

Contact network analysis of patients with Novel Coronavirus Pneumonia - Based on 237 cases in Shaanxi Province

Yang Zhangbo (✉ yang_zb@xjtu.edu.cn)

School of Humanities and Social Science and Institute for Empirical Social Science Research, Xi'an Jiaotong University

Research Article

Keywords: Novel Coronavirus, Novel Coronavirus Pneumonia, Digital Portraits of Patients, Virus Contact Network, Social Network Analysis

Posted Date: April 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-22062/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on March 1st, 2021. See the published version at <https://doi.org/10.1038/s41598-021-84428-x>.

Abstract

The spread of novel coronavirus is closely related to the structure of human social networks. Based on 237 cases in Shaanxi Province, using epidemiological retrospective statistics, data visualization, and social network analysis methods, this paper summarized characteristics of patients with new coronary pneumonia in Shaanxi Province, and analyzes these patients' dynamic contact network structure. The study found that there are many clustered infections through strong ties, about one-third cases are caused by relatives' infection. In the early stages of the epidemic, it was mainly imported cases, and in the later stages, it was mainly local infection cases. The infected people were mainly middle-aged men. Symptoms of imported cases occurred on average 3 days after they arrived, and medical measures were taken on average 5 days later. All cases showed symptoms in less than 2 days on average and were then taken to medical treatment. The virus contact network can be divided into multiple disconnected components. The largest component has 12 patients. The average degree centrality in the network is 0.987. The average betweenness degree is 0. The average closeness degree is 0.452. The average PageRank index is 0.0042. The number of contacts of patients is unevenly distributed in the network.

Introduction

Novel coronavirus pneumonia (SARS-Cov-2) is the third wave of coronavirus outbreaks in the 20th century 1. The outbreak of the epidemic in Wuhan, China, in early December 2019 caused concern among people all over the world. The World Health Organization has defined the epidemic as a public health emergency of international concern. By March 2020, the epidemic situation in China has been under control, but in many other countries it becomes gradually more serious, especially in Italy, Iran, and the United States.

Since outbreak, analyses of novel coronavirus pneumonia are mostly based on epidemiology, virology and medicine, involving case analysis of patients, construction of infection models, gene sequencing, clinical diagnosis, etc. 2–6. Epidemiological research is mainly based on patient cases to predict trends of the epidemic in order to better control the spread of the virus 7. Virology mainly analyzes the biological structure of coronavirus in preparation for vaccine development 3. Medicine research focuses on the diagnosis and treatment of disease symptoms 8. However, very few study analyzes the whole network structure and characteristics of virus transmission from the perspective of social networks.

In human society, the spread of disease needs two key factors: biological structure of the virus and social structure of the society. The virus's transmission route reflects the movement of people and the composition of their social networks. Virus carriers' contact and interaction with each other form a network that is called contact network. Nodes in the network represent people and edges represent various types of contact relationships, including both strong ties in families and between strangers 9 10. Through contact, virus can spread in human society. Analysis of virus contact network is the key to

understand the spread of disease, helping to clarify the transmission route of the virus, and is also an important basis for taking measures. Studies show that virus immune strategies based on contact networks are superior to random immune strategies 11. Most of current studies are based on epidemiological retrospective studies of patients with novel coronavirus pneumonia, and lack of analysis on dynamic contact networks of patients.

Many studies have analyzed the spread of various types of viruses among crowds from the perspective of social network, such as STD 9, AIDS 12, and Black Death¹⁰. These studies revealed that the contact network structure of different viruses is very different. The main reason is that virus is transmitted in different ways. For example, for HIV through sexual contact, blood contact and mother-to-child contact, the network structure is relatively simple and relatively sparse. But for viruses like the novel coronavirus, the contact network structure is very complex and dense.

The contact networks of many diseases are small-world network 13. This type of network has two characteristics. One is that there are many local connections which will form dense local clusters, the other is that there are occasional long-distance connections which can span regions and groups and connect different local clusters 14. In such a network structure, even if each person has limited number of contacts, virus can still quickly spread to the entire network, causing an outbreak of disease 9. This feature leads to the virus to spread beyond the original infection place, and cause synchrotron oscillation of disease outbreaks in different places 13. Unlike contact network, spatial network of viruses has distinct scale-free network features 15. Studies of contact networks for SARS have shown that only differences in network structure can significantly change the curve of the outbreak 16. For viruses with a basic regeneration number less than 2, changes in the network structure also have a significant impact on spread of the disease 17.

In epidemiological studies of novel coronavirus pneumonia, the Lancet first published 41 cases of novel coronary pneumonia, with a brief description of demographic characteristics of the patients 18.

The median age of those patients was 49, and 66% of them had seafood market contact history in Wuhan, China. There was a case of family cluster infection in their samples. Another analysis of 99 patients showed that average age of patients was 55 years old and the standard deviation was 13 years old, including 67 males and 32 females, 49 of whom had a history of South China seafood market exposure 19. Further analysis of 835 cases in Hubei Province of China showed that the average age of the patients was 49 years old, the male to female ratio was 2.7 to 1, and the fatality rate was 2.9% 7. Relevant modeling shows that the basic regeneration number of novel coronavirus is between 2–3, which is weaker than SARS 2 4. Summarizing the existing research, the early outbreak of the virus was based on the clustered infection in the South China Seafood Market in Wuhan. The group with the highest virus infection rate was men who were older than 55 years. The virus is mainly transmitted through droplets and contact, and has a strong ability to spread. Family-based cluster transmission is more common 2 5.

This study collected all confirmed cases published in Shaanxi Province of China from January 23, 2020 to February 16, 2020, a total of 237 cases. First, this paper analyzes the epidemiological characteristics

of patients with novel coronavirus pneumonia in Shaanxi province, and then studies the transmission route and contact network.

Methods

Methods. This paper mainly adopts two research methods. One is descriptive epidemiological research method, to digitally portrait the novel coronavirus pneumonia infected people in Shaanxi Province and analyze their epidemiological characteristics. The other is social network analysis method. I build and visualize the patient's contact network and calculate relevant network indicators, including degree centrality, closeness centrality, betweenness centrality and PageRank index. The dynamic change of the network has an important impact on the risk of infection. The phase change structure of the dynamic network is very different from the static network [20, 21]. But the research on coronavirus dynamic contact network is still less. I construct a dynamic network for corresponding analysis.

Data. All the samples of this study are from the officially announced cases of Health Committee of Shaanxi Province of China. The cases are from January 23 to February 16, 2020. After February 16th, the case growth in Shaanxi Province slowed down significantly. The paper analyzes and encodes the text of each reports.

All methods were carried out in accordance with relevant guidelines and regulations. The study received approval from the Ethics Committee of Xi'an Jiaotong University Health Science Center. The committee waived the need for informed consent as part of the study approval, since this was a retrospective data analysis.

Measures. First, code the demographic characteristics and case characteristics of each patient according to the description revealed by government, including gender, age, household registration, place of infection, time to arrive in Shaanxi, time when symptoms occur, time to visit hospital or quarantine.

Second, code the routes of infection no matter it is a stranger tie, strong tie, or weak tie. People's social networks consist of these three types of ties [22]. Strong ties refer to close friends, acquaintances and family members who have more daily contact, deeper affection intensity and high-level trust. Weak ties are those connections of lower contact frequency, lower affection intensity and lower level trust than strong contacts. Data from the Shaanxi Health and Medical Commission did not fully specify people infected by which kind of tie. I coded the data according to the following rules: If the case shows the infection was from their own family members or close contacts, then speculates that it is a strong tie. If the case is not clearly stated, we code it according to the household registration, work conditions, and travel conditions stated in the case. Therefore, the three routes of infection are not mutually exclusive in the data presentation. If it is possible to infect through one route, the code is 1, Otherwise 0. For example, if the patient has lived and worked in Hubei Province for a long time, and the place of infection is also in Hubei, the study speculates that she/he may be infected through three ways: strangers, weak ties and strong ties. But if the patient only stopped in Wuhan when the train returned to Shaanxi, the study

speculates that he is only likely to be infected by a stranger. In addition, the study also counted whether patients had a relative infection.

Third, the Shaanxi Health Commission's data lists the patient's contacts with each other. Based on the case data, we construct a patient contact matrix in chronological order, visualize the daily dynamic network, and calculate the corresponding network indicators.

Results

Basic characteristics of patients with novel coronavirus pneumonia in Shaanxi Province. Table 1 summaries the frequency and percentage of related variables, which can outline the basic situation of patients. Specifically, there are slightly more male patients and slightly more patients infected in Shaanxi Province. About 59% patients may be infected by strangers, and about 60% may be infected by weak ties such as general colleagues and friends. About 74% patients may be infected by strong ties such as close friends and relatives. 37% patients' relatives were also infected, which indicates that there are more clustered infections in the province.

Gender	Frequency	Percentage (%)	Infected place	Frequency	Percentage (%)
Female	108	45.57	Inside Shaanxi Province	124	52.32
Male	129	54.43	Outside Shaanxi Province	113	47.68
Total	237	100	Total	237	100
<hr/>					
Is there a possibility of being infected by a stranger?	Frequency	Percentage (%)	Is there a possibility of being infected by weak ties?	Frequency	Percentage (%)
Yes	140	59.07	Yes	143	60.34
No	97	40.93	No	94	39.66
Total	237	100	Total	237	100
<hr/>					
Is there a possibility of being infected by strong ties?	Frequency	Percentage (%)	Whether any relatives are infected?	Frequency	Percentage (%)
Yes	176	74.26	Yes	87	36.71
No	61	25.74	No	150	63.29
Total	237	100	Total	237	100

Table 1. Descriptive Statistics of patients with novel coronavirus pneumonia in Shaanxi Province, China.

Figure 1 shows the frequency distribution of patient age. It can be seen that it conforms to normal distribution. Among them, 48-year-olds have the most infections, and young people age 16 to 20 have the least infections. But the number of infections rise sharply above 22.

Figure 2 shows the average age of patients over time. It can be seen that the average age of infected persons increased significantly as time goes. By February 14, it becomes more than 80 years. This also shows that the epidemic control method has good effect. The infected people in later stages are older and weaker people who has weak transmission ability, and infection of those young and middle-aged with strong transmission ability was controlled.

Table 2 and Figure 3 show the average onset time of imported cases after arriving in Shaanxi, and the average interval of taking relevant medical measures after symptoms (such as cough, fever, etc.) in all cases. According to Table 2, the average age of the patients was 46 years old, the youngest was 3 years old, and the oldest was 89 (this case died in March; the only died case in Shaanxi). For imported cases from other regions in China, they developed symptoms on average 3 days after arriving in Shaanxi. The symptoms appeared as early as 5 days before arriving in Shaanxi. One case did not appear any symptoms until 19 days after arrived. After imported cases arrived in Shaanxi, they went to the clinic or were quarantined after an average of 5.4 days. The patients with the shortest interval had a history of visit doctors one day before arrive. The patient with the longest interval did not go to hospital or be quarantined until 17 days after he arrive. After the onset of symptoms, the average time to take relevant treatment was 1.6 days, indicating that the prevention and control measures in Shaanxi Province were timely and effective. The patient with the shortest time was quarantined 8 days before the onset of symptoms, and the patient with the longest time did not go to hospital until 14 days after the symptoms. No doubt the latter case has a higher risk of virus transmission.

Variables	Case number	Mean	S.E.	Minimum value	Maximum value
Age	237	45.90	16.58	3	89
Symptom onset date minus arriving in Shaanxi date (days)	94	3.489	4.560	-5	19
Diagnosis/quarantined date minus arriving in Shaanxi date (days)	86	5.488	3.846	-1	17
Diagnosis/quarantined date minus Symptom onset date (days)	178	1.607	2.973	-8	14

Table 2. Statistics related to the onset time of novel coronavirus patients in Shaanxi Province, China. It shows in figure 3 that with time goes, the average onset time of symptoms has a tendency to increase, which means that the later imported cases are often patients with a longer incubation period. Therefore, they were not detected in the early stage. At the beginning of the epidemic, Shaanxi Province has adopted measures such as quarantine for patients with short incubation periods. It can be seen that with the change of time, the average diagnosis time has a decrease trend, which means that the later prevention and control measures are taken in a timely manner. Many patients develop the disease during the quarantine period, which reduces the risk of spread caused by the virus incubation period.

The transmission route of novel coronavirus is mainly respiratory droplets and contact transmission. From the perspective of social network, infection occurs in three kinds of connection: strangers, weak ties

(such as ordinary friends, colleagues, etc.), strong ties (such as couples, family members, relatives, etc.). Figure 4 shows the types of contacts that patients may be infected with over time. In addition to the three main contacts, it also shows whether there is a relative infection of the patient. The change in type of ties was mainly related to the number of people infected. Our main concern is the proportion of each infection route. It can be seen that the strong ties infection route has always been relatively higher proportion than other routes, which shows that the spread of novel coronavirus in Shaanxi is mainly cluster infection. This also shows that the epidemic situation in Shaanxi has been effectively controlled, and has not caused a large number of stranger infections that are most likely to cause panic. However, there was a relatively high outbreak of stranger infections on February 7, mainly because of the cluster infection in Xi'an Duocai Shopping Center, where customers and businesses were infected, and many of them did not know each other. Correspondingly, it also shows that the clustered strong ties infection is the way that needs to be controlled in the epidemic prevention and control, which is basically consistent with the conclusions of various previous studies.

Dynamic Contact Network of Novel Coronavirus Pneumonia Patients Figure 5 is the dynamic contact network of patients with novel coronavirus pneumonia in Shaanxi Province. We intercepted three time points to present the network structure: early network (January 25), intermediate network (February 1), and later network (February 16).

The early contact network was relatively sparse, and most patients could not be identified the infection source. At this time, the largest cluster (component) was composed of three patients, number 9, 10, and 11, and their infection places were all in Wuhan. In the middle period, cluster-shaped infections have appeared, and several major clusters of infection formed. Case 26's cluster would expand into the largest cluster in later stage. The later network was divided into multiple clusters, the largest of one was a cluster of number 25 and 26 illustrated in the middle of the picture. They were a couple, natives of Shaanxi, who had symptoms after return from Wuhan by driving. They went to local hospitals 5 days after they had symptoms. The source of infection for case 160 at a later stage could also be traced to this cluster. However, there are fewer new clusters in the later period, which indicates that the control of virus transmission is better. In the later period, only cases 234–237 formed a fully-connected component. They belong to one family. There are still many unconnected cases in the network, most of which are imported cases. It is no longer possible to track their infection source outside the province.

Table 3 reports four centrality measurement of the contact network. Degree centrality expresses that, on average, how many other patients the focal patient has contact with, which is slightly larger than the basic regeneration number. Table 3 shows that the average degree of centrality is less than 1. The smallest degree is 0. The largest is 11, indicating that the patient (case 26 in figure 5) has contacted 11 other patients. According to the degree of centrality, it can be speculated that the basic regeneration number of novel coronavirus in Shaanxi Province is less than 1, which means that the spread of the virus is well controlled. Closeness centrality indicates the closeness of the patient with other patients. Higher values indicate faster transmission between patients and fewer intermediate patients. The average value is 0.452, which is a slightly higher closeness centrality, indicating that the infection is mostly cluster

infection. It can be seen from Figure 5 that over time, many aggregated sub-networks are formed, but the network is not fully connected under the action of prevention and control measures. Betweenness centrality indicates >the level the patient as an intermediary in spread of the virus. The average value is 0, which is very low, indicating that there are few chain transmissions. According to Figure 5, we can see the transmission mode is mostly one-to-many, that is, A-B, A-C transmission. The PageRank index measures the centrality of the patient's position in whole contact network. The average value is 0.0042, which is very low. But the maximum value is 0.0228, which indicates that the degree of connection is unevenly distributed among patients. Highly infectious persons can cause a major outbreak of the disease. The spread of Ebola virus is being associated with these super disseminators 23. From Figure 6, we know that a small number of people have a higher degree of centrality. But only three patients have a degree of centrality greater than 5. Most patients' degree centrality is zero. This shows that although the degree distribution of patients in Shaanxi Province is uneven, the highest number of contacts is low, and there is no super disseminator.

Variables	Case No.	Mean	S.D.	Minimum Value	Maximus Value
Degree Centrality	237	0.987	1.351	0	11
Closeness Centrality	237	0.452	0.440	0	1
Betweenness Centrality	237	0	0.0001	0	0.0012
PageRank	237	0.0042	0.0035	0.001	0.023

Table 3. Statistics of the contact network of novel coronavirus pneumonia patients in Shaanxi Province. The value of closeness centrality and betweenness centrality is in normalized form.

Discussion

The research conclusions of the thesis are as follows. In terms of characteristics of novel coronavirus pneumonia patients in Shaanxi Province of China, the susceptible population is mostly middle-aged men. There are many imported cases in the early stage of epidemic. Local

infection is the main part in the later stage. The infection route are mostly clustered infections based on strong ties such as relatives. In terms of virus contact network, the entire network is divided into multiple components and in a disconnected state. The degree distribution of the contact network is skewed. Most patients' degree centrality is zero. The highest patient's degree centrality is 11, indicating that there is no super disseminator in the network.

Based on the conclusions of this study, the epidemic prevention policy should pay special attention to cluster infections, prevent patients from forming a giant component in contact network, and avoid the network connected. The paper finds that infections are mostly based on strong ties. Therefore, for suspected infection, in-house quarantine is effective. But cluster infection may be caused in family. Centralized quarantine for patients with mild illness is the better way. It should also be emphasized that family members need to develop good hygiene habits.

Existing research on novel coronavirus pneumonia has not been analyzed from the perspective of contact network. This research enriches existing research and helps to better understand the social network transmission of novel coronavirus. Of course, due to time and data limitations, there are still some shortcomings in this study. For example, case analysis is limited to Shaanxi and there is no way to present a virus contact network across regions. I hope that there will be more studies on contact network of novel coronavirus pneumonia in the future based on detailed epidemic data.

Declarations

Acknowledgements

This work was supported by National Natural Science Foundation of China (Grant No. 71902155).

Data availability

All data and code are on an OSF data repository, see <https://osf.io/h9qmu/> . I made a video to depict the dynamic change of the contact network from January 23th to February 16th, see <https://youtu.be/JtU1sfgjup8> . All raw data (in Chinese) is available on the official website of Health Committee of Shaanxi Province of China, see <http://sxwjw.shaanxi.gov.cn/col/col9/index.html> . All data has been anonymized.

References

1. Munster, V. J., Koopmans, , van Doremalen, N., van Riel, D. & de Wit, E. A Novel Coronavirus Emerging in China—Key Questions for Impact Assessment. *New Engl. J. Med.* **382**, 692-694 (2020).
2. Chan, J. et al. A Familial Cluster of Pneumonia Associated with the 2019 Novel Coronavirus Indicating Person-To-Person Transmission: A Study of a Family Cluster. *The Lancet.* **395**, 514-523 (2020).
3. Zhu, et al. A Novel Coronavirus From Patients with Pneumonia in China, 2019. *New Engl. J. Med.* (2020).
4. Li, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New Engl. J. Med.* (2020).
5. Phan, L. T. et al. Importation and Human-To-Human Transmission of a Novel Coronavirus in Vietnam. *New Engl. J. Med.* **382**, 872-874 (2020).
6. Chinazzi, et al. The Effect of Travel Restrictions On the Spread of the 2019 Novel Coronavirus (COVID-19) Outbreak. *Science.* (2020).
7. Wang, , Horby, P. W., Hayden, F. G. & Gao, G. F. A Novel Coronavirus Outbreak of Global Health Concern. *The Lancet.* **395**, 470-473 (2020).

8. Chang, et al. Epidemiologic and Clinical Characteristics of Novel Coronavirus Infections Involving 13 Patients Outside Wuhan, China. *Jama*. (2020).
9. Bearman, P. S., Moody, & Stovel, K. Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks. *American journal of sociology*. **110**, 44-91 (2004).
10. Marvel, S. A., Martin, T., Doering, C. R., Lusseau, & Newman, M. E. The Small-World Effect is a Modern Phenomenon. *arXiv preprint arXiv:1310.2636*. (2013).
11. Salathé, et al. A High-Resolution Human Contact Network for Infectious Disease Transmission. *Proceedings of the National Academy of Sciences*. **107**, 22020-22025 (2010).
12. Jaffe, H. The Early Days of the HIV-AIDS Epidemic in the USA. *Nat. Immunol.* **9**, 1201-1203 (2008).
13. Kuperman, & Abramson, G. Small World Effect in an Epidemiological Model. *Phys. Rev. Lett.* **86**, 2909 (2001).
14. Watts, J. & Strogatz, S. H. Collective Dynamics of 'Small-World' Networks. *Nature*. **393**, 440- 442 (1998).
15. Eubank, et al. Modelling Disease Outbreaks in Realistic Urban Social Networks. *Nature*. **429**, 180- 184 (2004).
16. Meyers, L. A., Pourbohloul, B., Newman, E., Skowronski, D. M. & Brunham, R. C. Network Theory and SARS: Predicting Outbreak Diversity. *J. Theor. Biol.* **232**, 71-81 (2005).
17. Chen, et al. Highly Dynamic Animal Contact Network and Implications On Disease Transmission. *Sci. Rep.-UK*. **4**, 4472 (2014).
18. Huang, et al. Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. *The Lancet*. **395**, 497-506 (2020).
19. Chen, N. et al. Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study. *The Lancet*. **395**, 507-513 (2020).
20. Armbruster, , Wang, L. & Morris, M. Forward Reachable Sets: Analytically Derived Properties of Connected Components for Dynamic Networks. *Network Science*. **5**, 328-354 (2017).
21. Onaga, , Gleeson, J. P. & Masuda, N. Concurrency-Induced Transitions in Epidemic Dynamics On Temporal Networks. *Phys. Rev. Lett.* **119**, 108301 (2017).
22. Tian, F. & Lin, N. Weak Ties, Strong Ties, and Job Mobility in Urban China: 1978–2008. *Social Networks*. **44**, 117-129 (2016).
23. Lau, M. et al. Spatial and Temporal Dynamics of Superspreading Events in the 2014–2015 West Africa Ebola Epidemic. *Proceedings of the National Academy of Sciences*. **114**, 2337-2342 (2017).

Figures

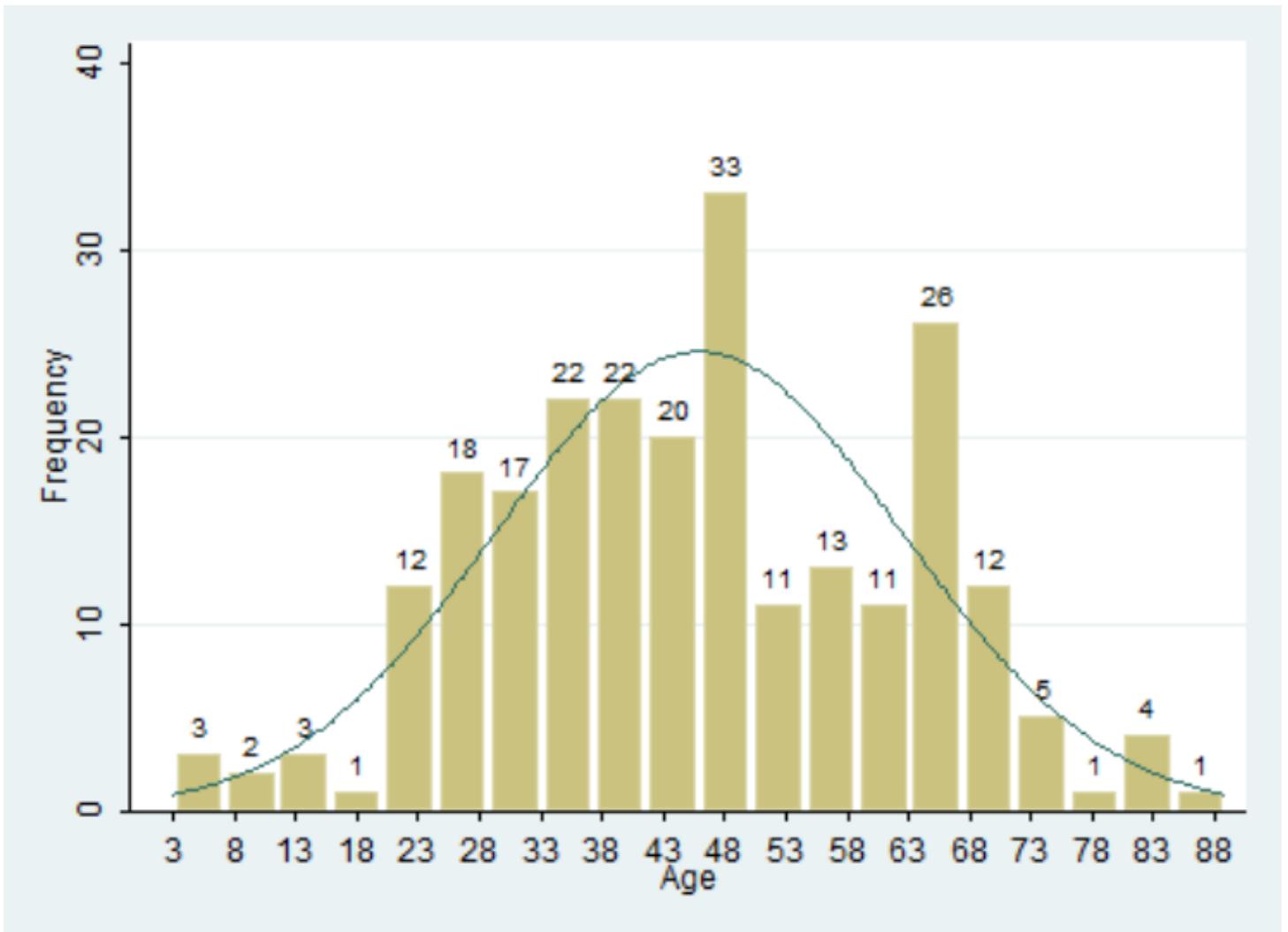


Figure 1

Age distribution of novel coronavirus pneumonia patients in Shaanxi Province, China.

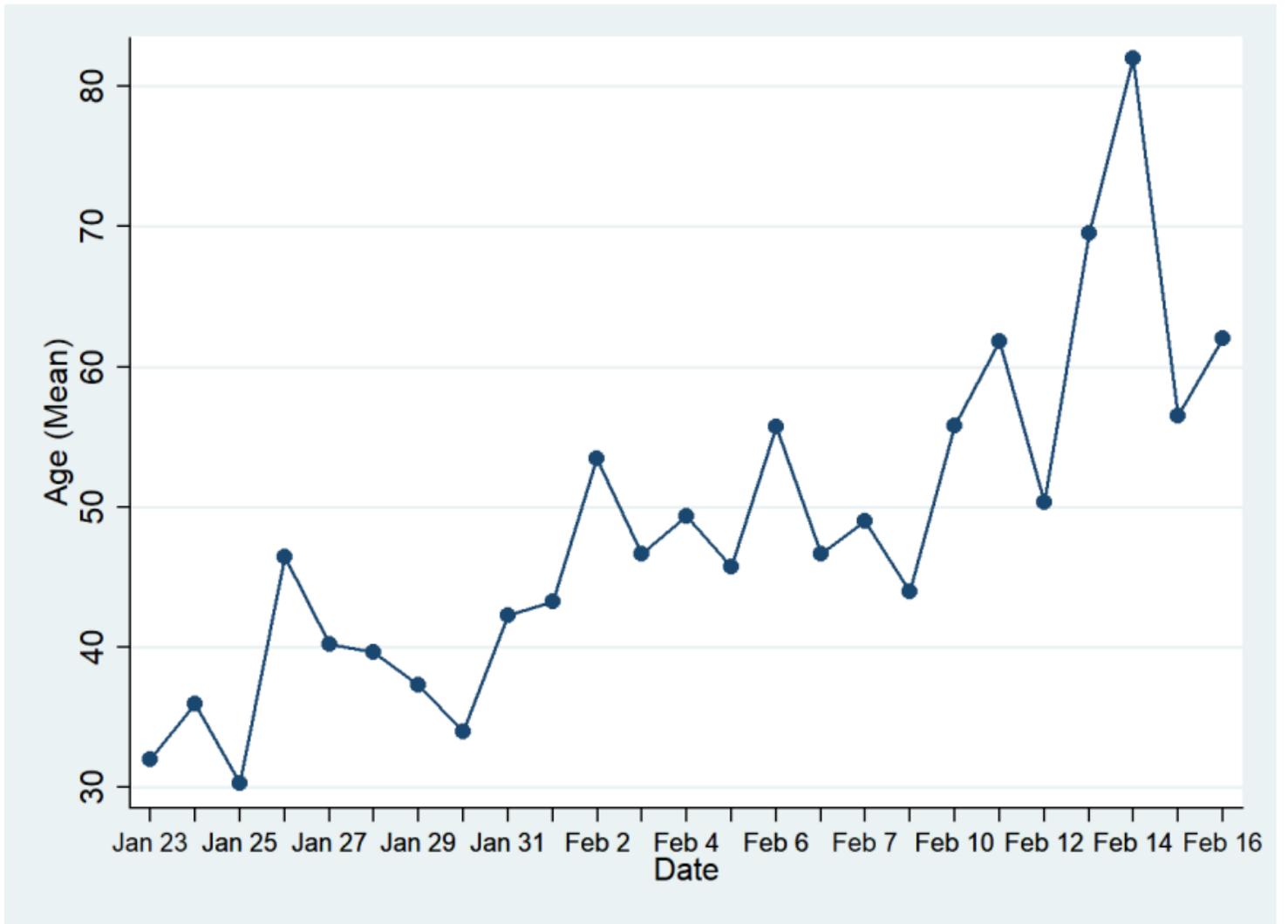


Figure 2

Mean Age of Novel Coronavirus Pneumonia Patients in Shaanxi Province, China.

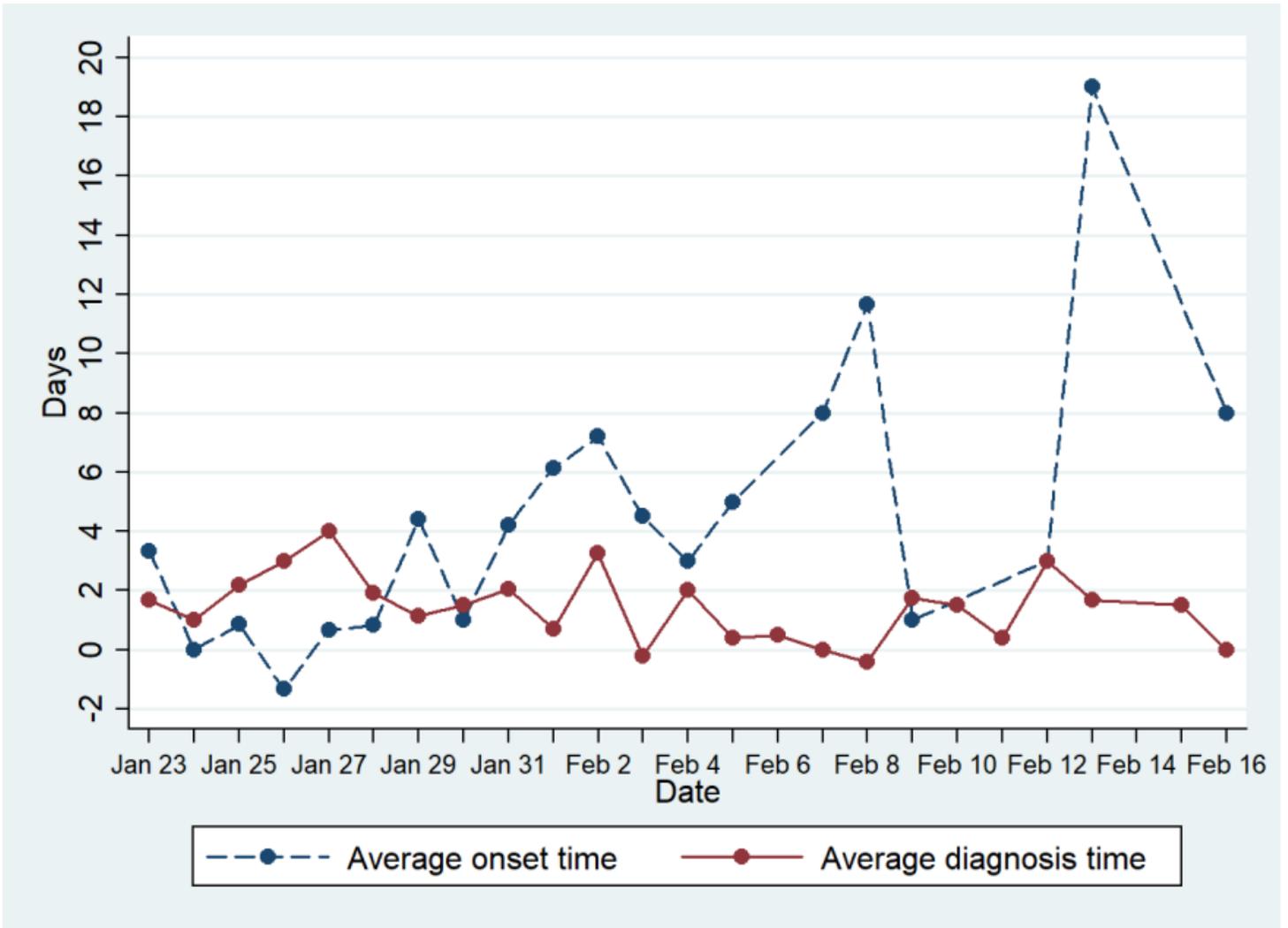


Figure 3

Changes in the average onset and diagnosis time of novel coronavirus pneumonia patients in Shaanxi Province. Average onset time refers to the time when symptoms occur in imported cases after they reach Shaanxi. Average diagnosis time refers to the time taken for all cases to take medical measures after symptoms appear. Both calculations are based on published cases daily.

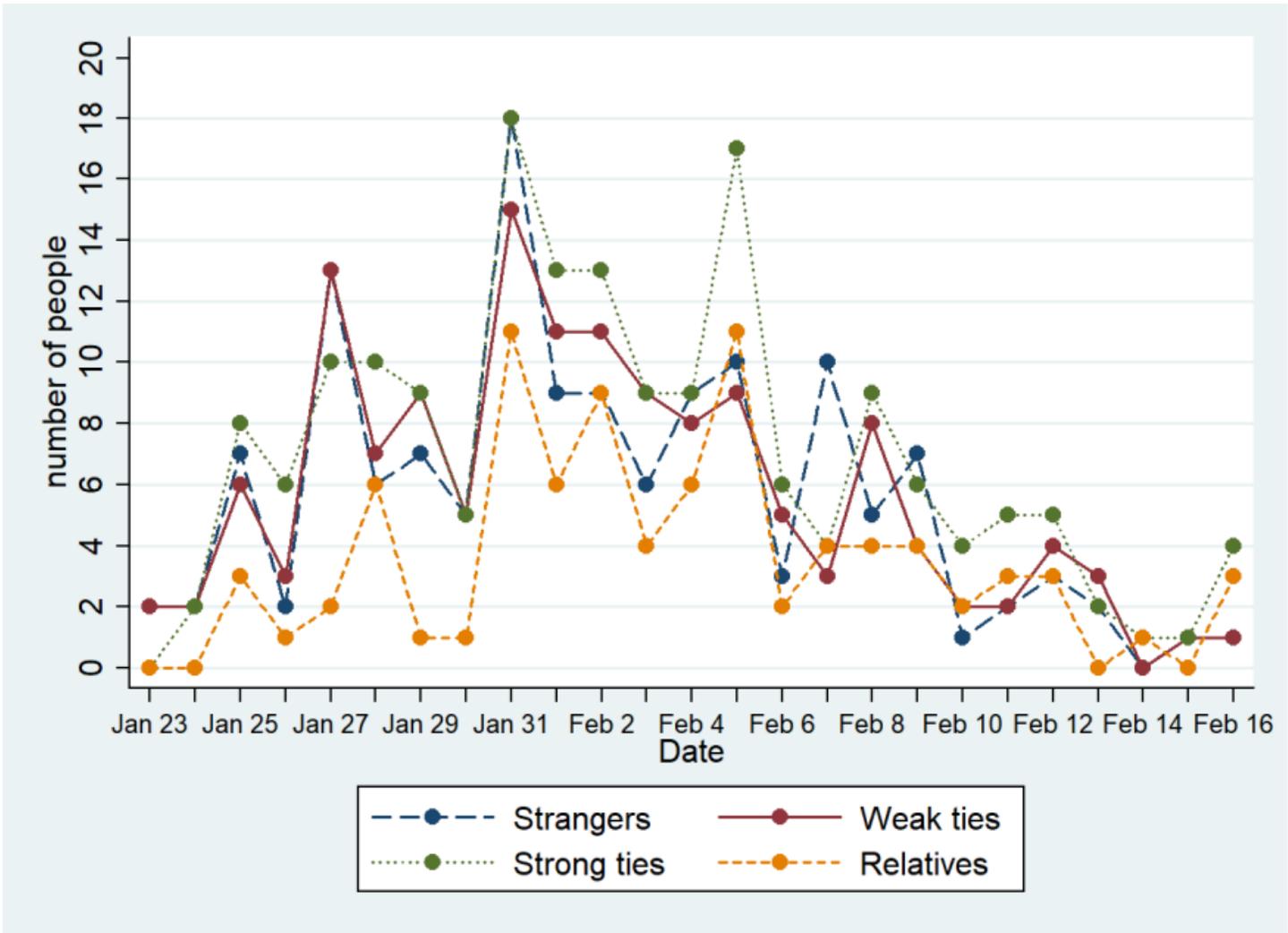


Figure 4

Social network infection route of patients with novel coronavirus in Shaanxi Province. Strangers refers to how many of the cases announced that day may have been infected by strangers. Weak ties indicate how many people may be infected by weak tie contacts. Strong ties refer to how many people may be infected by strong tie contacts. Relatives refer to whether any of the patient's relatives are infected.

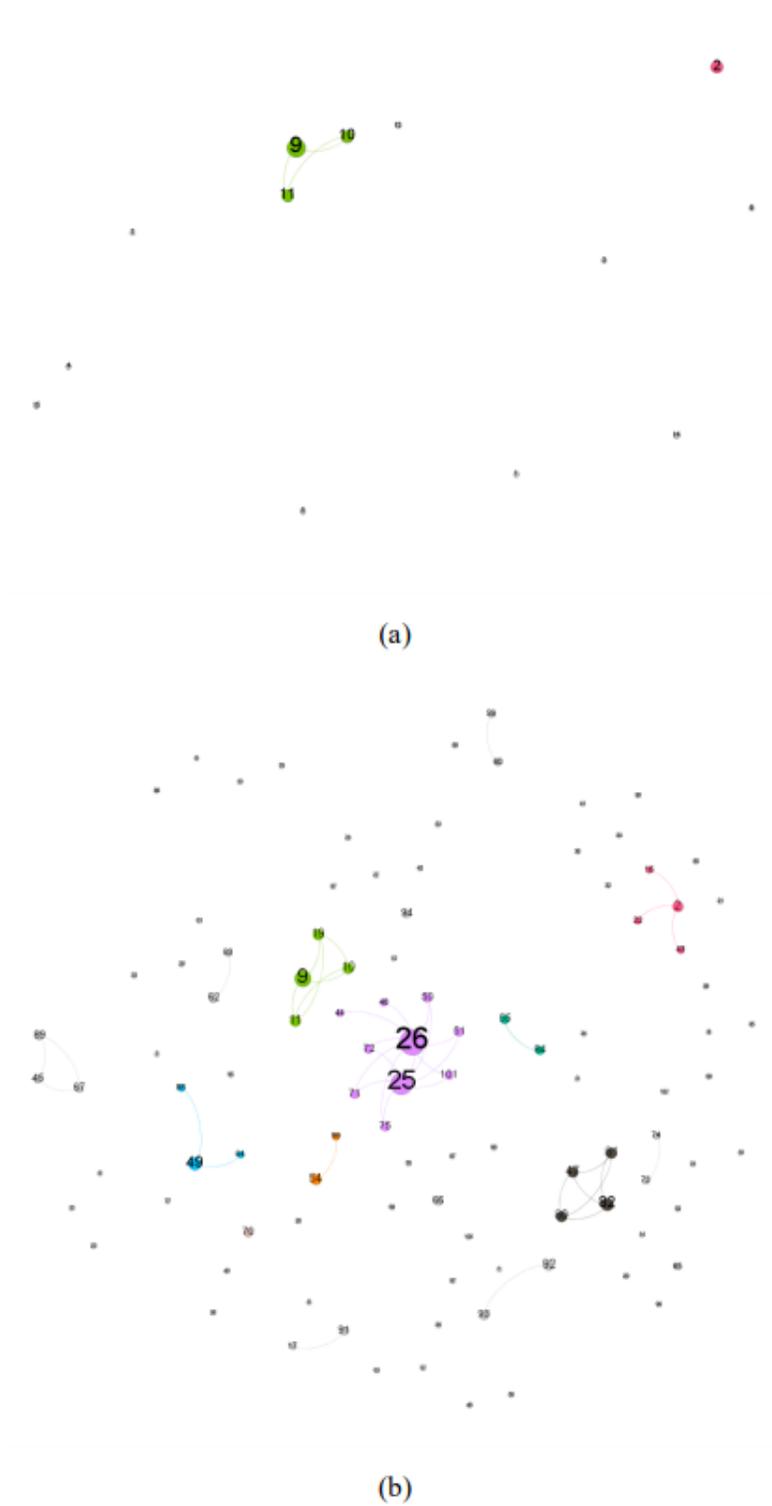


Figure 5

Dynamic contact network of patients with novel coronavirus pneumonia in Shaanxi Province. In the network, nodes represent patients, and edges represent transmission routes between patients. The larger the node size, the greater the number of the patient connected with. The number on the node is the case number. The larger the number, the later the patient is diagnosed. (a) Contact Network of patients with novel coronavirus pneumonia in Shaanxi Province of China on January 25, 2020. Node means patients,

edges means transmission routes. (b) Contact Network of patients with novel coronavirus pneumonia in Shaanxi Province of China on February 1, 2020. Node means patients, edges means transmission routes. (c) Contact Network of patients with novel coronavirus pneumonia in Shaanxi Province of China on February 16, 2020. Node means patients, edges means transmission routes.

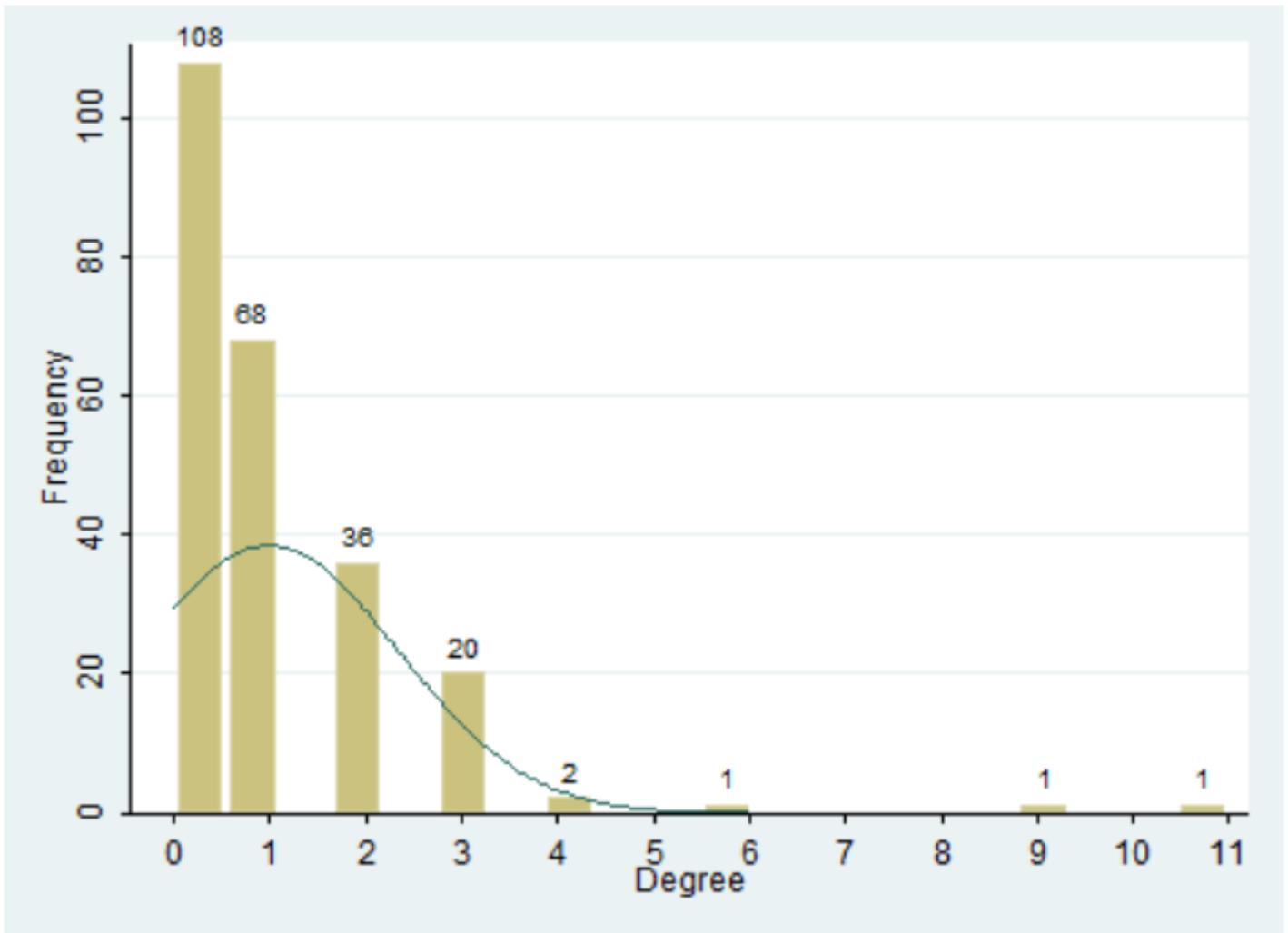


Figure 6

Distribution of degree centrality of contact network of novel coronavirus pneumonia patients in Shaanxi Province.