Investigating COVID-19 transmission in a tertiary hospital in Hanoi, Vietnam using social network analysis

Ngoc-Anh Hoang[1,2*], Thai Quang Pham[2,3*], Ha-Linh Quach[1,2], Ngoc Van Hoang[4], Khanh Cong Nguyen[2], Duc-Anh Dang[5**], Florian Vogt[1,6**]

[1] *National Centre for Epidemiology and Population Health, College of Health and Medicine, Australian National University, Canberra, Australia*

[2] *Department of Communicable Diseases Control, National Institute of Hygiene and Epidemiology, Hanoi, Vietnam*

[3] *Department of Biostatistics and Medical Informatics, School of Preventive Medicine and Public Health, Hanoi Medical University, Hanoi, Vietnam*

[4] *The General Department of Preventive Medicine, Ministry of Health, Hanoi, Vietnam*

[5] *National Institute of Hygiene and Epidemiology, Hanoi, Vietnam*

[6] *The Kirby Institute, University of New South Wales, Sydney, Australia*

*Equal first authors
**Equal last authors

ABSTRACT

Objectives: In March 2020, a COVID-19 outbreak in a major referral hospital in Hanoi, Vietnam led to 7,664 patients and staff being sent into lockdown for two weeks, and more than 52,200 persons across 49 provinces being quarantined. We assessed SARS-CoV-2 transmission patterns during this to-date largest hospital outbreak in Vietnam using social network analysis (SNA).

Methods: We constructed a directed relational network and calculated network metrics for 'degree', 'betweenness', 'closeness', and 'eigenvector' centrality to understand individual-level transmission patterns. We analysed network components and modularity to identify sub-network structures with disproportionately big effects.

Results: We detected 68 connections between 46 confirmed cases, of whom 27 (58.7%) were ancillary support staff, seven (15.2%) caregivers, six (13%) patients, and two (4.4%) nurses. Among the ten most important cases selected by each SNA network metric, transmission dynamics clustered in 17 cases, of whom 12 (70.6%) cases were ancillary support staff. Ancillary support staff also constituted 71.1% of cases in the dominant sub-network and 68.8% of cases in the three largest sub-communities.

Conclusions: We identified non-clinical ancillary support staff, who are responsible for room service and food distribution in hospital wards in Vietnam, as a group with disproportionally big impacts on transmission dynamics during this outbreak. Our findings call for a holistic approach to nosocomial outbreak prevention and response that includes both clinical and non-clinical hospital staff. Our work also shows the potential of SNA as a complementary outbreak investigation method to better understand infection patterns in hospitals and similar settings.

Keywords: nosocomial infection, hospital outbreak, social network analyses, COVID-19, Vietnam.

INTRODUCTION

Hospital staff and patients are at significantly higher risk of SARS-CoV-2 infection than the general population, and several COVID-19 outbreaks in hospitals have been reported around the world (1–3). In Vietnam, the first COVID-19 outbreak in a health care setting occurred in Bach Mai Hospital (BMH), a major tertiary referral hospital in Hanoi, when one of its staff was confirmed SARS-CoV-2 positive on 18 March 2020. The outbreak response gradually escalated to a hard lockdown of the entire hospital, and about 52,200 persons were quarantined across Vietnam as part of the containment efforts. A total of 46 cases were detected before the BMH outbreak was declared over after three weeks.

Standard epidemiological techniques like case investigation and contact tracing have been widely used during the COVID-19 pandemic. Additional techniques like social network analysis (SNA) are much less frequently used. The aim of SNA in outbreak investigation is to provide additional, in-depth information to elucidate detailed transmission patterns and sub-networks within the concerned populations (4–6). While SNA has, where applied, proved useful to explore individual-level transmission dynamics for diseases like tuberculosis, HIV, and E.coli outbreaks (7–9), it is not regularly used as part of the standard epidemiological outbreak investigation techniques in field epidemiology. While there have been a couple of attempts during the SARS-CoV-2 pandemic to use SNA to understand super-spreader events (10,11), and in a COVID-19 cluster in the community (12), the value of SNA for outbreak investigations of COVID-19 has not yet been sufficiently established. In particular, SNA has not yet been used to investigate SARS-CoV-2 transmission dynamics in closed settings like hospitals.

In this study, we used SNA techniques to gain an in-depth understanding of the transmission structure during the COVID-19 outbreak in the BMH, Hanoi, Vietnam in March 2020. Based on our findings, we also discuss the benefits of SNA as a supplementary tool in field epidemiology generally.

METHODS

Design

This study was a retrospective outbreak investigation using SNA. We calculated key network centrality metrics, in particular for 'degree', 'betweenness', 'closeness', and 'eigenvector' centrality, and constructed sociograms using contact tracing data collected during the BMH outbreak response to identify the transmission dynamics at the individual level. Appendix 1 contains detailed definitions and further explanations of SNA metrics used in this manuscript.

Setting

BMH is a major tertiary hospital in Hanoi, Vietnam. The Tropical Disease Centre of BMH was designated as one of the reference units to treat COVID-19 patients. On 18 March 2020, a nurse working at the COVID-19 screening office in this centre tested positive for SARS-CoV-2, thereby triggering the first nosocomial COVID-19 outbreak in Vietnam. Outbreak response was quickly escalated after active case finding and extensive testing discovered several cases in other departments, ultimately leading to a hard lockdown of BMH on 28 March 2020 for 7,664 persons, among them 6,258 staff, 793 patients, and 613 caregivers, within the hospital premises for two weeks. More than 52,200 persons were quarantined for 14 days across 49 provinces in Vietnam. About 15,000 tests were done among persons quarantined in the hospital and a further 30,000 in the community. A total of 46 laboratory-confirmed cases were

identified in seven units of BMH (Figure 1). Details about the investigation and response to the BMH outbreak using standard epidemiological methods are described in Appendix 2 as well as elsewhere (13–15).

Data collection

Data were collected in March 2020 during the outbreak investigation and response activities. For the SNA analysis, we compiled and extracted socio-demographic and occupational information on cases and their exposure histories from records of the Rapid Response Team (RRT) of the National Steering Committee for COVID-19 Prevention and Control in June 2020. During the BMH outbreak, the contact information of anyone testing positive for SARS-CoV-2 was provided to the RRT who interviewed all cases via phone. Collected information included demographic, occupational, and basic clinical information (age, sex, occupation, and comorbidity), date of symptom onset, as well as a detailed movement and contact history during the 14 days before the date of onset of symptoms or confirmation, whichever came first. Video surveillance inside the hospital was reviewed alongside the case interview to verify and complement the information provided by cases as part of outbreak response and to establish the most likely place of infection and source(s) of infection for each case.

Variables available for this analysis were the cases' age, sex, occupation, comorbidities, date of symptom onset, date of positive SARS-CoV-2 test result, place of presumed infection, and sources of infection for each case.

Data analysis

We created a relational network dataset with confirmed COVID-19 cases as "nodes" and the epidemiological contacts between them as "edges". Each case was provided with a unique identification number, and each case's unique epidemiological contacts with other cases were also identified and labelled. The relational data were analysed from a direct network perspective, which contained directionality of infection emanated from a "source case" towards the "target case" (16).

We calculated and compared the score rank of the social network centrality measures 'degree centrality', 'betweenness centrality', 'closeness centrality', and 'eigenvector centrality' to identify the key nodal cases that were particularly influential during this outbreak. We also calculated network components and modularity to identify sub-network structures that contributed disproportionately to the transmission of infection. Average path length and network diameter measures were analyzed to assess the dispersion level of infection within networks.

The relational data with nodes and edges were created in Microsoft Excel software and transferred into Gephi software (version 0.9.2) for network illustration and into R (version 4.1.1, package "igraph") for analyses (17,18). Network measures, components, and modularity were generated, and the outputs were derived as tables and graphs. The network measure tables from Gephi were exported to R for generating network summary measures. Network graphs were created to visualize relationship patterns between source and target cases. Yifan Hu layout and Fruchterman Reingold layout in Gephi software were used to visualize the networks (19). We also calculated frequencies and percentages to describe the distribution of cases' age, sex, occupation, comorbidity, place of presumed infection, and departments of recorded cases in the hospital using R (version 4.1.1).

Ethics

This manuscript was approved and exempted by the Internal Review Board, National Institute of Hygiene and Epidemiology, Vietnam as a part of national COVID-19 outbreak investigation and response activities.

RESULTS

Case characteristics

Overall, 46 cases were detected among ancillary support staff (27, 58.7%); caregivers (7, 15.2%); patients (6, 13%); nurses (2, 4.4%); and four secondary cases (8.7%) outside the hospital. Most cases were female (85%), between 35 and 64 years old (69.6%), and with comorbidities (71.7%). Forty (87%) cases acquired infection most likely inside the hospital, and the remaining cases through family and community contacts. Nearly half of cases (45.7%) had COVID-19 compatible symptoms during their infection period, whereas the rest (25, 54.3%) remained asymptomatic (Table 1).

Relationship analysis

There were a total of 68 epidemiological connections identified between the 46 laboratory-confirmed cases, of which 27 were source cases and 17 were target cases. Network measures showed that 29 cases (63%) had an in-degree centrality measure of zero and thus lacked any epidemiological link with a source case. Of the rest, seven cases (15.2%) had an in-degree measure of two, implying that two source cases were identified for each. A small number of cases had an in-degree measure of more than two (range 3 – 6). Nineteen (41.3%) of the total 46 cases had an out-degree centrality measure of zero and thus did not act as a source of infection to other cases. The remaining 27 (58.7%) cases were found to have an out-degree measure above one (range 1-6), thus presenting as sources of infection to a total of 17 target cases (Table 2).

The average path length between cases, indicating the average number of edges reflective of the length of the generational transmission, was 1.8. The network diameter, the largest path length between cases reflective of the highest number of transmission generations, was 4.

Based on degree centrality metrics, case P27 was the node with the highest number of links to other cases (10, 21.7% of the total cases in the network). Also, closeness, betweenness, and eigenvector network centrality metrics indicated that P27 was the most important node in the transmission network (closeness centrality = 0.0022, betweenness centrality = 335.9, eigenvector centrality = 0.8). Case P28, P18, and P32 also ranked high on the four centrality measures, and together these four cases contributed to 35 (51.5%) of all epidemiological connections in the network (Figure 2a and Table 3).

The betweenness centrality measure was zero for 30 cases (65.2%), implying that they did not play a role in the transmission of SARS-CoV-2 between any other pair of cases. The remaining 16 cases (34.8%) had a non-zero betweenness centrality value (range 1 to 30), indicating their involvement in the transmission route between other cases in the network. The three cases with the highest betweenness centrality were P27, P28, and P18 (Figure 2b).

Three cases (P20, P23, and P26) did not have many contacts with others (low degree centrality) and were not a bridge in the transmission route (low betweenness centrality); however, P26 was a key nodal case influential in transmitting SARS-CoV-2 infection due to short distances to all other cases in the outbreak (closeness centrality: P26 = 0.00216), and P20 and P23 influenced the most important cases in the network (eigenvector centrality: P20 = 0.8, P23 = 0.8) (Figure 2c and Figure 2d and Table 3).

Table 3 shows the measures of network centrality estimates for the 10 most important nodes selected by the four-network metrics. A total of 17 cases were identified based on the score rank of these network metrics. 12 of 17 (70.6%) were ancillary support staff, two (11.8%) were patients, and three (17.6%) were caregivers.

Two connected components (sub-networks) and five cases without epidemiological contacts were identified. The bigger of the two sub-networks ('giant components') accounted for 82.6% (38) of the total cases and 94% (64) of all transmission contacts of the network. Ancillary support staff contributed 71.1% (27 cases) in this component (Figure 3A).

We detected 10 communities (modules). The modularity score was 0.576, indicating the possible presence of community structure. The largest module contained 14 cases and 20 contacts, which accounted for 30.4% of total cases and 29.4% of total transmission contacts in the network. The two next second-largest communities contained nine cases each. The ancillary support staff contributed to 68.8% (22 cases) of the total cases in the three largest communities (Figure 3B).

DISCUSSION

We used SNA for a detailed understanding of the individual-level transmission dynamics during this first COVID-19 outbreak in the health care setting in Vietnam. Our findings highlight the importance of ancillary support staff, a non-clinical cadre of hospital staff that is often overlooked in the prevention of nosocomial infections, in driving the BMH outbreak. We also demonstrate the value of SNA as an additional method to understand the heterogeneity in transmission dynamics of COVID-19 outbreaks in closed settings in greater detail than standard outbreak investigation techniques allow.

We found a high degree of variation at the level of individual cases in SARS-CoV-2 transmission. Several epidemiological studies have since shown the importance of heterogeneity in SARS-CoV-2 transmission and the need to measure transmissions and variations at individual level (20–23). The out-degree centrality measures found in our study indicated that while nearly half (41.3%) of the cases in the network did not transmit infection, the rest (58.7%) were source cases that disproportionately transmitted to target cases. In-degree centrality measures in our analysis showed that no source of infection was found for 63% of cases and that 21.8% of cases had three or more sources of infection. In terms of the number of connected contacts (via both out-degree and in-degree centralities), cases P18, P27, P28, and P32 were the top 4 cases, of which P27 was the most influential case. P27 was an ancillary support staff who distributed food in all wards of BMH. Case P27 was detected six days after showing COVID-19-compatible symptoms after exhaustive testing for all persons present inside the BMH at that time. This case was epidemiologically closely linked to 7 other ancillary support staff who shared the same working and living space, and 3 patients treated in the BMH who had close contact during the infection period.

The importance of a case to others in a network is not only defined through its number of contacts but also by the interconnectivity of its secondary cases, which is measured through the betweenness metric. Betweenness centrality measures highlight that not all infections were directly transmitted from a few influential sources (with higher out-degree measures) to many target cases, but through influential bridging cases (with high betweenness measures) who acquired the infection from them. The high betweenness score of P27, P28, and P18 indicated that these three cases were not only connected to many other cases, but they also contributed as important pathways to further infection. However, the high degree of centrality but low betweenness score of P32, an ancillary support staff, indicated that while connecting many other cases, P32 was not a bridge of further transmission. In contrast, P39, another ancillary support staff, had a low degree centrality score but high betweenness score suggesting that although P39 did not have contact with many other cases, this case was critical in the spread of SARS-CoV-2 to others.

The path length metrics indicated that on average any source and target cases in the network could be reached by crossing 1.8 steps; the maximum was four steps between the farthest placed source and target cases. This highlights that the transmission of infection was not widely dispersed during the BMH outbreak through multiple generations; it was on average limited to less than two generations.

Our in-depth analysis of the transmission network of the BMH outbreak using SNA has relevant implications in two aspects. First, it could help public health personnel comprehend the key influential groups to prioritize early public health interventions to reduce the likelihood of nosocomial spread (24). In this outbreak, the network metrics (degree centrality, betweenness centrality, closeness centrality, eigenvector centrality) shed light on the potential importance of non-clinical hospital staff in driving transmission dynamics during nosocomial COVID-19 outbreaks. These were ancillary support staff who distributed food and cleaned wards in the hospital. This type of work meant they frequently moved across BMH and had contact with many people, including patients, caregivers, and health care workers. However, they were often overlooked during the initial response to COVID-19 outbreaks in hospitals due to not being directly involved in patient care. They were also exempted from the infection prevention control training at the beginning of the pandemic. We recommend that all types of hospital staff regardless of their involvement in patient care be considered as potential drivers of transmission during a nosocomial outbreak response. Focussing only on core clinical staff like doctors and nurses might delay successful outbreak containment.

Second, from a methodological perspective, integrating SNA alongside more established epidemiological investigation techniques during an outbreak could help target resources for outbreak response more efficiently (5). The identification of highly influential cases (with high out-degree and betweenness measures) and closely connected case sub-networks (components) could allow for prioritizing contact tracing activities in a more targeted way. Cases that disproportionately infect others and their networks could be identified quickly, and their contacts could be monitored more intensely to interrupt transmission faster. Nonetheless, one of the limitations of the practical use of SNA in field epidemiology is the time needed to collect the necessary data during a real-time outbreak response. Given the time-sensitive and volatile nature of infectious disease outbreaks, collecting data for network analyses alongside traditional outbreak containment activities might be complicated, especially when resources are stretched, and interfere with other response priority activities. Therefore, in a frontline setting, the use of SNA to support an outbreak investigation may have to be limited to unavoidable circumstances, e.g. contact tracing is not performed due to some reasons, e.g. rapid spread of the virus in the community and lack of human resources.

SNA to inform SARS-CoV-2 contact tracing has proved useful for supporting control transmission in South Korea; it was applied to support the efficiency of governmental responses to COVID-19 (10). A study from India used relational data to prioritize the most susceptible regions for target testing and quarantining (11). A study in Hong Kong used contact tracing data to identify the most important SARS-CoV-2 clusters in the community by constructing a SARS-CoV-2 network of transmission (23). Awareness of the value of SNA in COVID-19 outbreak response plus technical skills to perform such analyses should be increased among field epidemiologists, in particular in settings where timely genomic sequencing is absent.

We acknowledge several limitations of our study. First, this analysis was based on data from case interviews about movement and contact history over 14 days. The accuracy of some of the provided information could not be verified during the early outbreak response phase. For example, in our data, 63% of cases had no detected sources of infection, which was most likely due to recall inaccuracy. To support the accuracy of SNA, we recommend designing a standardized questionnaire to collect relational data and preparing trained interviewers for SNA data collection.

Second, because the data collection was not specifically designed for an SNA analysis, we acknowledge missing some meaningful information for a deeper understanding of SARS-CoV-2 transmission dynamics that could bias the network analysis, for example, information on the cases' behaviors to reduce the risk of infection or onwards transmission. Third, due to a lack of data for weighting directed edges, such as duration of contact or location of contact (e.g. in small, enclosed rooms with poor ventilation or in larger, well-ventilated areas), we had to treat all edges, representing contacts, as equal. Unweighted edges might lessen the importance of ancillary support staff in the transmission network, given their close working and living conditions. Fourth, we were not able to compare our results from SNA to genomic sequencing data as this was not available for our analysis. Though the outbreak was contained within three weeks (15), the initial source of infection for this outbreak was uncovered.

CONCLUSION

This study shows the successful application of SNA for an in-depth investigation of Vietnam's first COVID-19 outbreak in a hospital setting. We identified non-clinical ancillary support staff, who are responsible for room service and food distribution in hospital wards in Vietnam, as an important group with a disproportionally big impact on transmission dynamics during this outbreak. Outbreak response measures in hospitals often focus on patients and core medical staff such as doctors and nurses, while non-clinical hospital staff is overlooked during outbreak response in health facilities. Our findings call for a comprehensive approach to hospital infection prevention and control as well as nosocomial outbreak response that includes all professions of the hospital workforce. Our findings also point to the potential of SNA as a complementary method for field epidemiologists to better understand infection patterns in hospitals and similar settings and prioritize appropriate interventions accordingly.

REFERENCES

1. Xiang Y-T, Jin Y, Wang Y, Zhang Q, Zhang L, Cheung T. Tribute to health workers in China: A group of respectable population during the outbreak of the COVID-19. Int J Biol Sci. 2020;16(10):1739–40.

2. Chou R, Dana T, Buckley DI, Selph S, Fu R, Totten AM. Epidemiology of and Risk Factors for Coronavirus Infection in Health Care Workers: A Living Rapid Review. Ann Intern Med. 2020 Jul;173(2):120–36.

3. Reusken CB, Buiting A, Bleeker-Rovers C, Diederen B, Hooiveld M, Friesema I, et al. Rapid assessment of regional SARS-CoV-2 community transmission through a convenience sample of healthcare workers, the Netherlands, March 2020. Euro Surveill Bull Eur sur les Mal Transm = Eur Commun Dis Bull. 2020 Mar;25(12).

4. Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP, Bennett R, et al. Infection in Social Networks: Using

Network Analysis to Identify High-Risk Individuals. Am J Epidemiol [Internet]. 2005 Nov 15;162(10):1024–31. Available from: https://doi.org/10.1093/aje/kwi308

5. Vasylyeva TI, Friedman SR, Paraskevis D, Magiorkinisa G. Integrating molecular epidemiology and social network analysis to study infectious diseases: Towards a socio-molecular era for public health. Infect Genet Evol [Internet]. 2016;46:248–255. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5135626/

6. Wang X, An Q, He Z, Fang W. A Literature Review of Social Network Analysis in Epidemic Prevention and Control. Fuentes M, editor. Complexity [Internet]. 2021;2021:3816221. Available from: https://doi.org/10.1155/2021/3816221

7. Klovdahl AS. Social networks and the spread of infectious diseases: The AIDS example. Soc Sci Med [Internet]. 1985;21(11):1203–16. Available from: https://www.sciencedirect.com/science/article/pii/0277953685902692

8. Cook VJ, Sun SJ, Tapia J, Muth SQ, Argüello DF, Lewis BL, et al. Transmission Network Analysis in Tuberculosis Contact Investigations. J Infect Dis [Internet]. 2007 Nov 15;196(10):1517–27. Available from: https://doi.org/10.1086/523109

9. Devakumar D, Kitching A, Zenner D, Tostmann A, Meltzer M. Tracking sickness through social networks - the practical use of social network mapping in supporting the management of an E. coli O157 outbreak in a primary school in London. Epidemiol Infect [Internet]. 2013/02/27. 2013 Oct;141(10):2022–30. Available from: https://pubmed.ncbi.nlm.nih.gov/23445786

10. Jo W, Chang D, You M, Ghim G-H. A social network analysis of the spread of COVID-19 in South Korea and policy implications. Sci Rep [Internet]. 2021;11(1):8581. Available from: https://doi.org/10.1038/s41598-021-87837-0

11. Nagarajan K, Muniyandi M, Palani B, Sellappan S. Social network analysis methods for exploring SARS-CoV-2 contact tracing data. BMC Med Res Methodol [Internet]. 2020;20(1):233. Available from: https://doi.org/10.1186/s12874-020-01119-3

12. Kwan TH, Wong NS, Yeoh E-K, Lee SS. Mining relationships between transmission clusters from contact tracing data: An application for investigating COVID-19 outbreak. J Am Med Inform Assoc [Internet]. 2021 Oct 12;28(11):2385–92. Available from: https://pubmed.ncbi.nlm.nih.gov/34498059

13. Hoang V, Tran T, Nguyen T, Nguyen T, Nguyen Q. Successful containment of a COVID-19 outbreak in Bach Mai Hospital by prompt and decisive responses. Asian Pac J Trop Med [Internet]. 2021 May 1;14(5):234–5. Available from: https://www.apjtm.org/article.asp?issn=1995-7645

14. Tuan NQ, Anh HT, Yen NTH, Xuan DC, Van DT, Thanh D Van, et al. Prevent and control cross-transmission of COVID-19 in hospital settings: Lessons learned from a national hospital in Hanoi, Vietnam. J Glob Health [Internet]. 2021; Available from: http://www.jogh.org/documents/2021/jogh-11-03079.htm

15. Duy C, Nong VM, Van Ngo A, Doan TT, Nguyen TQ, Truong PT, et al. Nosocomial Coronavirus Disease Outbreak Containment, Hanoi, Vietnam, March–April 2020. Emerg Infect Dis J [Internet]. 2021;27(1):10. Available from: https://wwwnc.cdc.gov/eid/article/27/1/20-2656_article

16. SAGE. Basics of Social Network Analysis. In: Basics of Social Network Analysis [Internet]. SAGE; 2017 [cited 2022 Feb 23]. p. 3–23. Available from: https://us.sagepub.com/sites/default/files/upm-assets/78651_book_item_78651.pdf

17. Bastian M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In: Proceedings of the Third International Conference on Weblogs and Social Media. 2009.

18. Csardi, G. & Nepusz T. The igraph software package for complex network research. InterJournal, Complex Syst

[Internet]. 2006;1695:1–9. Available from:

https://www.bibsonomy.org/bibtex/bb49a4a77b42229a427fec316e9fe515#export

19. Gephi Tutorial Layouts [Internet]. 2011. Available from: https://gephi.org/tutorials/gephi-tutorial-layouts.pdf

20. Zhang Y, Li Y, Wang L, Li M, Zhou X. Evaluating Transmission Heterogeneity and Super-Spreading Event of COVID-19 in a Metropolis of China. Int J Environ Res Public Health [Internet]. 2020 May 24;17(10):3705. Available from: https://pubmed.ncbi.nlm.nih.gov/32456346

21. Arinaminpathy N, Das J, McCormick TH, Mukhopadhyay P, Sircar N. Quantifying heterogeneity in SARS-CoV-2 transmission during the lockdown in India. Epidemics [Internet]. 2021;36:100477. Available from: https://www.sciencedirect.com/science/article/pii/S1755436521000311

22. Endo A, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome open Res. 2020;5:67.

23. Adam DC, Wu P, Wong JY, Lau EHY, Tsang TK, Cauchemez S, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nat Med [Internet]. 2020;26(11):1714–9. Available from: https://doi.org/10.1038/s41591-020-1092-0

24. Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC. Network theory and SARS: predicting outbreak diversity. J Theor Biol. 2005 Jan;232(1):71–81.

Corresponding author: Ha-Linh Quach, Department of Communicable Diseases Control, National Institute of Hygiene and Epidemiology, Hanoi, Vietnam, and National Centre for Epidemiology and Population Health, College of Health and Medicine, Australian National University, Canberra, ACT, Australia. Phone +84 96 600 1080. Email u7061726@alumni.anu.edu.au
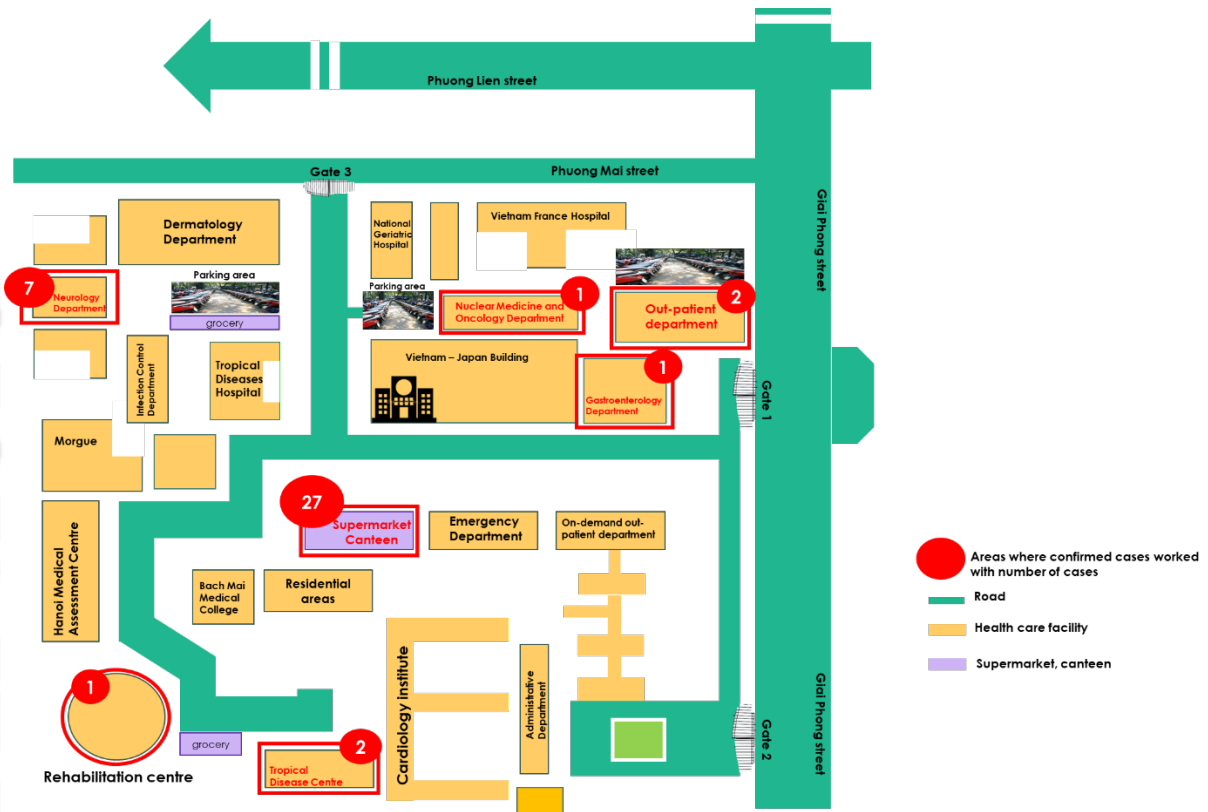
Figure 1. Map of infected areas in the Bach Mai hospital outbreak

Note: The map is not indicated four secondary cases outside the Hospital
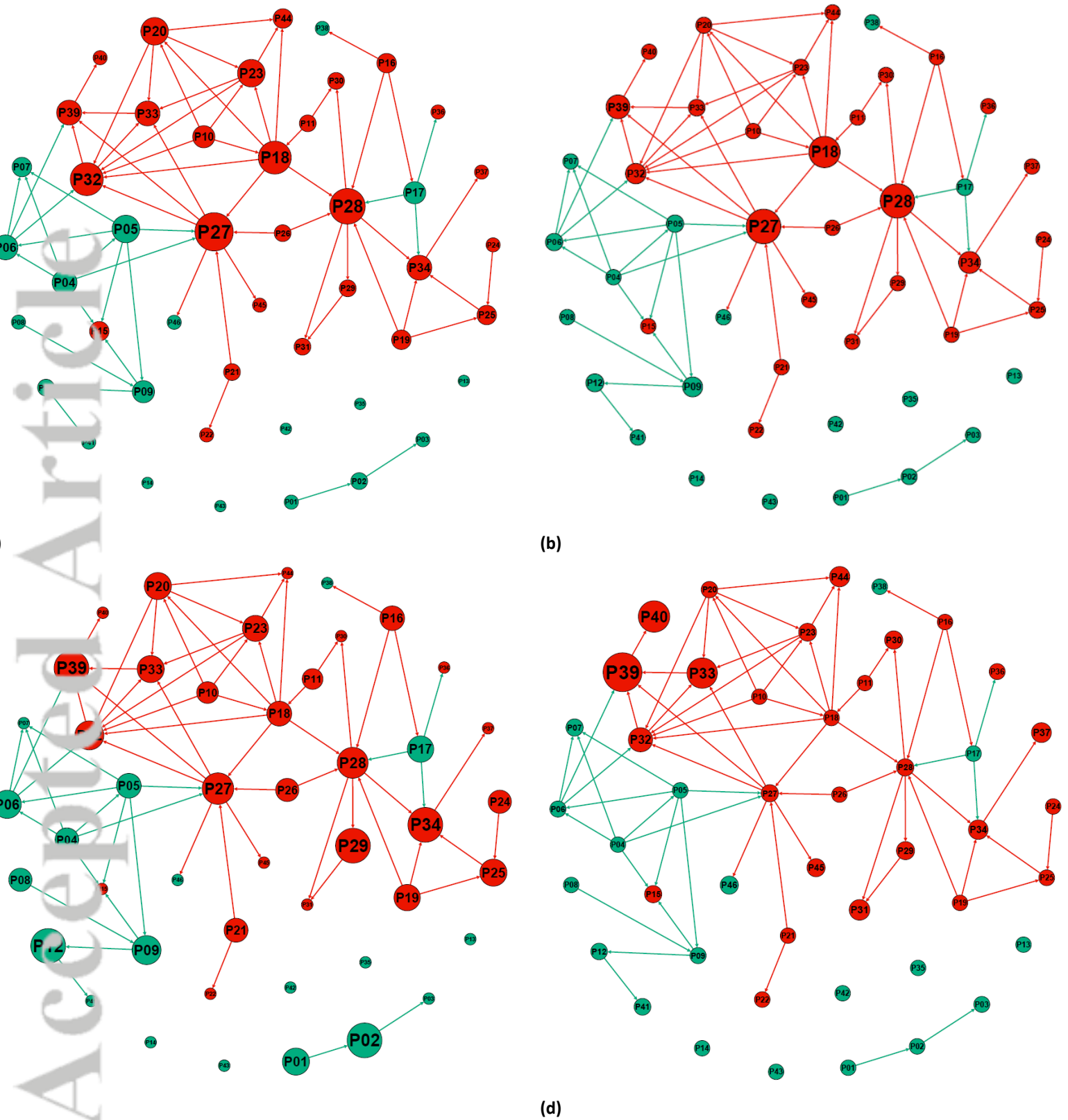
Figure 2. Network contact structure for the spread of SARS-CoV-2 infection in the Bach Mai Hospital.
Larger sized nodes imply higher score of (a) degree centrality, (b) betweenness centrality, c) closeness centrality, d) eigenvector centrality. Red circle indicated ancillary support staff and blue circle for other cases.
The colour of the edges was based on the colour of source case.

(a)

(b)
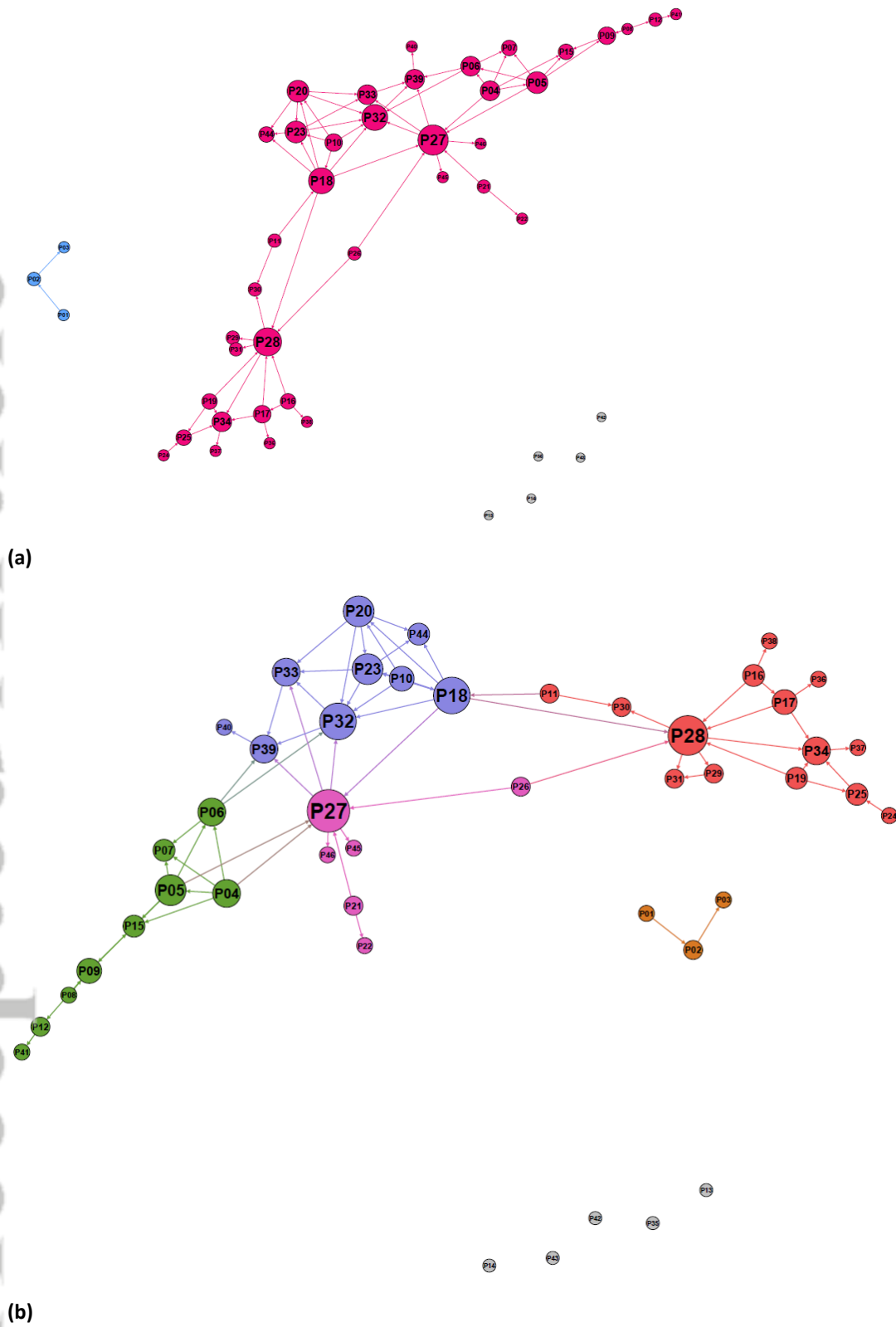
Figure 3. Connected components (a) and modularity (b) in the Bach Mai outbreak.

Larger sized nodes imply higher score of degree centrality. Different colours indicated different components (a) and

modules (b). Cases in the same component (a) and module (b) had the same colour.

Table 1. Case characteristics during the Bach Mai outbreak

Characteristics	N=46	%
Age groups		
<35	11	23.9
35 – 64	32	69.6
65+	3	6.5
Sex		
Female	39	84.8
Male	7	15.2
Occupation		
Ancillary support staff	27	58.7
Caregiver	7	15.2
Nurse	2	4.3
In-patient	4	8.7
Out-patient	2	4.3
Others	4	8.7
Departments of recorded cases		
Supermarket/Canteen	27	58.7
Neurology	10	21.7
Gastroenterology	2	4.3
Rehabilitation	1	2.2
Tropical Disease Centre	3	6.5
Nuclear Medicine & Oncology	1	2.2
Out patient	2	4.3
Comorbidity		
Yes	33	71.7
No	13	28.3
Place of presumed infection		
Hospital	40	87
Family	3	6.5
Community	1	2.2
Index cases	2	4.3
Having COVID-19 compatible symptoms		
Yes	21	45.7
No	25	54.3

Table 2. Case categorization based on degree centrality measures (n=46)

Degree centrality	No of cases	Percent	Case categorization[*]
In-Degree centrality measure			
0	29	63	No identified source of infection
2	7	15.2	Target case
3	4	8.7	Target case
4	3	6.5	Target case
5	2	4.3	Target case
6	1	2.2	Target case
Out-Degree centrality measure			
0	19	41.3	Not transmitted infection to anyone identified
1	10	21.7	Source case
2	5	10.9	Source case
3	5	10.9	Source case
4	3	6.5	Source case
5	3	6.5	Source case
6	1	2.2	Source case

[*] *Cases lacked source of infection if In-Degree centrality = 0; Target case had In-Degree centrality ≥1; Cases did not transmit SARS-CoV-2 to anyone had Out-Degree centrality = 0; Source case had Out-Degree centrality ≥1.*

Table 3. Summary of contact network metrics for the top 10 scores of COVID-19 cases (node (score)).

Score rank	Degree centrality	Betweenness centrality	Closeness centrality	Eigenvector centrality
1	P27* (10)	P27* (335.9)	P18* (0.00221)	P32* (1)
2	P28* (9)	P28* (329.0)	P27* (0.00220)	P18* (0.9)
3	P18* (8)	P18* (251.3)	P28* (0.00216)	P27* (0.8)
4	P32* (8)	P05 (152.3)	P32* (0.00216)	P20* (0.8)
5	P05 (6)	P09 (104.0)	P26* (0.00216)	P23* (0.8)
6	P20* (6)	P26* (92.0)	P05 (0.00211)	P33* (0.7)
7	P23* (6)	P34* (72.7)	P20* (0.00209)	P10* (0.6)
8	P04 (5)	P32* (65.8)	P23* (0.00209)	P39* (0.5)
9	P06 (5)	P17 (39.3)	P04 (0.00209)	P06 (0.4)
10	P33* (5)	P39* (38.3)	P06 (0.00208)	P44* (0.4)

Asterisks were ancillary support staff