-120, 2005.