A network analysis of 2019-nCoV epidemic in mainland China by k-core decomposition

Lei Qin, Yidan Wang, Qiang Sun, Xiaomei Zhang, Benchang Shia, Chengcheng Liu

Submitted to: Journal of Medical Internet Research
on: September 13, 2020

Table of Contents

A network analysis of 2019-nCoV epidemic in mainland China by k-core decomposition

Lei Qin[1*]; Yidan Wang[1*]; Qiang Sun[1*]; Xiaomei Zhang[1*]; Benchang Shia[2*]; Chengcheng Liu[3*]

[1]University of International Business and Economics Beijing CN
[2]Taipei Medical University Taiwan CN
[3]Capital University of Economics and Business Beijing CN
[*]these authors contributed equally

Corresponding Author:
Chengcheng Liu
Capital University of Economics and Business
No. 121 Huaxiang Zhangjia Road, Fengtai District
Beijing
CN

Abstract

Results: yield unexpected information on which are influential nodes and how important they are, as well as their geographic distribution and dynamic modes. Such a better understanding of how epidemic network form and function may help reduce the damaging effects of 2019-nCoV.

(JMIR Preprints 13/09/2020:24291)
DOI: https://doi.org/10.2196/preprints.24291

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

A network analysis of 2019-nCoV epidemic in mainland China by k-core decomposition

Abstract: Frequent interregional contacts and the high rate of infection spread catalyzed the formation of 2019-nCoV epidemic network. Identifying influential nodes and highlighting the hidden structural properties of the network is central for epidemic prevention and control. In this paper, we first construct the 2019-nCoV epidemic network among provinces in mainland China, after using the degree distribution to reveal some basic characteristics, the k-core decomposition method is employed to provide some static and dynamic evidence of figuring out the influential nodes and hierarchical structure, and then we exhibit the influence power of the above nodes and its evolution. Results yield unexpected information on which are influential nodes and how important they are, as well as their geographic distribution and dynamic modes. Such a better understanding of how epidemic network form and function may help reduce the damaging effects of 2019-nCoV.

Keywords: 2019-nCoV; epidemic network; prevention and control; k-core decomposition

1. Introduction

In December 2019, several cases of pneumonia unknown etiology were detected in Wuhan City, Hubei Province of China. Chinese authorities identified the causative agent was a novel coronavirus and then the World Health Organization (WHO) officially named it as 2019 novel coronavirus (2019-nCoV) on January 10, 2020. Compared with Severe Acute Respiratory Syndrome (SARS) in China in 2003, 2019-nCoV spreads faster and infects wider, furthermore it is more difficult to prevent and control. Considering the number of cases started increasing exponentially, the Chinese government imposed a lockdown in Wuhan on January 23, 2020, aiming to cut off the route of virus transmission through the traffic blockade. Hereafter, the 2019-nCoV clearly came under control. Since March 2020, this ongoing epidemic outbreak has now spread to more than 50 other countries, it undoubtedly casts a shadow over the global economy. In order to mitigate the impact of epidemics and ensure continuity of globally social development, an exploration of influential nodes and structural properties of the 2019-nCoV epidemic network is in urgent need.

There have been extensive researches on epidemiological transmission mechanisms from diverse perspectives, such as epidemiology, medical statistics, spatial information science, sociology and dynamic models [1-6]. Due to the wide spread, the epidemic data is often presented in the form of network. The application advantages of complex network theory are gradually highlighted [7-8]. Under the framework of complex network theory, k-core decomposition shows its utility in finding specific structural information of some computer generated and real-world networks [9-13].

While the existing literature is replete with explorations of epidemic network and applications of complex network, few studies involve the effective combination of the two. In the context of 2019-nCoV epidemic outbreak in mainland China, this paper applies k-core decomposition to the structural analysis of its network, with the purpose of getting some novel conclusions so as to be in good preparation for prevention and control. Our contribution is threefold: Firstly, the 2019-nCoV epidemic data of all provinces in mainland China is timely and unique. Secondly, we can get the related static and dynamic conclusions of influential nodes and hierarchical structure by applying k-core decomposition to 2019-nCoV epidemic network, furthermore, we also detect the common characteristics of the provinces represented by these important nodes. Lastly, the influence power of k-shell nodes and its evolution measured by out-strength and in-strength can promote our understanding of the role of provinces in epidemic transmission.

The remainder of this paper is structured as follows: Section 2 briefly introduces the construction and further analysis methods of 2019-nCoV epidemic network after describing the data used. In Section 3, we firstly summarize some basic structural properties of the epidemic network by means of degree distribution. Then k-core decomposition is applied to the constructed whole period and daily networks to conduct the static and dynamic investigation of its structure. Finally, the influence power (outgoing and incoming) of k-shell and its evolution are exhibited. Sections 4 and 5 offer discussions and conclusions, respectively.

2. Data and Methodology

2.1 Data

In view of the construction of 2019-nCoV epidemic network among provinces in mainland China, the cross-provincial traveling data (1690 observations) of confirmed patients were considered. In our study, the traveling extent mainly focuses on 31 provinces (except Taiwan, Hong Kong and Macao) in China, and the traveling options involves plane and train exclusively. The aforementioned data was obtained from the website http://2019ncov.ifacehub.com/api.html, which includes the records of confirmed official accounts of WeChat, Weibo and official website. In addition, some unverifiable data were eliminated. In total, 1615 observations (328 observations by plane and 1287 observations by train) were retained, and the period is from December 27, 2019 to February 25, 2020, covering 61 days. It should be noted that the connection between province A and province B can be established in 2019-nCoV epidemic network if the traveling data show a latent confirmed patient traveled from province A to province B and was diagnosed in province B.

Some variables at the provincial level, such as GDP per capita, population, volume of passenger transport, starting date of first-level response to major public health emergency, response time, and distance from Hubei province were also used in the follow-up study on the common characteristics of provinces represented by important nodes. Specifically, the first three and fourth variables were obtained from National Bureau of Statistics (NBS) and Provincial Health Committees in individual. Response time is calculated by average days between arrived dates and confirmed dates and spatial distances from Hubei province for others refer to Yu [14].

2.2 Methodology

Considering that the causative agent of 2019-nCoV is carried by humans, in other words, its transmission is mainly in the form of human to human, the cross-provincial traveling records of confirmed patients can directly depict the epidemic network to a large extent. Specific to the 2019-nCoV epidemic network $G=(V,E)$, province A and province B can be viewed as node V_A and node V_B respectively, there are corresponding directional edges E_{AB} or E_{BA} between nodes V_A and V_B if a confirmed patient traveled from province A to province B or from province B to province A. On this basis, the directionality information of edges is well considered in the following analysis of degree distribution and influence power, while not in the k-core decomposition.

As mentioned before, three methods are involved in the further structural analysis of 2019-nCoV epidemic network: degree distribution, k-core decomposition and influence power. Under the method framework of degree distribution, the degree of a node is defined as the number of its connected edges, and it can be divided into out-degree and in-degree according to the direction of edges. In addition, the cumulative degree distribution represents the probability distribution of nodes with degree not less than k. In Section 3, we will show both the cumulative distribution of out-degree and in-degree, aiming to reveal some basic characteristics of the epidemic network. The degree distribution provides some useful information on the network. However, it is limited by the revelation of complete structure, so other network methods should be applied, such as k-core decomposition.

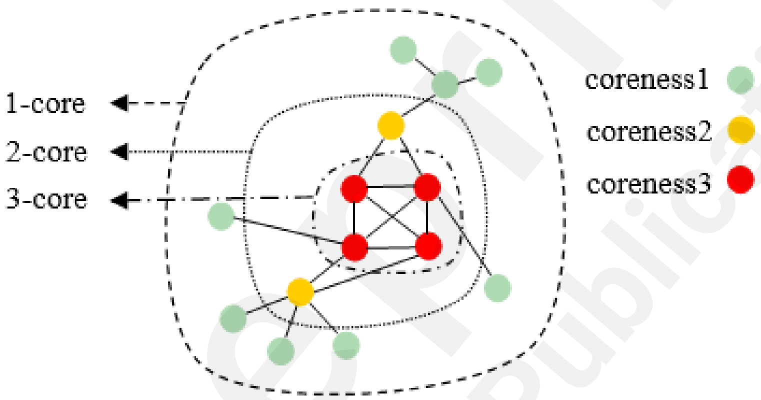


Fig. 1. Illustration of the k-core decomposition of a small network. The set of nodes belonging to the 1-core, 2-core and 3-core are enclosed by different types of lines. Different k-shells can also be distinguished by colors on the nodes.

The advantage of k-core decomposition is to detect the core and its surrounding shell of a complex network. The fundamental of this method is to decompose the network into multiple partitions, which relates to a straightforward procedure. Let $G=(V,E)$ be a graph with $n=|V|$ nodes and $e=|E|$ edges. The so-called k-core is a maximal connected subgraph of G in which the degree of all nodes is at least k. A node V_i has coreness $ks(V_i)=k$ if it belongs to the k-core instead of (k+1)-core. We note that the value of k is automatically learnt from the observed network data and also independent on our prior anticipation. Specifically, the k-core decomposition method can realize the k-shells classification of all nodes of G by removing them iteratively as follows. Firstly, we remove all nodes with degree $k=1$, and assign the coreness value $ks=1$ to the removed nodes. Secondly, a pruning process is repeated until there are only nodes with degree $k>1$. Next, we start repeating a similar pruning process for the nodes with degree $k=2$ and assign the corresponding coreness value $ks=2$. To the end, keep repeating the above procedure until all nodes of G are removed and assigned to one of the k-shells. Fig. 1 illustrates a simple k-core decomposition of a connected graph.

In order to identify the important nodes of 2019-nCoV epidemic network and reveal its hierarchical structure through k-core decomposition, a geospatial network topology map in the whole time period showing the coreness of each node (province) and their connections is plotted. From the dynamic angle, we also focus on the daily or weekly evolution of $k_{\max}$, and the number of nodes and edges. In addition, a group of scatter charts are drawn to describe the relationship between the coreness and characteristics of provinces, so that we can so that we can grasp something in common of important provinces in 2019-nCoV epidemic transmission. Most importantly, we can clearly present the hierarchical structure of epidemic network by week.

Ultimately, we introduce the method of influence power to measure the transmission effects (outgoing and incoming) among provinces.

3. Network Analysis

3.1 Degree Distribution

The method of degree distribution can offer a glimpse of network properties. In the study of network, the degree k of a node which is regarded as the number of its direct neighbors, can be measured by the number of connections with other nodes. Hence, the degree distribution $P(k)$ relates to the probability that a randomly chosen node has k connections. Considering the directivity, out-degree and in-degree are the number of outgoing and incoming connections. Correspondingly, the probability that a randomly chosen node has out-degree k_{out} and in-degree k_{in} are represented by $P(k_{out})$ and $P(k_{in})$.

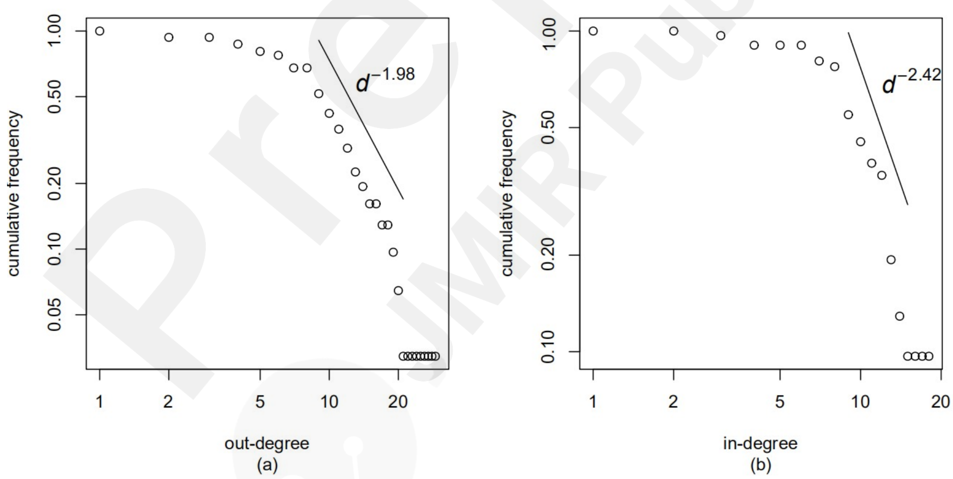


Fig. 2. Cumulative frequency graphs of out-degree and in-degree. (a): the cumulative distribution of out-degree follows an approximate power law with an exponent $\gamma=1.98$. (b): the cumulative distribution of in-degree follows an approximate power law with an exponent $\gamma=2.42$. The horizontal and vertical axes denote the number of connections (outgoing and incoming) and the respective cumulative frequency. Both axes are in logarithmic scale.

In terms of 2019-nCoV epidemic network, where links among 31 provinces are directed, the cumulative distributions of out-degree and in-degree are shown in Fig. 2. We can clearly see that they all follow the power law $P(k)\sim k^{-\gamma}$, and the values of exponent γ are 1.98 and 2.42 for the cumulative distributions of out-degree and in-degree. The above power-law distributions demonstrate that most provinces have low out-degrees, and only a small fraction of provinces maintain relatively strong outward epidemic transmission effects on others. Similarly, the overwhelming majority have low in-degrees, and provinces with stronger inward epidemic transmission effects only take a small proportion.

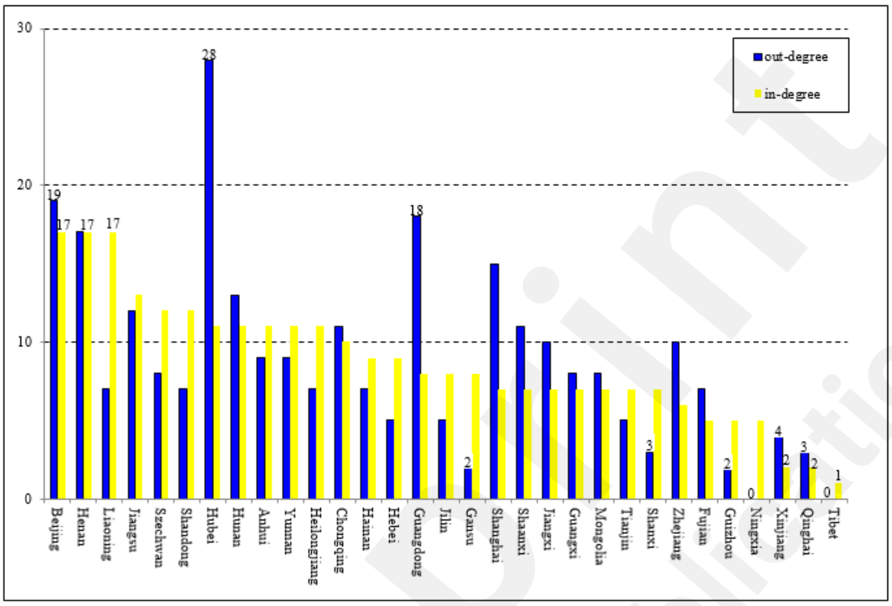


Fig. 3. Histogram describing the specific out-degree and in-degree of 31 provinces.

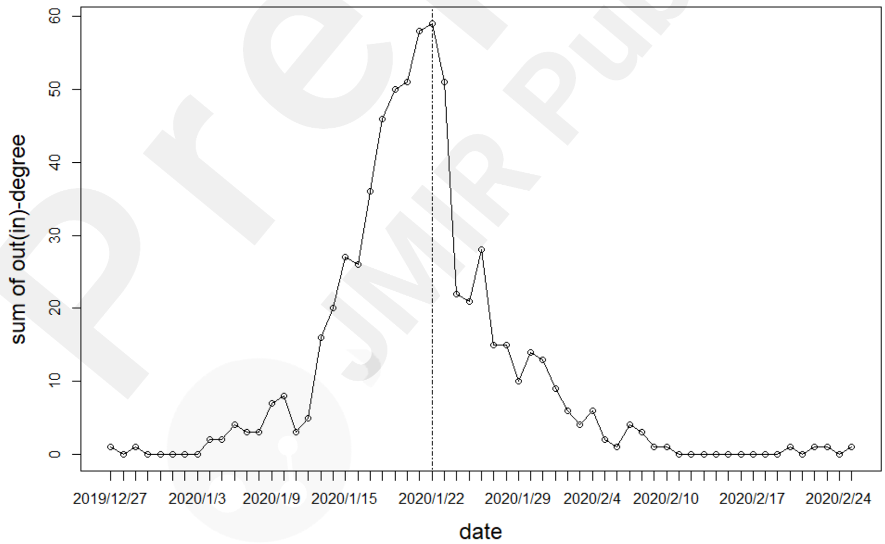


Fig. 4. Evolution of the sum of out (in)-degree in the daily 2019-nCoV epidemic networks. The out-degree of all nodes in a network is equal to its in-degree.

With the purpose of identifying the role of each province in the transmission of 2019-nCoV epidemic, a histogram describing the specific out-degree and in-degree of 31 provinces is shown in Fig. 3. In terms of out-degree, the corresponding values of 7 provinces (Xinjiang, Shanxi, Qinghai, Gansu, Guizhou, Ningxia and Tibet) are less than 5, what is more, the provinces of Ningxia and Tibet take 0. Provinces of Hubei, Beijing and Guangzhou rank in the top three, with 28, 19 and 18. Besides, in the case of in-degree, there are 3 provinces

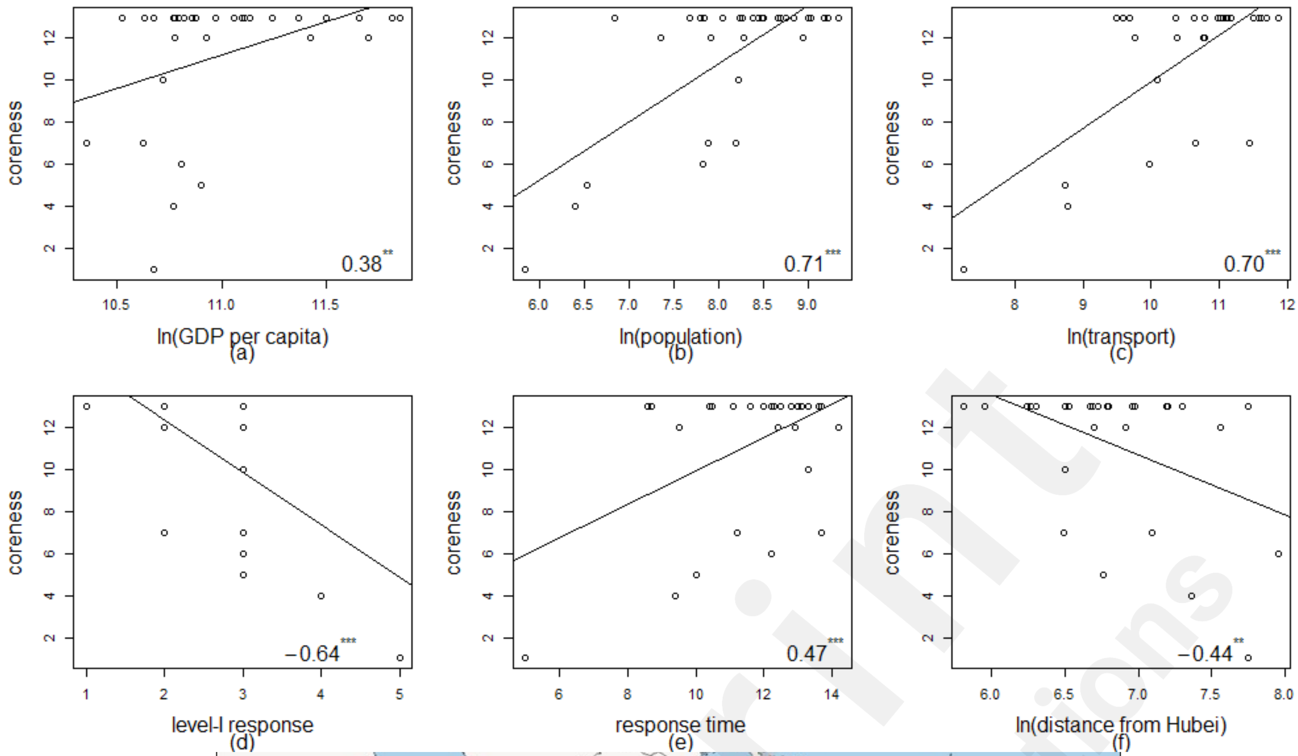
(Xinjiang, Qinghai and Tibet) less than 5, and the top three which take 17 are provinces of Beijing, Henan and Liaoning. We also attempt to characterize the intuitive time attribute of 2019-nCoV epidemic network, and present the evolution of sum of out (in)-degree in the daily networks in Fig. 4. The sum of out (in)-degree shows an inverted U-shape, and the maximum is achieved on January 22, 2020. It is regarded that the complexity of 2019-nCoV epidemic network is time-varying, and it stood out in January 22, 2020.

After grasping some basic characteristics of the epidemic network, it is reasonable to assume that small groups of nodes organize in a hierarchical manner into increasingly large groups. However, the method of degree distribution is lack of the cognition of which node belongs to which layer, and the differences between different layers are not clear enough. The k-core decomposition, which disentangles the hierarchical structure of networks by progressively focusing on their central cores, is of great use in obtaining the above detailed structural information.

3.2 K-core Decomposition

In the application of k-core decomposition, there is no necessity to take into consideration the directionality of 2019-nCoV epidemic network. Statically, the network relationship and the coreness of each node over the whole period (from December 27, 2019 to February 25, 2020) are displayed in Fig. 5. Here, what needs to be pointed out is that a visualization software called Gephi [15] is used to exhibit the topological image in geospatial space, so as to have a clear insight into the exact location of each node. We can see that there are 20 nodes with the highest coreness 13, accounting for 64.52% of 31 provinces, while the lowest coreness 1 appears in the remote Tibet. More specifically, all provinces adjacent to the outbreak area (Hubei province) have the highest coreness. In addition, Fig. 6 plotted the core size with respect to the coreness over the whole period. It can be seen that an increasing coreness usually results in a shrinking network. Combined with Fig. 5 and Fig. 6, we can see that the core size of coreness greater than 8 is 25, and the nodes in the innermost four layers account for 80.65% of the whole epidemic network. These results indicate that 2019-nCoV epidemic has affected the overwhelming majority of provinces in mainland China.

In order to investigate whether the characteristics of provinces are the key factors to determine the coreness of each node, six variables (GDP per capita, population, volume of passenger transport, starting date of first-level response to major public health emergency, response time, and distance from Hubei province) at the provincial level are introduced to draw the correlation diagram, as shown in Fig. 7. Here, a logarithmic transformation is applied to the indicators of GDP per capita, population, volume of passenger transport and distance from Hubei province. It can be deduced from the Fig. 7 that the starting date of first-level response to major public health emergency and distance from Hubei province are the significant negative correlation factors, while the others tend to be positively correlated. For example, we can reasonably believe that the starting date of first-level response to major public health emergency is related to the severity of 2019-nCoV epidemic. The earlier the first-level response starts, the stronger the infection spread of such a province is. Generally, people carried with coronavirus are more likely to appear in provinces with larger volume of passenger transport, which also determines the importance of these provinces for the epidemic spread.



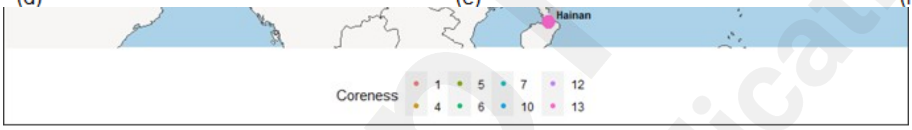


Fig. 5. Topological image in geospatial space depicting the network relationship and the coreness of each node over the whole period. The color of the node represents its coreness, which corresponds to the k-shell it belongs to.

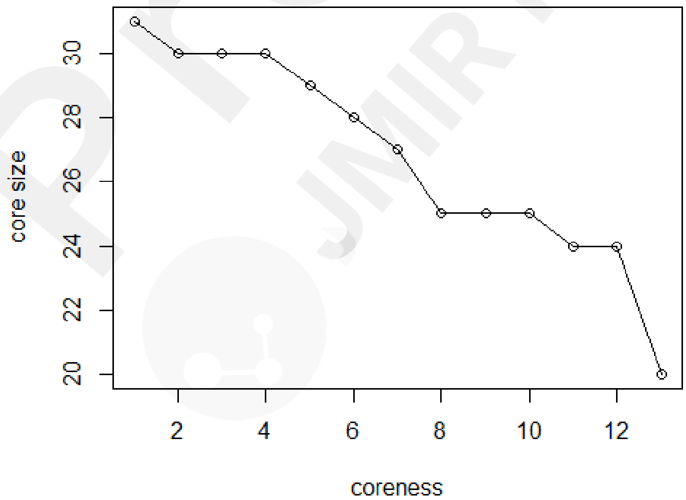
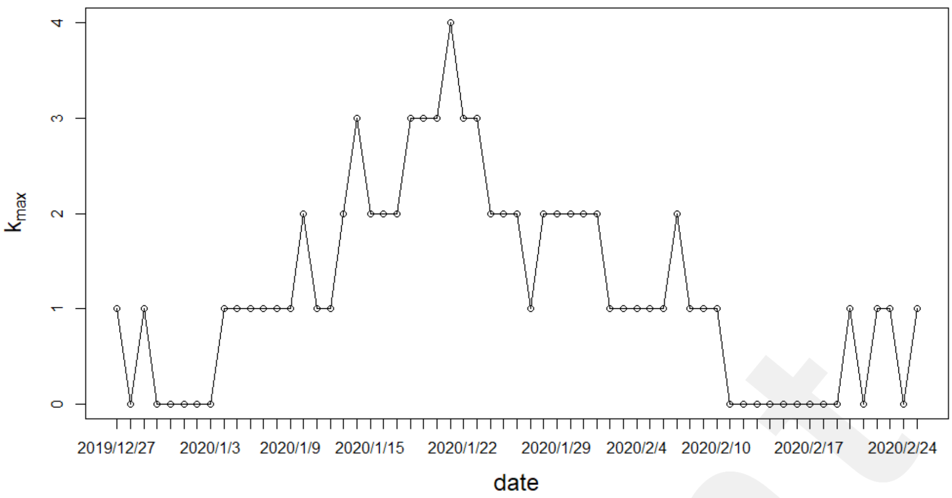


Fig. 6. Plot of the core size regarding the coreness over the whole period. The core size represented by the vertical axis means the number of nodes which their coreness is greater than or equal to that corresponding to the horizontal axis.

Fig. 7. Correlation diagram of the coreness and the characteristics of provinces. The horizontal and vertical axes denote the characteristic variables and the coreness respectively. Numbers in the lower right corner of each subgraph are the corresponding correlation coefficient, besides, ** and *** indicate the significance level of 5% and 1%, respectively.



In fact, the 2019-nCoV epidemic network is more likely to be dynamic rather than fixed over a period of time, and it is of physical necessity to investigate its time-varying maximum coreness k_{max}. Based on this, we first construct the daily epidemic networks, and Fig. 8 presents the evolution of k_{max}. It shows that there is an obvious trend of rising first and then falling for k_{max}, and the peak value appears on January 21, 2020. We can conclude that epidemic network from the end of January to the beginning of February is relatively the most complex. Considering the simple structure of the daily epidemic network, the experiential patterns of epidemic transmission summarized is limited. Furthermore, the whole period is divided into intervals, with a fixed window of 7 days.

Similar to Fig. 8, Fig. 9 shows the weekly evolution of k_{max}, node number and edge number. In terms of the trend commonness of maximum coreness k_{max}, the number of nodes and edges, there is a significant increase before the fourth week, and then a steep decline appears, thus the fourth week (January 17, 2020 to January 23, 2020) becomes its peak point. Considering the differences among the three, the number of nodes and edges goes up slightly after the eighth week. The above findings once again confirm that the fourth week is the critical period of the 2019-nCoV epidemic outbreak, and the network structure formed in the surrounding weeks is relatively complex.

Fig. 8. Evolution of the maximum coreness $k_{\max}$ in the daily 2019-nCoV epidemic networks. The horizontal and vertical axes correspond to date and the maximum coreness.

Fig. 9. Evolution of the maximum coreness $k_{\max}$, the number of nodes and edges in the weekly 2019-nCoV epidemic networks. The horizontal and vertical axes correspond to week and the relevant indicators.

In order to visualize the transmission process of 2019-nCoV epidemic dynamically, Fig. 10 describes the node composition and hierarchical structure of 2019-nCoV epidemic network by week. In addition, 31 provinces abbreviated by the node labels in Fig. 10 are shown in Table 1. In general, the epidemic network structure corresponding to the third to sixth weeks tends to be more complex, and the fourth week stands out. Chinese spring rush may give a reasonable explanation. Taking the period from the third week to the sixth week as an example, there is no doubt that Tibet has kept the lowest coreness all the way, while provinces of Anhui, Beijing and Guangdong show the highest coreness. There is a novel discovery about the changes of coreness in the epidemic outbreak area Hubei province, which is at the highest level from the third week to the fifth week, and suddenly dropped to the lowest in the sixth week. To some extent, the two-week-lagging effective control of 2019-nCoV epidemic transmission by lockdown measures imposed by Wuhan government on January 23, 2020 is verified.

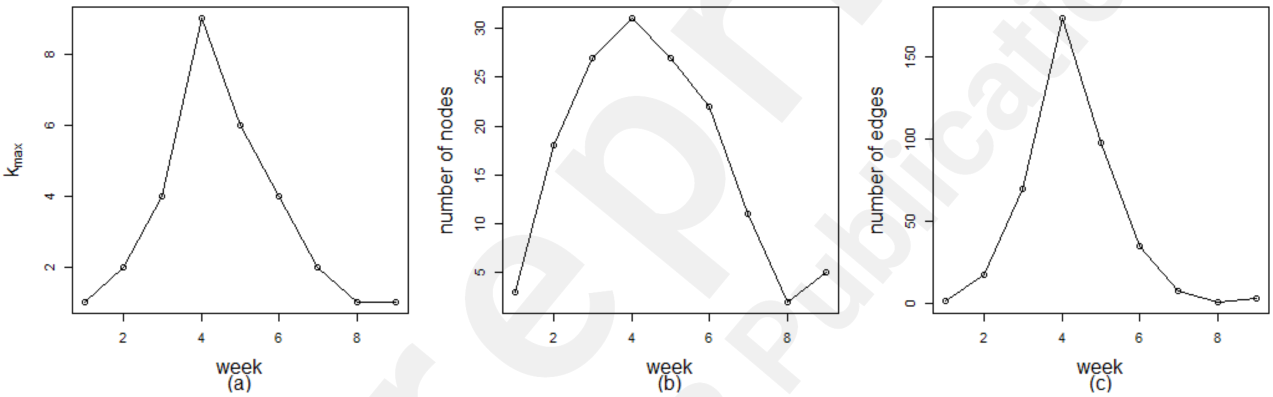


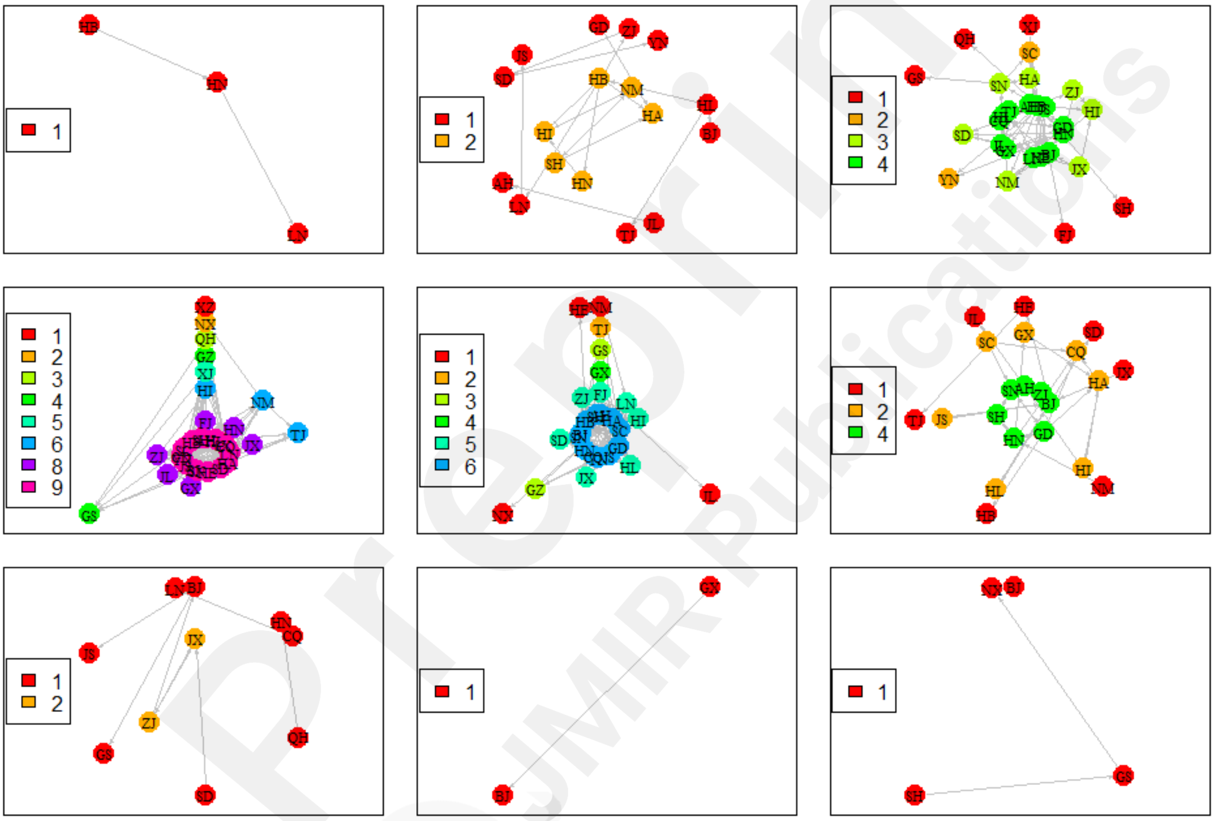
Table 1 **31 sample provinces in mainland China and their abbreviations**

Province	Anhui	Beijing	Chongqing	Fujian	Guangdong	Gansu	Guangxi	Guizhou
Abbreviation	AH	BJ	CQ	FJ	GD	GS	GX	GZ
Province	Henan	Hubei	Hebei	Hainan	Heilongjiang	Hunan	Jilin	Jiangsu
Abbreviation	HA	HB	HE	HI	HL	HN	JL	JS
Province	Jiangxi	Liaoning	Mongolia	Ningxia	Qinghai	Szechwan	Shandong	Shanghai
Abbreviation	JX	LN	NM	NX	QH	SC	SD	SH
Province	Shaanxi	Shanxi	Tianjin	Xinjiang	Tibet	Yunnan	Zhejiang	
Abbreviation	SN	SX	TJ	XJ	XZ	YN	ZJ	

Fig. 10. Dynamic network of 2019-nCoV epidemic uncovering the node composition and hierarchical structure. The color and label of each node denote the coreness and geographical province respectively. Subgraphs, from top to bottom, from left to right, correspond to the 9 weeks in order over the period from December 27, 2019 to February 25, 2020.

3.3 Influence Power

Epidemic, generally occurs in some regions first, and then spread out rapidly when the situation in these regions is more serious and out of control. Specific to its network structure, the so called outbreak or serious areas have high centrality and occupy a certain shell. Naturally, after applying the k-core decomposition to the epidemic network in a fixed period, to what degree each k-shell will influence the rest has become a more attractive question.



Here, the method of influence power can give a good answer. Under the framework of influence power, there are two indicators of out-strength and in-strength, which can respectively evaluate the power of any node influencing other nodes and influenced by other nodes. Correspondingly, the calculation depends on the number of outgoing and incoming links. In terms of 2019-nCoV epidemic network, if a large number of outgoing routes appears in one province, it usually has a great influence of transmission on other provinces. Similarly, in case a province has a large number of incoming routes, it is also greatly influenced by the epidemic transmission in others. Hence, in order to further quantify the influence power η_k of k-shell, the following formula can be used.

$$\eta_k = \frac{L_{k-shell}}{\sum_{k=1}^{k_{\max}} L_{k-shell}}$$

where η_k is the calculated ratio of influence power, and $L_{k\text{-}shell}$ relates to the number of links (outgoing or incoming) in each k-shell of concern.

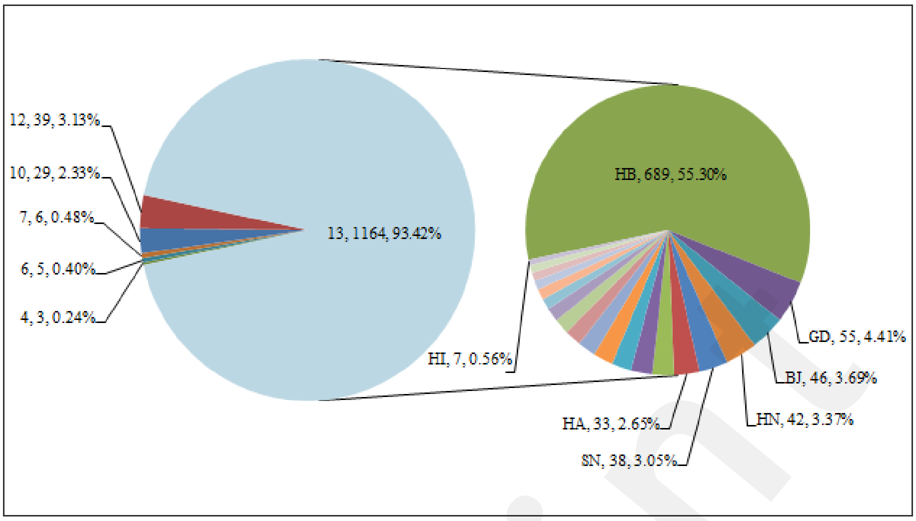


Fig. 11. Pie charts depicting the out-strength comparison of each shell and the innermost nodes based on 2019-nCoV epidemic network in the whole period. Considering the left, three groups of label numbers correspond to coreness, number of outgoing links and percentage in individual. They denote province, number of outgoing links and percentage on the right.

The results of influence power (out-strength and in-strength) of each shell and the innermost nodes measured based on 2019-nCoV epidemic network in the whole period are shown in Fig. 11 and Fig. 12 respectively. Due to some nodes (Ningxia and Tibet) of 2 shell don't have outgoing links, there are only 6 shells of the epidemic network in the whole period shown in Fig. 11. The maximum coreness of the 2019-nCoV epidemic network is 13, and almost all (93.42%) the outgoing links are from provinces with the highest coreness. Among these provinces, more than half (55.30%) of outgoing links come from Hubei province, which is the origin of this epidemic. Besides, Guangdong province and Beijing rank second and third, accounting for 4.41% and 3.69% respectively. Hainan province, which is in the same shell as those, has the lowest proportion (0.56%) of outgoing links. We can conclude that it is of physical importance to identify the key "outgoing" provinces for the prevention and control of 2019-nCoV epidemic.

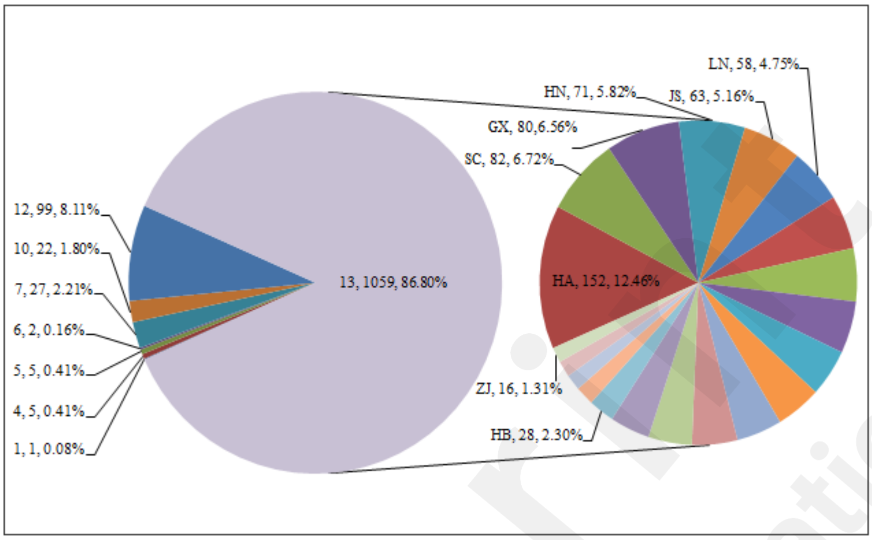


Fig. 12. Pie charts depicting the in-strength comparison of each shell and the innermost nodes based on 2019-nCoV epidemic network in the whole period. Considering the left, three groups of label numbers correspond to coreness, number of incoming links and percentage in individual. They denote province, number of incoming links and percentage on the right.

On the other hand, Fig. 12 can help us understand the “incoming” role of each shell and the innermost nodes in the transmission process of 2019-nCoV epidemic. Consistent with the findings in Fig. 5, the epidemic network in the whole period can be pruned recursively into 8 shells. In addition, each shell always has nodes with incoming links. The highest proportion of incoming links, up to 86.80%, appears in provinces with the highest coreness. Among these provinces, the top three are Henan province (12.46%), Szechwan province (6.72%) and Guangxi province (6.56%), while Zhejiang province (1.31%) takes the lowest proportion. Focusing on the epidemic breakout area of Hubei Province, there are only 28 incoming links (2.30%) compared with its largest number of outgoing links, which shows that it is less influenced by others. On the whole, the in-strength performance of provinces is different from out-strength, and the “incoming” roles of these provinces tend to be more equal.

In some cases, the dynamic influence of a shell consist of the most important nodes on the whole network is more worthy of attention. The above influence can be also measured by the method of influence power, depending on the time-varying set of the innermost nodes and all the related directional links. On the basis of daily 2019-nCoV epidemic networks, Fig. 13 presents the evolution of $\eta_{k_{\max}}$ from the perspectives of out-strength and in-strength, which is conducive to grasp the dynamic “outgoing” and “incoming” roles of those important provinces. It is notable that a relatively large gap between out-strength and in-strength exists on January 21, 2020, which echoes with the most complex daily epidemic network indicated in Fig. 8. We also observe that the out-strength of the innermost nodes is larger than its in-strength before January 27, 2020, which means that those corresponding provinces tend to have more influence on others rather than be influenced by others. After that, owning to the stricter control imposed by those provinces, their out-strength and the above contrast have been greatly weakened. To a certain extent, the above findings confirm the controlling effectiveness of Chinese government in 2019-nCoV epidemic.

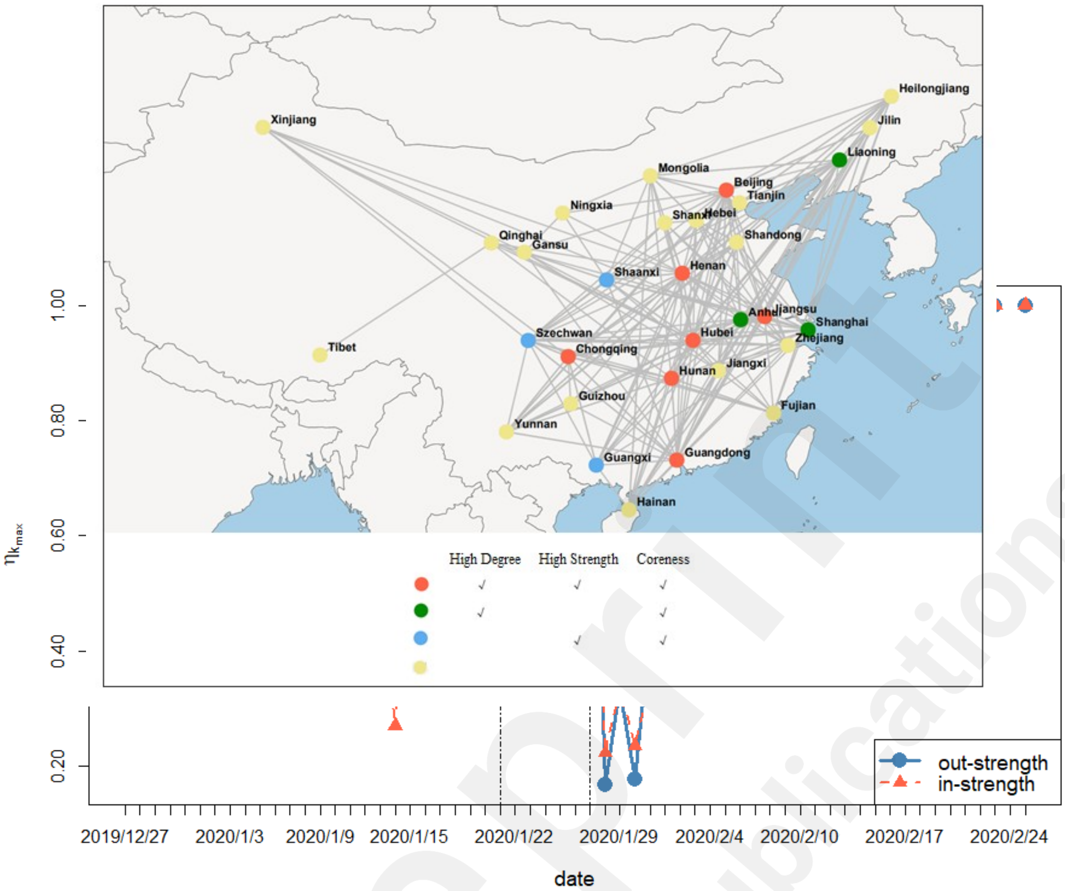


Fig. 13. Evolution of $\eta_{k_{max}}$ from the perspectives of out-strength and in-strength in the daily 2019-nCoV epidemic networks. The dotted lines correspond to January 21, 2020 and January 27, 2020 respectively.

4. Discussion

In economics, epidemic and many other fields, the increasing participants and data volume further complicates the formed network. It is critical to extract effective information from such a large and complex network. Hence, we need to identify and focus on the central nodes that drive the whole network instead of paying the same attention to all nodes. As mentioned above in our study, degree distribution is the simplest way to measure the centrality of each node in a network, which only involves the local structure around that node. Specifically, in a binary network, the degree distribution depends on the number of edges of the considered node. In a directed network, the connecting edges of a node may have two directions of outgoing and incoming, which correspond to out-degree and in-degree under the framework of degree distribution. Furthermore, the concept of degree has generally been extended to the sum of weights when analyzing a weighted directed network [16], and the strength (out-strength and in-strength) of nodes is proposed. Additionally, in order to acquire the more detailed information of the network structure, the k-core decomposition can be employed to disentangle the hierarchical structure of networks by progressively focusing on their central cores. In summary, the indicators of degree, coreness and strength we adopted above can provide different perspectives to understand nodes and structures of the network.

Fig. 14. Topological image in geospatial space depicting the network relationship and the centrality of each node over the whole period. The color of the node relates to the intensity of degree, strength and coreness.

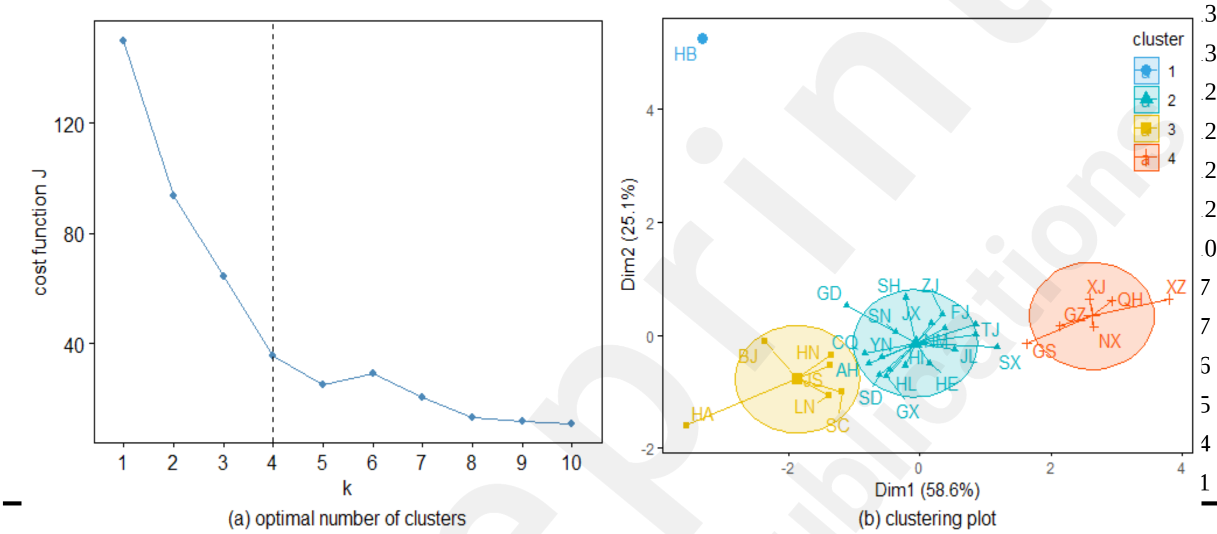
Taking the 2019-nCoV epidemic network in the whole time period as an example, we calculate the degree (out-degree and in-degree), strength (out-strength and in-strength) and coreness of all nodes (31 provinces) in Table 2, and plot the geospatial network topology map shown in Fig. 14. It should be emphasized that, considering the diversity of degree and strength of each node, the respectively top 10 nodes are defined as high degree and high strength after ranking in descending order. In Fig. 14, the color of each node is a description of the intensity of three indicators: degree, strength and coreness. The red nodes correspond to high degree, high strength and the highest coreness, the green nodes are representative for high degree and the highest coreness, the blue nodes denote high strength and the highest coreness, and the yellow nodes relate to other cases. Concerning the neighboring provinces of the epidemic outbreak area Hubei, three (Hunan, Henan and Chongqing) of them show the high degree and strength, as well as the highest coreness. Besides, in the same case of the highest coreness, the degree of Anhui province is high, the strength of Shaanxi province is high, while both of the degree and strength of Jiangxi province are not high. After expanding the considered objects nationwide, owning to the high performance of degree, strength and coreness, provinces of Beijing, Guangdong and Jiangsu can be regarded as the central nodes of 2019-nCoV epidemic network. The high-mobility population may give a more reasonable explanation by analyzing the commonness of these three provinces.

Some central properties of 31 provinces in 2019-nCoV epidemic are revealed in Table 2. Generally speaking, the maximum coreness of 2019-nCoV epidemic network is 13, and provinces with high degree and strength tend to take the highest coreness. What's more, it can be seen that high strength and low degree exist in some provinces of Szechwan, Shaanxi and Guangxi simultaneously, indicating that these provinces may have strong interaction with a few provinces. On the contrary, some provinces of Liaoning, Anhui and Shanghai have high degree and relatively low strength, which suggests that these provinces may have weak interaction with numerous provinces. The above results are more likely to be related to heterogeneity characteristics of those provinces, such as traffic, population, education and weather.

Table 2 Central properties of 31 provinces in 2019-nCoV epidemic

Province	Out-degree	In-degree	Degree	Out-strength	In-strength	Strength	Coreness
Hubei	28	11	39	689	28	717	13
Beijing	19	17	36	46	55	101	13
Henan	17	17	34	33	152	185	13
Guangdong	18	8	26	55	50	105	13
Jiangsu	12	13	25	26	63	89	13
Hunan	13	11	24	42	71	113	13

Liaoning	7	17	24	24	58	82	13
Shanghai	15	7	22	29	17	46	13
Chongqing	11	10	21	28	57	85	13
Szechwan	8	12	20	22	82	104	13
Anhui	9	11	20	26	56	82	13
Yunnan	9	11	20	18	43	61	13
Shandong	7	12	19	12	49	61	13
Heilongjiang	7	11	18	11	47	58	13
Shaanxi	11	7	18	38	49	87	13
Jiangxi	10	7	17	14	20	34	13
Hainan	7	9	16	7	49	56	13
Zhejiang	10	6	16	20	16	36	13



Note: 31 sample provinces in the table are arranged in descending order according to the indicators of coreness and degree.

Considering the direction of epidemic spread, we draw a K-means clustering graph based on indicators of out-degree, in-degree, out-strength, in-strength and coreness, as shown in Fig. 15. The left panel presents the optimal number of clusters by Elbow method and the right one visualizes the corresponding clusters. We can clearly see that the optimal clustering number of 31 provinces is 4, and these four clusters obviously exist. Specifically, Hubei province stand out and form a single cluster, which can be explained by its role of initiator in this epidemic. Provinces of Beijing, Henan, Hunan, Jiangsu, Liaoning and Szechwan are in the same group, and show relatively high values on the above indicators. In contrast, provinces with lower values, such as Gansu, Guizhou, Ningxia, Qinghai, Tibet and Xinjiang, belong to another cluster. Besides, the remaining are grouped together. The clustering here is similarly to Fig. 14 and Table 2, and the uniqueness lies in considering directional factors when examining the central properties of each province.

Fig. 15. The optimal number of clusters and clustering plot. The horizontal and vertical axis of the left panel correspond to the number of clusters and cost function. The horizontal and vertical axis of the right panel represent the first and second principal component respectively.

All in all, these findings in this section will help us to comprehensively detect the important nodes in 2019-nCoV epidemic network and grasp the spread path among provinces in mainland China. On this foundation, we can further identify some economic and social factors which determine the development of this epidemic, and finally the effective control can be achieved by imposing some public interventions. With more countries or regions involved in 2019-nCoV epidemic, the structure of the world epidemic network has become more complex, and it is more urgent to explore the corresponding central properties.

5. Conclusions

The 2019-nCoV epidemic outbreak associated with a new coronavirus of probable bat is spreading around the world, and more and more countries or regions have been involved in the epidemic network. How to find the most important nodes and hierarchical structure of the above network has become a priority. Focusing on China, the 2019-nCoV epidemic network is well constructed by the cross-provincial traveling records of confirmed patients, and then the methods of degree distribution, k-core decomposition and influence power are employed in the further structural analysis.

With regard to the empirical results of degree distribution, the fact of the power-law distribution suggests that most provinces have either low out-degrees or low in-degrees, and only a small fraction tend to have relatively strong outward or inward epidemic transmission effects. In terms of descending order, three provinces of Hubei, Beijing and Guangzhou rank in the top three of out-degree, and the top three of in-degree relate to provinces of Beijing, Henan and Liaoning.

The application of k-core decomposition has also resulted in some novel findings. Firstly, we verified the hierarchical structure of the 2019-nCoV epidemic network over the whole period, and more than half of 31 provinces are in the innermost core. Secondly, we consider the correlation of the characteristics and coreness of each province, and identify some significant negative and positive factors. In Addition, time-varying maximum coreness is investigated in two perspectives of daily and weekly. Respectively, the obvious trend of rising first and then falling appeared on January 21, 2020 and the fourth week. To be more specific, considering the dynamic transmission process of 2019-nCoV epidemic, three provinces of Anhui, Beijing and Guangdong always show the highest coreness from the third week to the sixth week, and Hubei province maintained the highest coreness until the fifth week, and suddenly dropped to the lowest in the sixth week.

In the follow-up, influence power (out-strength and in-strength) was introduced to measure the influence degree of each k-shell. It is observed that most outgoing links and incoming links are from provinces with the highest coreness. Moreover, we investigate the dynamic “outgoing” and “incoming” roles of those important provinces, and find that the out-strength of the innermost nodes is larger than its in-strength before January 27, 2020, while then there is a reversal.

It is worth mentioning that our study is limited to the transmission of 2019-nCoV epidemic network for 31 provinces in mainland China, and the k-core decomposition is applicable to the unweighted and undirected networks. More network structure analysis methods can be considered to explore the epidemic spread dynamics which involves more regions in China or the rest of the world in the future.

References

[1] Leung G M, Hedley A J, Ho L M, et al. The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: an analysis of all 1755 patients [J]. *Annals of Internal Medicine*, 2004, 141(09): 662-673.

[2] Anderson R M, Fraser C, Ghani A C, et al. Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic [J]. *Philosophical Transactions of the Royal Society B*, 2004, 359(1447): 1091-1105.

[3] Pang X, Zhu Z, Xu F, et al. Evaluation of control measures implemented in the severe acute respiratory syndrome outbreak in Beijing [J]. *JAMA*, 2003, 290(24): 3215-3221.

[4] Wang J F, Christakos G, Han W G, et al. Data-driven exploration of 'spatial pattern-time process-driving forces' associations of SARS epidemic in Beijing, China [J]. *Journal of Public Health*, 2008, 30(03): 234–244.

[5] Riley S, Fraser C, Donnelly C A, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions [J]. *Science*, 2003, 300(5627): 1961-1966.

[6] Shi Y. Stochastic dynamic model of SARS spreading [J]. *Chinese Science Bulletin*, 2003, 48(13): 1287-1292.

[7] Zheng X L, Zeng D, Sun A, et al. Network-based analysis of Beijing SARS data [J]. *Biosurveillance and Biosecurity*, 2008, 5354: 64-73.

[8] Adegboye O A, Elfaki F. Network analysis of MERS coronavirus within households, communities, and hospitals to identify most centralized and super-spreading in the Arabian Peninsula, 2012 to 2016 [J] *Canadian Journal of Infectious Diseases and Medical Microbiology*, 2018, 2018: 1-9.

[9] Alvarez-Hamelin J I, Dallasta L, Barrat A, et al. Large scale networks fingerprinting and visualization using the k-core decomposition [C]. *Neural Information Processing Systems*, 2005: 41-50.

[10] Zhang H, Zhao H, Cai W, et al. Using the k-core decomposition to analyze the static structure of large-scale software systems [J]. *The Journal of Supercomputing*, 2010, 53(02): 352-369.

[11] He X, Zhao H, Cai W, et al. Analyzing the structure of earthquake network by k-core decomposition [J]. *Physica A: Statistical Mechanics and Its Applications*, 2015, 421: 34-43.

[12] Fitzhugh S M, Butts C T. Patterns of co-membership: techniques for identifying subgraph composition [J]. *Social Networks*, 2018, 55: 1-10.

[13] Angelou K, Maragakis M, Argyrakis P, et al. A structural analysis of the patent citation network by the k-shell decomposition method [J]. *Physica A: Statistical Mechanics and Its Applications*, 2019, 521: 476-483.

[14] Yu Y. CHINA_SPATDWM: Stata module to provide spatial distance matrices for Chinese provinces and cities [J]. *Statistical Software Components*, 2009.

[15] https://gephi.org/.

[16] Barrat A, Barthelemy M, Pastorsatorras R, et al. The architecture of complex weighted networks [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(11): 3747-3752.

Supplementary Files