Prediction and Analysis of the Social Network Structure of Financial Institutions Based on the Dissemination of Public Opinion

Lingyi Li(✉) and Yuejiang Peng

School of Computer Science and Engineering, Sun Yat-Sen University, Guangdong, China
leialy2001@163.com

Abstract. Along with the rapid development of the Internet in China, the Internet has become an important and indispensable component of our modern life. At the same time, with the rapid and vigorous economic and social development after the reform and opening up, we have to admit that it has also brought some negative impacts, and the frequent emergence of financial emergencies on the Internet has made the network emergencies on the Internet a problem that cannot be ignored and needs to be solved. This paper collects information dissemination data of the "e-Speed Loan" online emergency as the basis of empirical research, and uses social network analysis to study the dissemination of public opinion after the occurrence of the financial institution's reputation security online emergency. In terms of practical significance, we can grasp the characteristics of online public opinion in the process of dissemination, provide countermeasures and suggestions for solving online emergencies, and provide reference significance for the handling of online emergencies.

Keywords: Public opinion dissemination · financial institutions · social network prediction method · online emergencies

1 Introduction

Usually, the stage division of emergency online public opinion is theoretically based on the life cycle theory of emergency management. According to this theory, the stages of public opinion evolution can be roughly divided into three stages, four stages or five stages [1]. In this paper, the stages of public opinion evolution in financial security emergencies are divided into three stages, namely, the gestation stage of public opinion, the diffusion stage of public opinion and the decline stage of public opinion [2]. In this paper, we will analyze the stages of public opinion on financial security emergencies by analyzing the characteristics of each stage of the development of public opinion on the "e-Speed Loan" incident as a typical case.

J. Yen et al. (Eds.): ICBIS 2023, AHCS 14, pp. 270–277, 2023.
https://doi.org/10.2991/978-94-6463-198-2_30

Table 1. Selected data samples

	NetEase Finance	Home of Online Lending	Sohu Finance	Wall Street News
NetEase Finance	0	1	0	0
Home of Online Lending	1	0	20	1
Sohu Finance	0	20	0	0
Wall Street News	0	1	0	0
Zero One Finance	0	0	0	1
Flush Finance	0	0	0	10
Yeo.com	0	0	0	17

2 Data Capture of the Public Opinion Network of the "E-Speed Loan" Incident

The data source of this paper requires a large amount of public opinion data, and the analysis of the relationships among network members requires the establishment of a matrix of their interactive relationships [3]. The snowball method will be used for node selection in this paper. In this paper, we select individuals and triggers with high activity in the "e-Speed Loan" incident and capture data from public opinion platforms including larger news networks. Similarly, data from personal opinion platforms (e.g., Weibo) will also be included.

2.1 The Way Data is Captured

In this paper, we select some propagation nodes, and then connect the nodes by following, retweeting and commenting among them. In this paper, the "interaction matrix" and the undirected network structure diagram of the public opinion communication network are used [4]. In the interaction matrix, the horizontal rows of the matrix represent the sources of information, the vertical columns represent the recipients of information, and the "real numbers" and "zeros" are used to indicate the existence and strength of the relationship between actors [5]. If member A retweets and follows member B's message, it means that there is a relationship between A and B. The strength of the relationship is expressed as the sum of the number of comments and likes [6], which makes this interaction matrix a weighted matrix. The details are as follows:

The following Tables 1 and 2 shows some of the data captured in this paper:

2.2 Rationalization of Data Crawling

The explanation for the plausibility of the data in this paper comes from three main points:

Table 2.

	NetEase Finance	Home of Online Lending	Sohu Finance
NetEase Finance	0	1	0
Home of Online Lending	1	0	20
Sohu Finance	0	20	0
Wall Street News	0	1	0
Zero One Finance	0	0	0
Flush Finane	0	0	0
Yeo.com	0	0	0

Firstly, since this paper does not stop at a particular media node to collect data, but adopts the "snowball sampling method", which can ensure that the data collection is not biased and can show the complete relationship between the nodes of the whole network.

Second, this paper is a comprehensive data collection for video websites, which can show the relationship of video website class nodes in a more comprehensive way.

3 Measurement and Analysis of the Overall Structure of the Public Opinion Network of the "E-Speed Loan" Incident

3.1 Network Density Measurements

According to Jun Liu's "Holistic Network Analysis", which defines network density as follows, network density is a measure of the level at which a member of the holistic network transmits messages to each other. In a social network, the numerical value of network density is calculated as the ratio of the value of the real existing connectedness to the maximum amount of connections generated by all actors in that overall network being connected to each other.

We first measured the overall network density and the individual network density using UCINET software, as follows.

(1) Measurement of "individual network density" in the overall network

When studying the structure of the overall network, it is important not to neglect the study of each member's composition. Measuring the network density of the overall network is first studied at the individual micro level, which can give us a clearer understanding of the importance of the role played by individuals in the overall network.

We used UCINET software to measure the individual network density in this network during the "e-Speed" event. Due to the relatively large number of nodes, this paper only shows the top ten nodes with high individual network density in Table 3, which can be observed in Table 3.

As you can visually see from the table, Eastern Fortune ranks first in all nodes. In the column of "size", the value of the node "Oriental Fortune" is 35, which indicates the size of the individual network of "Oriental Fortune" in the overall network, and

Table 3. Measurements of "individual network density" of the online opinion network of the "e-Speed Loan" incident

Nodes	Size	Relationships	Maximum number of relationships	Density
Oriental wealth network	35	94	1190	7.90%
NetEase Finance	23	40	506	7.91%
Peaceful Goose City	19	36	342	10.53%
Beijing Business News	14	22	182	12.09%
China Economic Network	14	18	182	9.89%
First Financial Daily	13	20	156	12.82%
Mutual Fund Network	13	16	156	10.26%
Wall Street News	12	20	132	15.15%
Sina Finance	12	18	132	13.64%
LeTV Video	12	18	132	13.64%

represents the number of members that have relationship with "Oriental Fortune" is 35. The second indicator "logarithm" is the number of relationships with the node "Oriental Fortune" (excluding self-self-relationships), and the total number of relationships for "Oriental Fortune" is 94. In the column of "Maximum number of pairs", the value of "Oriental Wealth" is 1190, which indicates the maximum total number of relationships that can exist for the individual "Oriental Wealth" in the individual network. Number. The "density" in the table represents the ratio between the total number of relationships and the maximum number of possible pairs of this node, i.e. $94/1190 = 7.90\%$.

(2) Measurement of the network density of the overall network

In contrast to the measure of individual networks, the network density measure of the overall network is a measure of the overall structure, specifically, the overall network network density is equal to the number of pairs of relationships actually contained in this overall network versus the number of relationships that may exist in this overall network from a theoretical perspective. The measured results are shown in Table 4 below:

From the table, we can see that there are a total of 108 nodes in the opinion network about the "e-Speed Loan" incident, but the number of real connections among the 108 nodes is only 394, so we can measure that the overall density of this network is only

Table 4. Results of the overall density measurement of public opinion network on the "e-Speed Loan" incident

Number of nodes	108
Total number of relationships	0394
Overall Density	0.1121

0.1121. According to Liu Jun, as described in his monograph "Integral Network Analysis", in real life, the number of relationships between nodes is usually smaller than the number of all relationships that may exist for all nodes in this network, and he used a random selection model to analyze that the maximum value of network density in reality does not exceed 0.3.So the network density of 0.1121 shows that the density between nodes in this network is tighter and the information flow conveys better density in this network.

3.2 Measurements of Network Clustering Coefficients for the Overall Network

The clustering coefficient is a parameter that reflects the level of convergence of all nodes in a network. In the overall network, if the clustering coefficient of a network is high, it means that the nodes in this network are in a densely connected state and all nodes have closer relationships. Specifically, the clustering coefficient is the ratio of the actual number of relationships between a node and other nodes in its proximity and the theoretical maximum possible number of relationships between this node and its proximity, and it takes a value between 0 and 1. The larger the value, the better the cohesiveness of the node. The measured results are shown in Table 4 below:

From the above table, we can see that the node with the highest clustering coefficient is "West.com", which reaches 1 in the opinion network of "e-Speed Loan". This means that 100% of the nodes around Hexun.com can communicate directly with it in this opinion network, which indicates that the cohesiveness of this node is better. It can be noted that "NetEase Finance" has the highest number of pairs, with 595 pairs, but the clustering coefficient is only 0.145. Since the clustering coefficient is a ratio relationship, this explains why the node "NetEase Finance" has a high logarithm of composition, but a low cohesion coefficient. The measured results are shown in Table 5 below (Table 6):

From the above table, we can see that the average clustering coefficient of the overall network is 0.127 in the opinion network of e-Speed Loan. In real life, this value is not high, which means that the overall cohesion coefficient of this public opinion network

Table 5. Clustering coefficients of the public opinion network of the "e-Speed Loan" incident

Western Network	1.000	6
Online Loan Storm	0.700	10
Online Loan Interconnection	0.667	6
Online Lending World	0.667	6
Beijing Business News	0.593	91
Panorama.com	0.528	36
CICC	0.333	15
Qingdao Licang Finance	0.303	66
MoneyGram.com	0.273	55
Zhongan Online	0.267	15

Table 6. Average data of clustering coefficients of the public opinion network on the "e quick loan" incident

Overall network clustering coefficient	0.127
Weighted average overall network clustering coefficient	0.128

is low, and the information spreads in the network mainly relies on indirect ways and has a narrow space for dissemination.

3.3 Reachability Measurements in Networks

The reachability measure measures the length of the diameter of information propagation between nodes in this opinion network and the average diameter of propagation of the overall network. In a network, the shorter the length of the diameter means that in this network, the information propagates through fewer links and the number of nodes through which the information passes.

4 The Centrality of Location in the Public Opinion Network of the "E-Speed Loan" Incident

Location centrality, as the name implies, is the degree to which a node is centrally located in this network. In general social network analysis, scholars often study the "power" of a node in the network, that is, the extent to which the node influences other nodes in the network.

4.1 Point Degree Centrality Measure

In our study of point degree centrality, we use two main metrics, one is point degree centrality and the other is point degree centrality potential. The point degree centrality indicates the ability of a node to transmit information directly with other nodes. The higher the point centrality of a node, the closer the node's position in the network is to the center of the entire network, and the more direct connections it has with other nodes, the more "power" it has.

4.2 Intermediary Centrality Measure

Mediation centrality, also known as intermediate centrality. Literally, the intermediate centrality measures the importance of the intermediate node between two nodes. For example, if there is a connection between A and B, but the connection must be made through point H, it means that the connection between A and B is dependent on point H for its existence. If a node is at the connection point of multiple interaction paths, it goes without saying that the node has the ability to control the other two directional nodes on the interaction path, and once the node misinterprets the information, it will affect the

reception of information from other nodes connected to that point, and even misinterpret the information of the whole network. Correspondingly, the intermediary-centric potential is concerned with the intermediary-centric dynamics of the entire network, rather than a specific point. If a network has a relatively high intermediary centrality potential, the greater the difference in centrality between the node representing the highest intermediary centrality in this network and the other nodes means that the network is relatively well interconnected.

5 Measurement and Analysis of the Position Role of the Public Opinion Network of the "E-Speed Loan" Incident

5.1 K-core Analysis

K-core analysis is an analytical method for studying cohesive subgroups in social network analysis. First of all, let's elaborate the concept of cohesive subgroups. When some nodes in a network are so closely related that they form a subgroup, we call such a group a cohesive subgroup. K-core analysis is the analysis of the cohesive subgroups in the overall network in which a particular node of this network is interconnected with at least K other nodes. When the number of connections is K, it also means that this node is directly connected to K point nodes in the subgroup.

5.2 Core-edge Analysis

The "core-edge" analysis is a metric that reveals which nodes are in the core and which nodes are in the edge of an overall network. This involves the core-edge structure, which is a structural pattern with dense nodes at the center and sparse nodes at the edge of the network. The "core-edge" analysis is based on such a model to compare the closeness of the nodes at the center of the overall network with the sparsity of the nodes at the external edges of the overall network.

6 Conclusions

Social network analysis provides a variety of parameters that allow us to analyze the structural characteristics of the overall network. According to the usage, these parameters can be roughly divided into two categories, one is a measure of the overall network structure and the other is a measure of the identification of key nodes. They describe the overall overview of the whole social network from macroscopic and microscopic perspectives, respectively. Specifically, the metrics that reflect the overall structure of the network are network density, clustering coefficients, and accessibility measures. On the other hand, there are many individuals or nodes in the network, and these nodes play different roles in the network. Some nodes hold more resources in the network and are at the core of the network, some nodes are at the edge of the network, while others play the role of a bridge in the network. Point-degree centrality, intermediary centrality and proximity centrality are important metrics for measuring the role of node locations, while K-core analysis and core-edge analysis are the main analytical tools for layering the network structure.

References

1. Zhang, Orange. Exploring the path and evolutionary mechanism of public opinion dissemination in public crisis in the era of full media[J]. Journal of Hubei Open Vocational College,2022,35(15):141-142.
2. Wang Zhe,He Fei. Research on the evolution and management mechanism of online public opinion crisis in colleges and universities [J]. Journal of Southwest University for Nationalities (Humanities and Social Sciences Edition),2022,43(08):157-162.
3. Zhou Shengchao. Analysis of the game of online public opinion dissemination in emergency situations in colleges and universities [J]. DOI:https://doi.org/10.13665/j.cnki.hzjjykj.2022.15.027.
4. Guo Jinguang, Song Chen. Volatility of online opinion dissemination and opinion control strategy--based on GARCH and gray Verhulst method[J]. Journal of Nantong University (Social Science Edition),2022,38(04):90-100.
5. Zi Guishuang. Analysis of the current situation and reflection on the network public opinion caused by the collapse of idol's persona under mass communication[J]. Zaizhi, 2022(07): 122-124.
6. Tang Yunfeng, Liu Tao, Wang Yanyan. Network stratification, microblog public opinion dissemination and virtual field group polarization[J]. Zhejiang Social Science,2022(07):78–84+159.DOI:https://doi.org/10.14167/j.zjss.2022.07.002.