

ABSTRACT

Title of Dissertation: UNCOVERING THE HIDDEN POPULATION
IN THE TRANSMISSION OF COVID-19:
ASYMPTOMATIC CASES

Weijun Yu, Doctor of Philosophy, 2023

Dissertation directed by: Associate Professor Quynh Camthi Nguyen
Department of Epidemiology and Biostatistics

This dissertation investigates the impact of international and local mobility of asymptomatic versus symptomatic COVID-19 cases on the pandemic in Hong Kong.

The first manuscript analyzed empirical data from 11,775 confirmed COVID-19 cases in Hong Kong from January 2020 to April 2021, building a retrospective cohort. The results indicated that COVID-19 asymptomatic airport or flight crew were ten times more likely to have inbound air travel history than symptomatic airport or flight crew (adjusted RR=10.00, 95% CI: 4.00–25.00), and the median flight duration of asymptomatic cases was 4.6 person-hours shorter than that of symptomatic cases ($p<0.01$).

The second manuscript presented a social network analysis study that build networks for the three peaks of COVID-19 diagnosis in Hong Kong. The results showed that asymptomatic cases were 1.33 times more likely to be presented in the inbound flight cabin or airport with other COVID-19 cases simultaneously than symptomatic cases (95%CI: 1.21-1.45) at the early stage of

the pandemic. Additionally, the study found that network percolation simulation targeted attacks were more efficient than random failures in dismantling networks with a low level of connectedness.

The third manuscript used geocoded COVID-19 cases' travel records in Hong Kong to conduct a spatial analysis study. The findings indicated that asymptomatic cases visited locations mostly clustered in the southern part of Hong Kong, while symptomatic cases visited locations mostly clustered in the middle and southern parts of Hong Kong. This study also found that Geographically Weighted Regression models performed better among symptomatic cases than asymptomatic cases, and median local travel time was higher ($p < 0.01$) among asymptomatic (68.09 person-minutes) than symptomatic cases (59.46 person-minutes) based on 19,568 Origin-Destination Cost Matrix least-cost paths.

Overall, this dissertation highlights the importance of promoting public health prevention strategies to contain future infectious disease pandemics at the early stage, regardless of the presence of symptoms. Moreover, it suggests that travel restriction may not be effective in dismantling networks with a low-level of connectedness. Local health authorities and policymakers should tailor detection and containment strategies based on spatial variability in different areas.

UNCOVERING THE HIDDEN POPULATION IN THE TRANSMISSION
OF COVID-19: ASYMPTOMATIC CASES

by

Weijun Yu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:

Associate Professor Quynh Nguyen, Chair
Associate Professor Thu Nguyen
Professor Hongjie Liu
Associate Professor Xin He
Professor Jie Chen
Dr. Cheryll Alipio

© Copyright by
Weijun Yu
2023

Acknowledgements

I would like to express my sincere gratitude towards my dissertation committee for their generous support during what was undoubtedly a challenging time. Specifically, I would like to extend my heartfelt thanks to my advisor and committee chair, Dr. Quynh Nguyen, for her unwavering guidance and support throughout my pursuit of a PhD. I am also deeply grateful to the committee members, including Drs. Hongjie Liu, Thu Nguyen, Xin He, Cheryll Alipio, and Jie Chen, for their insightful and invaluable advice on my dissertation.

Furthermore, I would like to acknowledge the contributions of Drs. Michelle Girvan and Daniel Serrano from the COMBINE program at the University of Maryland College Park, who provided me with technical support on network analysis. Also, I am thankful for the significant support and guidance of Drs. Jonathan Resop and Dr. Jack Ma from the Department of Geographical Science at the University of Maryland College Park, who provided me with assistance on spatial analysis and spatial network analysis, respectively.

Finally, I would like to express my appreciation to the Center for Health Protection at the Hong Kong Department of Health for making COVID-19 reports publicly available. Their contributions were essential to the success of this dissertation, and I am truly grateful for their support.

Table of Contents

Acknowledgements.....	ii
Table of Contents.....	iii
List of Tables	v
List of Figures	vi
List of Illustrations.....	viii
Chapter 1: Introduction.....	1
1.1 Background and Rationale.....	1
1.1.1 Literature Review (Chapter 3): Air Travel and COVID-19.....	5
1.1.2 Literature Review (Chapter 4): Network Analysis	8
1.1.3 Literature Review (Chapter 5): Spatial Analysis	10
1.2 Objective/Research Questions	14
1.3 Theoretical/Conceptual Framework.....	16
1.4 Innovation and Significance	21
Chapter 2: Methods.....	24
2.1 Study design and data source (Chapter 3).....	24
2.2 Study design and data source (Chapter 4).....	27
2.3 Study design and data source (Chapter 5).....	31
2.4 Ethics consideration	33
Chapter 3: Associations between COVID-19 Asymptomatic Status and Inbound Air Travel	35
Abstract	36
3.1 Introduction.....	38
3.2 Methods.....	41
3.3 Results.....	45
3.4 Discussion	49
3.5 Conclusion	52
Chapter 4: Social Network Analysis of COVID-19 Transmission among Asymptomatic and Symptomatic Cases in Inbound Air Travel	53
Abstract.....	54

4.1 Introduction	56
4.2 Methods.....	58
4.3 Results	64
4.4 Discussion	73
4.5 Conclusion	77
Chapter 5: Spatial Analysis of Local Mobility for COVID-19 Asymptomatic and Symptomatic Cases in Hong Kong	78
Abstract	79
5.1 Introduction	81
5.2 Methods.....	83
5.3 Results	90
5.4 Discussion	104
5.5 Conclusion	106
Chapter 6: Conclusion.....	108
Bibliography	110

List of Tables

Table 3.1 Selected Policies and Measures by Hong Kong's Government Across the Three Peaks.....	46-47
Table 3.2 Descriptive Statistics of Confirmed COVID-19 Cases in Hong Kong.....	47
Table 4.1 Temporal Stratification of the Three Peak Networks Analysis.....	59
Table 4.2 Descriptive Statistics of the Three Peaks Networks.....	63
Table 4.3 Descriptive Statistics of Nodes' Attributes in the Three Peak Networks...	69
Table 4.4 Baseline and Best-fit Exponential Random Graph Models.....	70
Table 4.5 Odds Ratios for Predictors in Best-fit EGRMs.....	72
Table 5.1 Statistics of Global Moran's I & Average Nearest Neighbor Analysis.....	99
Table 5.2 Explanatory Variables Included in the Final Models.....	100
Table 5.3 R-Squared Values of Best Fit Models: Geographically Weighted Regression vs. Ordinary Least Squares Regression.....	100

List of Figures

Figure 1.1 Conceptualization Directed Acyclic Graph of Manuscript One.....	17
Figure 1.2 Conceptualization Flowchart of Manuscript Three.....	21
Figure 2.1 Undirected Network and Directed Network.....	27
Figure 2.2 Routes for Cases' Travel Records in Hong Kong Road Network.....	32
Figure 3.1 Locally Estimated Scatterplot Smoothing for COVID-19 Cases Situation in Hong Kong (Jan 2020-Dec 2021).....	46
Figure 3.2 Directed Acyclic Graph for Association between COVID-19 Asymptomatic Status and Inbound Air Travel.....	48
Figure 4.1 Network of COVID-19 Cases Traveled to Hong Kong by Air during Peak One (January 2020-May 2020).....	66
Figure 4.2 Network of COVID-19 Cases Traveled to Hong Kong by Air during Peak Two (June 2020-September 2020).....	67
Figure 4.3 Network of COVID-19 Cases Traveled to Hong Kong by Air during Peak Three (November 2020-February 2021).....	68
Figure 4.4. Network Percolation of the Three Peak Networks.....	73
Figure 5.1 Hong Kong's Eighteen Districts.....	85
Figure 5.2 Hong Kong's 154 Tertiary Planning Units.....	86
Figure 5.3 Point Maps of Locations Visited by COVID-19 Cases.....	90-91
Figure 5.4 Choropleth Map of Cumulative Number of Point Features in 154 Tertiary Planning Units in Hong Kong.....	92
Figure 5.5 Kernel Density Map for Cumulative Mobility of Asymptomatic Cases and Point Map for the Highest Density Region.....	94
Figure 5.6 Kernel Density Map for Cumulative Mobility of Symptomatic Cases and Point Map for the Highest Density Region.....	95
Figure 5.7 Kernel Density Map for Mobility during Peak 1: Comparison between Asymptomatic and Symptomatic Cases.....	96

Figure 5.8 Kernel Density Map for Mobility during Peak 2: Comparison between Asymptomatic and Symptomatic Cases.....	97
Figure 5.9 Kernel Density Map for Mobility during Peak 3: Comparison between Asymptomatic and Symptomatic Cases.....	98
Figure 5.10 Choropleth Map of GWR Local R^2 Values.....	101
Figure 5.11 Choropleth Map of GWR Standard Residuals.....	102-103
Figure 5.12 Illustration of Origin-Destination Cost Matrix Analysis.....	104

List of Illustrations

Diagram 2.1 Create Two-Mode Matrix by Affiliation Matrices.....	29
Diagram 4.1 Convert Adjacency Matrix for MPLE.....	64

Chapter 1: Introduction

1.1 Background and Rationale

In December 2019, the first human-to-human transmissible coronavirus disease caused by SARS-CoV-2 (COVID-19) was reported in Wuhan, China (Li et al, 2020a). Since then, three major epidemics of coronavirus have occurred globally: severe acute respiratory syndrome (SARS), Middle East Respiratory Syndrome (MERS), and COVID-19. COVID-19 is reported to be less severe than previous epidemics of coronavirus diseases, such as SARS and MERS (Munster et al, 2020). The World Health Organization (WHO) defines an asymptomatic case as an individual who tested positive for SARS-CoV-2 infection but does not show any symptoms (WHO, 2021). During the SARS outbreak in 2003, few asymptomatic cases were reported (Chan-Yeung & Xu, 2003; Che et al, 2006), and the first reported asymptomatic SARS infection in Hong Kong was a registered nurse who worked in a local hospital that treated approximately one-third of confirmed SARS cases in Hong Kong (Lee et al, 2003a). However, research during the COVID-19 pandemic has shown that asymptomatic cases can transmit SARS-CoV-2 to others (Bai et al, 2020; WHO, 2020). Although only a small number of asymptomatic cases of SARS-CoV-2 become presymptomatic cases with mild to moderate symptoms after a certain period, most asymptomatic cases without comorbidity conditions do not develop symptoms at all (Yu et al, 2020; Nogrady, 2020).

According to Sah and colleagues (2021), who conducted a systematic review of more than 350 studies, asymptomatic cases account for more than one-third of confirmed COVID-19 cases (Sah et al, 2021). Due to the viral transmission of SARS-

CoV-2, accurately assessing the transmissibility of asymptomatic cases presents challenges (Nogady 2020). Without widespread mandatory or voluntary testing, identifying all asymptomatic cases in a given region may be impossible.

Hong Kong, a special administrative region of China, is a densely populated international hub, with approximately 200,000 individuals residing in small apartments less than 60 square feet in size (To & Yuen, 2020). However, despite this, Hong Kong reported one of the lowest prevalence rates of COVID-19 compared to other similarly densely populated developed regions (To and Yuen, 2020). One previous study suggested that neighborhoods with lower urban development, such as those that are purely residential with fewer crosswalks and sidewalks, have fewer reported COVID-19 cases (Nguyen et al 2020). Hong Kong health authority has experience dealing with severe coronavirus outbreaks, including SARS 2003 outbreak, and doctors and scientists in Hong Kong have conducted high-quality research on coronavirus since then (Lee et al, 2003b; Tsang et al, 2003). The successful containment of SARS 2003 allowed Hong Kong's local government and residents to respond quickly to COVID-19, even before the first local COVID-19 case was confirmed (Wong et al, 2020). Hong Kong citizens actively implemented public health prevention strategies, such as wearing masks, washing hands frequently, and practicing social distancing (Law, Leung, & Xu, 2020; Li et al, 2020b). It is worth noting that some individuals may become reinfected with COVID-19 and not present any symptoms. The first reported reinfection of COVID-19 globally was a Hong Kong resident who traveled between Europe and Hong Kong, recovered from his first infection with mild symptoms, and then tested positive again with high viral load but with no symptoms after 4.5 months

when he took a mandatory COVID-19 test required for everyone arriving in Hong Kong (Parry 2020).

Understanding the distribution of the COVID-19 cases in a population is essential, particularly when it comes to the frequency and patterns of asymptomatic cases, given their potential to spread SARS-CoV-2 and exacerbate the pandemic. However, despite the COVID-19 epidemic lasting for three years, the role of asymptomatic cases in the spread of the virus and their contribution to the pandemic remains incompletely understood. A significant obstacle of studying this population is the difficulty of tracking them due to the absence of symptoms. Fortunately, since the beginning of the pandemic, the Hong Kong Department of Health has been providing the public with detailed information on the symptomatic status of testing-confirmed cases (asymptomatic or symptomatic with a specific date of onset). This information also includes COVID-19 cases' local and international travel information in the 14 days preceding the diagnosis of SARS-CoV-2 infection (Center of Health Protection HK, 2022). Additionally, data from mandatory viral tests are available from all passengers arriving in Hong Kong since early 2020 (Center of Health Protection HK, 2020). These rich datasets provide an excellent opportunity to investigate the travel patterns of COVID-19 cases based on their symptomatic status.

Airborne infectious disease has been found to be linked to air travel activity (Mangili et al, 2015). The SARS outbreak in 2003 clearly demonstrated the potential for coronavirus to spread via international flights (WHO, 2003), infecting 8,096 individuals across 29 countries in Asia, Europe, North America, and South America (CDC, 2017). In 2012, MERS infected more than 2,442 individuals with a travel history

to the Arabian Peninsula (WHO, 2019) across 27 countries in the Middle East, South Asia, Africa, Europe, and North America (CDC, 2014). As of writing, COVID-19 has infected over 0.3 billion individuals worldwide, including in Antarctica (BBC,2020). Only two countries - Turkmenistan and Democratic People's Republic of Korea – have yet to report COVID-19 data (WHO, 2022). Ma et al (2021) conducted a meta-analysis study that included 95 studies and 29 million participants who took viral tests, suggesting that 2% of tested participants and 53% of positive participants who took a flight or cruise were asymptomatic.

Close contact with COVID-19 is defined as “someone who was less than six feet away from an infected person for a cumulative total of 15 minutes or more over a 24-hour period” (CDC, 2022). However, commercial flights typically last much longer than 15 minutes and can last several hours, especially for international flights. Therefore, flying with infected people increases the risk of COVID-19 infection, particularly when sitting next to a COVID-19 case (Khanh et al, 2020). Furthermore, the flight cabin is an indoor environment with poor ventilation and limited space between passengers. However, it is currently unknown whether the frequency of flying in airplanes differ between asymptomatic cases and symptomatic cases prior to diagnosis. Additionally, other public community settings, such as restaurants (Lu et al, 2020), apartments (CDC, 2021), and churches (James et al, 2020), have reported elevated risks of COVID -19 infection. However, no studies have yet examined the differences between asymptotic and symptomatic cases' travel histories.

This dissertation aims to address existing knowledge gaps by investigating several key factors related to COVID-19 infection. Firstly, it will explore potential

associations between COVID-19 asymptomatic status and inbound air travel history, as well as flight duration. Secondly, it will examine network structures of empirical networks to capture the temporal changes of asymptomatic and symptomatic COVID-19 cases in inbound air travel to Hong Kong, estimate the probability of links in the networks, and simulate the impact of travel restrictions on networks in flight cabins and airports. Finally, this dissertation will examine differences in spatial density distribution and travel time between asymptomatic and symptomatic cases' local mobility prior to diagnosis within Hong Kong territory and predict cases' local mobility using Hong Kong's spatial socioeconomic factors.

1.1.1 Literature Review (Chapter 3): Air Travel and COVID-19

The impact of air travel on the global COVID-19 pandemic is still a topic of debate. Despite a 43% reduction in air travel compared to passenger volumes in 2019 (Bielecki et al, 2021), daily COVID-19 incidence rates continued to rise globally, even with several countries implemented travel restrictions to China from January to February 2020 (Lau et al, 2020 a), and the lockdown of Wuhan, China, which may lead to temporary significant decrease in COVID-19 cases (Lau et al, 2020 c). In a mathematical modeling study, Thompson and team (2018) argued that modern travel can actually help increase immunity instead of exacerbating a pandemic, since it may increase population immunity by increasing viral infections. However, a study by Watts and colleagues (2020) examined human mobility through real-time GPS mobile device from Arizona to other regions in the USA, and their results suggested that human mobility contributes to the continuing COVID-19 pandemic in the US.

In response to the COVID-19 pandemic, various strategies have been implemented to ensure safe air travel and reduce the spread of SARS-CoV-2 in flight cabins (IATA, 2020). These includes measures such as mask wearing, hand hygiene, rapid testing, and social distancing, which have been shown to be effective in reducing infections on flights (Khatib et al, 2020; Bielecki et al, 2021). However, some approaches, such as temperature measuring and PCR tests prior to boarding, have been found to be less effective (Bielecki et al, 2021). For example, Hoehl and his team (2020) reported seven COVID-19 cases, including one asymptomatic case, among 102 passengers on a commercial flight from Tel Aviv, Israel to Frankfurt, Germany in March 2020. No prevention strategies, such as mask wearing, were applied during the flight (Hoehl et al, 2020), highlighting the importance of implementing such measures. Even with mandatory masking, indirect contact through shared spaces such as toilets can still be risky. Studies have shown that transmission can occur during mealtimes or when using toilets on flights (Bae et al, 2020).

Existing literature provides mixed findings on transmission of SARS-CoV-2 in flight cabins. While Jia et al. (2021) concluded that transmission of SARS-CoV-2 is unlikely to occur in flight cabins but rather within households, others have suggested that in-flight transmission of the virus is a real concern (Swadi et al, 2021; Toprani, Scheibler & Nagel, 2020). Barnett and Fleming (2020) estimated the risk of death from COVID-19 infection on a two-hour flight to be higher than the risks posed by air crashes or terrorism. Chen et al. (2020) found an overall attack rate of 4.8% for COVID-19 on a flight from Singapore to Hangzhou, China, while Khanh et al. (2020) identified in-flight transmission of the virus in a 10-hour flight from London to Hanoi. Similarly,

Pavli et al. (2020) found five probable cases of in-flight transmission of SARS-CoV-2 among 18 flights from Italy, Israel and UK to Greece from February to March 2020.

The existing literature on the risk of in-flight transmission through close physical proximity is inconclusive. For instance, Hertzberg and colleagues (2018) conducted a study on in-flight influenza transmission and found that the risk of infection was low for passengers who did not sit close to infected patients on ten transcontinental flights with a flight duration of less than 5 hours. However, Leitmeyer and team members (2016) concluded from a systematic review of fourteen publications that in-flight transmission of flu occurred even among passengers who sat further than two rows from the infected patients. Olsen and colleagues (2003) also reported instances of in-flight transmission of SARS among passengers who sat further than three rows from infected passengers.

Limited evidence exists regarding in-flight transmission of SARS-CoV-2, as only a small number of publications are available (Kelly, Bambury & Boland, 2021). Moreover, the flight duration of the flights covered in existing literature is sporadic, ranging from 2 hours to over 10 hours. Notably, no existing studies have explored the association between COVID-19 asymptomatic status and inbound air travel. Additionally, no research has collected empirical data on flight duration across different continents and examined the difference of flight duration between asymptomatic cases and symptomatic patients.

1.1.2 Literature Review (Chapter 4): Network Analysis

Network analysis has been increasingly applied to public health and air transportation studies, revealing findings that traditional epidemiological studies often miss. From sexually transmitted diseases such as syphilis (Rothenberg et al, 1998) and HIV (Klov Dahl et al, 1994) to issues such as obesity (Christakis & Fowler, 2007), drug abuse (Valente, Mouttapa & Gallaher, 2004), and smoking (Christakis & Fowler, 2008), network studies have proven to be valuable in identifying key factors in disease transmission. Network analysis has also been utilized to examine the resilience of air transportation networks (Sun & Wandelt, 2021), with studies showing that airports in East Asia are at high risk due to their high population density and flight volumes (Daon, Thompson & Obolski, 2020). In the context of COVID-19, network analysis has been used to evaluate human mobility during early outbreaks in China (Chen et al, 2020) and to suggest that connectivity of airports played a significant role in COVID-19 outbreaks among Pacific Small Island Developing States (Leal Filho et al, 2020).

Rewiring network is a method used to examine the robustness of a network, and one such approach is percolation. Percolation involves removing some nodes with connected links from a network, which can be done randomly or by targeting specific nodes with the highest degrees (Newman, 2018; Menczer, 2020; Barabási, 2016). This is known as random failure or targeted attack, respectively. Researchers have found that while US airport networks are robust to errors, they are very vulnerable to targeted attacks (Chi et al, 2004). In Pakistan, centrality measures such as degree, betweenness, and closeness have been used to identify important nodes in airport networks (Malik et al, 2019).

While interviewing participants is the main method of collecting social network data, some studies have used archival secondary data sources. For example, Breiger (1974) used nine months of newspaper reports on female afternoon tea socializing events in 1941 to construct a two-mode network. Harris et al (2009) analyzed 1,877 publications on secondhand smoke from 1965 to 2005 to study a 40-year citation network study.

Several studies have reported in-flight transmission of SARS-CoV-2 by symptomatic COVID-19 patients (Ng et al, 2020; Pavli et al, 2020), while only a few mentioned asymptomatic cases (Murphy et al, 2020; Yang et al, 2020; Bae et al, 2020). Two macro population-level studies have explored the correlation between flight volume and COVID-19 cases. A strong positive correlation ($r=0.8$) was found between international flight passenger volume and globally reported COVID-19 cases (Lau et al, 2020 b; Ma et al, 2021). Additionally, the number of COVID-19 cases is positively correlated with a nation's GDP, and transportation is the main cause of outbreaks in developed countries (Ma et al, 2021). However, individual-level studies are lacking to examine networks, particularly among asymptomatic cases.

The risk of COVID-19 infection due to seat proximity is controversial in existing literature. While some studies suggest that sitting within two rows of an infected person increase the risk of airborne infectious diseases transmission (Hertzberg et al, 2018; Bielecki et al, 2021; Chen et al, 2020), such as SARS (Olsen et al, 2003) and COVID-19 (Bielecki et al, 2021; Chen et al, 2020). In contrast, other research indicates that COVID-19 transmission can occur beyond two rows, as it did with SARS and influenza (Schwartz et al, 2020). Additionally, factors beyond seat

proximity, such as interactions in departure gates, movements within the cabins (Hoehl et al, 2020), and use of lavatories (Bae et al, 2020), can contribute to COVID-19 transmission. Hence, focusing solely on seat proximity leaves a critical knowledge gap in existing publications.

1.1.3 Literature Review (Chapter 5): Spatial Analysis

Spatial point data is crucial for exploring and understanding geographic patterns of disease transmission (Diggle, 2013). To obtain this data, geocoding is necessary, which assigns a physical address with a latitude and longitude (Waller & Gotway). Moran's I is a statistics measure that helps identify spatial autocorrelation and similarities between nearby point locations (Moran, 1950; Li, Calder & Cressie, 2007). Cluster analysis, another method for examining the relationships of point events (Edwards & Cavalli-Sforza, 1965), has been widely used in spatial epidemiology to identify patterns of variation in geographic data (Jacquez, 2008). Spatial analytical tools like Moran's I and cluster analysis have been combined with network analysis approaches to explore disease transmission (Firestone et al, 2011; Emch et al, 2012). Mapping spatial point data can also help researchers visualize clusters of geographic events associated with a particular disease (Waller & Gotway, 2004). Recent studies have utilized spatial autocorrelation analysis (Moran's I) to identify regional clusters of COVID-19 in Europe (Amdaoud, Arcuri & Levratto, 2021), as well as clusters of COVID-19 testing and result rates in New York City (Cordes & Castro, 2020).

Spatial analysis has a long history in epidemiology, particularly in infectious disease studies. One of the earliest examples of the application of spatial analysis in

epidemiology was the famous 1854 cholera outbreak study conducted by Dr. John Snow in London (Snow, 1855). Monmonier aggregated the cholera data from Snow's study to map the point locations of cholera deaths in three districts of London, revealing that the Broad Street cluster was less significant compared to other clusters (Monmonier, 2018). Maxcy (1967) examined data on food suppliers such as groceries and restaurants to demonstrate spatial patterns in Brill's disease outbreak in the southeastern United States, since these suppliers could be venues for the rodents to transmit typhus. Xiang and colleagues (2000) merged spatial and public health data from various sources to investigate associations between pesticide use and low birth weight in Colorado, using geographical information system (GIS) techniques. Mahay and Laumann (2004) mapped data from the Chicago Health and Social Life Survey collected from 1995 to 1997 and identified two potential bridge communities that connected sexual partners from different areas of Chicago to spread HIV.

Combining spatial analysis with network analysis has become increasingly popular in health and economic organizational studies. For instance, Ronald and colleagues conducted a longitudinal study on HIV transmission in communities in Nang Rong, Thailand since 1984 using social network analysis and geographical information system (GIS) methods (Morris, 2004). Another group of scientists employed spatial analysis by GIS to investigate potential barriers, such as rivers or streams, which could impede social and economic network ties among villages in the same area of Nang Rong, Thailand (Faust et al, 2000). In an organizational network study, Sailer and McCulloh (2012) found that spatial distance and office layouts' configuration are important in predicting social ties in an organization. Bolton and

colleagues (2012) suggested that contact in households and work environments are crucial to increase disease transmission risk. Waller and team members (1992) analyzed leukemia data with census tract in New York state and mapped the leukemia incidence in major cities and found that higher rates of leukemia cases occurred in the large cities.

Spatial analysis with network analysis has been increasingly utilized in existing COVID-19 literature. For instance, Jo and colleagues (2021) concluded that population density is associated with spread of COVID-19 in South Korea, and they employed transport Origin-Destination data to conduct network analysis. Pedrosa and Albuquerque (2020) examined hospital intensive care beds for COVID-19 treatment as spatial point events in Ceará, Brazil using Moran's I, and they identified positive spatial autocorrelation ($I=0.66$). Similarly, researchers from Finland also used Moran's I to identify spatial autocorrelation in Helsinki and globally, and they found that clusters of high risk for SARS-CoV-2 infection are associated with vulnerable socioeconomic neighborhoods in eastern Helsinki. They also concluded that socioeconomic factors, such as household income, and demographic factors, such as citizenship, play an important role in COVID-19 infection (Siljander et al, 2022).

Cluster analysis has proven to be a fruitful approach in infectious disease transmission research. For instance, Giebultowicz and colleagues (2011) found that cholera clustering is associated with household social clustering in rural Bangladesh. Similarly, Cordes and Castro (2020) showed that white population with high socioeconomic status was associated with a low rate of positive results for COVID-19, while black population with no medical insurance were associated with a high rate of

positive results. In Europe, one study found that senior citizens, unemployment rate, and socioeconomic factors are associated with a high COVID-19 related mortality rate, while increasing the number of medical employees and hospital access is associated with lower death rate (Amdaoud, Arcuri & Levratto, 2021).

Other spatial analysis approaches such as kernel density analysis and geographically weighted regression, have also been utilized in some infectious disease literature. Kernel density analysis estimates a probability density function, which qualifies the likelihood of observing an event at a specific geographic location (Waller & Gotway, 2004). Fu and Zhai (2021) used Geographical and Temporally Weighted Regression (GTWR) models to examine spatiotemporal relationship between sedentary behaviors and social vulnerability in New York City.

Origin-Destination (OD) Matrices have been used in a number of COVID-19 studies. For instance, Gómez and colleagues (2019) proposed that origin-destination information can be used to estimate an epidemic spreading in a population (Gómez et al, 2019), while Pepe and colleagues (2020) captured population mobility patterns among Italian provinces using weighted daily OD Matrices. In another study, a group of Italian scientists measured noise pollution in Rome's Road Network during the COVID-19 pandemic and used OD Matrices to simulate Floating Car Data (Aletta et al, 2020). This data source captures real-time traffic information by using GPS or cell phones to locate vehicles within the road network of interest (Altintasi, Tuydes-Yaman & Tuncay, 2017). In Japan, a group of scientists used OD Matrices to simulate a highly reliable and applicable prediction model for COVID-19 (Jiang et al, 2021). Another group of researchers also recommended OD Matrices to explore population movement

during the COVID-19 pandemic (Mansour, Kajtazi & Ghazawneh, 2022). In Mexico, Prieto and team members (2022) used OD Matrices to simulate population mobility and forecasted the spread of COVID-19 in Mexico City.

Most existing network studies have limited or no data regarding the geographic locations of network nodes, making it difficult to model social ties with spatial data (Nyerges, Couclelis & McMaster, 2011). Additionally, Rosensteel et al. (2021) identified significant spatial heterogeneity of influenza epidemics across different geographic regions and seasons, making predicting future regions of flu epidemics challenging.

1.2 Objective/Research Questions

This dissertation investigates how the international and local mobility of asymptomatic COVID-19 cases contributes differently to the pandemic in Hong Kong compared to symptomatic cases. This dissertation is structured into three specific aims:

Specific Aim 1 (Manuscript 1): This retrospective cohort study investigates the association between COVID-19 asymptomatic status and inbound air travel history, as well as flight duration. The study will answer the following three research requests:

- a. Are asymptomatic COVID-19 cases more likely to take inbound flights than symptomatic COVID-19 cases?
- b. Do asymptomatic and symptomatic COVID-19 cases who took inbound flights differ in flight duration?

- c. Do demographic characteristics and underlying health conditions of COVID-19 cases modify the association between asymptomatic status and inbound air travel history 14 days prior to diagnosis?

Specific Aim 2 (Manuscript 2): This social network analysis study investigates network structures of empirical networks in order to capture the temporal changes of asymptomatic and symptomatic COVID-19 cases in inbound air travel to Hong Kong. This study will address the following three research questions:

- a. Are there differences in network measures (i.e., degree and density) between asymptomatic and symptomatic cases in these networks?
- b. Do the links in the networks vary according to asymptomatic status, sex, age, or other factors?
- c. Do random failures and targeted attacks differ in their efficiency for dismantling COVID-19 cases networks in flight cabins and airports?

Specific Aim 3 (Manuscript 3): This spatial analysis study examines spatial differences between asymptomatic and symptomatic cases' historical local mobility records 14-day prior COVID-19 diagnosis in Hong Kong.

- a. Do locations visited by asymptomatic and symptomatic cases in Hong Kong follow complete spatial randomness?
- b. Are COVID-19 cases' local mobility spatially associated with population density, medical center locations, and transportation in Hong Kong?
- c. Do travel time estimated by least-cost paths by Origin-Destination (OD) Cost Matrices differ between asymptomatic and symptomatic cases?

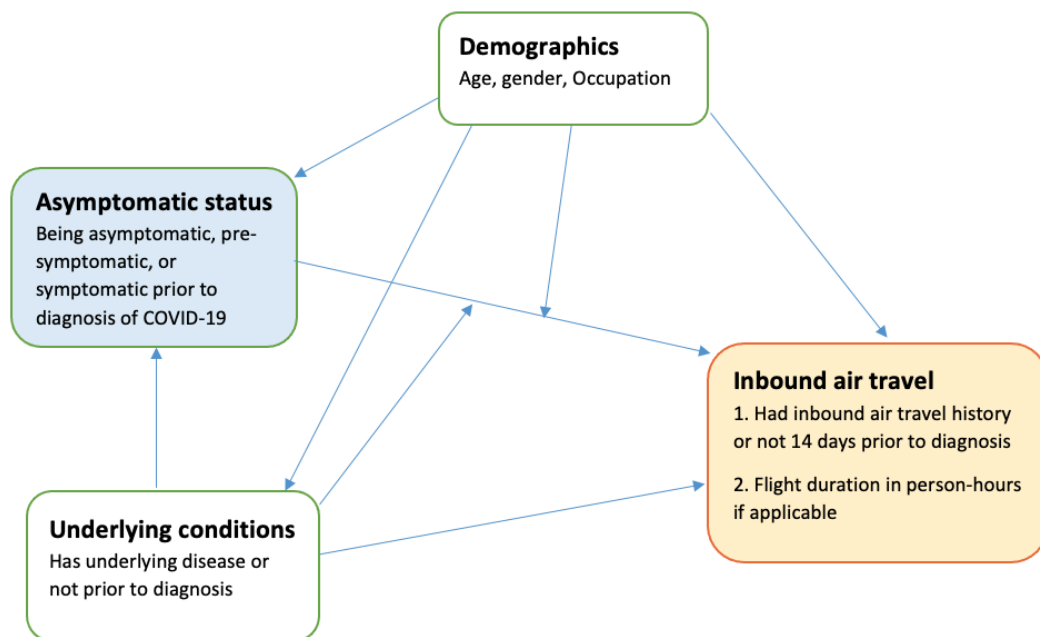
1.3 Theoretical/Conceptual Framework

Manuscript one (Chapter 3) aims to investigate the association between COVID-19 asymptomatic status and inbound air travel to Hong Kong. The hypothesis is that individuals who are asymptomatic are more likely to travel to Hong Kong via inbound flights prior to their diagnosis, and their flight durations are longer compared to their symptomatic counterparts. It is further hypothesized that demographic factors and underlying conditions of COVID-19 patients may modify or confound this association. For example, individuals without underlying conditions who are asymptomatic may be more likely to travel to Hong Kong compared to those who are symptomatic with underlying conditions. Age may also play a role, as younger individuals with asymptomatic status may be more likely to travel than older individuals. Moreover, occupation and underlying health conditions could potentially modify or confound the association of interest. For instance, individuals who work in specific occupations such as flight crews are more likely to be asymptomatic and associated with inbound air travel compared to the general population, since only healthy individuals without severe underlying conditions are recruited as flight crews.

Figure 1.1 presents the directed acyclic graph that conceptualizes of this study. To measure the association of interest, we utilize the risk ratio as suggested by theoretical considerations in previous studies. The risk ratio is “a ratio of the probability of an event occurring in the exposed group to the probability of the same event occurring in the non-exposed group” (Tenny & Hoffman, 2022). If the incidence of an outcome in a study population is higher than 10%, it is considered a common outcome (Zhang, 1998). In our study, inbound air travel history 14 days prior to diagnosis has a

prevalence of over 10% among the participants, making it a common outcome. Therefore, we will use risk ratio as our measure of the association, which is recommended for studies with common outcomes rather than the odds ratio generated by logistic regression, as the latter can exaggerate the association (McNutt, 2003; Zhang, 1998). Additionally, age, sex, and underlying conditions, such as obesity and cardiovascular disease, are known confounders that can distort the association between COVID-19 and other diseases, such as diabetes (Tadic, Cuspidi & Sala, 2020) and HIV (Lesko & Bengtson, 2021). In contrast, sex has been suggested as a modifier in a COVID-19 study (Haitao et al, 2020), and occupation, family income, and socioeconomic status have been identified as modifiers in a prospective cohort study exploring the association between climate and pediatric infectious disease in India (Singh et al, 2021).

Figure 1.1 Conceptualization Directed Acyclic Graph of Manuscript One (Chapter 3)



Manuscript two (Chapter 4) aims to address the gap in traditional epidemiological models of infectious disease (i.e., Susceptible-Infected-Recovered model or Susceptible-Infected model), which often ignore the social structure of a population when predicting epidemics (Barabási, 2017). Incorporating social network components into infectious disease models have been shown to be effective in optimizing public health infectious disease control (Dimitrov & Meyers, 2010) and predicting the impact of infectious disease pandemics (Read, Eames & Edmunds, 2008). For example, large-scale social network mobility models were used to simulate stochastic forecasting on global H1N1 influenza (Balcan et al, 2009; Tizzoni et al, 2012) and severe acute respiratory syndrome (SARS) (Brockmann & Helbing, 2013) pandemics.

A network consisting of two groups of nodes is referred to as a bipartite network ((Barabási, 2016; Newman, 2018; Menczer, 2020). In the past, a method to examine bipartite networks has been proposed, in which the bipartite network is broken down into two unipartite networks and analyzed separately (Robin 2015). However, this approach has limitations as it only considers the bipartite network as two separate unipartite networks and may miss the potential interactions between the two groups of nodes (Robin 2015). To fully leverage the benefits of both bipartite and unipartite networks, this study builds bipartite networks and examines unipartite structures.

Centrality measures are fundamental in social network analysis (Freeman, 1978; Borgatti & Everett, 2006). Degree, betweenness, and closeness are the most commonly used measures for centrality (Valente, 2010). While eigenvector centrality can be affected by missing data, degree is robust to such data and is therefore the most

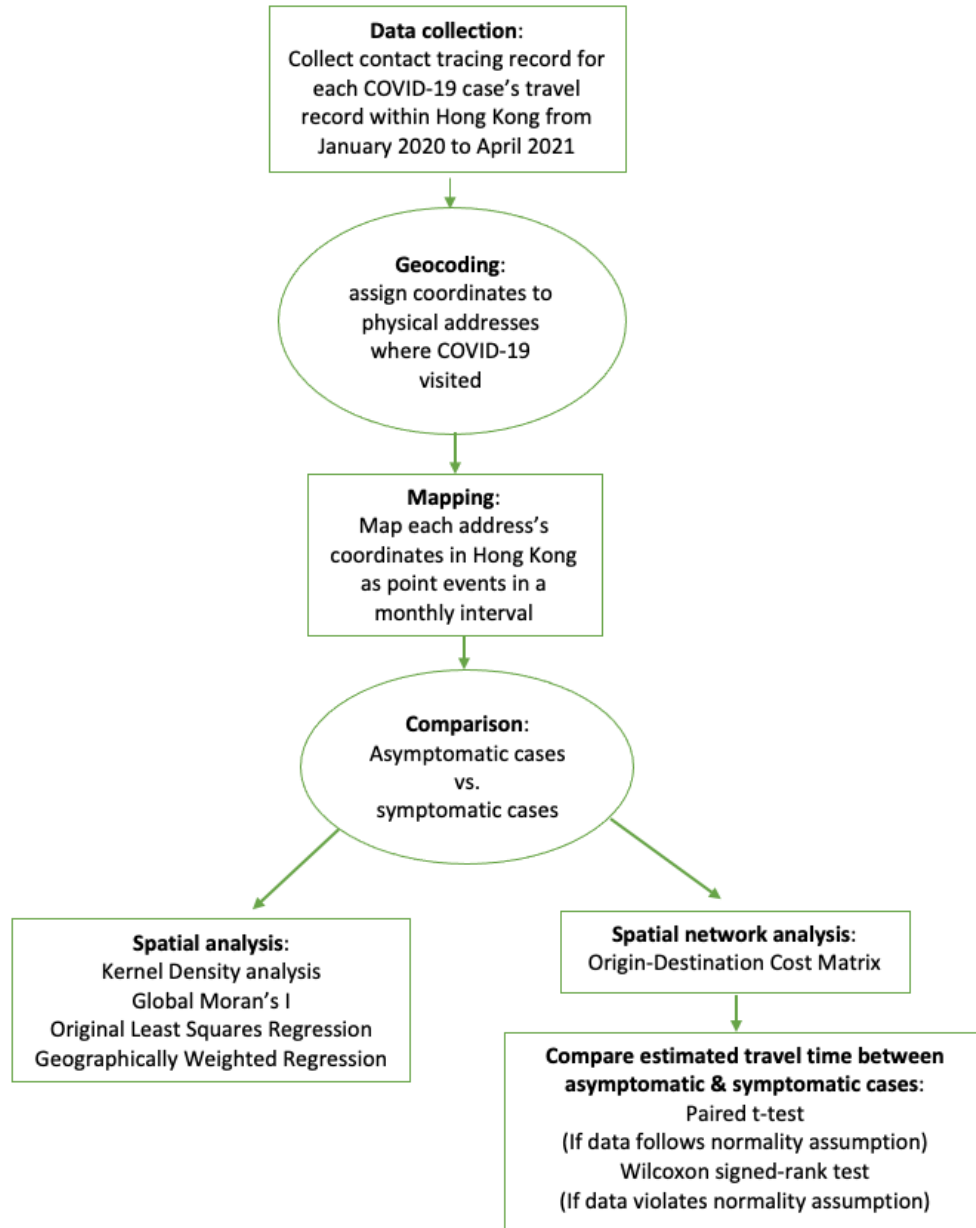
recommended measure in networks with missing data (Valente, 2010). Density reflects the level of clustering in a network, and such clustering can either facilitate or hinder the spread of infectious diseases within and between subgroups, respectively (Valente, 2010). Hence, examining density in social networks is crucial for studies on public health networks. Consistent with the theoretical guidance from previous infectious disease literature, this study focuses on degree and density of the networks.

The Maximum Pseudo-Likelihood Estimation (MPLE) approach is widely used in statistical analysis for network studies, and it's especially accurate when links in networks do not depend on each other (Valente, 2010). On the other hand, the Exponential Random Graph Model (ERGM) is one of the most popular statistical models for network studies (Valente, 2010; Schmid & Desmarais, 2017). To estimate the parameters for ERGM, the Maximum Likelihood method Markov Chain Monte Carlo (MCMC) is accurate (Valente, 2010). However, Maximum Pseudo-Likelihood Estimation (MPLE) has been shown to be an effective approach to estimate network links and is better suited for networks with a large number of nodes than Maximum Likelihood Estimation (MLE) using Markov Chain Monte Carlo (MCMC) (Schmid & Desmarais, 2017). Thus, this study will estimate links in the networks using MPLE. Estimating links is an important procedure for this study because it reveals whether an association exists between nodes due to asymptomatic status, sex, or age. To test the robustness of networks, network percolation has been applied, and previous studies have used random failures and targeted attacks to examine flight networks. For instance, to dismantle a transmission network, it may be more efficient to vaccinate individuals with the highest degree in the disease contact network than to vaccinate the

population totally at random (Newman, 2018). Therefore, this study will simulate and compare the impact of random failures and targeted attacks approach to dismantle the empirical networks.

Manuscript three (Chapter 5) is based on theoretical findings from existing literature and preliminary analysis. Previous studies have shown that public transport is considered high-risk for transmission of SARS-CoV-2 due to limited space (Edwards et al, 2021), with bus or metro riders in China being at high risk of COVID-19 infection (Luo et al, 2020; Shen et al, 2020; Zheng et al, 2020). As public transport is a dominant mode of commuting in Hong Kong (Tang & Lo, 2008), this study aims to examine the spatial differences between asymptomatic and symptomatic cases' historical local mobility records 14 days prior to COVID-19. **Figure 1.2** illustrates the conceptualization flowchart of this study. Additionally, the preliminary analysis using the Average Nearest Neighbor method indicated significant clustering of point locations visited by asymptomatic cases in southern Hong Kong, specifically in the districts of Yau Tsim Mong, Central and Western, and Wan Chai, which have high population density numerous restaurants, shopping malls, hospitals, and office buildings (Hong Kong Tourism Board, 2022). These preliminary findings suggest that point locations are unlikely to be randomly distributed and may be associated with the independent variables of interest. Furthermore, with 11,976 routes estimated by Origin-Destination (OD) Matrix and travel time calculated for each route, this study will use spatial analysis and spatial network approaches to examine COVID-19 cases' local mobility in Hong Kong, which is deemed feasible by the preliminary results.

Figure 1.2 Conceptualization Flowchart of Manuscript Three (Chapter 5)



1.4 Innovation and Significance

Manuscript one (Chapter 3) presents an innovative first study that extensively examines the association between COVID-19 asymptomatic status and transmissibility of asymptomatic cases during inbound flight to Hong Kong. This study utilizes

historical data from 11,775 COVID-19 cases, hundreds of flights, and several departure airports. The investigation of the association is based on empirical data and controlled for potential confounders, filling a knowledge gap left by simulation or modeling studies. In addition, this study expands the public's understanding of travel behaviors and flight duration between asymptomatic and symptomatic cases. The results of this study reveal whether the absence of symptoms encourage asymptomatic cases to travel or whether symptomatic patients continue to travel due to mistaking symptoms for other common respiratory illnesses prior to diagnosis. As this study highlights the heterogeneity of the airports globally, it examines not only flights departing from Asia but also other international flight hubs such as Europe and America. The findings of this study are significant and valuable for policymakers in designing effective health policies to curb COVID-19 or other respiratory infectious disease pandemic in the future. The exclusion of outbound flights from this study is due to the Hong Kong Department of Health's inability to collect individual-level data from all outbound passengers, resulting in a loss of follow-up.

Manuscript two (Chapter 4) is an innovative paper as it examines the international mobility of COVID-19 cases by analyzing empirical networks among cases on international flights and airports and explores temporal changes during the three pandemic peaks in Hong Kong. Network analysis captures relationships between cases at an individual-level, addressing gaps in traditional epidemiological studies. The study's focus on the network structure of agent-based COVID-19 cases' international mobility, and its role in explaining why travel restrictions failed to contain the global pandemic is both innovative and informative. Travel restrictions are simulated by

network percolation for the three peaks, providing decision-makers with valuable evidence to consider when implementing public health strategies and travel restrictions in future pandemics.

Manuscript three (Chapter 5) is an innovative study designed to reveal the local mobility of asymptomatic COVID-19 cases in Hong Kong 14 days prior to diagnosis and compare it with their symptomatic counterparts, using empirical agent-based data. This study's significance lies in its ability to generate informative findings regarding population mobility before COVID-19 diagnoses and spatial differences between asymptomatic and symptomatic cases, which is critical to better understand why we failed to contain COVID-19 at the early stage of the pandemic globally. Longer travel time implies a higher risk of spreading SARS-CoV-2 in communities. Moreover, this study captures the temporal change over the three peaks of the pandemic for one and half years in a densely populated urban region, Hong Kong. Therefore, the study's results provide critical suggestions to policymakers that spatial variation in different communities should be considered when implementing public health policies. These findings can be generalized to other similar transportation hubs in the world.

Chapter 2: Methods

This chapter provides a detailed description of the methods used in Chapter 3 to 5.

2.1 Study design and data source (Chapter 3)

This study is a retrospective cohort study that uses empirical COVID-19 surveillance data in Hong Kong to examine the association between COVID-19 asymptomatic status (asymptomatic or symptomatic) and air travel (inbound air travel 14 days prior to diagnosis and flight duration) in Hong Kong. The data was obtained from publicly available reports by the Center for Health Protection in Hong Kong Department of Health. These reports contained detailed, de-identified information about COVID-19 cases and were collected from two websites: the daily situation reports of COVID-19 from: <https://www.coronavirus.gov.hk> and the contact tracing reports of each case from: <https://wars.vote4.hk/en/cases>. The data were merged using each case's unique tracking ID.

Data retrieved from daily situation reports includes information on each case's *age*, *sex*, *asymptomatic status*: 1) asymptomatic or 2) symptomatic with a specific date of onset, *residency status*: 1) Hong Kong resident or 2) non-Hong Kong resident, *treatment status*: 1) hospitalized, 2) discharged, 3) no admission to hospitalization, 4) deceased, and *case classification*: 1) imported case, 2) local case, 3) epidemiologically connected with imported case, 4) epidemiologically connected with local case.

Data retrieved from the contact tracing reports includes information on each case's *brief medical history*: 1) with underlying condition, 2) with good health, 3) unknown, *occupation*: some cases' occupations were listed (e.g. doctor, domestic

worker, flight attendant, sailor etc.), *flight record* if the case took inbound flight 14 days prior to diagnosis: 1) flight number with airline initials for some cases, 2) Departed country or city if specific flight information is missing.

Eligibility Criteria

This study includes a total of 11,775 COVID-19 cases diagnosed between January 2020 and April 2021, during which there were three distinct peaks of cases diagnosis. Data beyond April 2021 were not included. A detailed LOESS figure illustrating these peaks is presented in Chapter 3. The analysis specifically focused on confirmed COVID-19 cases that had a history of inbound air travel to Hong Kong within 14 days prior to diagnosis during the specified time period.

Important Variables

Inbound air travel history to Hong Kong was measured by examining the travel history of all COVID-19 cases 14 days prior to diagnosis. Cases were labeled “1” for they had traveled to Hong Kong by air during that period and “0” if they had not.

Flight duration was measured in person-hours for cases that had traveled. Of the cases included in this variable, 79.5% had specific flight number and travel date records, 1.6% had departure cities records, and 18.9% have departure countries records.

Asymptomatic status of COVID-19 cases was determined by collecting daily situation reports from the Center for Health Protection in Hong Kong Department of Health and reviewing medical records provided by medical care providers. Cases were categorized as asymptomatic, symptomatic (with a known date of symptoms onset), or unknown.

Underlying health condition(s) were determined from cases' contact tracing reports and classified as either no underlying conditions or with underlying conditions such as cardiovascular disease or diabetes.

Age, collected by The Center for Health Protection in Hong Kong Department of Health, was the biological age of cases.

Sex was the biological sex of cases (female or male) and was collected by The Center for Health Protection in Hong Kong Department of Health.

Occupation was categorized as either airport or flight crew (e.g., flight attendant, pilot, airport staff) or non-airport or flight crew (all other jobs outside the airport or airline industry) based on cases' reported job titles in contact tracing reports.

Assessment of Potential Bias

The Center for Health Protection in Hong Kong Department of Health obtained records of all the cases' travel history 14 days prior to diagnosis from objective sources such as immigration department, airport, and airlines to avoid measurement issues caused by recall bias from cases' self-reported data. While measurement issues may not arise for cases with available flight number or departure city record, potential measurement issues are likely for cases with only departure country records. For such cases, flight duration can only be estimated rather than exactly recorded. For instance, if a case is recorded as traveling from the USA to Hong Kong without information about the flight number or departure city, the median flight duration between the longest (e.g., departed from New York City) and the shortest (e.g., departed from Seattle) historical flight itineraries departing from the USA and arriving in Hong Kong would be used to estimate the flight duration.

The study examined all the COVID-19 cases' demographic characteristics including age, sex, and occupation (e.g., flight crews or other occupations), as well as underlying conditions for potential interaction effects (effect modification). For variables not identified as potential modifiers, they were examined as potential confounders.

2.2 Study design and data source (Chapter 4)

This empirical social network analysis study aims to capture the temporal changes in COVID-19 cases networks on inbound flights during three peaks of COVID-19 in Hong Kong from March 2020 to April 2021. All networks constructed in this study are undirected networks, as depicted in **Figure 2.1**, with COVID-19 cases represented as nodes and unweighted links indicating shared flight cabins or airports. It is worth noting that the length or thickness of the links has no significance in this study, and the link's length is not based on Euclidean distance. Therefore, there is no difference in interpretation of longer or shorter links in this study.

Figure 2.1 Undirected Network and Directed Network



This study utilizes the same data source as the first manuscript (Chapter 3), with a total of 1,934 COVID-19 cases included for network analysis across the three

peaks. Detailed information regarding the three peaks and inclusion criteria of network nodes can be found in the methods section of Chapter 4.

Network Creation

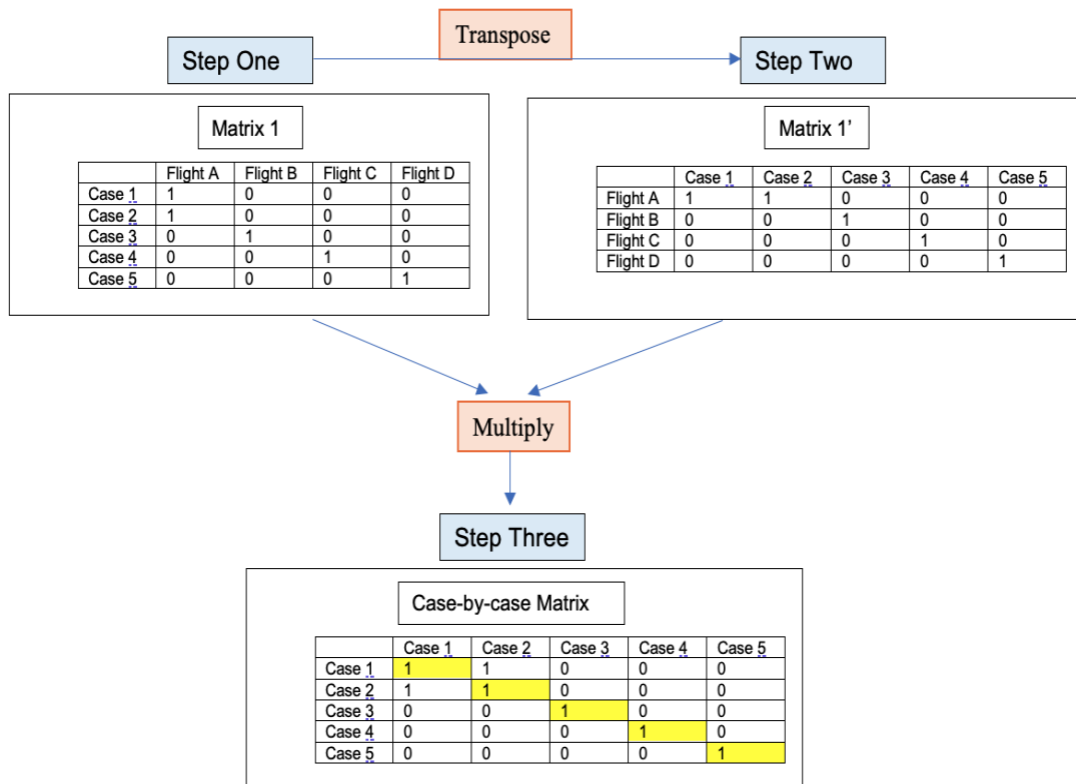
To create networks for this study, two-mode matrices were constructed in three steps. The first step involved building affiliation matrices, followed by transposing the affiliation matrices in the second step. The third step involved multiplying the affiliation matrices from step one with the transposed matrices from step two, resulting in two-mode matrices. **Figure 2.2** provides an illustration of steps two and three. For instance, case 1, 2, 3, 4, and 5 represent five confirmed COVID-19 cases with recent air travel to Hong Kong. Flights A, B, C, and D are four inbound flights that departed from different countries and arrived in Hong Kong in the same month.

1. Matrix 1 is a 5-by-4 affiliation matrix generated in step one, where each input “1” indicates that the case took the corresponding flight, and “0” indicates that the case did not take the flight. For example, Case 3 took Flight B, so a “1” is placed in the 3rd row and 2nd column cell. On the other hand, case 5 did not take Flight A, so a “0” is placed in the 5th row and 1st column cell.
2. In step two, Matrix 1 from step one is transposed, resulting in a 4-by-5 transposed Matrix 1'. The rows in Matrix 1' correspond to the columns in Matrix 1, while the columns in Matrix 1' correspond to the rows in Matrix 1. The input values in Matrix 1' retain the same indication as in Matrix 1, with each “1” indicating the case took the corresponding flight, and “0” indicating that the case did not take the flight.

3. In step three, Matrix 1 from step one is multiplied by Matrix 1', resulting in a 5-by-5 two-mode symmetric Case-by-case Matrix. Each “1” in the matrix indicates the corresponding cases stayed in the same flight cabin with another case, while “0” indicates that the cases did not share the same flight cabins with others. For example, both Case 1 and Case 2 took Flight A, so “1” is placed in the 1st row and 2nd column cell and the 2nd row and 1st column cells. Conversely, Case 3, 4, and 5 did not share flight cabins with others, so “0” is placed in their respective cells.

The Case-by-case Matrix is symmetric. The diagonal cells, highlighted in yellow in **Diagram 2.1**, have a value of “1” because cases are in the same flight cabin with themselves.

Diagram 2.1 Create Two-Mode Matrix by Affiliation Matrices



Important Variables

Link formation is a binary variable representing links, or “ties”, that connect nodes with recent air travel history who had traveled within the same flight cabins or were physically present at the same airports within one calendar day. This study disallows self-loops, meaning nodes cannot be connected to themselves. At least two nodes need to present in the same flight cabin to form a link. If a link exists between two nodes, *Link formation* is coded as “1”; otherwise, it is coded as “0”. Nodes with no link are considered isolates.

Asymptomatic status of COVID-19 cases is a binary variable measured through daily situation reports collected by The Center for Health Protection in Hong Kong Department of Health and medical records provided by medical care providers. Cases are stratified into one of three categories: 1) asymptomatic (exhibited no symptoms at all), 2) symptomatic (had a known date of onset with symptoms), or 3) unknown.

Occupation refers to the jobs reported in contact tracing reports. Cases are stratified into one of the two categories: 1) airport or flight crew (e.g., flight attendant, pilot, airport staff), or 2) non-airport or flight crew (jobs not related to airports or airlines).

Flight duration is a continuous variable measured in person-hours among cases with inbound air travel history 14 days prior to diagnosis. If a specific flight number and travel date are available, the historical record of flight duration measured in hours and minutes is recorded from the airline’s official website. When only the departure geographical location (e.g., country or city) is known, flight duration is estimated from similar historical flight itinerary records operated by major airlines. When only the departure country is known, the median flight duration is used to estimate the flight

duration. Of the cases included in this variable, 79.5% had specific flight number and travel date records, 1.6% had departure cities records, and 18.9% have departure countries records.

Departure continent is a categorical variable measured by the continent where the country from which the case departed is located. Cases' departure continents are stratified into one of the six categories: 1) Asia, 2) Africa, 3) Europe, 4) North America, 5) South America, and 6) Oceania.

2.3 Study design and data source (Chapter 5)

This study is a retrospective spatial analysis and network study that aims to explore spatial differences between asymptomatic and symptomatic cases' historical local travel records prior to their COVID-19 diagnosis in Hong Kong. The study uses the same datasets as those used in chapter 3 and 4. The Center for Health Protection in the Hong Kong Department of Health has been publicly posting reports containing de-identifiable information for COVID-19 cases since the first case was reported on January 23, 2020, on two websites: 1) daily situation reports of COVID-19 (<https://www.coronavirus.gov.hk>) and 2) contact tracing reports of each case (<https://wars.vote4.hk/en/cases>). This study retrieves spatial feature layers data on Hong Kong's population density, medical centers' and public transport entrances' geographical locations from Esri China (HK) (<https://opendata.esrichina.hk/>).

Eligible Criteria

This study includes a total of 1,282 COVID-19 cases (641 asymptomatic cases and 641 symptomatic cases) for spatial analyses. The cases were randomly selected

from all confirmed COVID-19 cases with group infection records related to superspreading events in Hong Kong during the three peaks. The random selection followed the proportion of asymptomatic and symptomatic cases for the three peaks among the entire study population. Each travel record is treated as a spatial point feature, and the coordinates of the point features are manually geocoded using Google Map. International coordinates outside Hong Kong are excluded as the study focuses on measuring patterns of local mobility.

Spatial Networks Creation

To create spatial networks for this study using ArcMap, two input feature layers are included. The first input feature layer is the road network of Hong Kong, retrieved from Esri China (HK), while the second input feature layer is a point feature layer containing the locations visited by COVID-19 cases in Hong Kong's territory 14 days before diagnosis. Two output feature layers are produced: the first is an Origin-Destination Cost Matrix network shapefile (refer to **Figure 5.12** in Chapter 5 for an example), and the second is a route layer shapefile (refer to **Figure 2.2** for an example).

Figure 2.2 Routes for Sample Cases' Travel Records within Hong Kong Road Network



Important Variables

COVID-19 cases' local mobility is measured as a continuous variable, representing the number of spatial point features that indicates the physical locations they traveled to within Hong Kong territory prior to diagnosis.

Population distribution in Hong Kong is measured by a group of spatial socioeconomic factor variables, including sex, education attainment, employment status, household composition, working hours per week, household income per month, and number of household rooms. These variables are retrieved from the open data pool provided by Esri China (HK) and the government census resource DATA.GOV.HK.

Clinics and health centers is a feature layer indicating the geographical location of Hong Kong Department of Health affiliated clinics and health centers.

COVID-19 community testing centers is a feature layer indicating the geographical location of testing centers for COVID-19 in Hong Kong.

Elderly centers is a feature layer indicating the geographical location of day care centers in Hong Kong for elderly residents.

Railway entrances is a feature layer indicating the geographical location of the entrances to the railway in Hong Kong.

Metered parking spaces is a feature layer indicating the geographical location of Hong Kong's metered parking spaces.

2.4 Ethics consideration

This dissertation uses de-identified COVID-19 cases data publicly available from the Center for Health Protection in Hong Kong Department of Health. To protect

the identities of the COVID-19 cases, unique tracking IDs were created by the Hong Kong Department of Health. For example, the first three confirmed COVID-19 cases were assigned tracking IDs 1, 2, and 3 respectively. The tracking IDs enabled researchers to monitor each case's hospitalization, discharge, deceased status, and contact tracing reports without disclosing their identities to the public. Since the data is publicly available and de-identified, this study was considered as a non-human subject research and was officially waived on March 22, 2022, by the Institutional Review Board at the University of Maryland College Park (IRB number 1882289-1).

Chapter 3: Associations between COVID-19 Asymptomatic Status and Inbound Air Travel

Abstract

Background:

Air travel had been posed as a risk for global spread of airborne infectious disease. Asymptomatic COVID-19 cases exhibited no symptoms prior to diagnosis. Revealing the transmission of SARS-CoV-2 among asymptomatic cases was crucial to containment of the virus. How asymptomatic cases contribute to the overall in-flight transmission has been understudied.

Methods:

In total, 11,775 confirmed COVID-19 cases in Hong Kong from January 2020 to April 2021 were included in the study. Log-binomial regression models were used to analyze the association between COVID-19 asymptomatic status and the binary outcome of having inbound air travel history or not. The median flight duration between asymptomatic and symptomatic cases was compared using the Wilcoxon rank-sum test.

Results:

Almost two thirds of COVID-19 cases with inbound air travel history 14 days prior to diagnosis were asymptomatic. There was a significant association between COVID-19 asymptomatic status and inbound air travel behavior, which was modified by cases' occupation ($p < 0.0001$). Asymptomatic airport or flight crew were ten times more likely to have inbound air travel history compared to symptomatic airport or flight crew (adjusted RR=10, 95% CI: 4.00–25.00), while asymptomatic cases who were not airport or flight crew were 1.4 times more likely to have inbound air travel history compared to symptomatic cases (adjusted RR=1.14, 95% CI: 1.12–1.16). The median

flight duration of asymptomatic cases was 4.6 person-hours shorter compared to symptomatic cases ($p<0.01$).

Conclusion:

Asymptomatic COVID-19 cases had a significantly higher risk of taking inbound air travel 14 days prior to diagnosis than symptomatic cases, and asymptomatic cases had a shorter median flight duration. These findings suggest that public health prevention strategies to contain future airborne infectious disease pandemics should be promoted to the general public regardless of symptoms.

3.1 Introduction

The global spread of infectious diseases has been greatly influenced by the dynamics of global connections (Coelho et al, 2020; Tuite 2020). With over three billion passengers travelling by air each year, there is a growing concern about potential transmission of contagions during air travel (Hertzberg et al, 2018). Airports in East Asia with high Fragile States Indexes have been estimated to have the highest risk for future global disease outbreaks due to high population density and flight volumes (Daon, Thompson & Obolski, 2020). The Fragile States Index is a measurement tool that highlights the vulnerabilities contributing to the risk of a geographical state or country to global disease outbreaks (the Fund for the Peace, 2013). Asian airport hubs are the most likely sites to continue the COVID-19 pandemic (Bogoch et al, 2020), while some North American and European regions with air connections to geographical regions with high Frigile State Indexes are also at risk (Lau et al, 2020 b; Tuite 2020). The severe acute respiratory syndrome (SARS) 2003 pandemic (Olsen et al, 2003; Mangili & Gendreu, 2005), tuberculosis, measles (Jungerman et al, 2017), and influenza epidemics (Hertzberg et al, 2018) have all been examples of air travel posing a risk for the global spread of airborne infectious diseases.

While there are limited published studies on COVID-19 transmission during air travel, it is important to note that this does not indicate that flying during the pandemic was safe (Freedman & Wilder-Smith, 2020). Researchers have cautioned that the risk of COVID-19 infection during air travel is real and persistent (Khanh et al, 2020; Khatib et al, 2020; Kong et al, 2020). Despite pre-boarding examination, N95 respirators, social distancing measures, and post-flight quarantine being mandatory for

the 299 passengers on an 11-hour flight from Milan, Italy to South Korea in March 2020, Bae and his team found six asymptomatic cases among them, highlighting the possibility of in-flight transmission (Bae et al, 2020). Evidence of transmission of SARS-CoV-2 in charter flights has also been reported, with 66.67% of the passengers (106 out of 159) testing positive for COVID-19 before entry on a charter flight from Iraq to China (Yue et al, 2021). Furthermore, a study conducted in Australia revealed that 11 COVID-19 cases identified on a domestic flight shared the identical virus strain with the nine cases from a recent cruise, which had not previously been identified in Australia (Speake et al, 2020; Swadi et al, 2021). These findings emphasize the need for continued caution when travelling during the pandemic.

Hong Kong, as a special administrative region of China, is one of the world's most densely populated international transport hubs. Since the first COVID-19 case was reported on January 23, 2020, the Center for Health Protection in Hong Kong Department of Health, along with other local authorities, has been actively tracing all confirmed COVID-19 cases. According to Lau and colleagues (2020), regions with low Healthcare Access and Quality Index (HAQ-Index) may underreport the number of COVID-19 cases, thus compromising containment of the virus. In contrast, Hong Kong is one of the world's healthiest regions and has a well-established healthcare system, which provides access to high-quality service for its residents from both private and public healthcare providers (Kong et al, 2015). The Hong Kong Department of Health's active surveillance of COVID-19 is a significant contribution to COVID-19 research. By using this publicly available dataset, Choi and team members identified four COVID-19 cases (one asymptomatic) among 294 passengers who traveled on a 15-

hour flight from Boston, USA to Hong Kong in March 2020. The four cases shared identical virus genetic sequences that had not been reported in Hong Kong previously, which provided strong evidence for transmission of SARS-CoV-2 in flight cabins (Choi et al, 2020).

The identification of asymptomatic cases is critical to prevent the transmission of SARS-CoV-2 (Ng et al, 2020). Asymptomatic individuals do not display any COVID-19 symptoms before diagnosis (Bae et al, 2020). Unfortunately, public COVID-19 datasets may not accurately reflect the number of infected cases due to underreporting of asymptomatic cases (Choi et al, 2020). Furthermore, not only passengers and flight crews, but also airport employees such as screening officers are at high risk of contracting COVID-19. In the United States, the Transportation Security Administration (TSA) has reported 20,946 employees infected with COVID-19 and 36 fatalities (TSA, 2022). At Los Angeles International Airport, for example, 672 screening officers and 58 non-screening employees had contracted COVID-19 as of December 31, 2021, with similar outbreaks occurring at other airport hubs in the country (TSA, 2022).

Although the contribution of asymptomatic cases to in-flight transmission of COVID-19 has not been widely studied understudied (Kong et al, 2020), this study aims to fill this knowledge gap by examining the association between COVID-19 asymptomatic status and air travel behavior prior to diagnosis. Previous studies have been limited by their focus on single flights and small sample sizes, making it difficult to investigate temporal changes. In contrast, this retrospective cohort study includes a larger sample size (N=11,745) and a two-years of data collection, allowing for a more

comprehensive understanding of the role of asymptomatic cases in COVID-19 transmission during air travel.

3.2 Methods

Data source

The Center for Health Protection in the Hong Kong Department of Health has publicly released reports containing detailed de-identifiable information for COVID-19 cases since the first case was reported on January 23, 2020, on two websites: 1) daily situation reports of COVID-19 from: <https://www.coronavirus.gov.hk/eng/index.html> and 2) contact tracing reports of each case from: <https://wars.vote4.hk/en/cases>. This study merged information from the two websites using each case's unique tracking ID. We utilized the Locally Weighted Scatterplot Smoothing (LOESS) technique to identify patterns in the daily incidence data of COVID-19 cases in Hong Kong. The LOESS curve allowed us to smooth out random fluctuations and noise present in the raw data, providing a clearer view of the important epidemic curves. This study included a total 11,775 cases from January 2020 to April 2021, which were identified by following the trend of the LOESS epidemic curve.

The data collected from the daily situation reports provides information on the *age*, *sex*, and *asymptomatic or symptomatic status* of each COVID-19 case. The contact tracing reports included *brief medical histories* of cases with information on whether they have underlying condition or good health history, as well as their *occupation* (e.g., flight attendant, pilot, airport staff, etc.). *Flight records* are also included if the case took an inbound flight 14 days prior to diagnosis. Flight duration data was obtained

from airlines' official websites for historical flight record. For instance, we retrieved the flight number CX906 for a specific COVID-19 case's contact tracing report and searched the historical record of CX906 on Cathay Pacific official website, collecting the flight duration of 2 hours 40 minutes in person-hours.

Human Subjects and Ethics Consideration

This study utilized publicly available data on COVID-19 cases from the Center for Health Protection in Hong Kong Department of Health. The data were de-identified, and COVID-19 cases were monitored using unique tracking IDs generated by the Hong Kong Department of Health. These tracking IDs were created to protect the identities of the COVID-19 cases while enabling researchers to follow-up on important updates regarding hospitalization, discharge, deceased status, and contact tracing reports without compromising the cases' identities. As the data was publicly available and de-identified, this study was classified as a non-human subject research and was formally waived by the Institutional Review Board at the University of Maryland College Park (IRB number 1882289-1).

Dependent Variables

Inbound air travel history to Hong Kong was measured by COVID-19 cases' travel history 14 days prior to diagnosis. Cases were labeled "1" for those who had inbound air travel history and "0" for those who did not. The travel history records were collected by the Center for Health Protection in Hong Kong Department of Health from objective sources such as the immigration department, airport, and airlines.

Flight duration was measured in person-hours and recorded from airlines' official websites when a specific flight number and travel date were available. When only the departure geographical locations (e.g., cities) were known, the flight duration was estimated from similar historical fly itinerary records by major operated airlines. This study estimated flight duration by the median flight duration when only the departure country was known. Specifically, for cases without information about flight number or departure city, this study estimated flight duration by the median flight duration between the longest and shortest historical flight itineraries departing from the country of origin and arriving in Hong Kong. Of the cases included in this variable, 79.5% had specific flight number and travel date records, 1.6% had departure cities records, and 18.9% have departure countries records.

Independent Variables

Asymptomatic status was determined by daily situation reports and medical records, and COVID-19 cases were classified into three categories: asymptomatic (exhibited no symptoms at all), symptomatic (had a date of onset with symptoms), or unknown.

Underlying health condition(s) were identified through contact tracing reports and divided into two categories: with underlying conditions (e.g., indicated having a known underlying health condition such as cardiovascular disease, diabetes and etc.), or with no underlying conditions (e.g., had a good health in the past).

Age was the biological age of cases and was collected by The Center for Health Protection in Hong Kong Department of Health.

Sex was the biological sex of cases (female or male) and was collected by The Center for Health Protection in Hong Kong Department of Health.

Occupation was the jobs that cases reported in contact tracing reports and divided into two categories: airport or flight crew (e.g., flight attendant, pilot, airport staff), or non-airport or flight crew (all other jobs outside the airport or airline industry).

Statistical methods

Descriptive analysis was conducted on all COVID-19 cases, calculating mean values and standard deviations for age and flight duration as continuous variables. Proportions were calculated for categorical variables, including asymptomatic status, inbound air travel history 14 days prior to diagnosis, sex, occupation, and underlying health condition(s).

To investigate the association between COVID-19 asymptomatic status and inbound air travel history, log-binomial regression models were used, and risk ratio was reported. The Hierarchically Well-Formulated Model principle was followed to select the best fit model. Firstly, all the independent variables, potential effect modifiers, and confounders were included in the full model. Then, the interaction effects of potential modifiers and potential confounders were examined in reduced models. The significance of the interaction terms was determined using the Wald test, and significant effect modifiers were retained in the best fit model. The 10% rule was applied to determine if variables were confounders by comparing crude and adjusted risk ratio. A variable was considered a confounder if there was a greater than 10% change between crude and adjusted estimates. The best fit model retained statistically

significant effect modifiers and controlled for confounders. Adjusted risk ratios were reported by stratification of the significant effect modifier.

To examine differences in flight duration between asymptomatic and symptomatic groups, a comparison test was conducted. Firstly, normality assumption was tested, but our data violated this assumption. Therefore, a nonparametric Wilcoxon rank-sum test was used to compare the two groups.

In this study, the missing data rate was less than 1%, and the pairwise deletion method was utilized to manage missing data. Only a quarter percent (0.25%) of the data (30 out of 11,775 cases) had missing information on cases' COVID-19 asymptomatic status and sex, and thus were deleted from data analysis. The statistical significance level was set to be 0.05, and a *p*-value smaller than 0.05 was considered statistically significant. The statistical analyses were conducted using SAS OnDemand for Academics (SAS Institute, Cary, NC), and the LOESS figure was plotted using the R package ggplot.

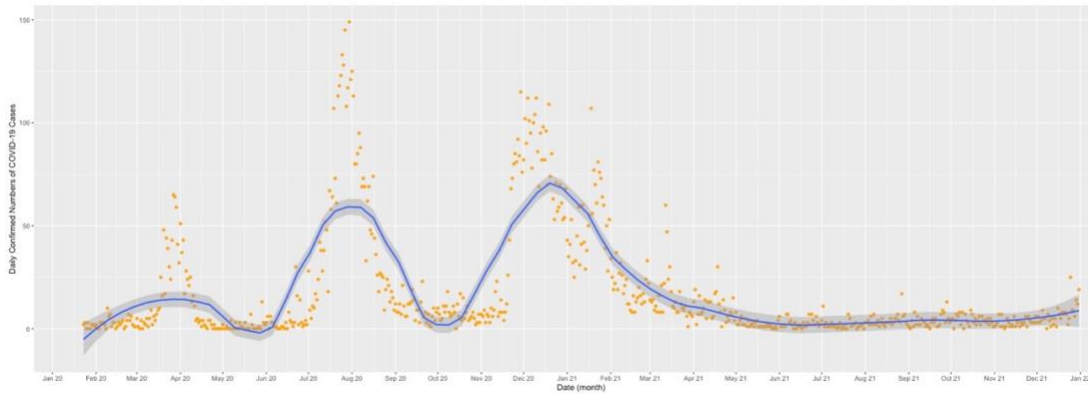
3.3 Results

Descriptive analysis

Figure 3.1 demonstrates the COVID-19 cases situation in Hong Kong from January 2020 to April 2021. The LOESS curve shows three peaks in the of diagnosis of COVID-19 cases during this period, while the curve flattened from May 2021 to December 2021. Therefore, data after May 2021 were excluded from this study, leaving a total 11,745 COVID-19 cases for analysis. The study population had a mean age of 44.2 ($\pm$ 19.3) years old, with over half ($n=6,115$) being female. Approximately 3% of

the study population (n=347) reported underlying health conditions, while 1% (n=104) reported occupations related to airports or airlines. About one third of the study population (n=3,606) were asymptomatic.

Figure 3.1 Locally Estimated Scatterplot Smoothing for COVID-19 Cases Situation in Hong Kong (Jan 2020-Dec 2021)



The Hong Kong government implemented various policies and measures during the three peaks, which are displayed in **Table 3.1**. These measures included border closure to non-Hong Kong residents, mandatory mask-wearing, tightened screening at airports, social distancing in public places and transportation.

Table 3.1 Selected Policies and Measures by Hong Kong's Government Across the Three Peaks

Peak	Period	Policies & Measures
One	Jan 2020-May 2020	- Required mask wearing for hospital visiting (Hong Kong's Center for Health Protection, 2020) - Tightened screening at airport (Leung, 2020) - Stopped visitors from Hubei province, Mainland China to visit Hong Kong (Hong Kong's Information Services Department, 2020a) - Closed border to all non-residents from overseas (Hong Kong's Information Services Department, 2020b)
Two	Jun 2020-Sep 2020	- Toughened social distancing and required mask wearing in public (Cheung & Ting, 2020) - Relaxation of social distancing and mandatory mask-wearing extended on public transportation (Government of the Hong Kong, 2020)
Three	Oct 2020-Apr 2021	- Prioritized mandatory COVID-19 testing for symptomatic cases, senior home staff, and taxi drivers (Hong Kong's Information Services Department, 2020c)

-
- Began mass vaccination (Hong Kong Free Press, 2021)
 - Arranged flights from United Kingdom, and arranged quarantine for Hong Kong resident arrivals from low-risk regions (Government of the Hong Kong, 2021)
-

Table 3.2 presents the descriptive statistics of the study population, including the asymptomatic and symptomatic cases. Asymptomatic cases had a slightly younger mean age (40.3 ± 19.9 years old) than symptomatic cases (45.9 ± 19.5 years old). The sex distribution of asymptomatic cases was similar to symptomatic cases. About two thirds of the cases who reported inbound air travel history 14 days prior to diagnosis of COVID-19 were asymptomatic, while 17% of the cases who reported underlying health conditions were asymptomatic. Moreover, 73% of the cases who worked as airport or flight crew were asymptomatic.

Table 3.2 Descriptive Statistics of Confirmed COVID-19 Cases in Hong Kong

	Symptomatic cases (n=8139)	Asymptomatic cases (n=3,606)
Age (years)		
Mean age ($\pm$ SD)	45.9 ($\pm$ 19.5)	40.3 ($\pm$ 19.9)
Sex [n (%)]		
Female	4,238 (52.1%)	1,877 (52%)
Male	3,901 (47.9%)	1,729 (48%)
Underlying condition [n (%)]		
Yes	288 (3.5%)	59 (1.6%)
No	7,851 (96.5%)	3,547 (98.4%)
Inbound air travel history [n (%)]		
Yes	936 (11.5%)	1,533 (42.5%)
No	7,203 (88.5%)	2,073 (57.5%)
Occupation classification [n (%)]		
Airport or flight crew	28 (0.3%)	76 (2.1%)
Non-airport or flight crew	8,111 (99.7%)	3,530 (97.9%)

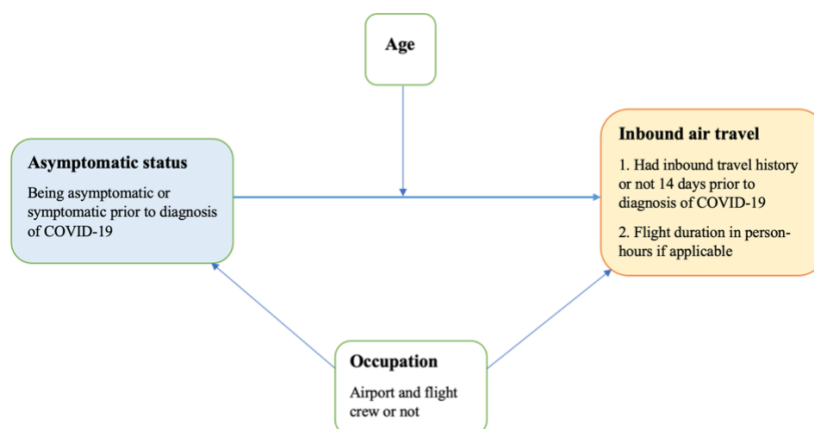
Data were retrieved from the Hong Kong Department of Health Daily Situation Reports as of April 30, 2021.

Association between COVID-19 Asymptomatic Status and Inbound Air Travel

The study's findings, obtained through log-binomial regression models, demonstrated an association between COVID-19 asymptomatic status and inbound air

travel behavior, and this association was modified by occupation (airport /flight crew or not) ($p < 0.0001$). Age was found to be a confounder, as evidenced by the change in crude and adjusted risk ratios (32.9% greater than 10%). **Figure 3.2** illustrates the significant effect modifier and confounder for the association of interest. After adjusting for age as a confounder, airport or flight crew's inbound air travel history was significantly associated with asymptomatic status before COVID-19 diagnosis (adjusted RR=10.00, 95% CI: 4.00 – 25.00). This indicated that COVID-19 asymptomatic airport or flight crew were ten times more likely to have inbound air travel history 14 days before diagnosis than symptomatic airport or flight crew. Among COVID-19 cases who did not work as airport or flight crew, the association of interest was also significant (adjusted RR=1.14, 95%CI: 1.12-1.16). This indicated that COVID-19 asymptomatic cases who did not work as airport or flight crew were 1.14 times more likely to have inbound air travel history 14 days before diagnosis than symptomatic cases. Sex and underlying health condition(s) were found not to be effect modifiers nor confounders after examining for interaction terms and confounding effect.

Figure 3.2 Directed Acyclic Graph for Association between COVID-19 Asymptomatic Status and Inbound Air Travel



Differences in Flight Duration: COVID-19 Asymptomatic Cases vs. Symptomatic Cases

We found that COVID-19 asymptomatic cases who had inbound travel history 14 days before diagnosis had a shorter median flight duration of 6.5 (25% Q1: 4.0 - 75% Q3:10.4) person-hours compared to symptomatic cases with a median of 11.1 (25% Q1: 5.0 - 75% Q3:11.7) person hours. The data of both asymptomatic and symptomatic cases was found to be not normally distributed through the normality test (Kolmogorov-Smirnov test $p < 0.01$). Therefore, a nonparametric Wilcoxon rank-sum test was conducted, which showed that flight duration of COVID-19 asymptomatic cases was significantly different from symptomatic cases ($p < 0.01$). COVID-19 case had a median flight duration that was 4.6 person-hours shorter than symptomatic cases ($p < 0.01$) if they had inbound air travel 14 days before diagnosis.

3.4 Discussion

Our findings indicate that asymptomatic cases are at a higher risk of engaging in inbound air travel behavior 14 days prior to a COVID-19 diagnosis compared to symptomatic cases. This aligns with our hypothesis that asymptomatic cases are more likely to travel than symptomatic cases due to the absence of symptoms. This finding is crucial for future airborne infectious disease control, particularly during the early stage of a pandemic when rapid testing and mandatory testing before boarding is not yet available. In such situations, individuals without symptoms are likely to continue with their travel plans. Hence, national and local health authorities should strongly advise the general population to implement effective public health prevention strategies such as wearing masks and maintaining social distancing while traveling, regardless of

their symptomatic status. It is challenging for the public without medical expertise to differentiate between various respiratory infectious diseases such as influenza, common cold, or COVID-19. Thus, practicing public health prevention strategies is the most cost-effective and efficient way to contain a potential pandemic.

Our findings revealed a unique pattern in flight duration, where asymptomatic cases had a shorter median flight duration than symptomatic cases. However, this pattern cannot be interpreted as a lower risk of COVID-19 infection for asymptomatic cases due to a shorter median flight duration, as two critical factors impacted this finding. Firstly, the diagnosis and viral testing of COVID-19 at the early stage of the pandemic focused only on individuals with symptoms, resulting in our data capturing flights with long durations mostly among symptomatic cases, thus missed asymptomatic cases during this period. Secondly, several countries, such as the United States and United Kingdom, which are geographically far from Hong Kong, implemented travel restrictions against travelers with symptoms as the COVID-19 pandemic spread globally. Therefore, most flights our data captured after the summer of 2020 departed from countries that are geographically close to Hong Kong, such as India, the Philippines, with short flight durations. Moreover, individuals with symptoms were discouraged from traveling by that time, resulting in most of the cases detected being asymptomatic.

The healthy worker effect (Last, 1993) was observed in our study population, meaning that workers had a lower mortality rate than the general population due to a higher likelihood of good health. In particular, a significant higher percentage (73%) of the air or flight crew in our study were asymptomatic before their COVID-19

diagnosis compared to the general study population (31%). This could be attributed to the fact that individuals with underlying health conditions, such as hypertension, cardiovascular disease, or obesity are less likely to be hired as flight crew, resulting in a healthier population. Only 17% of cases with underlying conditions were asymptomatic prior to COVID-19 diagnosis, indicating that individuals with a healthier history were more likely to be asymptomatic. Asymptomatic air or flight crew were found to be ten times more likely to take inbound flights 14 days before their diagnosis than symptomatic crew, highlighting the need for tailored pandemic containment strategies for this specific occupation to prevent hidden spread.

One strength of this study was that data on asymptomatic cases were collected from objective sources, such as lab testing results and flight history, minimizing the possibility of recall bias. This sets our study apart from others that relied on participants' self-reported data (Chen et al, 2020), where recall bias was more likely to occur. Hong Kong's status as an international transport hub allowed for a diverse population with different nationalities and races to be included in this study, making the results generalizable not only to Asia but also to other densely populated international hubs in Europe and America. Another strength of this study is the heterogeneity of the departed cities included in the inbound flights data. Inbound flights from all over the world were analyzed, making the findings more comprehensive. The use of ethical, de-identified individual-level data to compare asymptomatic cases and symptomatic patients on air travel is also a strong point.

This study has some limitations that need to be acknowledged. Firstly, it only examined inbound flights in Hong Kong, with outbound flights excluded due to loss of

follow-up. Among cases with inbound flight records, 70% were Hong Kong residents, 10% were non-Hong Kong residents, and 20% had unknown residency status. Thus, the findings are primarily relevant to the Hong Kong Department of Health to track inbound flights among this population. Secondly, as an observational study using secondary data, there was no direct access to COVID-19 cases to address gaps in the data source. Therefore, information regarding cases' underlying health conditions could not be thoroughly examined.

3.5 Conclusion

Asymptomatic cases had a significantly higher risk of taking inbound air travel to Hong Kong 14 days before the diagnosis of COVID-19, compared to symptomatic cases. On average, the flight duration among asymptomatic cases was two person-hours shorter than that among symptomatic cases, which suggests that there may have been underdetection of asymptomatic cases when implementing air travel restrictions during the early stage of the COVID-19 pandemic. These findings highlight the importance of promoting public health prevention strategies to contain future airborne infectious disease pandemics at an early stage. Effective screening and detection strategies should be implemented for all individuals traveling to high-risk areas, while also increasing public awareness about the importance of early detection and prevention measures, regardless of whether they are symptomatic or not.

Chapter 4: Social Network Analysis of COVID-19 Transmission among Asymptomatic and Symptomatic Cases in Inbound Air Travel

Abstract

Background:

Social network analysis is a powerful tool for examining relationships between individuals. In the case of SARS-CoV-2, which spreads within and across social networks, this type of research can provide valuable information for infectious disease control by examining network robustness. This study aims to examine empirical networks among COVID-19 cases on shared flights and airports during the three epidemic peaks of COVID-19 pandemic in Hong Kong.

Methods:

Our study included 1,934 COVID-19 cases in Hong Kong between January 2020 and February 2021. For each of the three epidemic peaks, we identified three networks that described passenger movement. The network nodes were COVID-19 cases that arrived in Hong Kong within 14 days before diagnosis, and network links connected cases (nodes) on the same inbound flight cabins or in the same arrival airport on the same calendar day. Descriptive analysis was conducted to estimate network size, density, degree, and centrality, while Exponential Random Graph Models (ERGMs) using Maximum Pseudo-Likelihood Estimation (MPLE) were utilized to estimate the likelihood of nodes forming links. We also employed network percolation techniques, such as random failures and targeted attacks, to dismantle the three peak networks.

Results:

The spread of SARS-CoV-2 in inbound flight cabins and airports was fastest during peak one, slowed during peak two, and significantly decreased during peak three. Peak one saw 1.34 times (95%CI: 1.22-1.46) higher likelihood of asymptomatic cases

forming links than symptomatic cases, while cases departing from Asia were 1.38 times (95%CI: 1.61-1.63) and 3.51 times (95%CI: 1.42-8.68) more likely to forming links in peak one and two, respectively, than cases departing from other continents. Asymptomatic cases in peak two showed no significant associated with link formation, while in peak three, they were less likely to form links (OR: 0.87, 95% CI: 0.78-0.87) compared to symptomatic cases. Age showed a significant negative association with link formation in peaks one (OR:0.996, 95%CI: 0.994-0.996), suggesting that younger cases were more likely than older cases to form links. Flight duration also showed a significant negative association with links in peaks one and three (OR:0.97, 95%CI: 0.96-0.98 and OR:0.98, 95%CI: 0.97-0.99, respectively), indicating that shorter flight durations were slightly more likely to be associated with the presence of links during these periods. Finally, Targeted attacks were more effective than random failures in dismantling networks.

Conclusion:

Our study underscores the significant role of asymptomatic cases in the early stages of the COVID-19 pandemic, particularly in inbound flight cabins and airports. Early intervention is crucial, and policymakers should prioritize public health strategies such as wearing masks and maintaining social distance. Travel restrictions may not be effective in containing the pandemic in fragmented networks like those found in inbound flight cabins and airports. Therefore, policymakers should continuously monitor the effectiveness of their interventions and adjust their policies over time to optimize their impact.

4.1 Introduction

The COVID-19 pandemic has resulted in the transmission of the SARS-CoV-2 virus both within and across personal networks. Among the various modes of viral transportation, air travel is a particularly important contributor to the spread of infectious diseases (Gilmore, 1998). Previous studies have established a link between the spread of diseases and commercial flights (Mangili et al, 2015). Despite the implementation of international travel restrictions early on the pandemic, the global spread of SARS-CoV-2 could not be effectively contained. To better understand how the virus was transmitted across borders, it is crucial to explore the intercontinental mobility patterns of infected individuals. Such analysis can provide valuable insights into the transmission dynamics of the virus.

Social network analysis is a powerful tool that enables researchers to examine the relationships between individuals in a social network. In this type of analysis, nodes represent individuals and links represent the relationships between them (Newman, 2018). When applied in public health research, social network analysis can provide valuable insights into the ways in which social relationships impact personal behaviors (Valente, 2010). In order to protect the anonymity of participants while analyzing their relationships, anonymous networks can be constructed using unique tracking numbers (Valente, 2010).

A robust network is one that maintains its function and density even when some of its nodes are removed (Menczer, 2020). Analyzing the robustness of a network can provide valuable insights into its resilience and potential vulnerabilities, which can be useful for developing interventions and strategies aimed at promoting health behaviors

and controlling the spread of infectious disease. Network intervention is an approach to enhancing network robustness and involves methods such as rewiring networks. This process involves rearranging the structure of networks to create new links or modify existing ones (Valente, 2010). Percolation is a technique that can be used to examine the robustness of networks and generate valuable information for infectious disease control. For example, vaccinating individuals the highest degree of connectivity in a disease contact network can be more efficient than vaccinating individuals at random, as it can dismantle the transmission network and reduce the spread of the disease (Newman, 2018).

Hong Kong International Airport is one of the busiest airports in the world, connecting the city to over 220 global destinations through approximately 120 airlines since its establishment in 1998, (Hong Kong International Airport (HKIA), 2021). Prior to the COVID-19 pandemic, the airport handles over 70 million passengers in 2019 (HKIA, 2021). However, like other international airport hubs, the pandemic had severe impact on air traffic movements in Hong Kong, with the number of passengers decreasing to 1.35 million by the end of 2020 (HKIA, 2021).

To address gaps in existing research, it is necessary to examine network changes among COVID-19 cases who took inbound flights to Hong Kong during the epidemic peaks of SARS-CoV-2 infection from 2020 to mid-2021. While some limited studies have used network approaches for COVID-19, no studies have examined agent-based networks in flight cabins and airports. Most studies have focused on macro-level city or airport networks for infectious disease transmission, rather than agent-based individual-level data (Colizza et al, 2006; Coelho et al, 2020; Daon, Thompson &

Obolski, 2020). Moreover, no studies have examined COVID-19 cases in flight cabins and airports networks during the peaks of the epidemic, nor have they examined the impact of travel restrictions on empirical networks using network percolation approaches. Hence, this study contributes significantly to COVID-19 research by addressing critical knowledge gaps. Specifically, we examine empirical networks among COVID-19 cases on shared flights and airports during the epidemic peaks of SARS-CoV-2 infection in Hong Kong and simulated the impact of travel restrictions on these cases using network percolation.

4.2 Methods

Date source

The Center for Health Protection in the Hong Kong Department of Health has publicly released reports containing detailed de-identifiable information for COVID-19 cases since the first case was reported on January 23, 2020, on two websites: 1) daily situation reports of COVID-19 from <https://www.coronavirus.gov.hk> and 2) contact tracing reports of each case from: <https://wars.vote4.hk/en/cases>. This study merged information from the two websites using each case's unique tracking ID. A total 1,934 COVID-19 cases were included in this network analysis study. For network visualization and statistical analysis, cases from a 4-month interval data collection for each peak (**Table 4.1**) were included. **Figure 3.1** demonstrates the LOESS curve of the three peaks of COVID-19 diagnosis in Hong Kong.

Table 4.1 Temporal Stratification of the Three Peak Networks Analysis

Time period	
Peak one network	Late January 2020 – May 2020
Peak two network	June 2020 – September 2020
Peak three network	November 2020 – February 2021

The data collected from the daily situation reports provides information on the *age*, *sex*, and *asymptomatic or symptomatic status* of each COVID-19 case. The contact tracing reports included *brief medical histories* of cases with information on whether they have underlying condition or good health history, as well as their *occupation* (e.g., flight attendant, pilot, airport staff, etc.). *Flight records* are also included if the case took an inbound flight 14 days prior to diagnosis. Flight duration data was obtained from airlines' official websites for historical flight record.

Human Subjects and Ethics Consideration

This study utilized publicly available data on COVID-19 cases from the Center for Health Protection in Hong Kong Department of Health. The data were de-identified, and COVID-19 cases were monitored using unique tracking IDs generated by the Hong Kong Department of Health. These tracking IDs were created to protect the identities of the COVID-19 cases while enabling researchers to follow-up on important updates regarding hospitalization, discharge, deceased status, and contact tracing reports without compromising the cases' identities. As the data was publicly available and de-identified, this study was classified as a non-human subject research and was formally waived by the Institutional Review Board at the University of Maryland College Park (IRB number 1882289-1).

Dependent Variable

Link formation was measured by connecting nodes (COVID-19 cases with inbound air travel history) who had traveled within the same flight cabin or were physically present at the same airports on the same calendar day. Self-loops were not permitted, meaning passengers could not be connected to themselves by links. To form a link, at least two passengers (nodes) are required to be in the same flight cabin or airport. For example, passengers A and B would form a link if they took the same flight from London to Hong Kong on the same date, regardless of their seating positions during the flight. Similarly, if passenger C flew from Tokyo to Hong Kong and arrived on the same day as passengers A and B, links would be established among the three passengers since they all arrived at the same airport on the same calendar day. The dependent variable, *link formation*, was coded as “1” if a link existed between two nodes and “0” if it did not. Nodes without any connected link were treated as isolates.

Independent Variable

Asymptomatic status of COVID-19 cases was determined based on daily situation reports from the Center for Health Protection in Hong Kong Department of Health and medical records provided by medical care providers. Cases were categorized as asymptomatic (no symptoms exhibited), symptomatic (onset data with symptoms), or unknown.

Age was treated as a continuous variable in this study, calculated as the difference between cases’ biological age collected by the Center for Health Protection in Hong Kong Department of Health. For instance, if case A was 18 years old and case B was 35 years old, their age difference of 17 years was recorded in the vector data.

Sex, recorded as the biological sex of cases (female or male), was collected by the Center for Health Protection in Hong Kong Department of Health. The homophily of Each case based on sex was noted in vectors, such that if case A and case B were both female, their vector value for sex homophily was “1”.

Underlying health condition(s) were measured by cases’ contact tracing reports, which were classified into two categories 1) no underlying conditions (good health history), and 2) with underlying conditions (underlying health condition such as cardiovascular disease, diabetes etc.).

Occupation was the jobs that cases reported in contact tracing reports and divided into two categories: airport or flight crew (e.g., flight attendant, pilot, airport staff), or non-airport or flight crew (all other jobs outside the airport or airline industry).

Flight duration was measured in person-hours and recorded from airlines’ official websites when a specific flight number and travel date were available. When only the departure geographical locations (e.g., cities) were known, the flight duration was estimated from similar historical fly itinerary records by major operated airlines. This study estimated flight duration by the median flight duration when only the departure country was known. Specifically, for cases without information about flight number or departure city, this study estimated flight duration by the median flight duration between the longest and shortest historical flight itineraries departing from the country of origin and arriving in Hong Kong. Of the cases included in this variable, 79.5% had specific flight number and travel date records, 1.6% had departure cities records, and 18.9% have departure countries records.

Departure continent was measured by the continent where the country of departure was located. Cases' departure continents were classified into six categories: 1) Asia, 2) Africa, 3) Europe, 4) North America, 5) South America, and 6) Oceania.

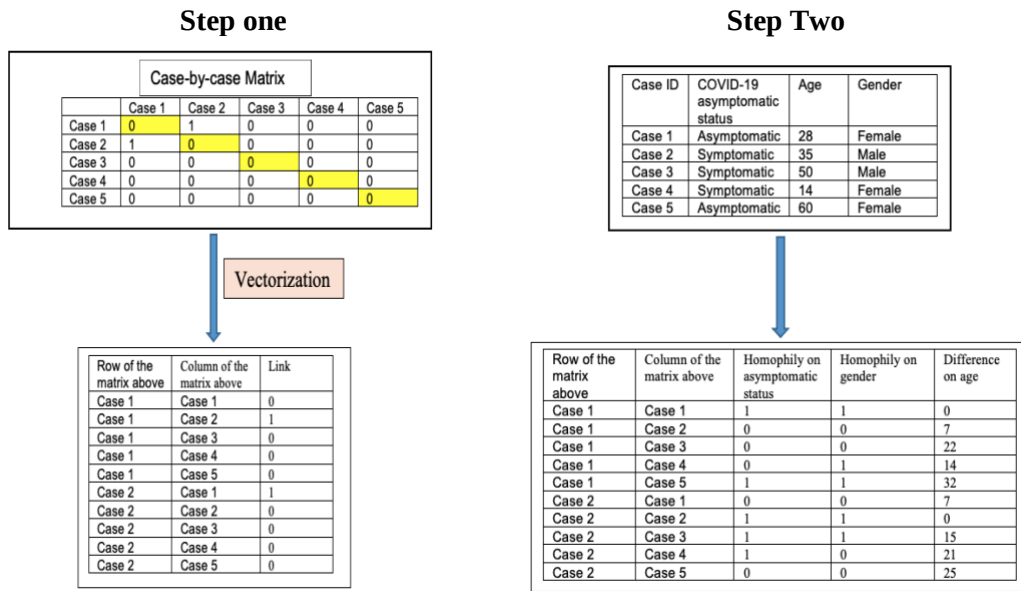
Statistical methods

In this study, the network nodes were defined as COVID-19 cases that had traveled to Hong Kong by air within 14 days before diagnosis. Network links were defined as connections between cases on the same inbound flight cabins or present at the same airport. At least two passengers (nodes) needed to be in the same flight cabin or airport at the same time to have an unweighted link presence, regardless of their seating arrangements. Link length or thickness was not significant, and it was not based on Euclidean distance. Descriptive statistics analysis was performed to calculate the *network size* and *density* were. *Network size* (N) is the number of nodes in each network, and *Density* is the fraction of links to the number of connections in each network. For centrality measurement, the *average degree*, *betweenness*, and *eigenvector centrality* of each network were calculated. *Average degree* is the average number of nodes connected to each node. *Betweenness centrality* measures node importance by examining the number of shortest paths that pass through it. *Eigenvector centrality* measures node importance by examining the importance of neighboring nodes.

The Exponential Random Graph Model (ERGM), using the Maximum Pseudo-Likelihood Estimation (MPLE) method, was utilized to access whether link formation was associated with seven factors, including homophily on COVID-19 cases' asymptomatic status, age differences, sex homophily, occupation homophily,

underlying health condition homophily, departure continent homophily, and flight duration differences. To build the ERGM, we followed a two-step process. Firstly, we vectorized the adjacency matrices for each network, with the diagonal set to “0” as self-loop links were not permitted in this study. Secondly, we transformed each COVID-19 case node’ attributes (such as their asymptomatic status) into network data. **Diagram 4.1** provides an illustration of the data processing steps. Additionally, odds ratios were reported to interpret the association between link formation and node attributes.

Diagram 4.1 Convert Adjacency Matrix for Maximum Pseudo-Likelihood Estimation



We used network percolation to examine the impact of removing nodes on the network during each epidemic peak. Two types of node removal were simulated and compared: 1) *Random failure* and 2) *targeted attack*. *Random failure* involves randomly removing nodes from the network, while *targeted attack* involves removing nodes with the highest degrees first, followed by lower degree nodes. We compared

the breakdown thresholds of each network between the results of random failure and targeted attack.

Missing data in this study was managed using a pairwise deletion method due to the low missing data rate of 0.98%. It's important to note that nodes with missing information on COVID-19 asymptomatic status were included in the network visualization but excluded from network statistical analysis. A statistical significance level of 0.05 was set, where p value smaller than 0.05 were considered statistically significant. In addition, Monte Carlo empirical p -value were used to assess the goodness-of-fit of the model, where p -value smaller than 0.05 indicated a poor fit of the model. The R package igraph was used for network analysis and network visualization was plotted using Gephi 0.10.1. Python 3.11.1 was used for network percolation.

4.3 Results

Network descriptive analysis and visualization

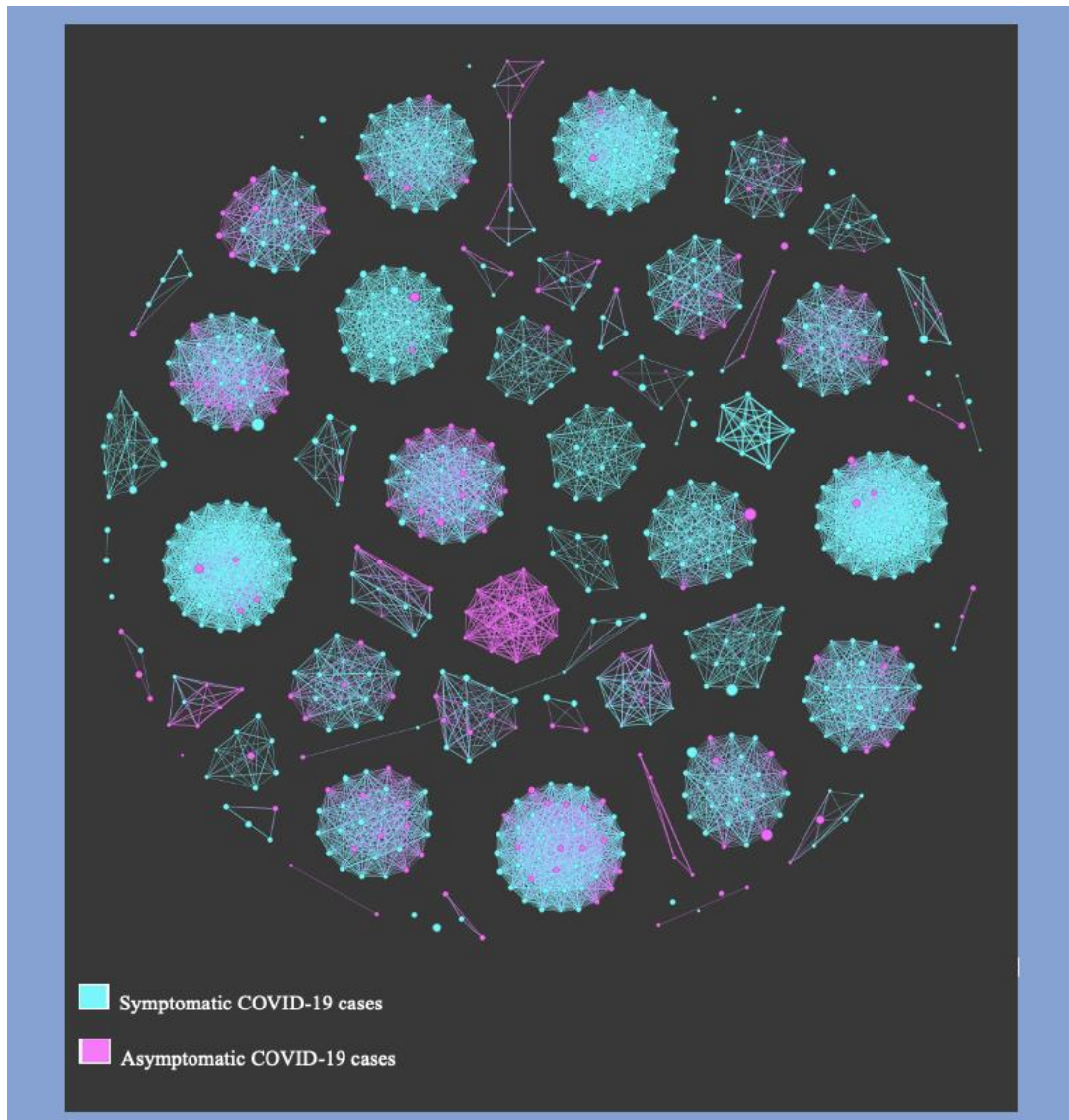
Table 4.2 displays the descriptive statistics for the three peaks. Peak one had the largest number of nodes, links, density, and average degree among the three networks. In contrast, peak two had the highest average betweenness and eigenvector centrality, indicating that the clusters in this network were better connected to other clusters than those in the peak one and peak three networks. Peak three had the lowest value for all statistics among the three. These results suggest that the spread of SARS-CoV-2 in inbound flight cabins and airports was fastest during peak one, slowed during peak two, and significantly decreased during peak three.

Table 4.2 Descriptive Statistics of the Three Peak Networks

	Peak one network	Peak two network	Peak three network
Number of nodes (n=)	765	612	557
Number of links (n=)	8735	3178	1568
Density	0.03	0.02	0.01
Average degree	46	21	11
Average betweenness	0.04	0.35	0.05
Average eigenvector centrality	0.0088	0.0131	0.0092

Figures 4.1-4.3 present the network topology for the three peaks. These undirected networks exhibit a fragmented pattern, where nodes are connected in multiple separate clusters, and clusters do not frequently connect to other clusters. Asymptomatic COVID-19 cases are represented in pink nodes, while symptomatic COVID-19 cases are represented in aqua nodes. During peak one (**Figure 4.1**), most clusters were composed of symptomatic cases, and several densely connected clusters were present. In peak two (**Figure 4.2**), the number of asymptomatic cases dramatically increased, while the number of symptomatic case and densely connected clusters decreased. Peak three (**Figure 4.3**) still had a majority of asymptomatic case present in the clusters, and the number of isolates increased while the number of densely connected clusters reduced. Based on the network visualization, it appears that during the initial stages of the pandemic, symptomatic cases likely played a greater role in the spread of SARS-CoV-2 through inbound air travel to Hong Kong. As Hong Kong implemented quarantine for symptomatic cases and mandatory COVID-19 testing, more asymptomatic cases were detected and mainly contributed to the spread of SARS-CoV-2 in inbound flight cabins and airports during peak two and three.

**Figure 4.1 Network of COVID-19 Cases Traveled to Hong Kong by Air during Peak
One (January 2020-May 2020)**



**Figure 4.2 Network of COVID-19 Cases Traveled to Hong Kong by Air during
Peak Two (June 2020-September 2020)**

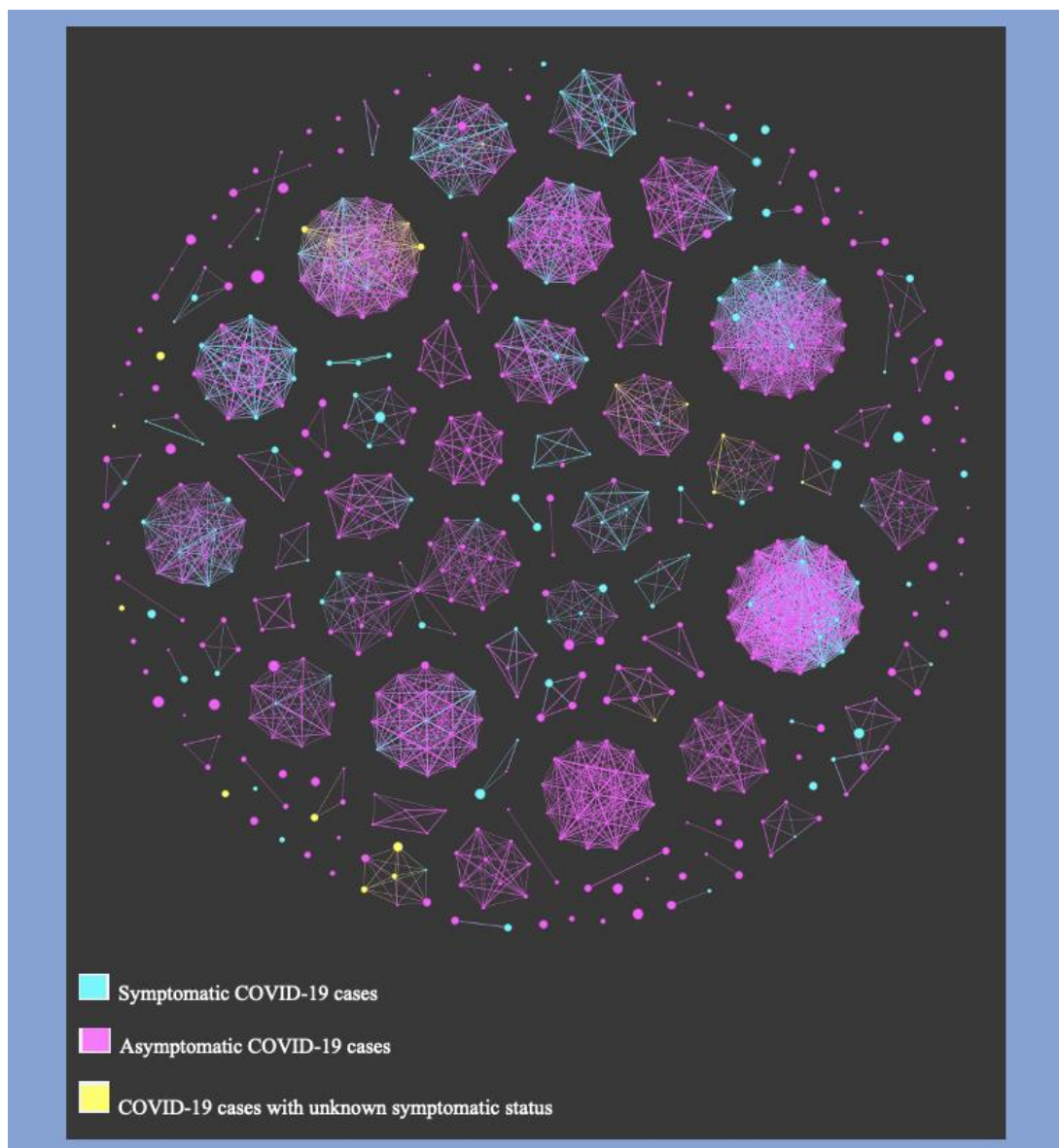
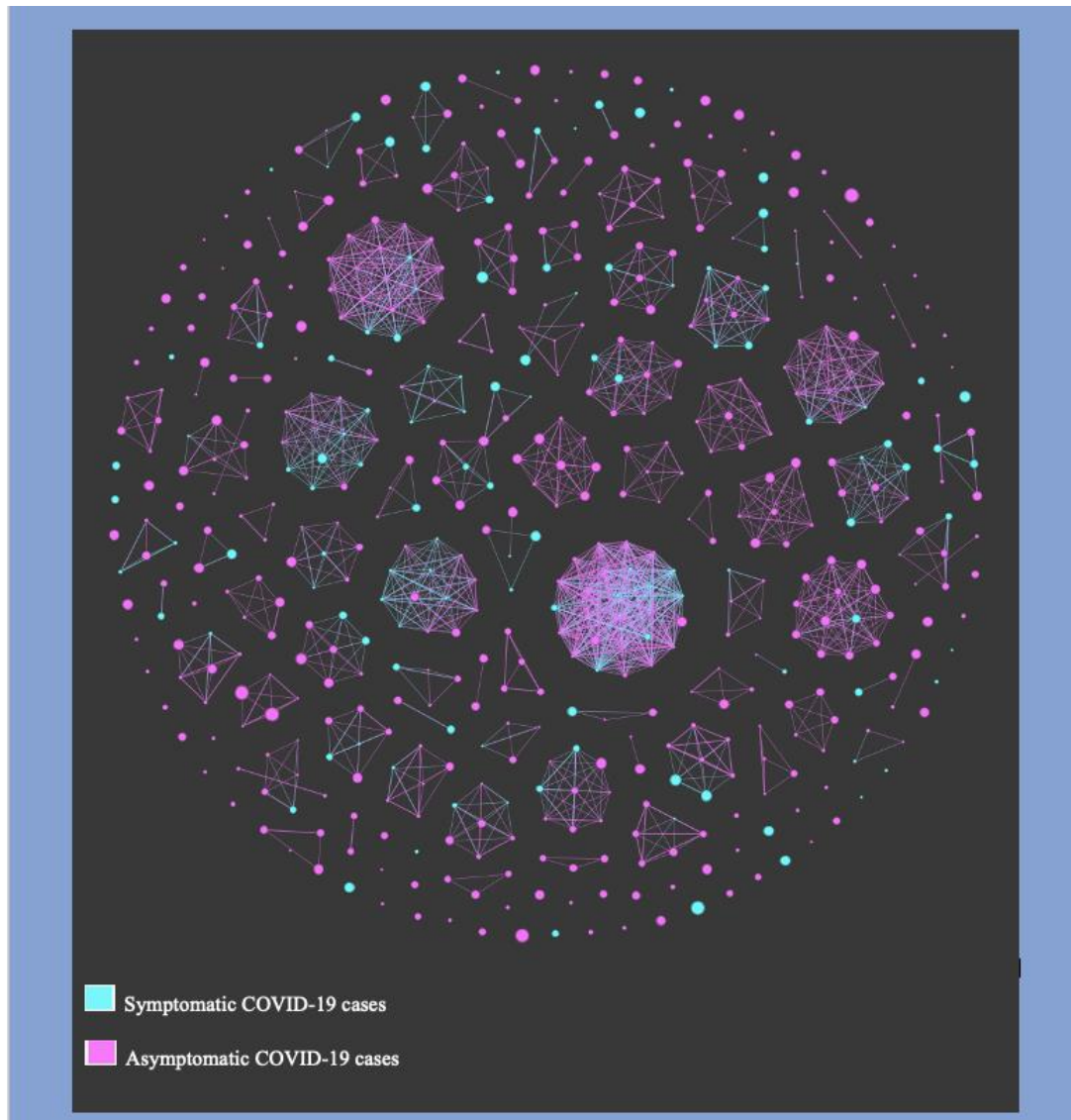


Figure 4.3 Network of COVID-19 Cases Traveled to Hong Kong by Air during Peak Three (November 2020-February 2021)



Associations between Network Nodes Attributes and Link Formation

Table 4.3 presents descriptive statistics of nodes' attributes in this study. The average age of passenger nodes was similar among the three peaks networks, around mid-thirty years old. The proportion of female nodes and nodes in airport or flight crew category increased from peak one to peak three, while nodes with underlying health

conditions decreased over time. The proportion of inbound flights departed from Asia increased from peak one to peak three, coinciding with a dramatic increase in the proportion of asymptomatic nodes during the same period.

Table 4.3 Descriptive Statistics of Nodes' Attributes in the Three Peak Networks

	Peak one network	Peak two network	Peak three network
Mean age (years)	34.86	32.79	35.31
Sex	45% female 55% male	51% female 49% male	54% female 46% male
Occupation	1.5% airport or flight crew 98.5% non-airport or flight crew	6% airport or flight crew 94% non-airport or flight crew	7% airport or flight crew 93% non-airport or flight crew
Underlying condition	8% with underlying condition 92% with good health history	3% with underlying condition 97% with good health history	0.2% with underlying condition 99.8% with good health history
Departure continent	26% Asia 2.5% Africa 58% Europe 12% North America 0.5% South America 1% Oceania	87% Asia 2% Africa 8% Europe 2.5% North America 0.2% South America 0.3% Oceania	66.5% Asia 1% Africa 24% Europe 8% North America 0.5% South America
COVID-19 asymptomatic status	25% asymptomatic 75% symptomatic	76% asymptomatic 24% symptomatic	78% asymptomatic 22% symptomatic
Average flight duration (person-hour)	10.78	5.63	8.15

Table 4.4 displays statistics from the best-fit Exponential Random Graph Models for the three peaks' networks. A comparison of the predictors included in baseline models (null models) and best-fit models were conducted based on AIC values, with Monte Carlo empirical p values reported for each predictor. The best-fit-

models had lower AIC values compared to the baseline models, suggesting that the predictors included in the best-fit models were effective in predicting the presence of links. The goodness-of-fit Monte Carlo empirical p -values for the predictors included in best-fit-models were all larger than 0.05, indicating that all the predictors were a good fit.

Table 4.4 Baseline and Best-fit Exponential Random Graph Models (ERGMs)

	Baseline models		Best-fit models		
	Predictors	AIC value	Predictors	AIC value	Goodness-of-fit Monte Carlo empirical p -value
Peak one network	Links	78532	Links + age + flight duration + COVID-19 asymptomatic status + departure continent	77423	Links: 0.82 Age: 0.90 Flight duration: 0.90 Asymptomatic status: 0.84 Departure continent: 0.84
Peak two network	Links	32202	Links+ age + flight duration + COVID-19 asymptomatic status + departure continent	31420	Links: 0.62 Age: 0.70 Flight duration: 0.62 Asymptomatic status: 0.36 Departure continent: 0.60
Peak three network	Links	17525	Links+ age + flight duration + COVID-19 + departure continent	17265	Links: 0.90 Age: 0.88 Flight duration: 0.96 Asymptomatic status: 0.72 Departure continent: 0.80

Note: Smaller AIC is better. Goodness-of-fit Monte Carlo empirical p -value >0.05 means good fit of the predictor.

During the first peak, link formation was significantly negatively associated with age (OR: 0.995, 95%CI: 0.994-0.996) and flight duration (OR: 0.97, 95%CI: 0.96-0.98), suggesting that for each additional year of age, the odds of having link formation decrease by 0.5%, and for each additional hour of flight duration, the odds of having link formation decrease by 3%. For example, the probability of two cases with a 60-year-old and 50-year-old forming a link was 0.55%, while this probability increased to

0.79% for two cases with a 10-year-old and a 20-year-old. Asymptomatic cases had a positive and significant association with the likelihood of link formation ($p < 0.01$), with asymptomatic cases being 1.34 times (95%CI: 1.22-1.46) more likely to form a link than symptomatic cases. Departure continent was also significantly associated with link formation ($p < 0.01$), with cases departing from Asia had higher odds of forming a link compared to cases departing from other continents (OR: 1.38, 95%CI: 1.16-1.63) and cases departing from Europe (OR: 2.15, 95%CI: 1.83-2.54) and North America (OR: 2.56, 95%CI: 2.19-2.99) also having higher odds of forming a link.

During peak two, flight duration was positively associated with link formation (OR: 1.03, 95%CI: 1.02-1.05), suggesting that for each additional hour of flight duration, the odds of having link formation increase by 3%. For example, the probability of two cases taking a 6-person-hour inbound flight to form a link was 0.08%, while this probability increased to 0.2% for two cases who took a 20-person-hour flight. Asymptomatic cases were not significantly associated with link formation during peak two, but departure continent remained significant. Cases departing from Asia had the highest odds of forming a link compared to cases departing from other continents (OR: 3.51, 95%CI: 1.42-8.68), while cases departing from Europe also had higher odds of forming a link (2.65 times (95%CI: 1.61-4.34)). Cases departing from North America slightly lower odds of forming a link (OR: 2.07, 95%CI: 1.19-3.59).

In the third peak, flight duration was negatively associated with link formation (OR: 0.98, 95%CI: 0.97-0.99), meaning for each additional hour of flight duration, the odds of having link formation decrease by 2%. For instance, the probability of two cases taking a 3-person-hour inbound flight to form a link was 0.07%, while this

probability decreased to 0.05% for two cases who took a 15-person-hour flight. Asymptomatic cases were less likely to form a link compared to symptomatic cases (OR: 0.87, 95%CI: 0.78-0.87). Age and departure continent were not significantly associated with link formation. **Table 4.5** displays the odds ratios for predictors in best-fit Exponential Random Graph Models across the three peaks.

Table 4.5 Odds Ratios for Predictors in Best-fit Exponential Random Graph Models

Predictors	Peak 1		Peak 2		Peak 3	
	Odds ratios	<i>p</i> -value	Odds ratios	<i>p</i> -value	Odds ratios	<i>p</i> -value
Age (years)	0.996	<0.01*	0.994	<0.01*	1.001	0.35
Flight duration (person-hours)	0.977	<0.01*	1.034	<0.01*	0.983	0.03*
Asymptomatic for COVID-19	1.336	<0.01*	1.084	0.69	0.869	0.01*
Departure-Asia	1.376	<0.01*	3.507	0.01*	0.821	0.47
Departure-Europe	2.152	<0.01*	2.646	<0.01*	1.566	0.07
Departure-North America	2.562	<0.01*	2.072	0.01*	1.204	0.47

** denotes statistical significance because *p*-value is less than 0.05.*

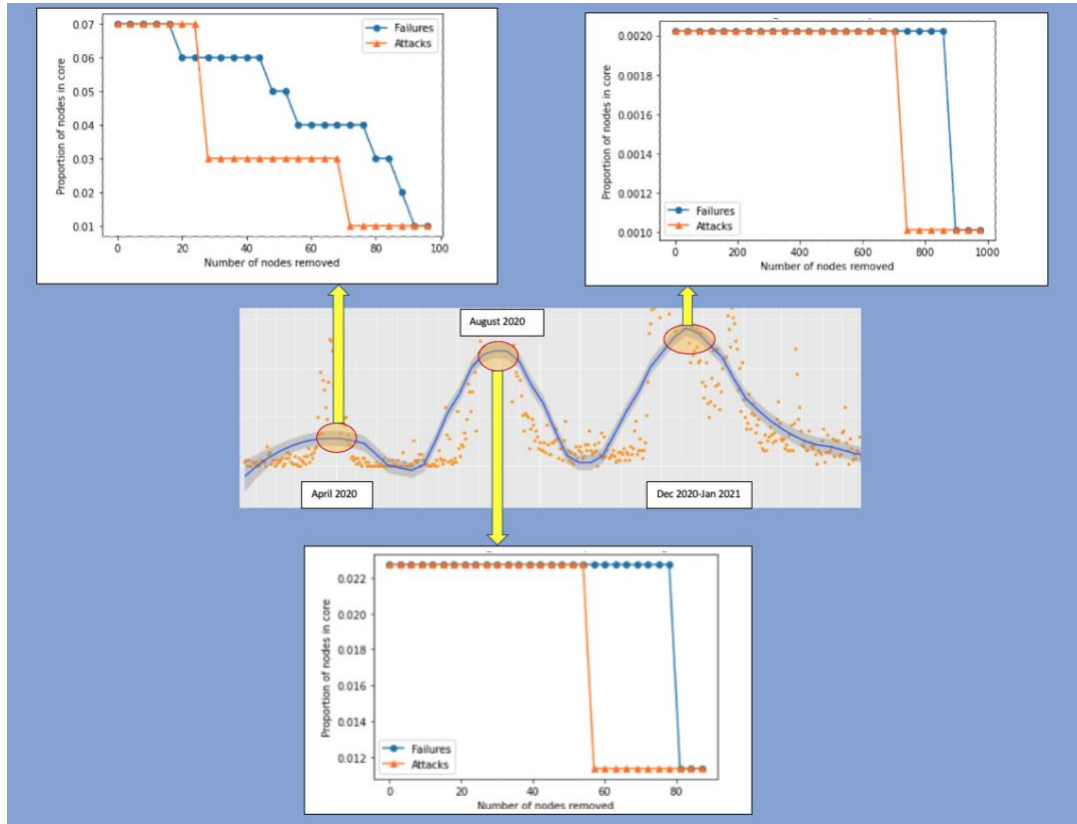
Network percolation

The dismantling of the networks of the three peaks through random failures and targeted attacks were simulated, and both methods showed a similar pattern, as illustrated in Figure 4.4. The blue line with circle-shaped dots represents random failures, while orange line with triangle-shaped dots represents targeted attacks. The x-axis shows the number of nodes removed during the simulations, while the y-axis represents the proportion of the most important and influential nodes in the network (known as the “core”) that were affected by nodes removal.

Targeted attacks were found to be more efficient than random failures in dismantling the three networks. This was evident from the illustration that number of

nodes removed by targeted attacks needed to reach a disproportionate impact on the overall connectivity of the network was less than that by random failures across the three peaks. Therefore, targeted attacks were better suited to stopping the spread of SARS-CoV-2 in inbound flight cabins and airports.

Figure 4.4. Network Percolation of the Three Peak Networks



4.4 Discussion

Network data exhibits a unique nature, with highly correlated observations that differ from traditional data, where observations are usually independent. Therefore, to interpret our results, we require both network visualization and quantitative analysis. Relying solely on network visualization for peak one (**Figure 4.1**) could lead to the erroneous conclusions that symptomatic cases were primarily responsible for the early-

stage spread of SARS-CoV-2 through inbound air travel to Hong Kong, but this conclusion could be biased due to underdetection of asymptomatic cases. Conversely, our ERGM results show that asymptomatic cases are more likely to connect than symptomatic cases during peak one, despite there being more symptomatic cases visualized. This critical message highlights to policymakers that asymptomatic cases did not contribute less than symptomatic cases at the early stage of the pandemic.

As the pandemic proceeded into peak two and three, asymptomatic cases became the majority of nodes in our network visualization (**Figure 4.2-4.3**), but with an increasing number of isolates. This increase is reflected on our ERGM results, during peak two, asymptomatic status was not significantly associated with the likelihood of forming a link. However, in peak three, despite the symptomatic cases having fewer nodes displayed in network visualization, asymptomatic cases had a lower likelihood of forming a link than symptomatic cases due to a large number of isolates. The combined results from network visualization and ERGM analysis suggest that during peak two and three, asymptomatic cases played a larger role in contributing the pandemic than symptomatic cases, primarily due to public health restrictions reducing the likelihood of symptomatic individuals travelling, and were more likely to contribute to the pandemic in a less connected manner.

Our research findings suggest that the optimal timing to contain the spread of SARS-CoV-2 in inbound flight cabins and airports was at the early stage of the pandemic, known as peak one. During this period, successfully dismantling the network required the removal of fewer than 80 nodes through targeted attacks or fewer than 100 through random failures. As the pandemic progressed, the network became

less connected, making it increasingly challenging to dismantle the fragmented network. By peak three, it was necessary to remove all nodes through targeted attacks or random failures to dismantle the network completely.

The fragmented nature of the three networks shown in **Figures 4.1-4.3** highlights the difficulty in stopping the transmission of SARS-CoV-2 in flight cabins and airports. Unlike well-connected networks, fragmented networks were much harder to deal with as clusters within these networks do not connect well with each other. Therefore, strategies focused on random failure, such as travel restrictions, may not be efficient in containing the COVID-19 pandemic.

Our findings suggest that targeted attacks were more efficient than random failures to dismantle the networks. However, implementing such strategies in real life is unrealistic, as authorities are unlikely to know where the highest degree large component of a network is located prior to a pandemic, meanwhile large component in a network is likely to be dynamic instead of static. Travel restrictions often targeted individuals from specific geographical regions, similar to targeting a random cluster in our networks. However, banning individuals from a specific region is akin to random attacks in our network percolation, which only destroys the connectedness of one cluster while allowing the rest to continue spreading SARS-CoV-2.

Based on the key findings of our study, we suggest that policymakers implement the following implications for future epidemics or pandemics. First and foremost, early intervention is crucial. Prioritize should prioritize targeted measures during the early stages of outbreaks to prevent viruses from spreading, as it becomes more challenging to control later on. Additionally, Policymakers should instead

prioritize evidence-based public health strategies, such as wearing masks and maintaining social distance, to contain future pandemics. Furthermore, policymakers should be aware of the connectedness of the network in high-risk setting and take this into account when implementing measures to prevent the spread of viruses.

Finally, policy interventions should evolve over time. Policymakers should monitor the effectiveness of their interventions and adjust their policies according. This approach will allow them to implement evidence-based strategies that are tailored to the unique challenges of each outbreak, epidemic or pandemic.

The flight record data collected in this study are of good quality, as they are collected by contact tracers in Hong Kong and supported by objective data sources from airlines and the immigration department, rather than relying on self-reported data from cases which could be vulnerable to recall bias. The use the unique tracking IDs protected participants' identities while providing us with valuable information on cases' age, sex, and flight record. Our evaluation of travel restrictions using empirical individual-level networks provided critical findings that may have been missed by modeling studies using simulated networks. However, there were limitations to this study. We could not build directed networks to determine the source of infection, which limited our ability to apply certain measurements. Our study was limited by missing information on outbound air travel, which prevented us from examining whether similar patterns were found. Future studies should address these gaps through international collaboration and data collection, which are essential for better understanding the spread of infectious diseases and developing effective policies to control pandemics. By prioritizing collaboration and data collection, researchers and

policymakers can identify new insights and refine existing strategies, paving the way for more effective interventions that can safeguard public health and minimize the social and economic impacts of pandemics.

4.5 Conclusion

Our study underscores the significant role that asymptomatic cases played in the early stages of the COVID-19 pandemic, particularly in inbound flight cabins and airports. Early intervention is crucial, and policymakers should prioritize public health strategies such as wearing masks and maintaining social distance, as our findings suggest that travel restrictions may not be effective in containing the pandemic in fragmented networks like those found in inbound flight cabins and airports. We recommend that policymakers monitor the effectiveness of their interventions and adjust their policies over time to optimize their impact. By prioritizing evidence-based strategies and policy evolution over time, policymakers can better prepare for future outbreaks and pandemics, and safeguard the well-being of populations around the world.

Chapter 5: Spatial Analysis of Local Mobility for COVID-19 Asymptomatic and Symptomatic Cases in Hong Kong

Abstract

Background:

Learning COVID-19 cases' mobility is crucial for predicting and managing pandemics. Factors such as socioeconomic status, population density, and demographic have been linked to disease spread in certain regions.

Methods:

This study included 1,282 randomly selected COVID-19 cases (641 asymptomatic, 641 symptomatic) in Hong Kong. Local mobility was analyzed using point, Kernel density, and choropleth maps created from 2,850 geocoded coordinates from cases' local travel history. Spatial clustering patterns were explored using Global Moran's I and Average Nearest Neighbor. Geographically Weighted Regression (GWR) and Ordinary Least Squares Regression (OLS) were used to examine the relationship between local mobility and socioeconomic factors, health and elderly centers, and transportation locations. We estimated 19,568 least-cost-paths using Origin-Destination Cost Matrix. The travel time differences between asymptomatic and symptomatic cases were compared using Wilcoxon rank-sum test.

Results:

According to the kernel density maps, locations visited by asymptomatic cases were mostly clustered in the southern part of Hong Kong, while symptomatic cases visited locations clustered in the middle and southern parts. Density of locations changed similarly for both asymptomatic and symptomatic cases during the three peaks, with the highest concentration of locations visited by cases near Victoria Harbor posing the highest risk of COVID-19 infection. In predicting local mobility, GWR outperformed

OLS in predicting local mobility (adjusted R^2 : 40% vs. adjusted R^2 : 35%), with a better performance among symptomatic cases (symptomatic: standardized residues z-score: 1.13, p -value: 0.26 vs. asymptomatic: z-score: 1.72, p -value: 0.09). Local travel time was higher among asymptomatic cases (median: 68.1 person-minutes; 25% Q1: 50.2-75% Q3: 98.2) than symptomatic cases (median: 59.5 person-minutes; 25% Q1: 42.3-75% Q3: 77.8, $p < 0.0001$). The presence of health and elderly centers, transportation characteristics, and socioeconomic factors in a geographical region did not show significant prediction on the local mobility of asymptomatic cases.

Conclusion:

Asymptomatic COVID-19 cases exhibited higher local mobility prior to diagnosis than symptomatic cases, indicating a higher risk of spreading SARS-CoV-2 via mobility. Both groups' mobility showed significant clustering. To address future infectious disease outbreaks at early stages, local health authorities and policymakers should tailor detection and containment strategies by considering the spatial variability in different areas. Public health officials can prioritize resources such as testing and contact tracing in areas with higher local mobility and implement measures to limit mobility in these areas to reduce the risk of transmission.

5.1 Introduction

Fully enumerating all people that have been exposed to SARS-CoV-2 in local communities can be ideal but unrealistic. Robins suggested that the contact diary is a useful tool to understand contact structures of airborne infectious disease transmission in communities. Specifying places that infected people have visited is a practical alternative to identify contact individuals, and still can provide critical knowledge about how the disease was transmitted (Robins, 2015). Learning COVID-19 cases' mobility at the local is extremely important to understanding and predicting changes in the pandemic at the country-level (Jiang et al, 2021; Mansour, Kajtazi & Ghazawneh, 2022).

Spatial networks, particularly transportation networks, are useful in uncovering community mobility during the COVID-19 pandemic. Transportation, especially urban commuter networks, provide information on the movements of populations in urban communities (Barthélemy, 2011). Previous studies have analyzed road networks and social networks to identify clusters of pediatric diarrheal disease in Bangladesh (Perez-Heydrich et al, 2013), and community mobility in transport stations, retails, parks, and working sites in India (Saha, Mondal & Chouhan, 2021). In spatial networks, nodes refer to geographic entities and links represent roads that connect these entities (Doytsher, Galon & Kanza, 2010). The Origin-Destination (OD) Matrix, is spatial network approach, has been widely used to explore population mobility during COVID-19 pandemic.

Socioeconomic status, population density, and demographic factors have been identified as associated with spread of disease in specific geographic regions. For

instance, McCaw et al (2010) found that households characteristics, such as household size, were significantly associated with the composition of encounter networks for airborne disease transmission. Spatial point event analysis has been utilized in a few epidemiological studies, including English et al. (1999)'s investigation of pediatric asthma data from California's Medicaid program to examine the associations between traffic flow and pediatric asthma in patients with and without asthma.

Finding a spatial database that precisely matches public health research interests is challenging (Waller & Gotway, 2004). Obtaining detailed data about population movements during COVID-19 pandemic is difficult. Some researchers propose using social media data as a proxy to detect mobility (Aljeri, 2020). In addition, it is difficult to explore the spatial distribution of COVID-19 due to limited GIS resources (Pedrosa & Albuquerque, 2020). Only a few epidemiological studies have examined the spatial scales of behaviors or diseases (Waller & Gotway, 2004).

No previous studies have investigated the agent-based mobility and spatial patterns of COVID-19 cases in Hong Kong, nor have any studies estimated travel time using real travel records of COVID-19 cases. This study aims to fill these critical knowledge gaps by examining potential determinants (e.g., population density, medical facilities, and transportation options) that influence COVID-19 cases' mobility within the Hong Kong territory 14-day prior diagnoses. Additionally, we will explore differences between asymptomatic and symptomatic cases.

5.2 Methods

Date source

The Hong Kong Department of Health's Center for Health Protection began sharing publicly accessible reports containing de-identifiable information about COVID-19 cases on two websites since the first case was reported on January 23, 2020. Data were gathered from 1) daily situation reports of COVID-19 at <https://www.coronavirus.gov.hk/eng/index.html> and 2) contact tracing reports of each case at <https://wars.vote4.hk/en/cases>. Information from both websites was merged using each case's unique tracking ID. **Figure 3.1** displays the LOESS curve of the three peaks of COVID-19 diagnoses in Hong Kong.

Physical addresses of cases' travel records within Hong Kong prior to their COVID-19 diagnosis were obtained from the aforementioned daily situation and contact tracing reports. Data from daily situation reports included the asymptomatic status of each COVID-19 case: 1) asymptomatic, or 2) symptomatic with a specific date of onset. Spatial feature layers data related to Hong Kong's geographical division, population density, health centers, and locations of transportation were collected from Esri China (HK) (<https://opendata.esrichina.hk/>).

This study included a sample of 1,282 COVID-19 cases for spatial analysis, consisting of 641 asymptomatic cases and 641 symptomatic cases, who were selected at random from a total of 2,379 COVID-19 cases with group infection records across the three peaks, as these individuals were associated with super-spreader events. Initially, 641 asymptomatic cases with group infection records within Hong Kong were included in the study, and then, to match the sample size of asymptomatic cases, 641

randomly selected symptomatic cases with group infection records within Hong Kong were included for spatial analysis.

A total of 2850 physical addresses (1260 for asymptomatic cases, 1588 for symptomatic cases) were collected from the 1,282 COVID-19 cases' travel record within Hong Kong before their diagnosis. These 2850 physical addresses were geocoded into coordinates and saved as point features. Point features were projected to Hong Kong territory shapefiles for spatial analysis. Two geographical division methods were used for Hong Kong territory shapefiles: one divided Hong Kong into eighteen districts (Home Affairs Department) (**Figure 5.1**), while the other divided the territory into 154 Tertiary Planning Units (Planning Department, 2011) (**Figure 5.2**).

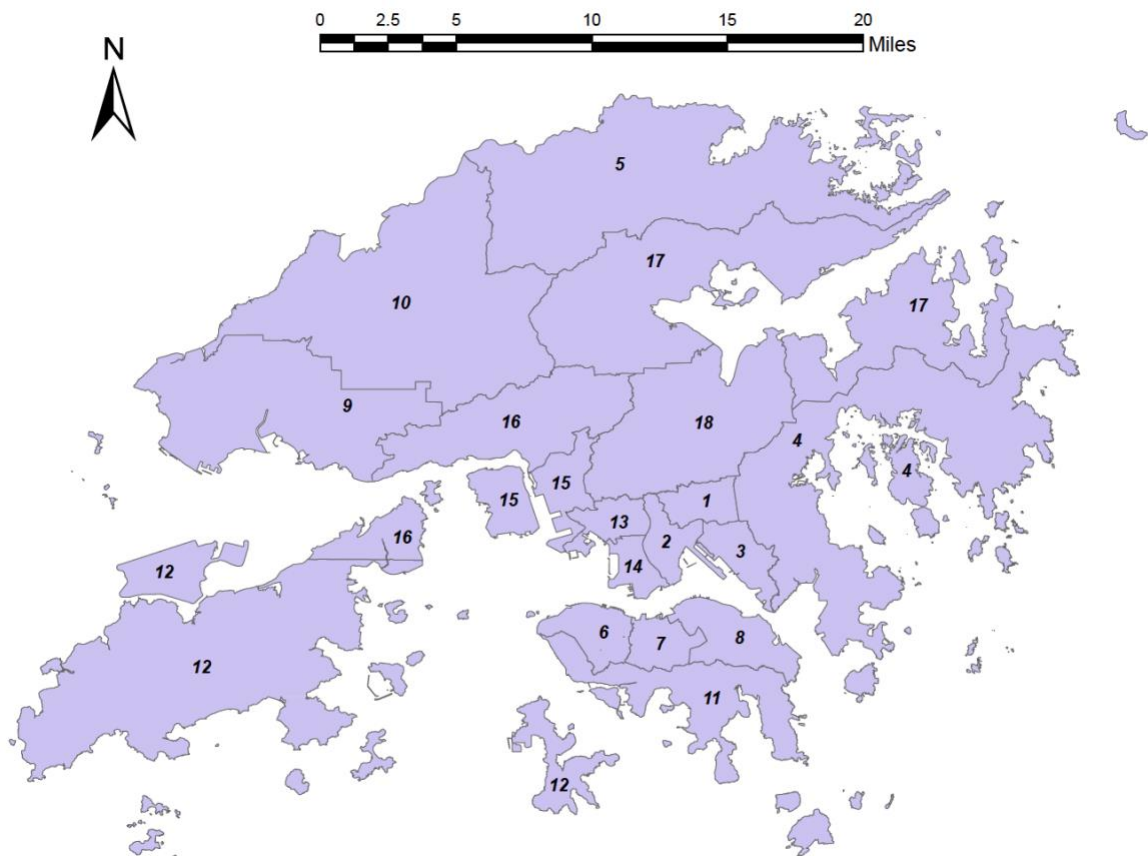
To visualize the spatial distribution of COVID-19 cases' local mobility in Hong Kong, three types of maps were made. Point maps were created to display cumulative geographic locations visited by asymptomatic and symptomatic cases within Hong Kong's 18 districts before diagnosis. Kernel density heat maps were generated to visualize the density hubs of asymptomatic and symptomatic cases' local mobility, capturing temporal changes in concentration of locations across the three peaks. Choropleth maps were created to continuously visualize cases' mobility within Tertiary Planning Units.

Human Subjects and Ethics Consideration

This study utilized publicly available, de-identified COVID-19 case data provided by the Center for Health Protection in Hong Kong Department of Health. Cases were tracked using unique tracking IDs created by the department, such as the first three confirmed COVID-19 cases with IDs 1, 2, and 3. These tracking IDs

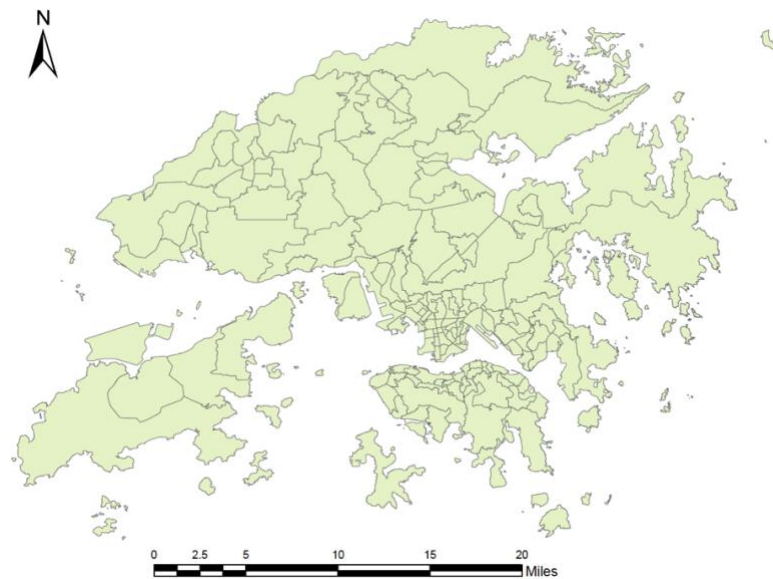
safeguarded the identities of the COVID-19 cases from the public while allowing researchers to monitor updates on hospitalization, discharge, or deceased status and contact tracing reports without revealing the cases' identities. Due to the publicly available and de-identification of the data, the University of Maryland College Park's Institutional Review Board deemed this study as non-human subject research and granted a formal waiver (IRB number 1882289-1).

Figure 5.1 Hong Kong's Eighteen Districts



1	Wong Tai Sin	10	Yuen Long
2	Kowloon City	11	Southern
3	Kwun Tong	12	Islands
4	Sai Kung	13	Sham Shui Po
5	North	14	Yau Tsim Mong
6	Central & Western	15	Kwai Tsing
7	Wan Chai	16	Tsuen Wan
8	Eastern	17	Tai Po
9	Tuen Mun	18	Sha Tin

Figure 5.2 Hong Kong's 154 Tertiary Planning Units



Outcome Variable

COVID-19 cases' local mobility is a continuous variable measured by the number of spatial point features representing physical locations visited by cases within Hong Kong territory before diagnosis.

Explanatory Variable

Population distribution by sex: A feature layer indicating Hong Kong labor population distribution by 154 Tertiary Planning Units in 2016, categorized by sex (male, or female).

Population distribution by education attainment: A feature layer indicating Hong Kong population distribution by 154 Tertiary Planning Units in 2016, categorized by residents' education attainment (no schooling, primary education, or post-secondary).

Population distribution by employment status: A feature layer indicating Hong Kong population distribution by 154 Tertiary Planning Units in 2016, categorized by residents' employment status (employee, employer, self-employed, or unemployed).

Population distribution by household composition: A feature layer indicating Hong Kong population distribution by 154 Tertiary Planning Units in 2016, categorized by household composition (composed of couples, couples with unmarried children, or single person).

Population distribution by working hours: A feature layer indicating Hong Kong population distribution by 154 Tertiary Planning Units in 2016, categorized by working hours per week (less than 16 hours per week, 35-40 hours, or greater than 60 hours).

Population distribution by household income: A feature layer indicating Hong Kong population distribution by 154 Tertiary Planning Units in 2016, categorized by household income per month (less than 6k HK dollars, 10k-19k HK dollars, or greater than 30k HK dollars).

Population distribution by number of household rooms: A feature layer indicating Hong Kong population distribution by 154 Tertiary Planning Units in 2016, categorized by the number of rooms in domestic households (single-room, two-rooms, or greater than three rooms).

Clinics and health centers: A feature layer indicating the geographical locations of Hong Kong Department of Health-affiliated clinics and health centers.

COVID-19 community testing centers: A feature layer indicating the geographical locations of COVID-19 testing centers in Hong Kong communities.

Elderly centers: A feature layer indicating the geographical locations of day care centers in Hong Kong for elderly residents.

Railway entrances: A feature layer indicating the geographical locations of entrances to Hong Kong's railway system, referred to as metro or underground in other regions.

Metered parking spaces: A feature layer indicating the geographical locations of Hong Kong's metered parking spaces.

Statistical methods

Kernel density analysis was conducted to visualize the concentration of point features in this study. We set the search radius to 0.025 km for optimal visualization. Output raster data was created and visualized using heat maps to display the concentration of point features across the 18 districts of Hong Kong.

To investigate clustering patterns of spatial distribution for point features, we employed two spatial statistics tools: Global Moran's I (spatial autocorrelation) and Average Nearest Neighbor. Global Moran's I analysis assesses spatial autocorrelation based on feature location and values, while Average Nearest Neighbor analysis examines the average geographical distance between features and their nearest neighboring features.

Geographically Weighted Regression (GWR) and Ordinary Least Squares (OLS) Regression models were performed to examine spatial heterogeneity in our data and to specifically explore the spatially varying relationship between our outcome variable (COVID-19 cases' local mobility) and explanatory variables (socioeconomic factors, health and senior centers, and transportation locations). Explanatory variables were transformed from feature layers to numeric variables and spatially joined with outcome variable for regression analysis with 154 features. Adjusted R-squared value

(0 to 100%) was used to measure model goodness of fit, and AIC value was utilized to compare model performances. A greater adjusted R-squared value indicates a better model fit, while a smaller AIC value indicates a better balance between goodness of fit and model complexity. The best-fit-models were selected by comparing AIC and adjusted R-squared values.

Spatial networks were created for the Origin-Destination Cost Matrix using Hong Kong Road Networks and feature layers containing point features. Least-cost-paths between origins and destinations were estimated by the Origin-Destination Cost Matrix and measured in person-minutes. In total, 200 coordinates (100 for asymptomatic, 100 for symptomatic) were randomly selected from the geocoded 2850 coordinates as origins and destinations. Origins and destinations were set to not self-loop in this study. There were 19,568 least-cost paths (9,688 for asymptomatic, 9,880 for symptomatic) estimated by the Origin-Destination Cost Matrix. Travel time for the 19,568 least least-cost paths were retrieved, and normality test was used to examine if the travel time data among cases was normally distributed. As our travel time data violated the normality assumption, a nonparametric Wilcoxon rank-sum test was conducted to determine if the median local travel time between asymptomatic and symptomatic cases significantly differed.

As part of this study, 0.7% of data (16 out of 2379 cases) had missing information on cases' local mobility in 14 days before diagnosis and were therefore excluded from data collection. The statistical significance level in this study was set at 0.05, with *p*-values smaller than 0.05 considered statistically significant. Spatial analysis, spatial statistical analysis, and map creation were conducted using ArcGIS

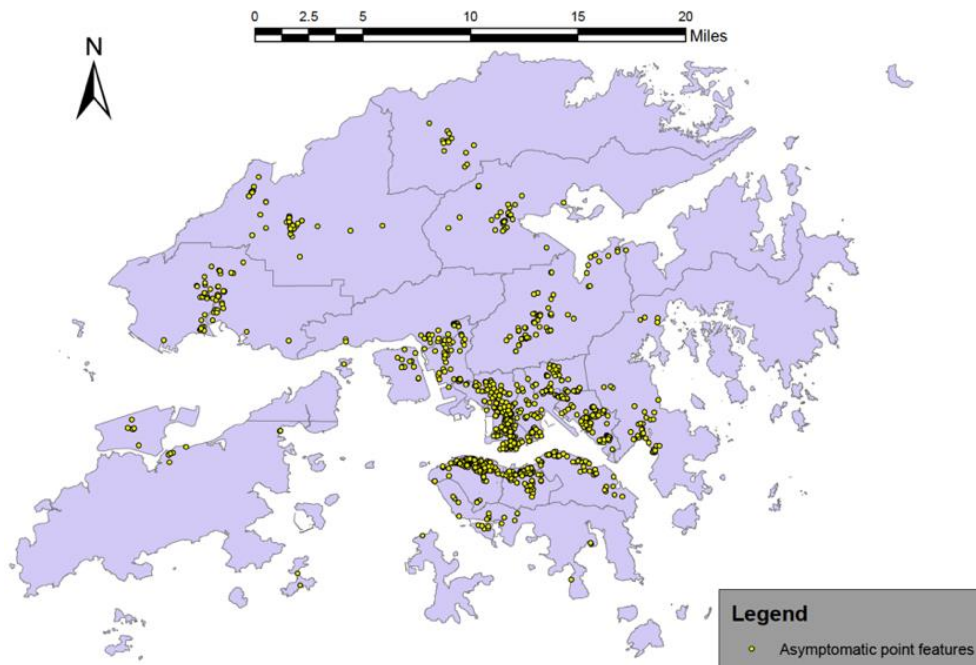
10.8.2, while the statistical comparison test was conducted using SAS OnDemand for Academics.

5.3 Results

COVID-19 Cases' Local Mobility Visualization

The figures presented in this study illustrate the geographical distribution of COVID-19 cases' local mobility within Hong Kong. **Figure 5.3** displays point maps of cumulative point locations visited by asymptomatic and symptomatic cases in Hong Kong's 18 districts. The maps use yellow points to represent locations visited by asymptomatic cases and red points to represent symptomatic cases. The distribution points between asymptomatic and symptomatic cases appears to be quite similar.

Figure 5.3 Point Maps of Locations Visited by COVID-19 Cases



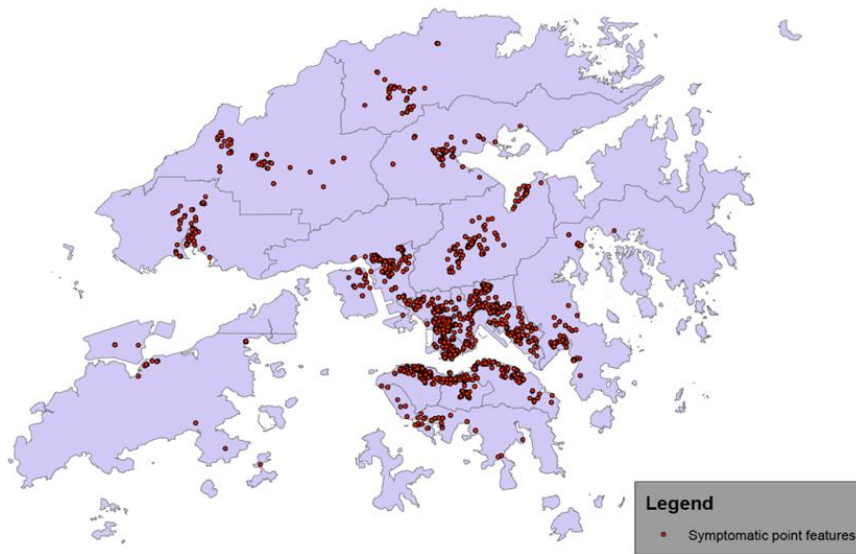
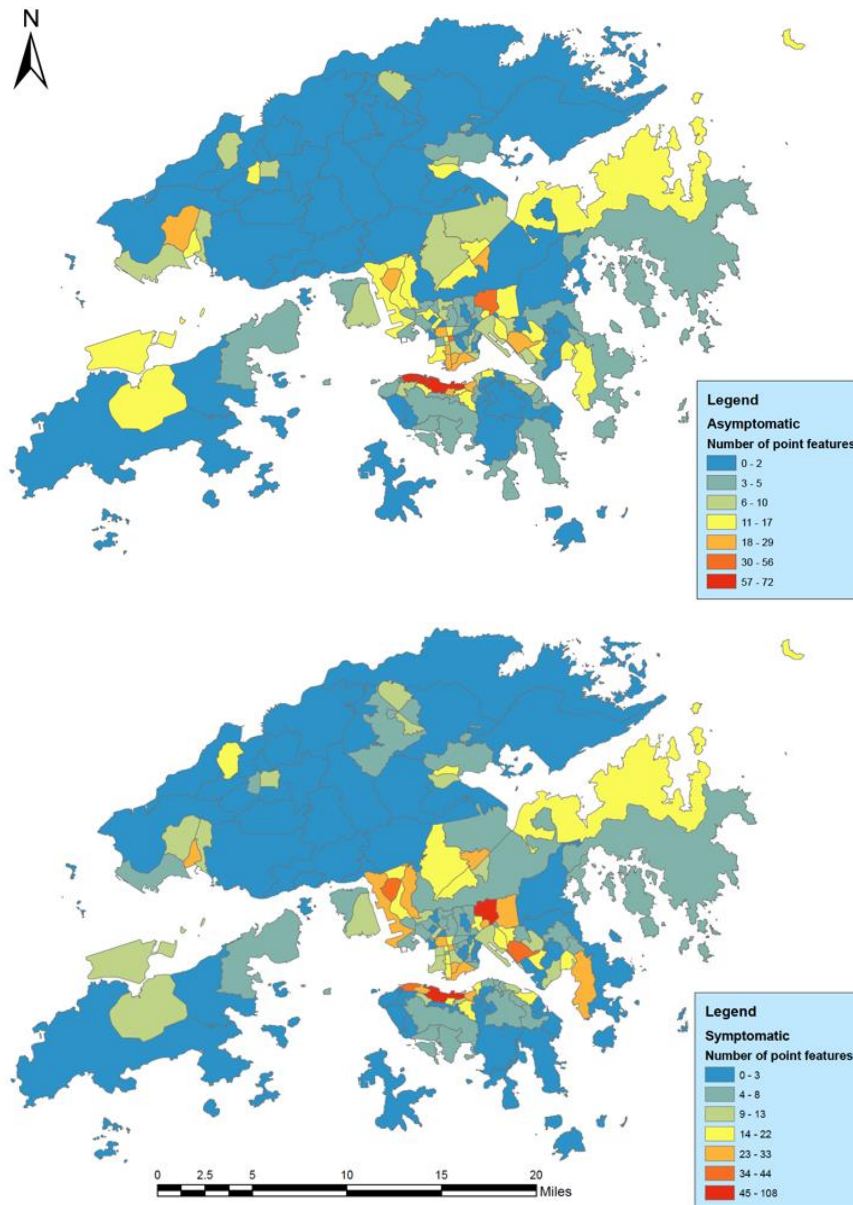


Figure 5.4 shows choropleth maps of cumulative numbers of point locations visited by asymptomatic and symptomatic cases in Hong Kong's 154 tertiary planning units. In this map, point features are represented as count numbers in each planning unit. The tertiary planning units with the lowest number of point features are shown in blue and are mostly located in the North. The units with the highest number of point features are shown in red and are mostly located in the South.

Kernel density heatmaps reveal clustering of COVID-19 cases' local mobility in Hong Kong. The heatmaps use colors to represent concentrations, with red indicating the highest concentration and yellow representing lowest. **Figure 5.5** and **Figure 5.6** display the cumulative concentration of locations visited by asymptomatic and symptomatic cases, respectively, across the three peaks. Additionally, each figure includes a point map of the region with the highest concentration of visited locations within Hong Kong Road Network. **Figure 5.7-5.9** shows areas of concentration of locations for asymptomatic and symptomatic cases' mobility during each peak.

Figure 5.4 Choropleth Map of Cumulative Number of Point Features in 154 Tertiary Planning Units in Hong Kong



Locations visited by asymptomatic cases were mostly clustered in the southern part of Hong Kong, particularly in Yau Ysim Mong, Central & Western, Wong Tai Sin, and Wan Chai (**Figure 5.5** up-map). The highest density districts for asymptomatic cases are characterized by yellow point features distributed along the Hong Kong Road Networks in urban areas (**Figure 5.5** down-map), rather than forests or mountain areas. In contrast, symptomatic cases mostly visited locations clustered in the middle and

southern parts of Hong Kong, particularly in Wong Tai Sin, Yau Ysim Mong, Wan Chai, and the intersection area between Kawi Tsing and Tsuen Wan (**Figure 5.6** up-map). The highest density districts for symptomatic cases are characterized by red point features distributed within the Hong Kong Road Networks (**Figure 5.6** down-map). Overall, the point features of both asymptomatic and symptomatic cases show a strong association with the Hong Kong Road Networks.

Temporal change in density show a similar trend between asymptomatic and symptomatic cases. In peak one, most cases visited the southern districts near Victoria Harbor (**Figure 5.7**). During peak two, the density hub shifted northward to areas near Wong Tai Sin District (**Figure 5.8**). In peak three, cases were distributed in both northern and southern districts between Victoria Harbor (**Figure 5.9**), and the density hubs were larger than the previous peaks. The findings from kernel density maps suggest that regions near Victoria Harbor pose the highest risk of COVID-19 infection, as they have the highest concentration of locations visited by COVID-19 cases.

Spatial Distribution Patterns of Point Features

Table 5.1 displays the statistics of Global Moran's I and Average Nearest Neighbor analyses. The results from the Global Moran's I analysis suggest that the spatial distribution of point features for locations visited by both asymptomatic (z-score: 4.999, p -value<0.001) and symptomatic cases (z-score: 2.032, p -value: 0.042) in this study are spatially clustered. Furthermore, the results from the Average Nearest Neighbor analysis indicate that the point features for both asymptomatic (z-score: -49.451, p -value <0.001) and symptomatic cases (z-score: -57.718, p -value <0.001) are also significantly clustered.

Figure 5.5 Kernel Density Map for Cumulative Mobility of Asymptomatic Cases and Point Map for the Highest Density Region

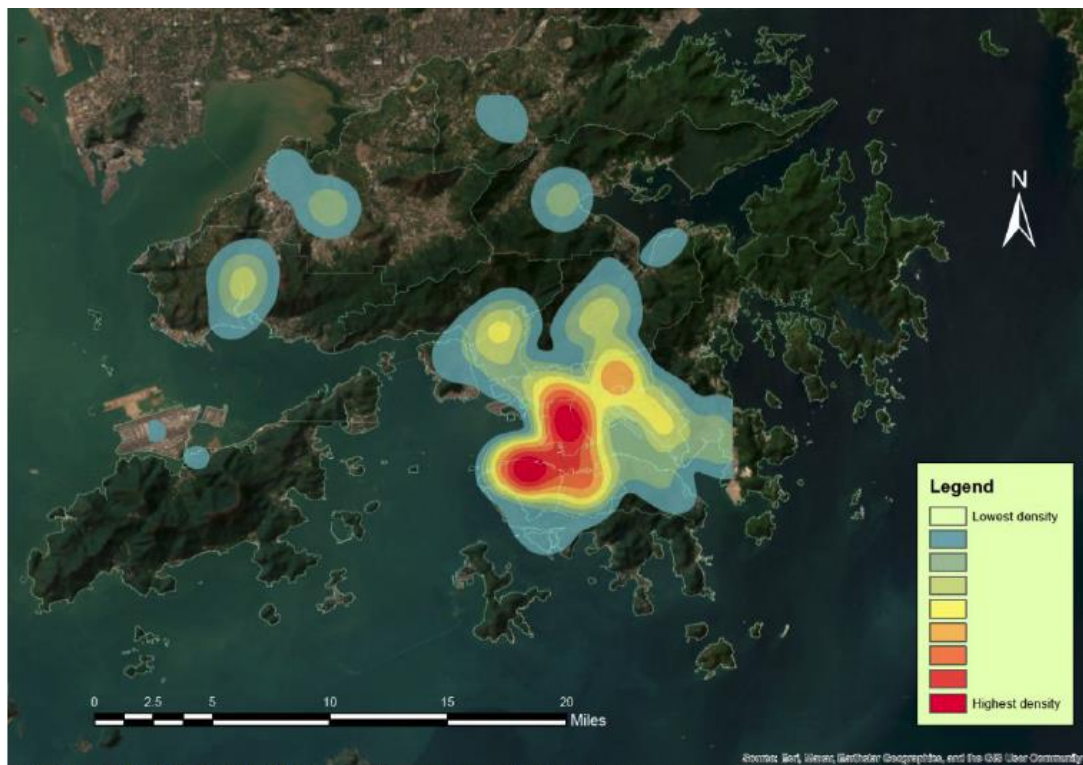


Figure 5.6 Kernel Density Map for Cumulative Mobility of Symptomatic Cases and Point Map for the Highest Density Region

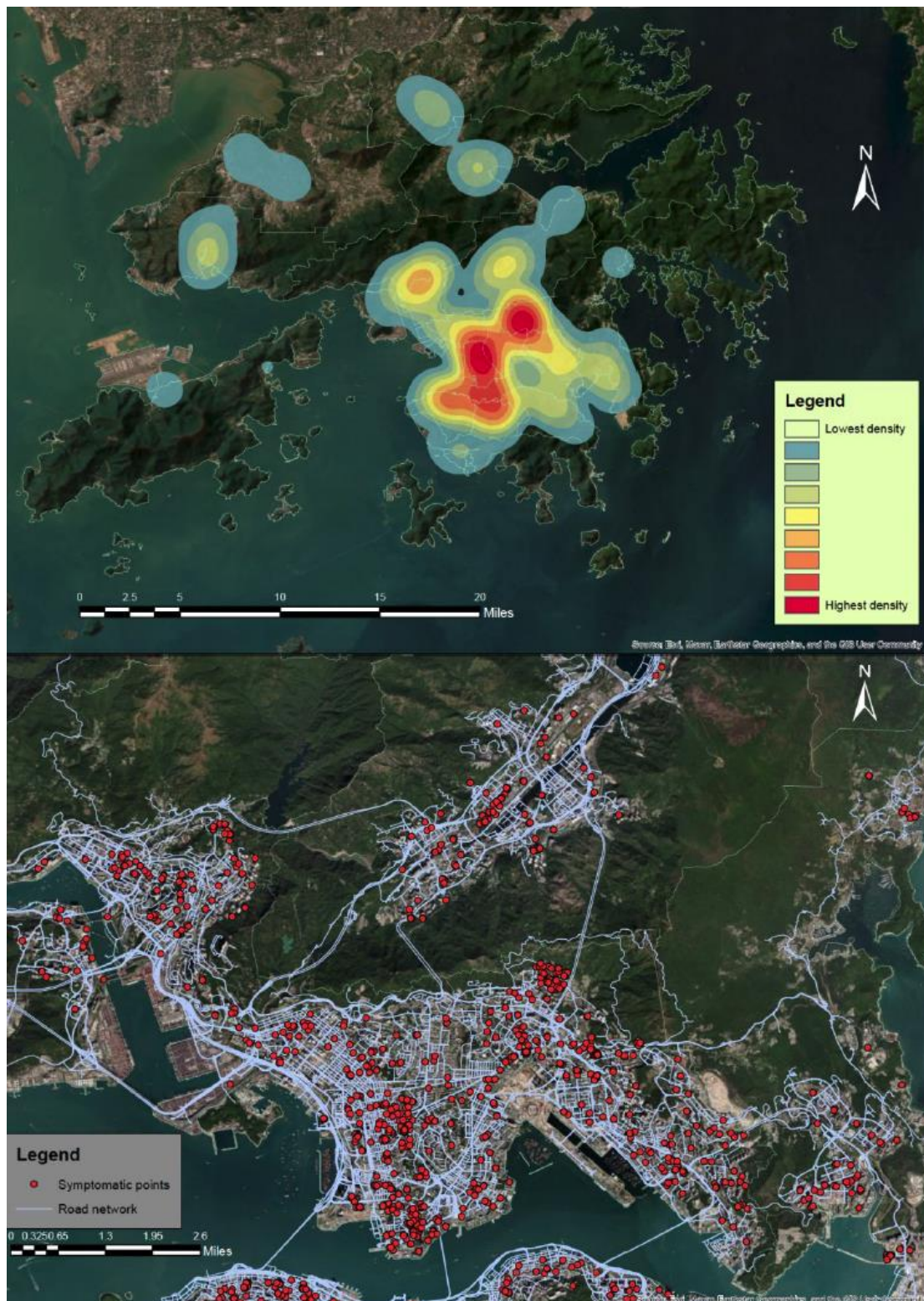


Figure 5.7 Kernel Density Map for Mobility during Peak 1: Comparison between Asymptomatic and Symptomatic Cases

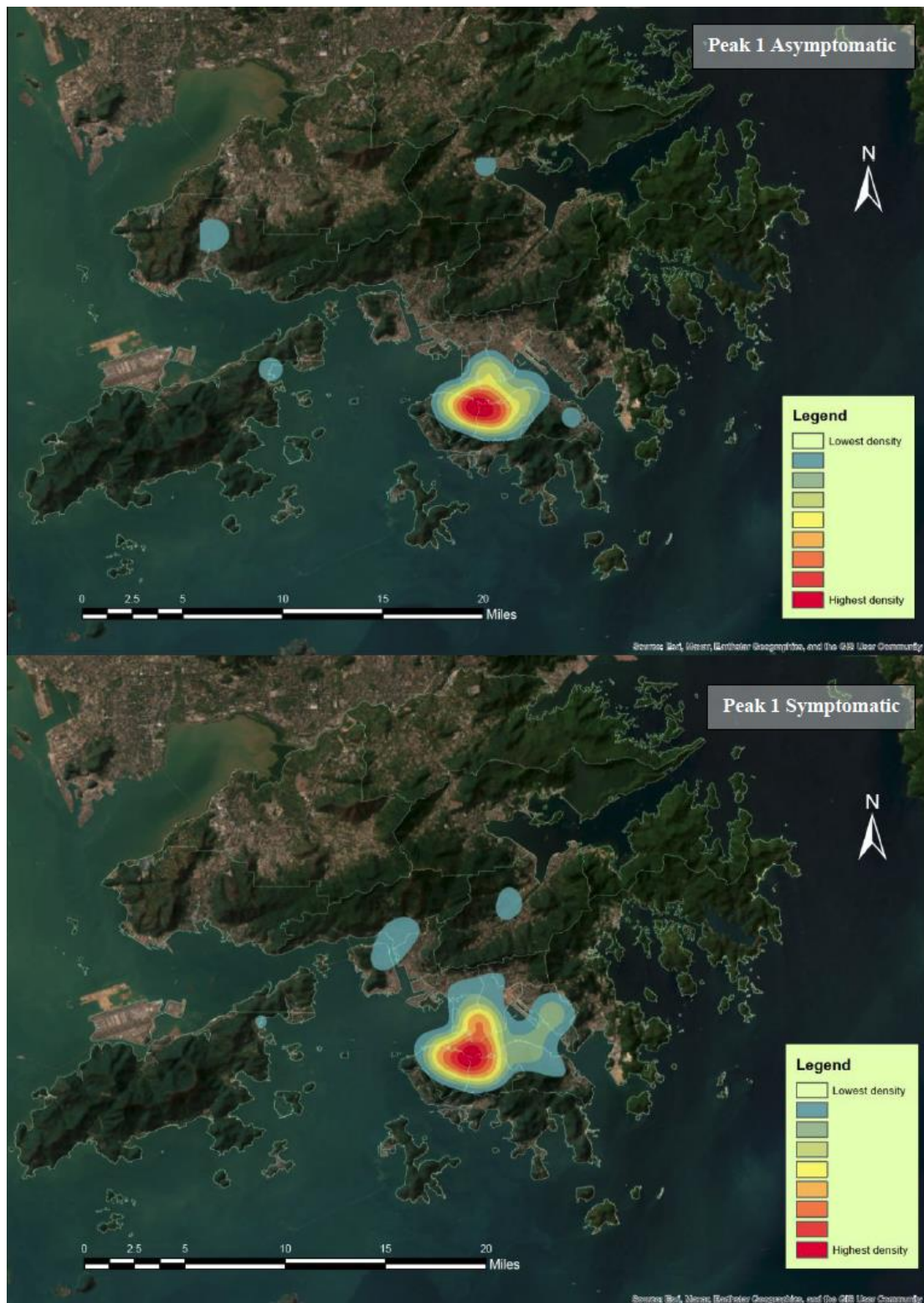


Figure 5.8 Kernel Density Map for Mobility during Peak 2: Comparison between Asymptomatic and Symptomatic Cases

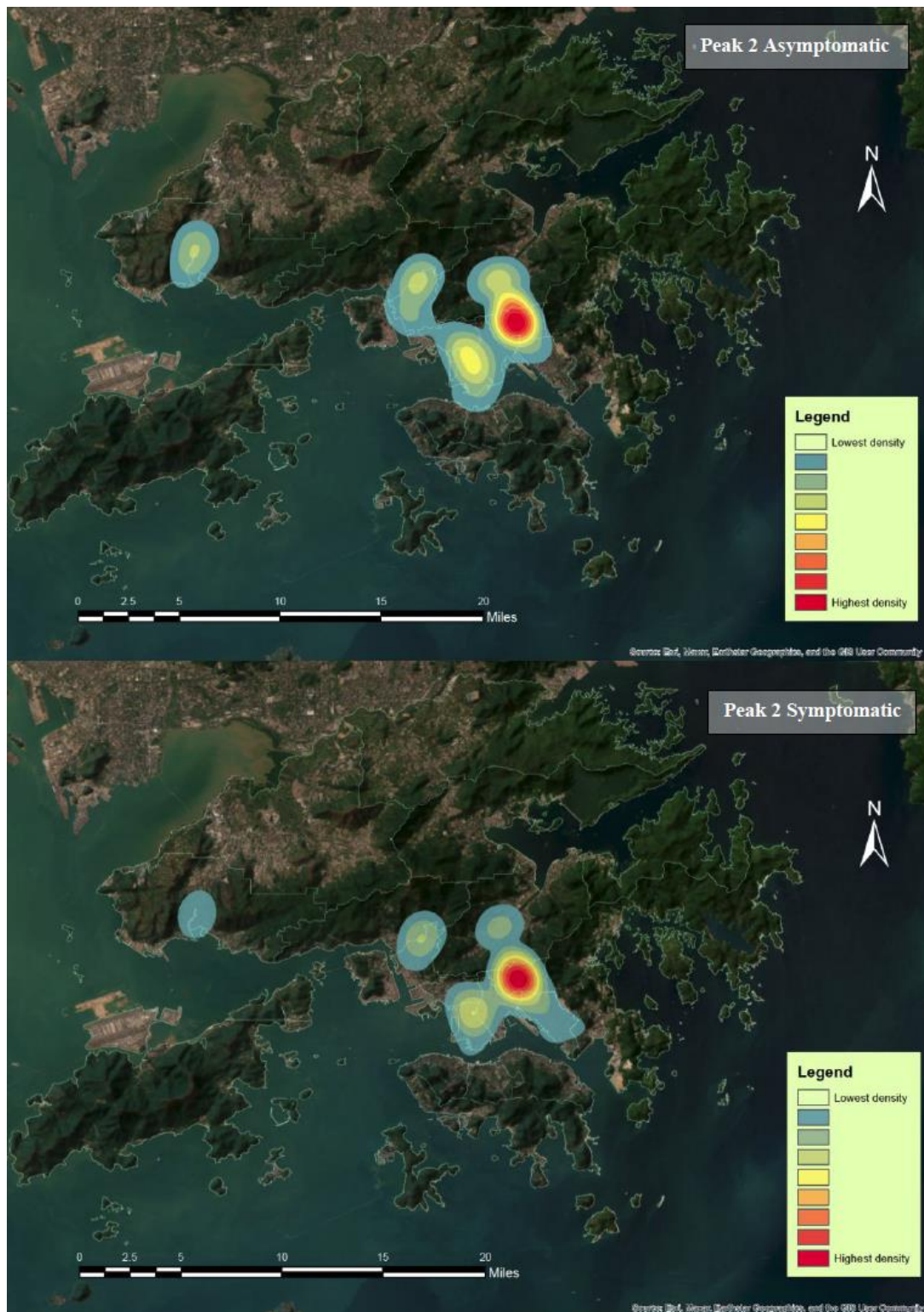


Figure 5.9 Kernel Density Map for Mobility during Peak 3: Comparison between Asymptomatic and Symptomatic Cases

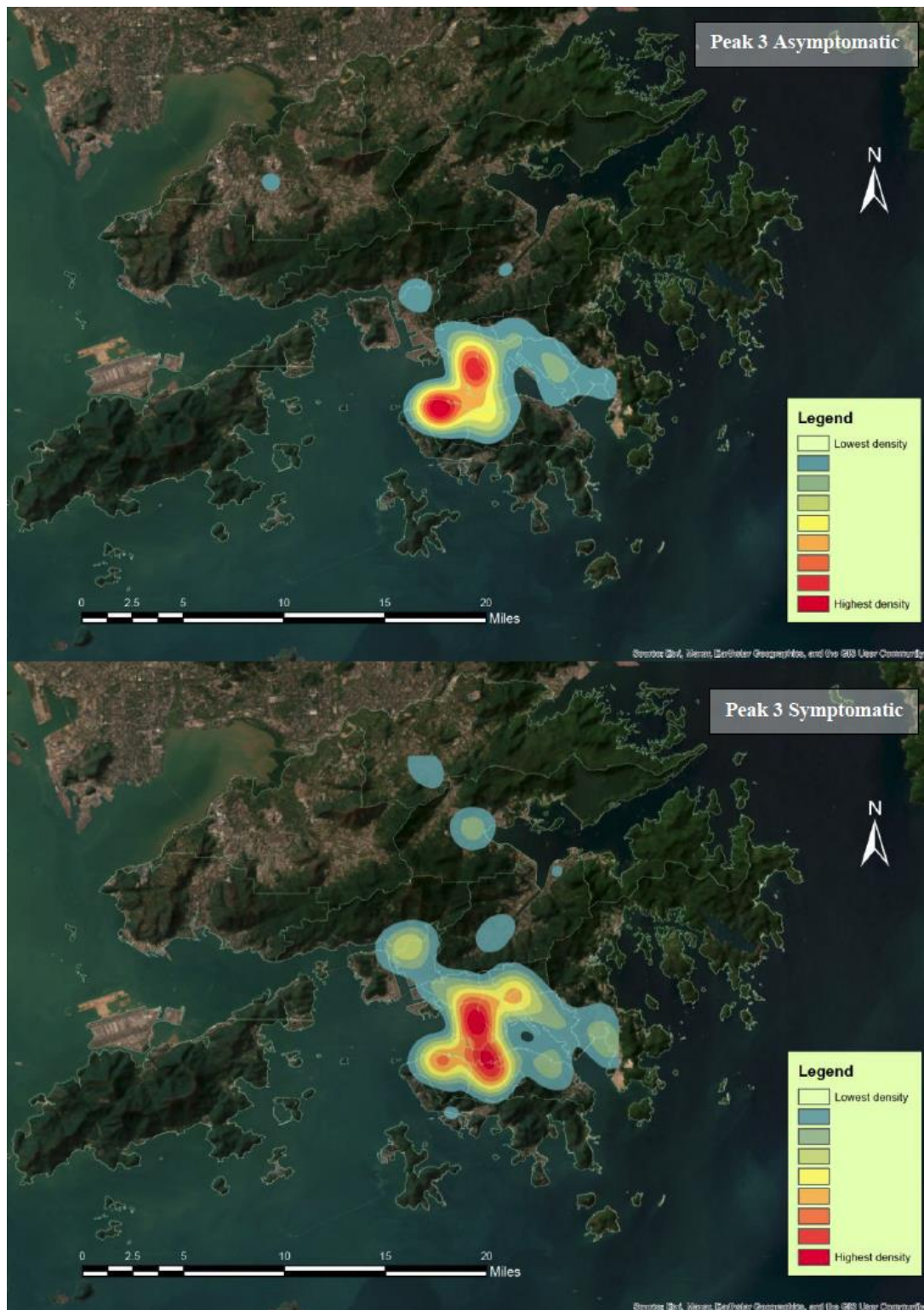


Table 5.1 Statistics of Global Moran's I and Average Nearest Neighbor Analysis

		Asymptomatic point features	Symptomatic point features
Global Moran's I	Distance method	Euclidean	Euclidean
	Moran's index	0.076	0.027
	z-score	4.999	2.032
	p-value	0.000001*	0.0421*
Average Nearest Neighbor	Distance method	Euclidean	Euclidean
	Nearest neighbor ratio	0.271	0.242
	z-score	- 49.451	- 57.718
	p-value	0.0001*	0.0001*

* denotes statistical significance because p-value is less than 0.05.

Relationship between COVID-19 cases' Local Mobility & Neighborhood Characteristics

The final models of asymptomatic and symptomatic cases included different explanatory variables from those examined in this study. **Table 5.2** displays the explanatory variables included in the final models, selected to best fit the outcome variable of the number of locations visited by COVID-19 cases. Population distribution by socioeconomic factors such as employment status and household income, and locations of health and senior centers improved the prediction on COVID-19 cases' local mobility, as evidenced by a reduction in AIC values and an increase in adjusted R-squared values of the models. The locations of transportation, such as railway entrances, improved the prediction of local mobility of symptomatic cases but not for asymptomatic cases. The population distribution by sex did not improve the performance of the models and was therefore excluded from the final analysis.

Geographically Weighted Regression (GWR) outperformed Ordinary Least Squares Regression (OLS) in predicting outcomes in this study, with higher adjusted R-squared values for both asymptomatic and symptomatic cases (**Table 5.3**). The best-fit GWR model explained 39.97% of the variance for asymptomatic cases and 42.7% for symptomatic cases. **Figure 5.10** shows the local R-squared values of the

best-fit GWR model for the 154 Tertiary Planning Unit (TPU) in Hong Kong, with dark blue polygons representing TPUs with the lowest local R-squared values, and red polygons representing those with highest. The best-fit GWR model worked best for the TPUs in the south and west for asymptomatic cases and in the east for symptomatic cases.

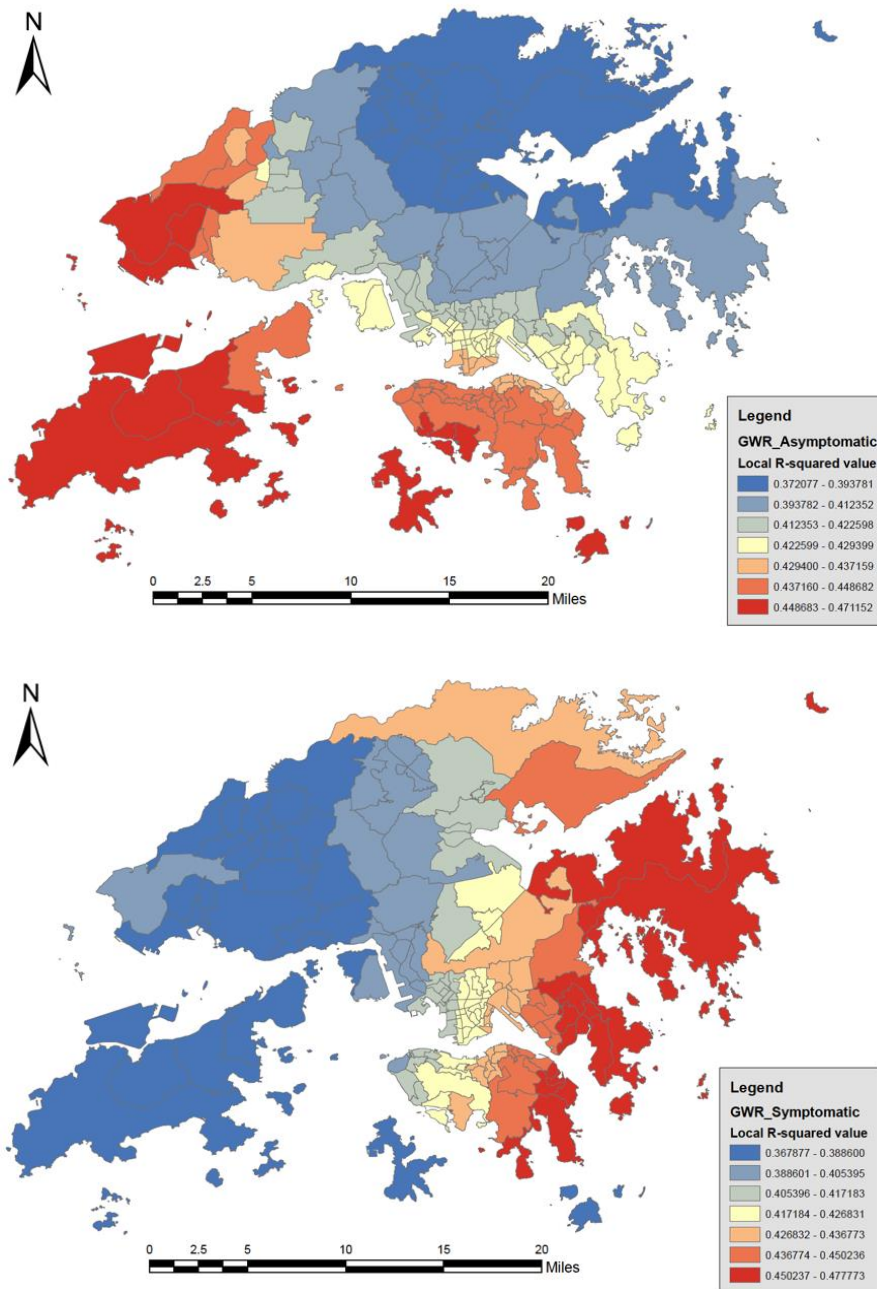
Table 5.2 Explanatory Variables Included in the Final Models

Final model of asymptomatic cases	Final model of symptomatic cases
1. Population distribution by education attainment (pre-primary education)	1. Population distribution by employment status (employees)
2. Population distribution by employment status (Unpaid family workers)	2. Population distribution by household composition (Composed of couple/one-person)
3. Population distribution by household composition (composed of couple/ couple and one child / one-person)	3. Population distribution by household income (HK\$10k-19k per month)
4. Population distribution by working hours (<18 hours per week)	4. Clinics and health centers
5. Population distribution by household income (<HK\$6k per month)	5. Elderly centers
6. Population distribution by number of household rooms (single room)	6. Railway entrances
7. Clinics and health centers	7. Metered parking spaces
8. COVID-19 community testing centers	
9. Elderly centers	

Table 5.3 R-Squared Values of Best Fit Models: Geographically Weighted Regression vs. Ordinary Least Squares Regression

	Geographically Weighted Regression		Ordinary Least Squares Regression	
	R² (%)	Adjusted R²(%)	R²(%)	Adjusted R²(%)
Best-fit-model for asymptomatic cases	48.6	39.97	39.9	35.3
Best-fit-model for symptomatic cases	48.8	42.7	36.97	33.5

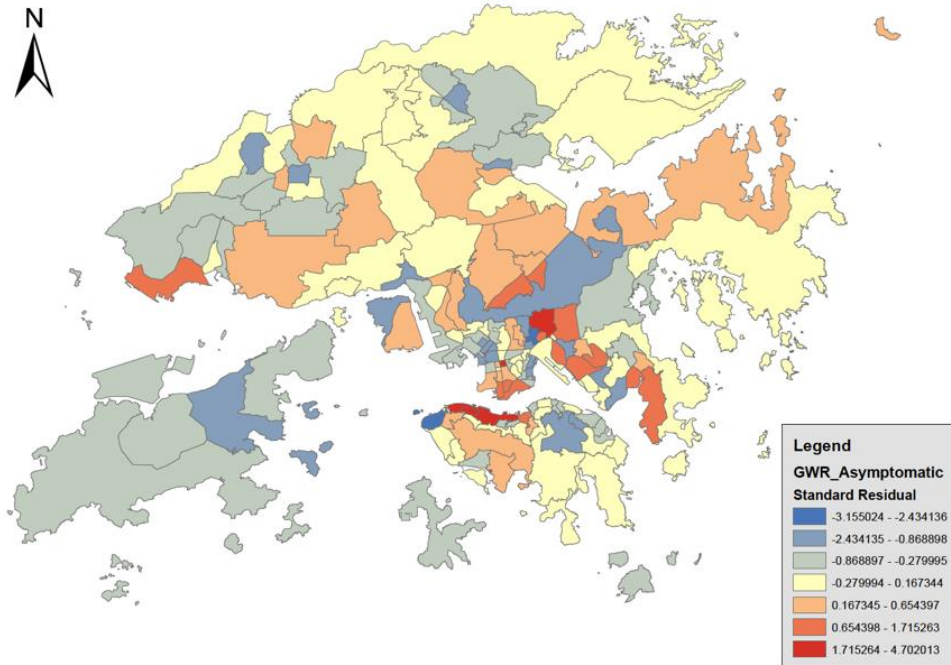
Figure 5.10 Choropleth Map of Geographically Weighted Regression Local R^2 Values

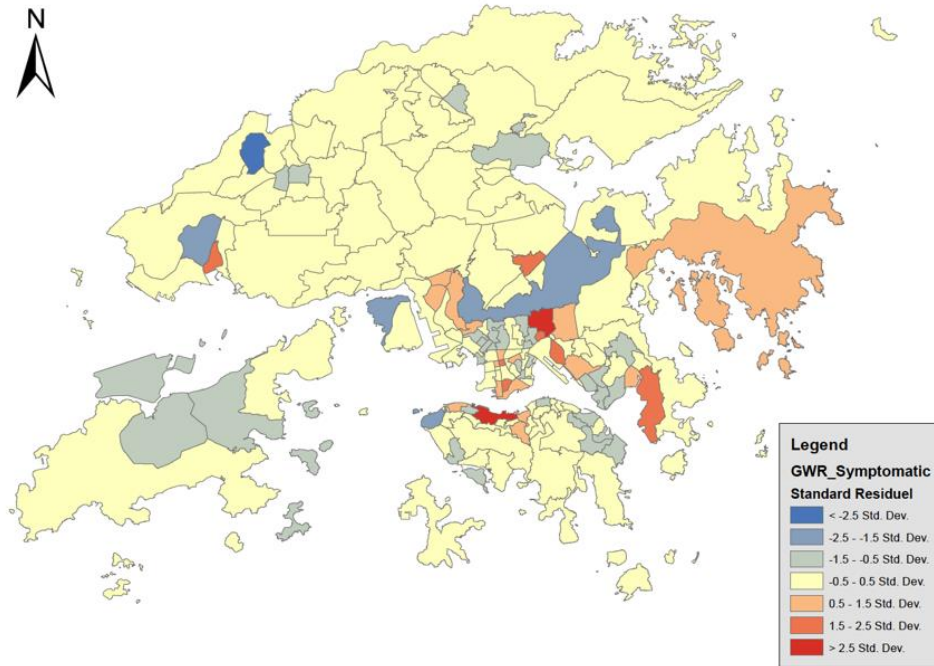


Global Moran's I was used to assess the spatial autocorrelation of standard residues from the best-fit GWR models for asymptomatic and symptomatic cases. The symptomatic GWR model showed randomly distribution of standardized residuals (z -score: 1.13, p -value: 0.26), while the asymptomatic GWR model displayed a non-

significant spatial clustering trend of standardized residuals (z-score: 1.72, p -value: 0.09). A random distribution of standardized residuals indicates that the model has capture all the important spatial patterns in the data and suggests that the model is a good fit for the data. The symptomatic GWR in this study serves as an example of a relatively good fit for the data. In **Figure 5.11**, dark blue polygons indicate underestimated Tertiary Planning Units, and red polygons indicate overestimated Tertiary Planning Units, meaning underestimation and overestimation were less frequent for symptomatic cases compared to asymptomatic cases. When the GWR model underestimates or overestimates the outcome for a location, it suggests that the model has not accurately captured the underlying spatial patterns of the data for that particular location.

Figure 5.11 Choropleth Map of Geographically Weighted Regression Standard Residuals





Local Travel Time Analysis: Asymptomatic vs. Symptomatic COVID-19 Cases

Figure 5.12 illustrates an example of the Origin-Destination Cost Matrix analysis. In this example, a COVID-19 case who visited 10 locations in Hong Kong prior to diagnosis was randomly selected. The locations were geocoded and represented by yellow points. The analysis estimated the least-cost paths (shown as yellow lines) along Hong Kong Road Network (blue lines within purple polygons) and calculated the travel time in person-minutes for each path.

In this study, we estimated the local travel time for COVID-19 cases using Origin-Destination Cost Matrix analysis. For asymptomatic cases ($n=9,688$ least-cost-paths), the median local travel time was 68.1 (25% Q1: 50.2–75% Q3: 98.2) person-minutes, while for symptomatic cases ($n=9,880$ least-cost-paths), it was 59.5 (25% Q1: 42.3–75% Q3: 77.8) person-minutes. However, local travel time data for both asymptomatic and symptomatic cases violated the assumption of normality (Kolmogorov-Smirnov test $p<0.01$). Therefore, we used a nonparametric Wilcoxon

rank-sum test to compare the local travel time between asymptomatic and symptomatic cases. The findings indicated that the median local travel time in person-minutes was statistically higher among asymptomatic cases compared to symptomatic cases ($p < 0.01$).

Figure 5.12 Illustration of Origin-Destination Cost Matrix Analysis



5.4 Discussion

The current study makes a significant contribution to the literature by using novel data and analysis techniques to more precisely explore COVID-19 cases' local mobility patterns. One of the strengths of this study is the use of agent-based data. Instead of using dot density maps, this study utilized point maps that accurately display the locations visited by the selected COVID-19 cases in Hong Kong. This is important because dot density map can be misleading since the points represent the value of an

attribute in a region, not the actual location of an event (Waller & Gotway, 2004). Another strength of this study is the visualization of the temporal change in concentrations of point features by kernel density heat maps for the three peaks. As a result, this study produced informative findings and enabled the readers to observe the temporal change of COVID-19 cases' local mobility in Hong Kong.

In our study, we utilized both Average Nearest Neighbor analysis and Global Moran's I to detect the spatial distribution of point features, and both tools led to the same conclusion of clustering pattern. Although the two tools operate differently, with Average Nearest Neighbor analysis being designed to measure point features and their neighbors by physical location (Esri, 2023) and Global Moran's I treating point features as continuous variable by measuring the number of point features in the 154 Tertiary Planning Units, they both indicate that COVID-19 cases tend to visit local destinations in specific regions rather than at random. This clustering pattern suggests that Local health authorities should be aware of this trend and optimize surveillance and containment strategies in those regions with high mobility for future airborne infectious disease outbreaks or pandemics.

We examined the relationship between the number of locations visited by COVID-19 cases and various neighborhood characteristics, including socioeconomic factors, health and senior centers, transportation locations. By doing so, these neighborhood characteristics were treated multidimensionally. We used Geographically Weighted Regression (GWR) model to analyze the spatial variability of the data. The standard residuals from the symptomatic GWR model were randomly distributed, indicating that the model successfully captured the spatial variability of the

data. However, the standard residuals from asymptomatic GWR model displayed a clustering distribution, which was not statistically significant, suggesting that the model may have failed to capture some unobserved factors related to asymptomatic cases' local mobility. These findings highlight the challenges of predicting the local mobility of asymptomatic cases based on neighborhood sociodemographic characteristics, health and senior centers, and transportation characteristics compared to symptomatic cases.

There are some limitations to this study. Although our Geographically Weighted Regression (GWR) models explained about half of the variance of the COVID-19 cases' local mobility, we could not achieve high adjusted R^2 values ($>70\%$) even after multiple models tuning. To improve the accuracy of model prediction, future studies should explore other potential factors. Another limitation is that the travel records were based on contacting tracing reports, which may have resulted in recall bias or underreported travel records. Unfortunately, as we used secondary data provided by the local health department, we were not able to access the cases to improve these measurement issues. Additionally, the travel time estimated by Origin-Destination Cost Matrix is also limited since the simulated routes are not necessarily the actual routes taken by the COVID-19 cases, resulting in travel time estimates that may not reflect the actual travel time of the cases.

5.5 Conclusion

Asymptomatic COVID-19 cases exhibited higher local mobility prior to diagnosis than symptomatic cases, indicating a higher risk of spreading SARS-CoV-2

via mobility. Both groups' mobility showed significant clustering. To address future infectious disease outbreaks at early stages, local health authorities and policymakers should tailor detection and containment strategies by considering the spatial variability in different areas. Public health officials can prioritize resources such as testing and contact tracing in areas with higher local mobility and implement measures to limit mobility in these areas to reduce the risk of transmission. A better understanding of the relationship between local mobility and infectious disease transmission can aid in the development of more effective public health interventions to protect communities from the spread of disease.

Chapter 6: Conclusion

This dissertation adds important knowledge to the existing literature on the local and international mobility of asymptomatic COVID-19 cases. By using historical data from 11,775 COVID-19 cases, several departure airports, and hundreds of flights, the study finds an association between COVID-19 asymptomatic status and the transmissibility of asymptomatic cases on inbound flight to Hong Kong. Specifically, the results indicate that asymptomatic cases are more likely to take inbound air travel 14 days prior to COVID-19 diagnosis compared to symptomatic cases. Additionally, the study reveals that average flight duration among asymptomatic cases is shorter than that among symptomatic cases. These findings expand the public's understanding of travel behaviors and flight duration between asymptomatic cases and symptomatic cases, providing evidence to support assumptions regarding the impact of absence of symptoms on travel behavior. These findings are valuable for policymakers to design effective health policies to curb COVID-19 or other airborne infectious disease pandemic in the future by considering the potential significant contribution of asymptomatic cases.

Among the three peaks of COVID-19 in Hong Kong, peak one saw the quickest spread of SARS-CoV-2 in inbound flight cabins and airports, which started to slow down in peak two and dramatically decreased in peak three. Asymptomatic cases were found to be 1.33 times more likely than symptomatic cases to be presented in the inbound flight cabin or airport with other COVID-19 cases simultaneously. The study's agent-based networks of inbound air travel to Hong Kong revealed a fragmented pattern with a low-level of connectedness, making it extremely challenging to be

dismantled these networks. The findings suggest that strategies such as travel restriction are inefficient in successfully containing the spread of SARS-CoV-2 through COVID-19 cases' international mobility. Instead, public health strategies such as wearing masks and social distancing should be prioritized regardless of presence of symptoms to contain future pandemics, instead of relying on travel restrictions at the early stage of any outbreaks or endemics.

Asymptomatic COVID-19 cases exhibited higher local mobility prior to diagnosis than symptomatic cases, increasing the risk of spreading SARS-CoV-2 through mobility in Hong Kong. While socioeconomic factors, medical centers, and transportation locations, can provide some insight, they have limited predictive power for COVID-19 cases' local mobility. In comparison to symptomatic cases, predicting the local mobility of asymptomatic cases is more challenging based on population density, medical centers, and transportation. To effectively manage future infectious disease outbreaks or pandemics, local health authorities and policymakers should tailor their detection and containment strategies by considering the spatial variability across different areas.

Bibliography

Aletta, F., Brinchi, S., Carrese, S., Gemma, A., Guattari, C., Mannini, L., & Patella, S. M. (2020). Analysing urban traffic volumes and mapping noise emissions in Rome (Italy) in the context of containment measures for the COVID-19 disease. *Noise Mapping*, 7(1), 114-122.

Aljeri, M. (2020, November). The Impact of COVID-19 Confinement on Regional Mobility of Spatial-Temporal Social Networks. In *Proceedings of the 18th ACM Symposium on Mobility Management and Wireless Access* (pp. 29-35).

Altintasi, O., Tuydes-Yaman, H., & Tuncay, K. (2017). Detection of urban traffic patterns from Floating Car Data (FCD). *Transportation research procedia*, 22, 382-391.

Amdaoud, M., Arcuri, G., & Levratto, N. (2021). Are regions equal in adversity? A spatial analysis of spread and dynamics of COVID-19 in Europe. *The European Journal of Health Economics*, 22(4), 629-642.

Bae, S. H., Shin, H., Koo, H. Y., Lee, S. W., Yang, J. M., & Yon, D. K. (2020). Asymptomatic transmission of SARS-CoV-2 on evacuation flight. *Emerging infectious diseases*, 26(11), 2705.

Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D. Y., Chen, L., & Wang, M. (2020). Presumed asymptomatic carrier transmission of COVID-19. *Jama*, 323(14), 1406-1407.

Balcan, D., Hu, H., Goncalves, B., Bajardi, P., Poletto, C., Ramasco, J. J., ... & Vespignani, A. (2009). Seasonal transmission potential and activity peaks of the new influenza A (H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC medicine*, 7(1), 1-12.

Barabási, A. L & Pósfai, M. (2016). *Network Science*. Cambridge: Cambridge University Press. ISBN 978-1-107-07626-6

Barabási, A. L., Loscalzo, J., & Silverman, E. K. (Eds.). (2017). *Network Medicine: Complex Systems in Human Disease and Therapeutics*. Harvard University Press.

Barnett, A., & Fleming, K. (2020). Covid-19 risk among airline passengers: Should the middle seat stay empty?. *MedRxiv*.

Barthélemy, M. (2011). Spatial networks. *Physics reports*, 499(1-3), 1-101.

BBC News (2020). Coronavirus spreads to Antarctic research station. Available from: <https://www.bbc.com/news/world-latin-america-55410065>

Bielecki, M., Patel, D., Hinkelbein, J., Komorowski, M., Kester, J., Ebrahim, S., ... & Schlagenhauf, P. (2020). Reprint of: Air travel and COVID-19 prevention in the pandemic and peri-pandemic period: A narrative review. *Travel medicine and infectious disease*, 38, 101939.

Bogoch, I. I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M. U., & Khan, K. (2020). Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. *Journal of travel medicine*, 27(2), taaa008.

Bolton, K. J., McCaw, J. M., Forbes, K., Nathan, P., Robins, G., Pattison, P., Nolan, T. & McVernon, J. (2012). Influence of contact definitions in assessment of the relative importance of social settings in disease transmission risk. *PloS one*, 7(2), e30893.

Borgatti, S. P., & Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Social networks*, 28(4), 466-484.

Breiger, R. L. (1974). The duality of persons and groups. *Social forces*, 53(2), 181-190.

Brockmann, D., & Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *science*, 342(6164), 1337-1342.

Centers for Disease Control and Prevention. (2014). Middle East Respiratory Syndrome (MERS). Available from: <https://www.cdc.gov/coronavirus/mers/index.html>

Centers for Disease Control and Prevention. (2017). Severe Acute Respiratory Syndrome (SARS). Available from: <https://www.cdc.gov/sars/index.html>

Centers for Disease Control and Prevention. (2021). Considerations for Owners and Operators for Multifamily Housing Including Populations at Increased Risk for Complications for COVID-19. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/community/multifamily-housing.html>

Centers for Disease Control and Prevention. (2022). COVID-19 Quarantine and Isolation. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/your-health/quarantine-isolation.html>

Centre for Health Protection, Department of Health, the Government of Hong Kong Special Administrative Region (2022). Latest local situation of COVID-19. Available from: https://www.chp.gov.hk/files/pdf/local_situation_covid19_en.pdf

Centre for Health Protection, Department of Health, the Government of Hong Kong Special Administrative Region (2020). Compulsory Testing for Certain Persons. Available from: <https://www.coronavirus.gov.hk/eng/compulsory-testing.html#arriving>

- Chan-Yeung, M., & Xu, R. H. (2003). SARS: epidemiology. *Respirology*, 8, S9-S14.
- Che, X. Y., Di, B., Zhao, G. P., Wang, Y. D., Qiu, L. W., Hao, W., ... & Yuen, K. Y. (2006). A patient with asymptomatic severe acute respiratory syndrome (SARS) and antigenemia from the 2003–2004 community outbreak of SARS in Guangzhou, China. *Clinical Infectious Diseases*, 43(1), e1-e5.
- Chen, H., He, J., Song, W., Wang, L., Wang, J., & Chen, Y. (2020). Modeling and interpreting the COVID-19 intervention strategy of China: A human mobility view. *PloS one*, 15(11), e0242761.
- Chen, J., He, H., Cheng, W., Liu, Y., Sun, Z., Chai, C., ... & Chen, Z. (2020). Potential transmission of SARS-CoV-2 on a flight from Singapore to Hangzhou, China: An epidemiological investigation. *Travel medicine and infectious disease*, 36, 101816.
- Cheung, E., & Ting, V. (2020). Hong Kong third wave: Covid-19 total exceeds 3,000 with 118 new infections, one related death as new hotline service launched. Available from: <https://www.scmp.com/news/hong-kong/healthenvironment/article/3095115/hong-kong-third-wave-elderly-woman-latest-covid>
- Chi, L. P., & Cai, X. (2004). Structural changes caused by error and attack tolerance in US airport network. *International Journal of Modern Physics B*, 18(17n19), 2394-2400.
- Choi EM, Chu DKW, Cheng PKC, Tsang DNC, Peiris M, Bausch DG, Poon LLM, Watson-Jones D. In-Flight Transmission of SARS-CoV-2. *Emerg Infect Dis*. 2020 Nov;26(11):2713-2716.
- Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4), 370-379.
- Christakis, N. A., & Fowler, J. H. (2008). The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21), 2249-2258.
- Coelho, M. T. P., Rodrigues, J. F. M., Medina, A. M., Scalco, P., Terribile, L. C., Vilela, B., ... & Dobrovolski, R. (2020). Exponential phase of covid19 expansion is driven by airport connections. *MedRxiv*.
- Colizza, V., Barrat, A., Barthélemy, M., & Vespignani, A. (2006). The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences*, 103(7), 2015-2020.

- Cordes, J., & Castro, M. C. (2020). Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spatial and Spatio-temporal Epidemiology*, 34, 100355.
- Daon, Y., Thompson, R. N., & Obolski, U. (2020). Estimating COVID-19 outbreak risk through air travel. *Journal of travel medicine*, 27(5), taaa093.
- Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. CRC press.
- Dimitrov, N. B., & Meyers, L. A. (2010). Mathematical approaches to infectious disease prediction and control. In *Risk and optimization in an uncertain world* (pp. 1-25).
- Doytsher, Y., Galon, B., & Kanza, Y. (2010, November). Querying geo-social data by bridging spatial networks and social networks. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks* (pp. 39-46).
- Edwards, A. W., & Cavalli-Sforza, L. L. (1965). A method for cluster analysis. *Biometrics*, 362-375.
- Edwards, N. J., Widrick, R., Wilmes, J., Breisch, B., Gerscheffs, M., Sullivan, J., ... & Espinoza-Calvio, A. (2021). Reducing COVID-19 airborne transmission risks on public transportation buses: An empirical study on aerosol dispersion and control. *Aerosol Science and Technology*, 55(12), 1378-1397.
- Emch, M., Root, E. D., Giebultowicz, S., Ali, M., Perez-Heydrich, C., & Yunus, M. (2012). Integration of spatial and social network analysis in disease transmission studies. *Annals of the Association of American Geographers*, 102(5), 1004-1015.
- English, P., Neutra, R., Scalf, R., Sullivan, M., Waller, L., & Zhu, L. (1999). Examining associations between childhood asthma and traffic flow using a geographic information system. *Environmental health perspectives*, 107(9), 761-767.
- Esri. (Accessed by Feb 1, 2023). Average Nearest Neighbor (Spatial Statistics). Available from: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/average-nearest-neighbor.htm>
- Faust, K., Entwisle, B., Rindfuss, R. R., Walsh, S. J., & Sawangdee, Y. (2000). Spatial arrangement of social and economic networks among villages in Nang Rong District, Thailand. *Social networks*, 21(4), 311-337.
- Firestone, S. M., Ward, M. P., Christley, R. M., & Dhand, N. K. (2011). The importance of location in contact networks: Describing early epidemic spread using spatial social network analysis. *Preventive veterinary medicine*, 102(3), 185-195.

Freedman, D. O., & Wilder-Smith, A. (2020). In-flight transmission of SARS-CoV-2: a review of the attack rates and available data on the efficacy of face masks. *Journal of Travel Medicine*, 27(8), taaa178.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, 1(3), 215-239.

Fu, X., & Zhai, W. (2021). Examining the spatial and temporal relationship between social vulnerability and stay-at-home behaviors in New York City during the COVID-19 pandemic. *Sustainable cities and society*, 67, 102757.

Giebultowicz, S., Ali, M., Yunus, M., & Emch, M. (2011). The simultaneous effects of spatial and social networks on cholera transmission. *Interdisciplinary perspectives on infectious diseases*, 2011.

Gile, K. J., & Handcock, M. S. (2010). 7. Respondent-driven sampling: An assessment of current methodology. *Sociological methodology*, 40(1), 285-327.

Gilmore, J. B. (1998). *In cold pursuit: Medical intelligence investigates the common cold*. Stoddart Pub.

Gómez, S., Fernández, A., Meloni, S., & Arenas, A. (2019). Impact of origin-destination information in epidemic spreading. *Scientific reports*, 9(1), 1-9.

The Government of the Hong Kong Special Administrative Region (2020). Government relaxes certain social distancing measures. Available from: <http://www.info.gov.hk/gia/general/202209/30/P2022093000818.htm>

The Government of the Hong Kong Special Administrative Region (2021). Government announces arrangements for designated flights from United Kingdom to Hong Kong and quarantine arrangements for arrivals from low-risk places. Available from: <http://www.info.gov.hk/gia/general/202104/01/P2021040100696.htm>

Harris, J. K., Luke, D. A., Zuckerman, R. B., & Shelton, S. C. (2009). Forty years of secondhand smoke research: the gap between discovery and delivery. *American journal of preventive medicine*, 36(6), 538-548.

Hertzberg, V. S., Weiss, H., Elon, L., Si, W., Norris, S. L., & FlyHealthy Research Team. (2018). Behaviors, movements, and transmission of droplet-mediated respiratory diseases during transcontinental airline flights. *Proceedings of the National Academy of Sciences*, 115(14), 3623-3627.

Hoehl, S., Karaca, O., Kohmer, N., Westhaus, S., Graf, J., Goetsch, U., & Ciesek, S. (2020). Assessment of SARS-CoV-2 transmission on an international flight and among a tourist group. *JAMA network open*, 3(8), e2018044-e2018044.

Home Affairs Department, the Government of Hong Kong Special Administrative Region. (Accessed by February 1, 2023). My Districts. Available from: https://www.had.gov.hk/en/18_districts/my_map.htm

Hong Kong's Center for Health Protection (2020). How to wear a surgical mask properly. Available from: <https://www.chp.gov.hk/en/index.html>

Hong Kong Free Press (2021). Covid-19: Hong Kong's vaccination program begins as three centers are switched to China-made vaccine. Available from: <https://hongkongfp.com/2021/02/26/covid-19-hong-kongs-vaccination-programme-begins-as-three-more-centres-announced-to-meet-strong-demand/>

Hong Kong's Information Services Department (2020a). Further measure to fight virus. Available from: https://www.news.gov.hk/eng/2020/01/20200128/20200128_182923_787.html

Hong Kong's Information Services Department (2020b). HK to ban non-resident arrivals. Available from: https://www.news.gov.hk/eng/2020/03/20200323/20200323_164827_699.html

Hong Kong's Information Services Department (2020c). Mandatory virus test for 3 groups. Available from: https://www.news.gov.hk/eng/2020/11/20201120/20201120_143357_046.html

Hong Kong International Airport. (Accessed by 2023 Jan 15). Hong Kong: Facts and Figures, HKIA at a Glance. Available from: <https://www.hongkongairport.com/en/the-airport/hkia-at-a-glance/fact-figures.page>

Hong Kong Tourism Board. Explore unique Hong Kong experience. Available online: <https://www.discoverhongkong.com/us/explore.html> (accessed on 1 July 2022)

Huisman, M. (2009). Imputation of missing network data: Some simple procedures. *Journal of Social Structure*, 10(1), 1-29.

International Air Transport Association (2020) IATA air travel and COVID-19. Available from: <https://www.iata.org/en/youandiata/travelers/health/>

Jacquez, G. M. (2008). Spatial cluster analysis. *The handbook of geographic information science*, 395(416).

James, A., Eagle, L., Phillips, C., Hedges, D. S., Bodenhamer, C., Brown, R., ... & Kirking, H. (2020). High COVID-19 attack rate among attendees at events at a church—Arkansas, March 2020.

- Jia, H. L., Li, P., Liu, H. J., Zhong, J. Y., Qin, P. Z., Su, W. Z., ... & Ni, M. (2021). Genomic elucidation of a COVID-19 resurgence and local transmission of SARS-CoV-2 in Guangzhou, China. *Journal of clinical microbiology*, 59(8), e00079-21.
- Jiang, R., Wang, Z., Cai, Z., Yang, C., Fan, Z., Xia, T., ... & Shibasaki, R. (2021, September). Countrywide Origin-Destination Matrix Prediction and Its Application for COVID-19. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 319-334). Springer, Cham.
- Jo, Y., Hong, A., & Sung, H. (2021). Density or connectivity: what are the main causes of the spatial proliferation of COVID-19 in Korea?. *International Journal of Environmental Research and Public Health*, 18(10), 5084.
- Jungerman, M. R., Vonnahme, L. A., Washburn, F., & Alvarado-Ramy, F. (2017). Federal travel restrictions to prevent disease transmission in the United States: An analysis of requested travel restrictions. *Travel medicine and infectious disease*, 18, 30-35.
- Kelly, D., Bambury, N., & Boland, M. (2021). In-flight transmission of wild-type SARS-CoV-2 and the outbreak potential of imported clusters of COVID-19: a review of published evidence. *Globalization and health*, 17(1), 1-5.
- Khanh NC, Thai PQ, Quach HL, Thi NH, Dinh PC, Duong TN, Mai LTQ, Nghia ND, Tu TA, Quang N, Quang TD, Nguyen TT, Vogt F, Anh DD. Transmission of SARS-CoV 2 During Long-Haul Flight. *Emerg Infect Dis*. 2020 Nov;26(11):2617-2624.
- Khatib, A. N., Carvalho, A. M., Primavesi, R., To, K., & Poirier, V. (2020). Navigating the risks of flying during COVID-19: a review for safe air travel. *Journal of travel medicine*, 27(8), taaa212.
- Klov Dahl, A. S., Potterat, J. J., Woodhouse, D. E., Muth, J. B., Muth, S. Q., & Darrow, W. W. (1994). Social networks and infectious disease: The Colorado Springs study. *Social science & medicine*, 38(1), 79-88.
- Kong, D., Wang, Y., Lu, L., Wu, H., Ye, C., Wagner, A. L., ... & Pan, H. (2021). Clusters of 2019 coronavirus disease (COVID-19) cases in Chinese tour groups. *Transboundary and Emerging Diseases*, 68(2), 684-691.
- Kong, X., Yang, Y., Gao, J., Guan, J., Liu, Y., Wang, R., Xing, B., Li, Y. and Ma, W., 2015. Overview of the health care system in Hong Kong and its referential significance to mainland China. *Journal of the Chinese Medical Association*, 78(10), pp.569-573.
- Koskinen, J. H., Robins, G. L., Wang, P., & Pattison, P. E. (2013). Bayesian analysis for partially observed network data, missing ties, attributes and actors. *Social networks*, 35(4), 514-527.

- Krieger, N., Waterman, P., Lemieux, K., Zierler, S., & Hogan, J. W. (2001). On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. *American journal of public health*, 91(7), 1114.
- Last J. (1995). *A Dictionary of Epidemiology*. 3rd ed. Oxford, UK: Oxford University Press.
- Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Ichii, H., Zacharski, M., ... & Khosrawipour, T. (2020). The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak. *Journal of Microbiology, Immunology and Infection*, 53(3), 467-472.
- Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Ichii, H., Schubert, J., ... & Khosrawipour, T. (2020). Internationally lost COVID-19 cases. *Journal of Microbiology, Immunology and Infection*, 53(3), 454-458.
- Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J., & Khosrawipour, T. (2020). The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *Journal of travel medicine*, 27(3), 1-7.
- Law, S., Leung, A. W., & Xu, C. (2020). Severe acute respiratory syndrome (SARS) and coronavirus disease-2019 (COVID-19): From causes to preventions in Hong Kong. *International Journal of Infectious Diseases*, 94, 156-163.
- Leal Filho, W., Lütz, J. M., Sattler, D. N., & Nunn, P. D. (2020). Coronavirus: COVID-19 transmission in pacific small island developing states. *International Journal of Environmental Research and Public Health*, 17(15), 5409.
- Lee, H. K., Tso, E. Y., Chau, T. N., & Tsang, O. T. (2003). Asymptomatic severe acute respiratory syndrome-associated coronavirus infection. *Emerging infectious diseases*, 9(11), 1491.
- Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G. M., ... & Sung, J. J. (2003). A major outbreak of severe acute respiratory syndrome in Hong Kong. *New England Journal of Medicine*, 348(20), 1986-1994.
- Leitmeyer, K., & Adlhoch, C. (2016). Influenza transmission on aircraft: a systematic literature review. *Epidemiology (Cambridge, Mass.)*, 27(5), 743.
- Leung, H. (2020). What to know about the Wuhan Pneumonia Outbreak. Available from: <https://time.com/5759289/wuhan-pneumonia-outbreak-disease/>
- Li, A. C. M., Cheung, P. C. G., & Lee, K. C. G. (2020). The situation in Hong Kong during the COVID-19 pandemic. *Psychological Trauma: Theory, Research, Practice, and Policy*, 12(6), 576.

- Li, H., Calder, C. A., & Cressie, N. (2007). Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. *Geographical analysis*, 39(4), 357-375.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., ... & Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England journal of medicine*, 382(13):1199-1207.
- Lu, J., Gu, J., Li, K., Xu, C., Su, W., Lai, Z., ... & Yang, Z. (2020). COVID-19 outbreak associated with air conditioning in restaurant, Guangzhou, China, 2020. *Emerging infectious diseases*, 26(7), 1628.
- Luo, K., Lei, Z., Hai, Z., Xiao, S., Rui, J., Yang, H., ... & Chen, T. (2020, October). Transmission of SARS-CoV-2 in public transportation vehicles: a case study in Hunan Province, China. In *Open forum infectious diseases* (Vol. 7, No. 10, p. ofaa430). US: Oxford University Press.
- Ma, L., Yu, Z., Jiao, Y., Lin, L., Zhong, W., Day, S. W., ... & Wang, G. (2021). Capacity of transportation and spread of COVID-19—an ironical fact for developed countries. *Environmental Science and Pollution Research*, 28(28), 37498-37505.
- Ma, Q., Liu, J., Liu, Q., Kang, L., Liu, R., Jing, W., ... & Liu, M. (2021). Global Percentage of Asymptomatic SARS-CoV-2 Infections Among the Tested Population and Individuals With Confirmed COVID-19 Diagnosis: A Systematic Review and Meta-analysis. *JAMA network open*, 4(12), e2137257-e2137257.
- Mahay, J., & Laumann, E. O. (2004). Neighborhoods as sex markets. *The sexual organization of the city*, 69-92.
- Malik, H. A. M., Mahmood, N., Usman, M. H., Rziwan, K., & Abid, F. (2019). Analysis of airport network in Pakistan utilizing complex network approach. *network*, 10(1).
- Mangili, A., & Gendreau, M. A. (2005). Transmission of infectious diseases during commercial air travel. *The Lancet*, 365(9463), 989-996.
- Mangili, A., Vindenes, T., & Gendreau, M. (2016). Infectious risks of air travel. *Infections of Leisure*, 333-344.
- Mansour, O., Kajtazi, M., & Ghazawneh, A. (2022). Using Mobile Data for Understanding Population Movement and Disease Transmission during Covid-19 Outbreak in the Nordics. In *Hawaii International Conference on System Sciences (HICSS)*, Maui, Hawaii, USA, 4-7 January, 2022 (pp. 7151-7160). HICSS.

- Maxcy, K. F. (1967). An epidemiological study of endemic typhus (Brill's disease) in the southeastern United States with special reference to its mode of transmission. *American Journal of Epidemiology*, 85(3), 333-355.
- McCaw, J. M., Forbes, K., Nathan, P. M., Pattison, P. E., Robins, G. L., Nolan, T. M., & McVernon, J. (2010). Comparison of three methods for ascertainment of contact information relevant to respiratory pathogen transmission in encounter networks. *BMC infectious diseases*, 10(1), 1-11.
- McMillen, D. P. (2003). Spatial autocorrelation or model misspecification?. *International Regional Science Review*, 26(2), 208-217.
- McNutt, L. A., Wu, C., Xue, X., & Hafner, J. P. (2003). Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American journal of epidemiology*, 157(10), 940-943.
- Menczer, F., Fortunato, S., & Davis, C. (2020). *A First Course in Network Science*. Cambridge: Cambridge University Press.
- Monmonier, M. (2018). *How to lie with maps*. University of Chicago Press.
- Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2) 17-23.
- Morris, M. (Ed.). (2004). *Network epidemiology: a handbook for survey design and data collection*. New York, NY: Oxford University Press.
- Munster, V. J., Koopmans, M., van Doremalen, N., van Riel, D., & de Wit, E. (2020). A novel coronavirus emerging in China—key questions for impact assessment. *New England Journal of Medicine*, 382(8), 692-694.
- Murphy, N., Boland, M., Bambury, N., Fitzgerald, M., Comerford, L., Dever, N., ... & O'Connor, L. (2020). A large national outbreak of COVID-19 linked to air travel, Ireland, summer 2020. *Eurosurveillance*, 25(42), 2001624.
- Newman, M. (2018). *Networks*. Oxford university press.
- Ng, O. T., Marimuthu, K., Chia, P. Y., Koh, V., Chiew, C. J., De Wang, L., ... & Lee, V. J. (2020). SARS-CoV-2 infection among travelers returning from Wuhan, China. *New England Journal of Medicine*, 382(15), 1476-1478.
- Nguyen, Q. C., Huang, Y., Kumar, A., Duan, H., Keralis, J. M., Dwivedi, P., ... & Tasdizen, T. (2020). Using 164 million Google Street View images to derive built environment predictors of COVID-19 cases. *International journal of environmental research and public health*, 17(17), 6359.

- Nogrady, B. (2020). What the data say about asymptomatic COVID infections. *Nature*, 587(7835), 534-536.
- Nyerges, T., Couclelis, H., & McMaster, R. (2011). *The SAGE handbook of GIS and society*. Sage Publications.
- Olsen, S. J., Chang, H. L., Cheung, T. Y. Y., Tang, A. F. Y., Fisk, T. L., Ooi, S. P. L., ... & Dowell, S. F. (2003). Transmission of the severe acute respiratory syndrome on aircraft. *New England Journal of Medicine*, 349(25), 2416-2422.
- Parry, J. (2020). Covid-19: Hong Kong scientists report first confirmed case of reinfection. *BMJ* 2020;370:m3340.
- Pavli, A., Smeti, P., Hadjianastasiou, S., Theodoridou, K., Spilioti, A., Papadima, K., ... & Maltezou, H. C. (2020). In-flight transmission of COVID-19 on flights to Greece: an epidemiological analysis. *Travel medicine and infectious disease*, 38, 101882.
- Pedrosa, N. L., & Albuquerque, N. L. S. D. (2020). Spatial Analysis of COVID-19 cases and intensive care beds in the State of Ceará, Brazil. *Ciência & Saúde Coletiva*, 25, 2461-2468.
- Pepe, E., Bajardi, P., Gauvin, L., Privitera, F., Lake, B., Cattuto, C., & Tizzoni, M. (2020). COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Scientific data*, 7(1), 1-7.
- Perez-Heydrich, C., Furgurson, J. M., Giebultowicz, S., Winston, J. J., Yunus, M., Streatfield, P. K., & Emch, M. (2013). Social and spatial processes associated with childhood diarrheal disease in Matlab, Bangladesh. *Health & place*, 19, 45-52.
- Planning Department, the Government of Hong Kong Special Administrative Region, 2011, Tertiary Planning Unit: Planning Department, Hong Kong Special Administrative Region.
- Prieto, K., Chávez-Hernández, M. V., & Romero-Leiton, J. P. (2022). On mobility trends analysis of COVID-19 dissemination in Mexico City. *Plos one*, 17(2), e0263367.
- Read, J. M., Eames, K. T., & Edmunds, W. J. (2008). Dynamic social networks and the implications for the spread of infectious disease. *Journal of The Royal Society Interface*, 5(26), 1001-1007.
- Robins, G. (2015). *Doing social network research: Network-based research design for social scientists*. Sage.

- Rosensteel, G. E., Lee, E. C., Colizza, V., & Bansal, S. (2021). Characterizing an epidemiological geography of the United States: influenza as a case study. *medRxiv (preprint)*.
- Rothenberg, R. B., Sterk, C., Toomey, K. E., Potterat, J. J., Johnson, D., Schrader, M., & Hatch, S. (1998). Using social network and ethnographic tools to evaluate syphilis transmission. *Sexually transmitted diseases*, 154-160.
- Sah, P., Fitzpatrick, M. C., Zimmer, C. F., Abdollahi, E., Juden-Kelly, L., Moghadas, S. M., ... & Galvani, A. P. (2021). Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. *Proceedings of the National Academy of Sciences*, 118(34).
- Saha, J., Mondal, S., & Chouhan, P. (2021). Spatial-temporal variations in community mobility during lockdown, unlock, and the second wave of COVID-19 in India: A data-based analysis using Google's community mobility reports. *Spatial and Spatio-temporal Epidemiology*, 39, 100442.
- Sailer, K., & McCulloh, I. (2012). Social networks and spatial configuration—How office layouts drive social interaction. *Social networks*, 34(1), 47-58.
- Schmid, C. S., & Desmarais, B. A. (2017, December). Exponential random graph models with big networks: Maximum pseudolikelihood estimation and the parametric bootstrap. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 116-121). IEEE.
- Schwartz, K. L., Murti, M., Finkelstein, M., Leis, J. A., Fitzgerald-Husek, A., Bourns, L., ... & Yaffe, B. (2020). Lack of COVID-19 transmission on an international flight. *Cmaj*, 192(15), E410-E410.
- Shen, Y., Li, C., Dong, H., Wang, Z., Martinez, L., Sun, Z., ... & Xu, G. (2020). Community outbreak investigation of SARS-CoV-2 transmission among bus riders in Eastern China. *JAMA internal medicine*, 180(12), 1665-1671.
- Sherpa, S. F., & Shirzaei, M. (2022). Country-wide flood exposure analysis using Sentinel-1 synthetic aperture radar data: Case study of 2019 Iran flood. *Journal of Flood Risk Management*, 15(1), e12770.
- Siljander, M., Uusitalo, R., Pellikka, P., Isosomppi, S., & Vapalahti, O. (2022). Spatiotemporal clustering patterns and sociodemographic determinants of COVID-19 (SARS-CoV-2) infections in Helsinki, Finland. *Spatial and Spatio-temporal Epidemiology*, 41, 100493.
- Singh, N., Mall, R. K., Banerjee, T., & Gupta, A. (2021). Association between climate and infectious diseases among children in Varanasi city, India: A prospective cohort study. *Science of The Total Environment*, 796, 148769.

Snow, J. (1856). On the mode of communication of cholera. *Edinburgh medical journal*, 1(7), 668.

Speake, H., Phillips, A., Chong, T., Sikazwe, C., Levy, A., Lang, J., ... & McEvoy, S. P. (2020). Flight-associated transmission of severe acute respiratory syndrome coronavirus 2 corroborated by whole-genome sequencing. *Emerging infectious diseases*, 26(12), 2872.

Stadtfeld, C., Snijders, T. A., Steglich, C., & van Duijn, M. (2020). Statistical power in longitudinal network studies. *Sociological Methods & Research*, 49(4), 1103-1132.

Sun, X., & Wandelt, S. (2021). Robustness of air transportation as complex networks: Systematic review of 15 years of research and outlook into the future. *Sustainability*, 13(11), 6446.

Swadi, T., Geoghegan, J. L., Devine, T., McElnay, C., Sherwood, J., Shoemack, P., ... & de Ligt, J. (2021). Genomic evidence of in-flight transmission of SARS-CoV-2 despite predeparture testing. *Emerging infectious diseases*, 27(3), 687.

Tadic, M., Cuspidi, C., & Sala, C. (2020). COVID-19 and diabetes: is there enough evidence? *The Journal of Clinical Hypertension*, 22(6), 943-948.

Tang, S., & Lo, H. K. (2008). The impact of public transport policy on the viability and sustainability of mass railway transit—The Hong Kong experience. *Transportation Research Part A: Policy and Practice*, 42(4), 563-576.

Tenny, S., Hoffman, M. R. Relative Risk. [Updated 2022 Mar 26]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK430824/>

The Fund of Peace (2013) Fragile States Index: Measuring Fragility: Risk and Vulnerability in 179 countries. Available from: <https://fragilestatesindex.org/>

The Transportation Security Administration (2022) Coronavirus (COVID-19) information: TSA Confirmed COVID-19 Cases. Available from: <https://www.tsa.gov/coronavirus>

Thompson, R. N., Thompson, C. P., Pelerman, O., Gupta, S., & Obolski, U. (2018). Increased frequency of travel may act to decrease the chance of a global pandemic. *bioRxiv*, 404871.

Tizzoni, M., Bajardi, P., Poletto, C., Ramasco, J. J., Balcan, D., Gonçalves, B., ... & Vespignani, A. (2012). Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC medicine*, 10(1), 1-31.

- To, K. K., & Yuen, K. Y. (2020). Responding to COVID-19 in Hong Kong. *Hong Kong Med J*, 26(3), 164-6.
- Toprani, S. M., Scheibler, C., & Nagel, Z. D. (2020). Interplay between air travel, genome integrity, and COVID-19 risk vis-a-vis flight crew. *Frontiers in Public Health*, 8, 886.
- Tsang, K. W., Ho, P. L., Ooi, G. C., Yee, W. K., Wang, T., Chan-Yeung, M., ... & Lai, K. N. (2003). A cluster of cases of severe acute respiratory syndrome in Hong Kong. *New England Journal of Medicine*, 348(20), 1977-1985.
- Tuite, A. R., Bhatia, D., Moineddin, R., Bogoch, I. I., Watts, A. G., & Khan, K. (2020). Global trends in air travel: implications for connectivity and resilience to infectious disease threats. *Journal of travel medicine*, 27(4), taaa070.
- Valente, T. W. (2010). *Social networks and health: Models, methods, and applications*. Oxford University Press.
- Valente, T. W., Gallaher, P., & Mouttapa, M. (2004). Using social networks to understand and prevent substance use: A transdisciplinary perspective. *Substance use & misuse*, 39(10-12), 1685-1712.
- Van Duijn, M. A., Gile, K. J., & Handcock, M. S. (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social networks*, 31(1), 52-62.
- Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data*. John Wiley & Sons.
- Waller, L. A., Turnbull, B. W., Clark, L. C., & Nasca, P. (1992). Chronic disease surveillance and testing of clustering of disease and exposure: application to leukemia incidence and TCE-contaminated dumpsites in upstate New York. *Environmetrics*, 3(3), 281-300.
- Watts, A., Au, N. H., Thomas-Bachli, A., Forsyth, J., Mayah, O., Popescu, S., & Bogoch, I. I. (2020). Potential for inter-state spread of Covid-19 from Arizona, USA: analysis of mobile device location and commercial flight data. *Journal of travel medicine*, 27(8), taaa136.
- Wong, S. Y., Kwok, K. O., & Chan, F. K. (2020). What can countries learn from Hong Kong's response to the COVID-19 pandemic?. *Cmaj*, 192(19), E511-E515.
- World Health Organization. (2003). Severe Acute Respiratory Syndrome (SARS). Available from: https://www.who.int/health-topics/severe-acute-respiratory-syndrome#tab=tab_1

World Health Organization. (2019). Worldwide reduction in MERS cases and deaths since 2016. Available from: <https://www.who.int/news/item/08-07-2019-worldwide-reduction-in-mers-cases-and-deaths-since-2016>

World Health Organization. (2020). Q&A: How is COVID-19 transmitted? Available from: <https://www.who.int/vietnam/news/detail/14-07-2020-q-a-how-is-covid-19-transmitted>

World Health Organization. (2021). Coronavirus disease (COVID-19): How is it transmitted? Available from: <https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted>

World Health Organization. (2022). Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>

Xiang, H., Nuckols, J. R., & Stallones, L. (2000). A geographic information assessment of birth weight and crop production patterns around mother's residence. *Environmental research*, 82(2), 160-167.

Yang, N., Shen, Y., Shi, C., Ma, A. H. Y., Zhang, X., Jian, X., ... & Qian, G. (2020). In-flight transmission cluster of COVID-19: a retrospective case series. *Infectious Diseases*, 52(12), 891-901.

Yu, C., Zhou, M., Liu, Y., Guo, T., Ou, C., Yang, L., ... & Zou, Z. (2020). Characteristics of asymptomatic COVID-19 infection and progression: A multicenter, retrospective study. *Virulence*, 11(1), 1006-1014.

Yuan, H. Y., & Blakemore, C. (2022). The impact of multiple non-pharmaceutical interventions on controlling COVID-19 outbreak without lockdown in Hong Kong: A modelling study. *The Lancet Regional Health-Western Pacific*, 20, 100343.

Yue, Y., Chen, Y., Du, X., Jin, Y., Hu, M., Jiang, X., ... & Tuo, X. (2021). A survey of a COVID-19 cluster of charter flight importation. *Public Health*, 199, 107-109.

Zhang, J., & Kai, F. Y. (1998). What's the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes. *Jama*, 280(19), 1690-1691.

Zheng, R., Xu, Y., Wang, W., Ning, G., & Bi, Y. (2020). Spatial transmission of COVID-19 via public and private transportation in China. *Travel medicine and infectious disease*, 34, 101626.