Contents lists available at ScienceDirect

EClinicalMedicine

Research paper

Uncovering transmission patterns of COVID-19 outbreaks: A region-wide comprehensive retrospective study in Hong Kong

Yang Liu, Zhonglei Gu, Jiming Liu*

Department of Computer Science, Hong Kong Baptist University, Hong Kong Special Administrative Region, People's Republic of China

ARTICLE INFO

ABSTRACT

Background: Given the dynamism and heterogeneity of COVID-19 transmission patterns, determining the most effective yet timely strategies for specific regions remains a severe challenge for public health decision-makers.

Methods: In this work, we proposed a spatiotemporal connectivity analysis method for discovering transmission patterns across geographic locations and age-groups throughout different COVID-19 outbreak phases. First, we constructed the transmission networks of the confirmed cases during different phases by considering the spatiotemporal connectivity of any two cases. Then, for each case and those cases immediately pointed from it, we characterized the corresponding cross-district/population transmission pattern by counting their district-to-district and age-to-age occurrences. By summating the cross-district/population transmission patterns of all cases during a given period, we obtained the aggregated cross-district and cross-population transmission patterns.

Findings: We conducted a region-wide comprehensive retrospective study in Hong Kong based on the complete data report of COVID-19 cases, covering all 18 districts between January 23, 2020, and January 8, 2021 (https://data.gov.hk/en-data/dataset/hk-dh-chpsebcddr-novel-infectious-agent). The spatiotemporal connectivity analysis clearly unveiled the quantitative differences among various outbreak waves in their transmission scales, durations, and patterns. Moreover, for the statistically similar waves, their cross-district/population transmission patterns could be quite different (e.g., the cross-district transmission of the fourth wave was more diverse than that of the third wave, while the transmission over age-groups of the fourth wave was more concentrated than that of the third wave). At an overall level, super-spreader individuals (highly connected cases in the transmission networks) were usually concentrated in only a few districts (2 out of 18 in our study) or age-groups (3 out of 11 in our study).

Interpretation: With the discovered cross-district or cross-population transmission patterns, all of the waves of COVID-19 outbreaks in Hong Kong can be systematically scrutinized. Among all districts, quite a few (e.g., the Yau Tsim Mong district) were instrumental in spreading the virus throughout the pandemic. Aside from being exceptionally densely populated, these districts were also social-economic centers. With a variety of situated public venues, such as restaurants and singing/dancing clubs, these districts played host to all kinds of social gathering events, thereby providing opportunities for widespread and rapid transmission of the virus. Thus, these districts should be given the highest priority when deploying district-specific social distancing or intervention strategies, such as lockdown and stringent mandatory coronavirus testing for identifying and obstructing the chain of transmission. We also observed that most of the reported cases and the highly connected cases were middle-aged and elderly people (40- to 69-year-olds). People in these age-groups were active in various public places and social activities, and thus had high chances of being infected by or infecting others.

Funding: General research fund of the Hong Kong research grants council.

---

* Corresponding author.
  E-mail address: jiming@comp.hkbu.edu.hk (J. Liu).

Research in context

Evidence before this study

The spread of COVID-19 is a complex process governed by many disease-related factors, so the transmission patterns within different countries or regions are highly heterogeneous and vary over time. As a result, it remains unclear how to identify the most effective yet timely strategies for specific regions. To effectively mitigate the spread of COVID-19, it is critical to answer the following question: how can the transmission patterns of outbreaks among different age-groups within/across populated geographic locations throughout different outbreak phases be accurately characterized?

Added value of this study

We showed that the proposed spatiotemporal connectivity analysis can accurately characterize the cross-district and cross-population interactions, and thus provide clear explanations of the underlying and heterogeneous disease transmission patterns. Through a comprehensive retrospective study of Hong Kong, we provided a scientifically grounded method of discovering and quantitatively characterizing the transmission dynamics of COVID-19 in terms of age-group, district, and periods in a typical metropolis. The proposed spatiotemporal analysis method can help guide efforts to control and prevent the spread of the disease not only in Hong Kong but also in other densely populated international metropolises.

Implications of all the available evidence

This work touches upon an essential problem at a critical moment in time: determining the most effective and timely strategies for halting the spread of COVID-19 in specific regions, which remains a serious challenge for public health decision-makers. By revealing transmission patterns using spatiotemporal connectivity analysis, we gained insights into the dynamism and heterogeneity of COVID-19 transmission, and into the optimization and implementation of common pharmaceutical (e.g., the vaccination of high-risk populations) and nonpharmaceutical (e.g., the lockdown of high-risk areas) intervention strategies.

1. Introduction

The COVID-19 pandemic continues. As of January 8, 2021, the daily increase in new cases has exceeded 0.5 million, resulting in over 87 million confirmed cases with more than 1.9 million deaths across 191 countries/regions [1–3]. To mitigate the spread and outbreak of COVID-19, many countries have implemented various intervention and social distancing strategies, such as quarantining suspected and confirmed cases, implementing school closures and workplace lockdowns, and reducing public transportation. The effects of different strategies have been modeled and evaluated [4–15]. It remains a serious challenge for public health decision-makers to identify the most effective yet timely strategies for specific regions. The transmission of COVID-19 is a complex process governed by many factors, such as demographics, population mobility, and environment [16]. Therefore, the transmission patterns within different countries or regions are highly heterogeneous and vary over time.

To effectively mitigate the spread of COVID-19, answering the following research question is critical: how can the transmission patterns of outbreaks among different age-groups within/across populated geographic locations throughout different outbreak phases

be accurately characterized, especially in densely populated cities/regions with high transmission risk? To answer this question, we must answer three sub-questions:

(1) Are there any underlying or hidden transmission networks of COVID-19 outbreaks throughout the different phases and, if so, how can they be uncovered?
(2) What are the structures of such transmission networks and what do they imply?
(3) Do these transmission networks reveal the cross-district or cross-population transmission patterns of outbreaks and, if so, what are the implications for public health policymaking?

To achieve the aforementioned goal, we proposed a spatiotemporal connectivity analysis method for discovering the transmission networks of COVID-19 outbreaks [17,18]. We conducted a fine-grained retrospective study of one of the world's most densely populated regions, Hong Kong, at both the individual and meta-population levels throughout different outbreak phases. As an international metropolitan city with a population of 7.5 million, Hong Kong has over 7,000 people per square kilometer. In an environment with such a dense population and high connectivity, residents have higher chances of coming into close contact with others. Inevitably, this can trigger and catalyze the rapid spread of infectious disease. Through a comprehensive retrospective study of Hong Kong, we provided a scientifically grounded method of discovering and quantitatively characterizing the transmission dynamics of COVID-19 in terms of age-group, district, and periods in a typical metropolis. The proposed spatiotemporal analysis method can help guide efforts to control and prevent the spread of the disease not only in Hong Kong but also in other densely populated international metropolises.

2. Methods

2.1. Data sources and collection

We conducted the study and analysis based on the complete data report of COVID-19 cases in Hong Kong, covering all 18 districts between January 23, 2020, and January 8, 2021. The data were publicly released by the Department of Health, Hong Kong Special Administrative Region Government [19]. (Note that The Government of the Hong Kong Special Administrative Region is the owner of the intellectual property rights to all content available on DATA.GOV.HK, including but not limited to all data.) For each confirmed case, the data attributes included age, gender, residency, onset date, report date, whether the individual was asymptomatic, the individual's building of residence, and the buildings that the individual visited during the 14 days before the date of case confirmation. We used the latitudes and longitudes of the buildings, identified from the buildings' names by the Google Geocoding API, to group single buildings that appeared under different names (due to inconsistencies in the collected data) as single locations. For simplicity, the term "building" in this study refers to the locations (9,305 locations out of 10,247 buildings) after pre-processing.

2.2. Construction of transmission networks via spatiotemporal connectivity

We constructed the transmission networks of the confirmed cases during different phases by considering the spatiotemporal connectivity of any two cases. Each case represented a node in the network. For the nodes of two cases, we defined the node of the first case as being connected to the node of the second case if (1) the onset date of the first case was within the 14 days before the onset date of the second case (referred to as "temporal connectivity") and (2) the two cases visited the same building within the 14 days before their

confirmed diagnosis (referred to as "spatial connectivity"). Based on the above definition, the transmission network can be constructed as a directed graph. The graph properties of the transmission network were defined as follows:

- Degree of a node: the number of connections (links) a node has to other nodes, denoted as k.
- Out-degree of a node: the number of connections that point from a node to other nodes.
- In-degree of a node: the number of connections that point to a node from other nodes.
- Maximum degree: the maximum value among all nodes' degrees, denoted as k_{max}.
- Average degree: the average value of all nodes' degrees, denoted as k_{aver}.
- Eccentricity of a node: the maximum distance from that node to all other nodes in the network.
- Degree distribution: the probability that a randomly selected node in the network has a degree of k.
- Power-law-like distribution and degree exponent: a power-law-like distribution has a degree distribution of $k^{-\gamma}$, where γ is the degree exponent.

More technical details of the procedure used to construct the transmission networks were provided in the Supplementary Materials section.

2.3. Discovery of cross-district and cross-population transmission patterns

For each node (case) and those nodes immediately pointed from it, we characterized the corresponding cross-district transmission pattern by counting their district-to-district occurrences. This quantified how frequently the virus was transmitted from one district to another by a specific patient. By summating the cross-district transmission patterns of all of the cases during a given period, we obtained an aggregated cross-district transmission pattern. This pattern indicated the intensity of disease transmission from one district to another over time. Similarly, we can construct the cross-population transmission pattern for each case and those cases immediately pointed from it, as well as the aggregated cross-population transmission pattern. Together, these patterns measured the intensity of the disease transmission dynamics from one age-group to another. More technical details of the procedure used to uncover the cross-district and cross-population transmission patterns were provided in the Supplementary Materials section.

2.4. Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, writing of the Article, or the decision to submit for publication. All authors had full access to all the data in the study and were responsible for the decision to submit the Article for publication.

2.5. Adherence to the RECORD guidelines

The manuscript adheres strictly to RECORD guidelines.

3. Results

3.1. Transmission network patterns

Between January 23, 2020, and January 8, 2021, 9153 cases of COVID-19 infection were reported in Hong Kong. To explore the structures and patterns of the transmission dynamics of COVID-19

outbreaks in Hong Kong at the individual level, we constructed the transmission networks for the major waves of outbreaks using the methods introduced in the Data and Methods section and in the Supplementary Materials section. Fig. 1(A) showed the daily new cases. Fig. 1(B) illustrated the transmission networks of the second, third, and fourth waves of outbreaks. The illustrations of the transmission networks were created using Gephi 0.9.2 [20]. Note that we did not emphasize Hong Kong's first wave of COVID-19 outbreaks, which occurred between late January and mid-February 2020. Retrospectively, the first wave of outbreaks was much less significant in terms of the scale and duration of COVID-19 infections. Therefore, we focused on the second, third, and fourth waves. In our analysis, the durations of the second, third, and fourth waves were from March 17 to April 12, 2020, July 5 to September 21, 2020, and November 19, 2020, to January 8, 2021, respectively. The start date of a wave was the date from which the number of daily new cases was consistently larger than (or equal to) 10 and increased quickly. The end date of the wave was the date from which the number of daily new cases remained below 10 for at least two weeks and the trend stabilizes. The start and end dates of each wave defined here were consistent with the government's announcements [21,22].

As Fig. 1(B) shows, three individual-level transmission networks were constructed for the second, third, and fourth waves, respectively. All three networks showed patterns of densely connected clusters, indicating the existence of transmission hotspots. Moreover, the node degree of all three transmission networks displayed the power-law-like distributions. Fig. S1 provided more details. We also showed the eccentricity distributions of the transmission networks of these three waves in Fig. S2. Despite these commonalities, the three transmission networks also demonstrated clear quantitative differences in their transmission scales, durations, and patterns. First, the maximum degree k_{max} and the average degree k_{aver} were quite different for these three transmission networks: for the second wave, $k_{max} = 64$ and $k_{aver} = 3.77$; for the third wave, $k_{max} = 112$ and $k_{aver} = 6.00$; and for the fourth wave, $k_{max} = 127$ and $k_{aver} = 6.34$. This indicated the much greater severity of the third and fourth waves of outbreaks than the second wave in terms of transmission intensity (measured by the degree of the transmission networks). Moreover, the degree exponent γ was also very different for the three transmission networks: for the second, third, and fourth waves, $\gamma = 1.59$, 2.18, and 1.89, respectively. This demonstrated that the degree distributions of the transmission networks of the third and fourth waves were steeper than that of the second wave, indicating the existence of highly connected nodes in the third and fourth waves. These highly connected nodes may represent super-spreader individuals who deserve the utmost attention in epidemiological investigations. Furthermore, despite the fourth wave having a lower total number of infections (3,673 cases) than the third wave (3,780 cases), the virus seems more transmissible than it was during the third wave, as reflected by the maximum and average degrees of the transmission matrices. This observation was also consistent with Fig. 1(B), which showed the more densely connected clusters formed in the fourth wave than in the third wave.

In order to further explore the roles of different building categories in disease transmission, we divided the buildings into residential ones and non-residential ones. Fig. S3 showed the transmission networks in residential and non-residential buildings of the three major waves of COVID-19 outbreaks in Hong Kong. We can observe that, although there was not much difference in the number of residential buildings (4,294) and that of non-residential buildings (5,953), the majority of the spatiotemporal connections (35,973 out of 50,789) appeared in non-residential buildings, indicating that non-residential scenarios dominated the disease transmission and contributed most of the highly connected nodes in all three waves. The reason might be that the scale of transmission in the household is generally small, while that in non-residential places such as clubs or restaurants could
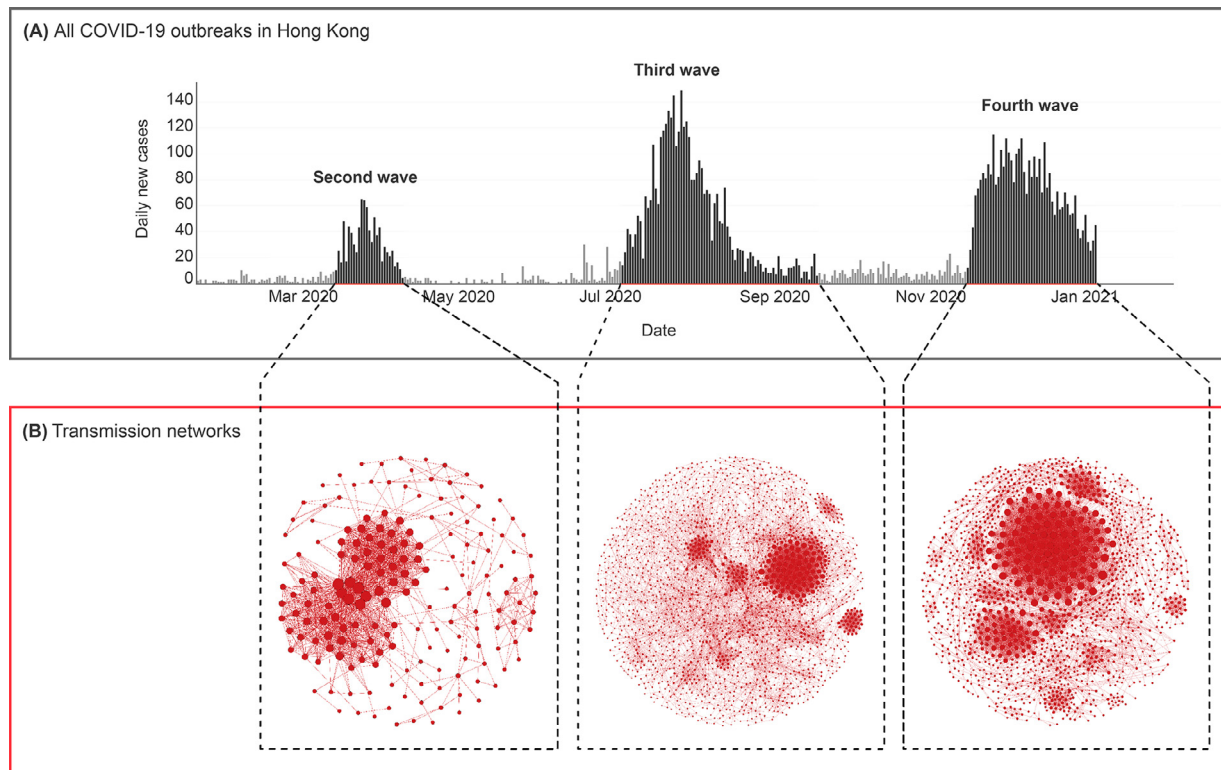
Fig. 1. The transmission networks of the major waves of COVID-19 outbreaks in Hong Kong between January 23, 2020, and January 8, 2021. **(A)** The number of daily new cases of COVID-19 in Hong Kong from January 2020 to January 2021. **(B)** The transmission networks (constructed based on the proposed spatiotemporal connectivity analysis method) of the second, third, and fourth waves of COVID-19 outbreaks in Hong Kong. The transmission networks were visualized using Gephi 0.9.2 [20]. Gephi is an open-source and multiplatform software distributed under the dual license CDDL 1.0 and GNU General Public License v3. **Note**: Hong Kong's first wave of COVID-19 outbreaks between late January and mid-February 2020 was also shown in (A) but not emphasized in this study. Retrospectively, the first wave of outbreaks was much less significant in terms of the scale and duration of COVID-19 infections than the other waves. Therefore, we focused on the second, third, and fourth waves.

be very large. Therefore, effective surveillance and tracing of the contacts and visits in the non-residential buildings, especially the public places, are of vital important in controlling the disease spread.

To make our investigation more comprehensive, we explored various ways to establish the temporal connectivity. In the current analysis, we included both symptomatic and asymptomatic cases to ensure data integrity. For asymptomatic cases, the onset date was replaced by the report date. However, this substitution may not be that accurate as the asymptomatic cases did not develop symptoms. To make the analysis more complete, we also considered an alternative way to construct the temporal connectivity by excluding all asymptomatic cases. The comparison between the degree distribution of the transmission network with asymptomatic cases and that without asymptomatic cases was shown in Fig. S4(A). We can observe that the degree distributions of both transmission networks demonstrated similar trends. Furthermore, we tested an alternative way to establish the temporal connectivity. Instead of setting a 14-day cutoff, we assumed that the transmission probability with respect to a specific length of serial interval followed a gamma distribution [17]. Then the temporal connectivity of two nodes was determined by this probability distribution. Fig. S4(B) showed the degree distributions of the transmission networks (with/without asymptomatic cases) constructed by the gamma distribution-based connectivity. Moreover, we conducted a sensitivity analysis on the degree distributions with respect to the number of cutoff days (from 1 to 14 days) and the number of days before symptom onset of index cases (from 0 to 5 days) for possible transmission. From Figure S5, we can see that the degree distributions with various settings were relatively robust, indicating the consistency of the proposed spatiotemporal analysis.

To examine in more detail the transmission patterns of the outbreaks, we magnified the transmission network of the fourth wave

(Fig. 2). The size of a node indicated its degree of connectivity and the different colors indicated the different case clusters identified and reported by the Hong Kong government [19]. For clarity of illustration, we only colored the 11 major clusters in the figure. The confirmed cases clearly formed different clusters/groups. The largest cluster was the Singing and Dancing cluster (colored orange), which included more than 700 confirmed cases from over 20 singing and dancing venues/clubs distributed across 7 of Hong Kong's 18 districts [23]. More importantly, the figure indicated that many other clusters and individual cases were connected with the Singing and Dancing cluster. For instance, the square node of case #6346 was a connector that linked two large clusters: the Singing and Dancing cluster and the Construction Sites at LOHAS Park/Kai Tak cluster (colored teal). The longest transmission path over the network (colored blue) also supported this: it passed through 12 cases (#5446, #5585, #5550, #5588, #6544, #5891, #6827, #7478, #8394, #8579, #8687, and #8824), 5 of which (#5585, #5550, #5588, #6544, and #5891) belonged to the Singing and Dancing cluster. The remaining seven cases belonged to other clusters. All of the observations from the figure were consistent with the Singing and Dancing cluster being the largest super-spreader group triggering Hong Kong's fourth wave of COVID-19 infections [24]. Identifying densely connected clusters can guide the accurate implementation of lockdowns or stringent mandatory coronavirus testing in high-risk groups/areas [25,26], so as to obstruct the chain of transmission in a timely manner.

3.2. Cross-district and cross-population transmission patterns throughout different outbreak phases

To facilitate the accurate deployment of control strategies in high-risk locations or age-groups for the effective mitigation of disease
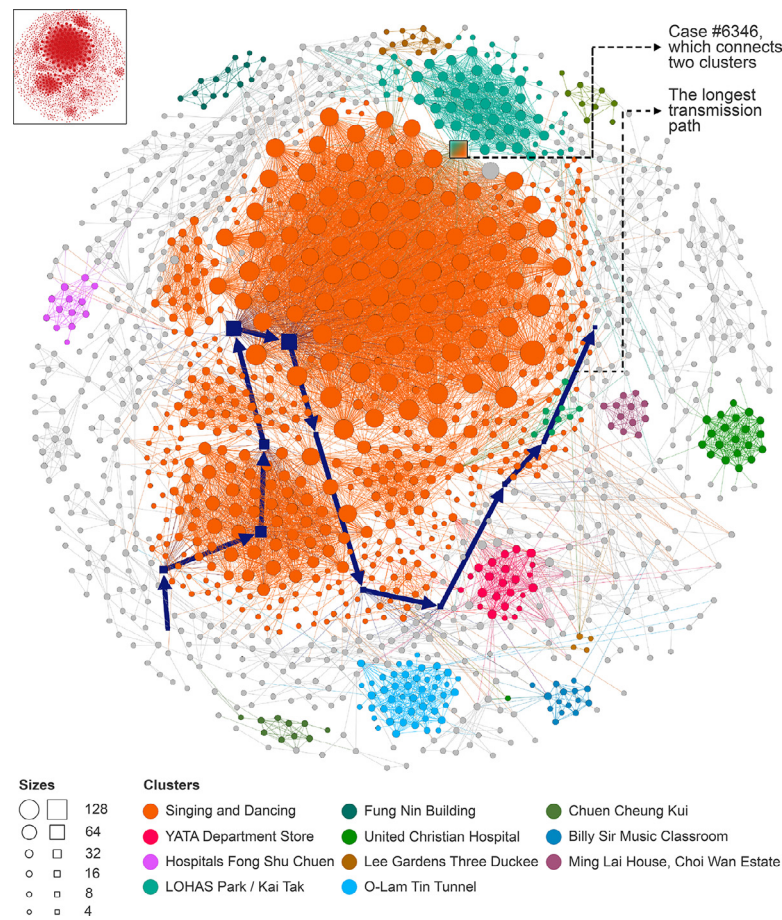
Fig. 2. The detailed transmission network of the fourth wave of COVID-19 outbreaks in Hong Kong, visualized using Gephi 0.9.2 [20]. Nodes (cases) with different degrees of connectivity were shown in different sizes. Different clusters were shown in different colors. A representative node that connected two major clusters and the longest transmission path over the transmission network were also highlighted.

spread, it is necessary to discover the cross-district and cross-population transmission patterns throughout different outbreak phases. We divided the entire Hong Kong region into 18 administrative districts (the names and indices of which are listed in Table S1) and its entire population into 11 age-groups (the details of which are listed in Table S2). Using the methods described in the Supplementary Materials section, we then calculated the aggregated cross-district and cross-population transmission patterns of the second, third, and fourth waves of COVID-19 outbreaks in Hong Kong. Fig. 3(B,C) illustrated the heat maps of the constructed cross-district and cross-population transmission patterns, respectively. The construction procedure and heat maps showed that the matrices were not symmetrical, indicating that our method can appropriately reflect directional transmission from one district (or age-group) to another.

For the second wave of outbreaks, the spread among districts (D1: Central & Western, D4: Wan Chai, and D5: Yau Tsim Mong) and the spread among age-groups (A3–A6: 20–59) were relatively concentrated. Thus, the epidemic was relatively easy to control. In the third wave, the spread among districts remained relatively concentrated, especially in D7 (Kowloon City) and D9 (Wong Tai Sin). However, transmission occurred across almost all age-groups. In this situation, it would be more appropriate to make district-specific control policies (e.g., strengthening social distancing strategies in severely affected districts) than age-specific control policies. For the fourth wave, although the statistical distribution of the daily case numbers was somewhat similar to that of the third wave, the transmission dynamics over districts (and age-groups) were quite different. The cross-district transmission of the fourth wave was more diverse than

that of the third wave, with D5 (Yau Tsim Mong) at the center of transmission. Meanwhile, the transmission over age-groups of the fourth wave was more concentrated than that of the third wave. This indicated that implementing age-specific intervention or control strategies (e.g., prioritization of vaccination for high-risk age-groups, or closure or access control of places for age-specific population gathering such as kindergartens, dancing clubs, or elderly care homes) during the fourth wave might have been more effective than doing so during the third wave.

3.3. Overall statistics for and general characteristics of the transmission patterns

To establish an overall understanding of the transmission connectivity among all confirmed cases during the past year, we summarized the distribution of cases in terms of their level of connectivity with other cases (i.e., degrees) in Tables 1 and 2. The out-degree and in-degree distributions were summarized according to gender, age, whether the individual was asymptomatic, residency, case classification, the connection with other cases, the time from onset to case confirmation, and district. To provide a more intuitive picture of the relationships between the out-degree/in-degree distributions and the abovementioned eight variables, we visualized all variable-degree histograms in Fig. S6.

The gender distribution showed more cases among people of the female gender. The age distribution showed that the reported cases were mainly distributed among individuals 20 to 69 years old (boxed in blue) in all out-degrees and in-degrees. This was consistent with
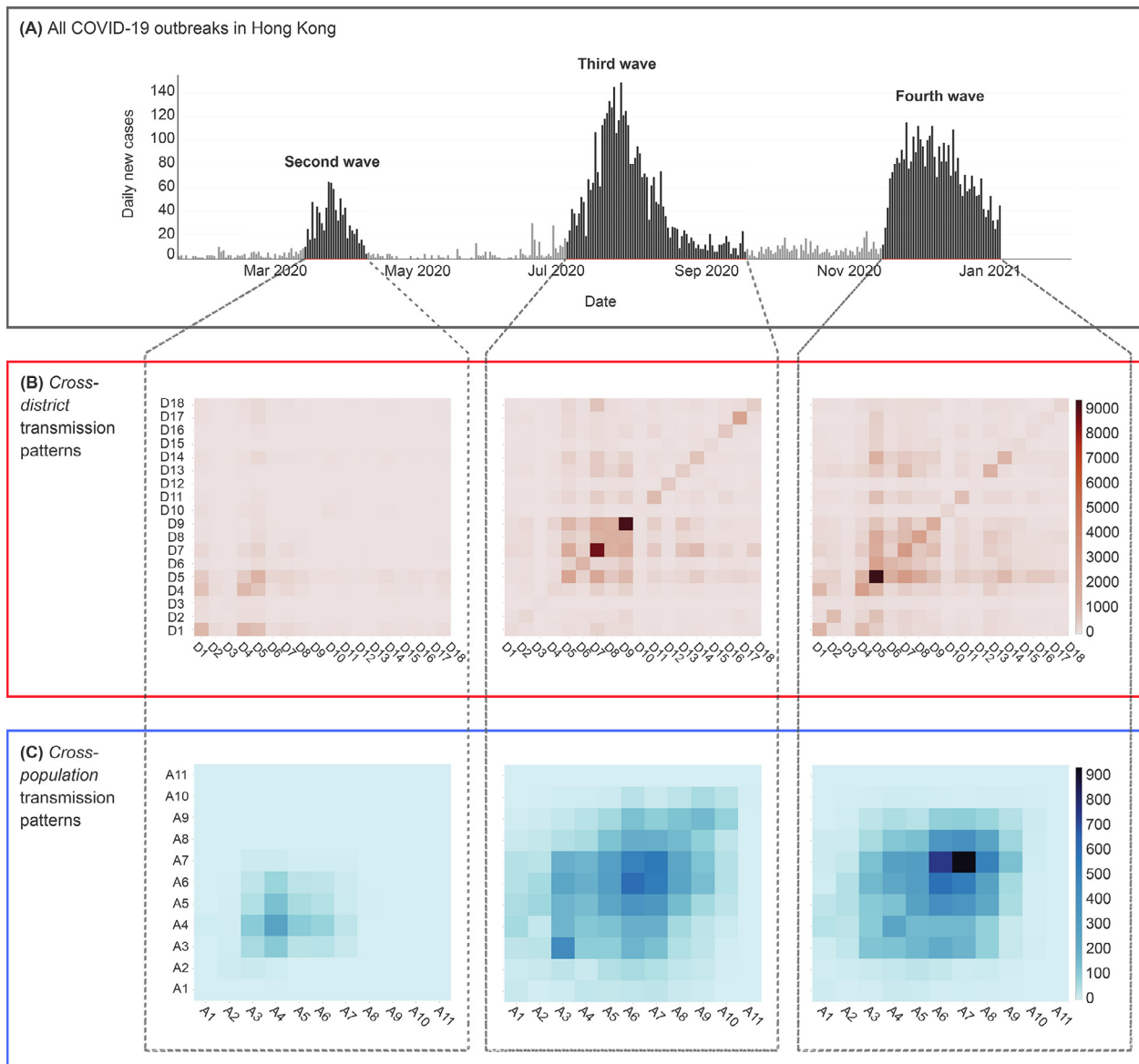
Fig. 3. The cross-district and cross-population transmission patterns throughout different outbreak phases. **(A)** The number of daily new cases of COVID-19 in Hong Kong from January 2020 to January 2021. **(B)** The heat maps of the cross-district transmission patterns of the second, third, and fourth waves of outbreaks, where D1 to D18 denote the indices of Hong Kong's 18 districts. More details were provided in Table S1. **(C)** The heat maps of the cross-population transmission patterns of the second, third, and fourth waves of outbreaks, where A1 to A11 denote the indices of the population's 11 age-groups. More details were provided in Table S2.

the fact that most people who were active in public places fall into this age range and thus had a higher probability of spreading the virus (corresponding to the out-degree) and becoming infected (corresponding to the in-degree) than children/youths (0–19 years old) and the elderly (≥ 70 years old). Most of the high out-degree and in-degree cases were distributed among 40- to 69-year-olds. The fact that people in these age-groups tended to gather in crowds or participated in activities involving many participants, such as dancing, singing, and Yum Cha (morning tea), explained this finding. Regarding classification, most cases were local or epidemiologically linked with local cases (boxed in green), whereas the impact of imported cases was relatively limited. The district distribution of the cases showed that Yau Tsim Mong and Wong Tai Sin contributed most of the high-degree cases (boxed in red). As this indicated these two districts as high-risk regions during outbreaks, they should be considered high-priority regions for disease control.

4. Discussion

We investigated the transmission patterns and dynamics of COVID-19 outbreaks in Hong Kong using spatiotemporal network analysis at both the individual and meta-population levels. The results offered some implications for the assessment of transmission risk and the deployment of policies and strategies for preventing disease spread.

First, the majority of confirmed cases were local or epidemiologically linked with local cases. Compared with the limited impact of imported cases, local cases dominated the outbreaks. Therefore, timely detection and treatment of the latest local cases become the key to preventing COVID-19 outbreaks, especially when the external/international transportation was already under strict control.

Second, among all the districts in our analysis, only quite a few (e. g., the Yau Tsim Mong district) played a key role in spreading the virus throughout the pandemic, especially during the fourth wave. In

Table 1

Comparison of cases with different out-degrees on the transmission network of COVID-19 in Hong Kong between January 23, 2020, and January 8, 2021. The data were represented in the format n (%), where n denotes the number of cases in each category and the value in brackets denotes the corresponding percentage. Furthermore, k denotes the out-degree of cases (nodes) and P-values were calculated based on the Pearson's chi-square test and the Fisher's exact test. Note that in the table, the case counts for districts could be duplicated if a case has visited multiple districts.

		$0 \leq k < 2^1$ $n = 6551\,(71.57)$	$2^1 \leq k < 2^2$ $n = 1124\,(12.28)$	$2^2 \leq k < 2^3$ $n = 752\,(8.22)$	$2^3 \leq k < 2^4$ $n = 371\,(4.05)$	$2^4 \leq k < 2^5$ $n = 200\,(2.19)$	$2^5 \leq k < 2^6$ $n = 120\,(1.31)$	$2^6 \leq k < 2^7$ $n = 35\,(0.38)$	P-value
Gender	F	3336 (50.92)	578 (51.42)	396 (52.66)	215 (57.95)	111 (55.5)	62 (51.67)	19 (54.92)	< 0.001
	M	3215 (49.08)	546 (48.58)	356 (47.34)	156 (42.05)	89 (44.5)	58 (48.33)	16 (45.71)	
Age	A1: 0–9 years	291 (4.44)	38 (3.38)	9 (1.2)	1 (0.27)	1 (0.5)	1 (0.83)	0 (0)	< 0.001
	A2: 10–19 years	472 (7.21)	58 (5.16)	25 (3.32)	4 (1.08)	2 (1)	2 (1.67)	0 (0)	
	A3: 20–29 years	1025 (15.65)	119 (10.59)	92 (12.23)	30 (8.09)	30 (15)	9 (7.5)	2 (5.71)	
	A4: 30–39 years	1117 (17.05)	176 (15.66)	113 (15.03)	44 (11.86)	24 (12)	15 (12.5)	0 (0)	
	A5: 40–49 years	1036 (15.81)	181 (16.1)	122 (16.22)	62 (16.71)	18 (9)	19 (15.83)	8 (22.86)	
	A6: 50–59 years	1017 (15.52)	218 (19.4)	141 (18.75)	79 (21.29)	45 (22.5)	29 (24.17)	6 (17.14)	
	A7: 60–69 years	940 (14.35)	200 (17.79)	161 (21.41)	84 (22.64)	37 (18.5)	24 (20)	13 (37.14)	
	A8: 70–79 years	425 (6.49)	84 (7.47)	55 (7.31)	37 (9.97)	21 (10.5)	14 (11.67)	5 (14.29)	
	A9: 80–89 years	178 (2.72)	39 (3.47)	29 (3.86)	26 (7.01)	16 (8)	6 (5)	1 (2.86)	
	A10: 90–99 years	48 (0.73)	11 (0.98)	5 (0.66)	4 (1.08)	6 (3)	1 (0.83)	0 (0)	
	A11: ≥ 100 years	2 (0.03)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
Asymptomatic	False	4376 (66.8)	963 (85.68)	651 (86.57)	332 (89.49)	183 (91.5)	103 (85.83)	34 (97.14)	< 0.001
	True	2175 (33.2)	161 (14.32)	101 (13.43)	39 (10.51)	17 (8.5)	17 (14.17)	1 (2.86)	
Residency	HK resident	6293 (96.06)	1103 (98.13)	722 (96.01)	363 (97.84)	199 (99.5)	114 (95)	33 (94.29)	< 0.001
	Non-HK resident	258 (3.94)	21 (1.87)	30 (3.99)	8 (2.16)	1 (0.5)	6 (5)	2 (5.71)	
Classification	Epidemiologically linked with imported case	37 (0.56)	2 (0.18)	2 (0.27)	0 (0)	0 (0)	0 (0)	0 (0)	< 0.001
	Epidemiologically linked with local case	3049 (46.54)	614 (54.63)	404 (53.72)	268 (72.24)	184 (92)	96 (80)	32 (91.43)	
	Local case	1607 (24.53)	425 (37.81)	284 (37.77)	94 (25.34)	15 (7.5)	24 (20)	3 (8.57)	
	Epidemiologically linked with possible local case	38 (0.58)	9 (0.8)	11 (1.46)	4 (1.08)	0 (0)	0 (0)	0 (0)	
	Imported case	1728 (26.38)	65 (5.78)	50 (6.65)	5 (1.35)	0 (0)	0 (0)	0 (0)	
	Possible local case	92 (1.4)	9 (0.8)	1 (0.13)	0 (0)	1 (0.5)	0 (0)	0 (0)	
Connection	Unlinked	3294 (50.28)	0 (0)	1 (0.13)	0 (0)	0 (0)	0 (0)	0 (0)	< 0.001
	Linked	3257 (49.72)	1124 (100)	751 (99.87)	371 (100)	200 (100)	120 (100)	35 (100)	
Time from onset to confirmation	0–7 days	5811 (88.7)	922 (82.03)	626 (83.24)	320 (86.25)	164 (82)	107 (89.17)	30 (85.71)	< 0.001
	8–14 days	610 (9.31)	188 (16.73)	117 (15.56)	38 (10.24)	29 (14.5)	13 (10.83)	4 (11.43)	
	> 14 days	130 (1.98)	14 (1.25)	9 (1.2)	13 (3.5)	7 (3.5)	0 (0)	1 (2.86)	
District	D1: Central & Western	680 (10.38)	193 (17.17)	123 (16.36)	71 (19.14)	57 (28.5)	31 (25.83)	1 (2.86)	< 0.001
	D2: Eastern	642 (9.8)	148 (13.17)	113 (15.03)	33 (8.89)	5 (2.5)	14 (11.67)	3 (8.57)	
	D3: Southern	275 (4.2)	48 (4.27)	19 (2.53)	11 (2.96)	7 (3.5)	1 (0.83)	0 (0)	
	D4: Wan Chai	626 (9.56)	217 (19.31)	145 (19.28)	81 (21.83)	109 (54.5)	117 (97.5)	3 (8.57)	
	D5: Yau Tsim Mong	1218 (18.59)	392 (34.88)	270 (35.9)	191 (51.48)	163 (81.5)	141 (117.5)	68 (194.29)	
	D6: Sham Shui Po	719 (10.98)	177 (15.75)	179 (23.8)	114 (30.73)	38 (19)	18 (15)	13 (37.14)	
	D7: Kowloon City	1404 (21.43)	406 (36.12)	300 (39.89)	222 (59.84)	66 (33)	48 (40)	6 (17.14)	
	D8: Kwun Tong	852 (13.01)	275 (24.47)	349 (46.41)	184 (49.6)	64 (32)	22 (18.33)	17 (48.57)	
	D9: Wong Tai Sin	617 (9.42)	244 (21.71)	297 (39.49)	232 (62.53)	161 (80.5)	191 (159.17)	34 (97.14)	
	D10: Islands	258 (3.94)	43 (3.83)	38 (5.05)	13 (3.5)	6 (3)	1 (0.83)	0 (0)	
	D11: Kwai Tsing	540 (8.24)	182 (16.19)	119 (15.82)	68 (18.33)	38 (19)	8 (6.67)	5 (14.29)	
	D12: North	212 (3.24)	61 (5.43)	65 (8.64)	29 (7.82)	13 (6.5)	0 (0)	0 (0)	
	D13: Sai Kung	427 (6.52)	157 (13.97)	100 (13.3)	49 (13.21)	34 (17)	28 (23.33)	6 (17.14)	
	D14: Sha Tin	681 (10.4)	240 (21.35)	166 (22.07)	107 (28.84)	49 (24.5)	11 (9.17)	12 (34.29)	
	D15: Tai Po	299 (4.56)	66 (5.87)	67 (8.91)	11 (2.96)	11 (5.5)	4 (3.33)	1 (2.86)	
	D16: Tsuen Wan	336 (5.13)	141 (12.54)	99 (13.16)	35 (9.43)	5 (2.5)	8 (6.67)	2 (5.71)	
	D17: Tuen Mun	592 (9.04)	158 (14.06)	111 (14.76)	74 (19.95)	44 (22)	11 (9.17)	4 (11.43)	
	D18: Yuen Long	556 (8.49)	101 (8.99)	81 (10.77)	23 (6.2)	25 (12.5)	4 (3.33)	1 (2.86)	

Table 2

Comparison of cases with different in-degrees on the transmission network of COVID-19 in Hong Kong between January 23, 2020, and January 8, 2021. The data were represented in the format n (%), where n denotes the number of cases in each category and the value in brackets denotes the corresponding percentage. Furthermore, k denotes the in-degree of cases (nodes) and P-values were calculated based on the Pearson's chi-square test and the Fisher's exact test. Note that in the table, the case counts for districts could be duplicated if a case has visited multiple districts.

		$0 \leq k < 2^1$ $n = 6362$ (69.51)	$2[1] \leq k < 2^2$ $n = 1259$ (13.76)	$2[2] \leq k < 2^3$ $n = 791$ (8.64)	$2[3] \leq k < 2^4$ $n = 396$ (4.33)	$2[4] \leq k < 2^5$ $n = 215$ (2.35)	$2[5] \leq k < 2^6$ $n = 104$ (1.14)	$2[6] \leq k < 2^7$ $n = 26$ (0.28)	P-value
Gender	F	3220 (50.61)	662 (52.58)	430 (54.36)	220 (55.56)	117 (54.42)	50 (48.08)	18 (69.23)	< 0.001
	M	3142 (49.39)	597 (47.42)	361 (45.64)	176 (44.44)	98 (45.58)	54 (51.92)	8 (30.77)	
Age	A1: 0–9 years	242 (3.8)	64 (5.08)	32 (4.05)	1 (0.25)	1 (0.47)	1 (0.96)	0 (0)	< 0.001
	A2: 10–19 years	435 (6.84)	86 (6.83)	34 (4.3)	5 (1.26)	2 (0.93)	1 (0.96)	0 (0)	
	A3: 20–29 years	957 (15.04)	189 (15.01)	84 (10.62)	36 (9.09)	32 (14.88)	9 (8.65)	0 (0)	
	A4: 30–39 years	1079 (16.96)	200 (15.89)	121 (15.3)	53 (13.38)	23 (10.7)	12 (11.54)	1 (3.85)	
	A5: 40–49 years	1039 (16.33)	196 (15.57)	107 (13.53)	60 (15.15)	30 (13.95)	14 (13.46)	0 (0)	
	A6: 50–59 years	1020 (16.03)	202 (16.04)	157 (19.85)	80 (20.2)	41 (19.07)	24 (23.08)	11 (42.31)	
	A7: 60–69 years	968 (15.22)	191 (15.17)	137 (17.32)	87 (21.97)	47 (21.86)	21 (20.19)	8 (30.77)	
	A8: 70–79 years	409 (6.43)	89 (7.07)	63 (7.96)	34 (8.59)	25 (11.63)	17 (16.35)	4 (15.38)	
	A9: 80–89 years	173 (2.72)	33 (2.62)	43 (5.44)	30 (7.58)	11 (5.12)	4 (3.85)	1 (3.85)	
	A10: 90–99 years	39 (0.61)	8 (0.64)	13 (1.64)	10 (2.53)	3 (1.4)	1 (0.96)	1 (3.85)	
	A11: ≥ 100 years	1 (0.02)	1 (0.08)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
Asymptomatic	False	4559 (71.66)	991 (78.71)	554 (70.04)	303 (76.52)	152 (70.7)	69 (66.35)	14 (53.85)	< 0.001
	True	1803 (28.34)	268 (21.29)	237 (29.96)	93 (23.48)	63 (29.3)	35 (33.65)	12 (46.15)	
Residency	HK resident	6103 (95.93)	1245 (98.89)	755 (95.45)	389 (98.23)	211 (98.14)	99 (95.19)	25 (96.15)	< 0.001
	Non-HK resident	259 (4.07)	14 (1.11)	36 (4.55)	7 (1.77)	4 (1.86)	5 (4.81)	1 (3.85)	
Classification	Epidemiologically linked with imported case	33 (0.52)	7 (0.56)	1 (0.13)	0 (0)	0 (0)	0 (0)	0 (0)	< 0.001
	Epidemiologically linked with local case	2573 (40.44)	879 (69.82)	561 (70.92)	324 (81.82)	199 (92.56)	85 (81.73)	26 (100)	
	Local case	1903 (29.91)	278 (22.08)	176 (22.25)	60 (15.15)	16 (7.44)	19 (18.27)	0 (0)	
	Epidemiologically linked with possible local case	35 (0.55)	13 (1.03)	9 (1.14)	5 (1.26)	0 (0)	0 (0)	0 (0)	
	Imported case	1723 (27.08)	75 (5.96)	44 (5.56)	6 (1.52)	0 (0)	0 (0)	0 (0)	
	Possible local case	95 (1.49)	7 (0.56)	0 (0)	1 (0.25)	0 (0)	0 (0)	0 (0)	
Connection	Unlinked	3295 (51.79)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	< 0.001
	Linked	3067 (48.21)	1259 (100)	791 (100)	396 (100)	215 (100)	104 (100)	26 (100)	
Time from onset to confirmation	0–7 days	5435 (85.43)	1141 (90.63)	721 (91.15)	361 (91.16)	201 (93.49)	97 (93.27)	24 (92.31)	< 0.001
	8–14 days	778 (12.23)	108 (8.58)	62 (7.84)	30 (7.58)	12 (5.58)	7 (6.73)	2 (7.69)	
	> 14 days	149 (2.34)	10 (0.79)	8 (1.01)	5 (1.26)	2 (0.93)	0 (0)	0 (0)	
District	D1: Central & Western	754 (11.85)	170 (13.5)	83 (10.49)	61 (15.4)	68 (31.63)	20 (19.23)	0 (0)	< 0.001
	D2: Eastern	622 (9.78)	150 (11.91)	114 (14.41)	41 (10.35)	25 (11.63)	6 (5.77)	0 (0)	
	D3: Southern	285 (4.48)	50 (3.97)	14 (1.77)	5 (1.26)	5 (2.33)	2 (1.92)	0 (0)	
	D4: Wan Chai	693 (10.89)	199 (15.81)	112 (14.16)	106 (26.77)	132 (61.4)	53 (50.96)	3 (11.54)	
	D5: Yau Tsim Mong	1255 (19.73)	433 (34.39)	280 (35.4)	124 (31.31)	185 (86.05)	131 (125.96)	35 (134.62)	
	D6: Sham Shui Po	728 (11.44)	233 (18.51)	119 (15.04)	109 (27.53)	50 (23.26)	17 (16.35)	2 (7.69)	
	D7: Kowloon City	1490 (23.42)	402 (31.93)	274 (34.64)	150 (37.88)	90 (41.86)	34 (32.69)	12 (46.15)	
	D8: Kwun Tong	910 (14.3)	319 (25.34)	292 (36.92)	175 (44.19)	44 (20.47)	20 (19.23)	3 (11.54)	
	D9: Wong Tai Sin	625 (9.82)	307 (24.38)	297 (37.55)	250 (63.13)	129 (60)	146 (140.38)	22 (84.62)	
	D10: Islands	254 (3.99)	56 (4.45)	33 (4.17)	11 (2.78)	3 (1.4)	1 (0.96)	1 (3.85)	
	D11: Kwai Tsing	533 (8.38)	188 (14.93)	124 (15.68)	84 (21.21)	19 (8.84)	9 (8.65)	3 (11.54)	
	D12: North	242 (3.8)	46 (3.65)	31 (3.92)	45 (11.36)	13 (6.05)	0 (0)	3 (11.54)	
	D13: Sai Kung	468 (7.36)	139 (11.04)	87 (11)	47 (11.87)	39 (18.14)	21 (20.19)	0 (0)	
	D14: Sha Tin	686 (10.78)	303 (24.07)	164 (20.73)	71 (17.93)	31 (14.42)	10 (9.62)	1 (3.85)	
	D15: Tai Po	301 (4.73)	66 (5.24)	46 (5.82)	19 (4.8)	24 (11.16)	2 (1.92)	1 (3.85)	
	D16: Tsuen Wan	392 (6.16)	102 (8.1)	79 (9.99)	32 (8.08)	14 (6.51)	5 (4.81)	2 (7.69)	
	D17: Tuen Mun	649 (10.2)	119 (9.45)	119 (15.04)	67 (16.92)	33 (15.35)	5 (4.81)	2 (7.69)	
	D18: Yuen Long	546 (8.58)	131 (10.41)	67 (8.47)	29 (7.32)	15 (6.98)	2 (1.92)	1 (3.85)	

fact, the population density in these areas is extremely high. Taking the Yau Tsim Mong district as an example: it is one of the most densely populated districts in Hong Kong, with over 48,000 persons per square kilometer. Moreover, these areas generally possess a large number of public venues for social gatherings, such as restaurants and singing/dancing clubs, thereby providing opportunities for the widespread and rapid transmission of the virus. Thus, it should be given the highest priority when deploying district-specific social distancing or intervention strategies. This implication was consistent with the Hong Kong government's recent lockdown and stringent mandatory coronavirus testing policies in the densely populated areas of Jordan, Mong Kok, and Yau Ma Tei in the Yau Tsim Mong district for identifying and obstructing the chain of transmission [25,26]. Note that the data we used in our analysis were between January 23, 2020 to January 8, 2021 while the government's policies were implemented on January 23, 2021, which is two weeks later than our analysis. The consistency between our analysis results and the implemented policies indicated that the proposed spatiotemporal connectivity analysis enables us to provide an early warning of the high-priority regions or locations for intervention, demonstrating the potential of the analysis in predicting the trend of the outbreak. Furthermore, the cross-district and cross-population transmission matrices constructed from the spatiotemporal connectivity analysis can be regarded as a generalization of the next-generation matrices of compartmental models for future epidemic forecasting.

Third, the reported cases and highly connected cases were mainly middle-aged and elderly people (e.g., 40- to 69-year-olds). Compared with young people, these individuals may be more active in various social activities and likely to gather in large public spaces, such as restaurants and singing/dancing clubs, and thus have higher chances of being infected by or infecting others. Therefore, the current policy prohibiting group gatherings implemented by the government can effectively help reduce the transmission risk. Obviously, the advantage of implementing age-specific control strategies is that it enables the decision-makers to take into consideration the age-specific susceptibility, transmissibility, mobility patterns, and social habits in the risk estimation as well as the corresponding intervention strategies, so that the high-risk age-groups can be identified and treated in a timely and effective way. However, compared to the location-specific strategies (e.g., lockdown and mandatory testing in a specific region), the age-specific strategies are not that easy to be implemented due to the mobility and uncertainty of the population in or between different regions/locations. One possible way to benefit from the advantages of both age-specific and location-specific strategies would be to implement the age-specific control or interventions in a location-specific manner. For example, to address the risk associated with the gathering of middle-aged and elderly people (40- to 69-year-olds), we could implement the strategy of social gathering ban in places where such high-risk age-groups are most likely to gather, such as singing/dancing clubs. By doing so, the age-specific transmission could be contained in a relatively more informed and effective way.

Finally, all of the waves of COVID-19 outbreaks in Hong Kong, even those with similar statistical distributions (Figs. (2,3)), can be systematically scrutinized, characterized, and differentiated with respect to their spatiotemporal and population-specific transmission patterns. To achieve effective control of the disease spread, tailor-made strategies must be developed and deployed according to the transmission dynamics rather than a static or overall statistic. By revealing the transmission patterns of COVID-19 among different age-groups within/across populated geographic districts/locations throughout different outbreak phases, we provided objective and quantifiable guidance for the control and prevention of COVID-19 outbreaks not only in Hong Kong but also in other international metropolises with similar population density, connectivity, and mobility characteristics.

The proposed spatiotemporal analysis method is general at the methodological level and instructive in understanding the COVID-19 transmission patterns among different age-groups within/across populated geographic locations throughout different outbreak phases. However, it should be noted that, when applying the method to other cities or different time periods, the location/time-specific settings, such as the data scale and availability, the testing capacity, the case categories, and the distribution of age-specific populations and mobility patterns should be taken into consideration, thus making the analysis results more situation-specific and informative. For example, compared to other large cities, Hong Kong had relatively few cases of COVID-19 during its various waves and fairly complete data information. Moreover, different cities/regions, during different time periods, may also present heterogeneities in other aspects related to disease transmission, e.g., the testing capacity, the percentage of the indigenous cases and imported cases, and the mobility patterns and social habits of different age-groups. Therefore, the numerical results reported in this study may not be directly applicable to other situations. To ensure that the analysis is tailor-made, and the results are context-aware, specific data, information, and settings in other cities/regions or different time periods should be adaptively used to guide the location/time-specific parameterization of the model.

Note that in this work, the information of activities/clusters (e.g., the singing/dancing clubs) that may lead to the transmission was directly obtained from the data website of the Hong Kong government [27], not inferred from our analysis. What we have done was integrating the spatiotemporal connectivity method with the obtained activity/cluster-related information to conduct a retrospective analysis. In fact, the reported information on activities or contacts that led to the transmission could be uncertain because the incubation period for COVID-19 ranges from 1 to 14 days [28], making the accurate determination of infection sources difficult. Therefore, instead of providing a deterministic inference on the activities/contacts that led to each transmission, the analysis conducted in this paper aimed to offer a better understanding of the spatiotemporal dynamics and trends of the disease spread, so as to indicate the potential scenarios that may have a relatively high risk of transmission for targeted control and intervention.

5. Data sharing

The modeling/analytics tools of this study will be made publicly available on: http://aic.hkbu.ai/.

CRediT authorship contribution statement

Yang Liu: Conceptualization, Visualization, Data curtion, Formal analysis, Writing – original draft. **Zhonglei Gu:** Data curtion, Formal analysis, Writing – original draft. **Jiming Liu:** Conceptualization, Visualization, Data curtion, Formal analysis, Writing – original draft.

Declaration of Interests

We declare no competing interests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.eclinm.2021.100929.

References

[1] World Health Organization, (WHO), WHO coronavirus disease (COVID-19) dashboard, https://covid19.who.int/.

[2] Dong E, Du H, Gardner L. An interactive Web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;20:533–4.

[3] Johns Hopkins coronavirus resource center, COVID-19 map, https://coronavirus.jhu.edu/.

[4] Gibney E. Whose coronavirus strategy worked best? Scientists hunt most effective policies. Nature 2020;581:15–6.

[5] Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, Stephenson AB, Leech G, Altman G, Mikulik V, Norman AJ, Monrad JT, Besiroglu T, Ge H, Hartwick MA, Teh YW, Chindelevitch L, Gal Y, Kulveit J. Inferring the effectiveness of government interventions against COVID-19. Science 2020. doi: 10.1126/science.abd9338.

[6] Liu Y, Gu Z, Xia S, Shi B, Zhou X-N, Shi Y, Liu J. What are the underlying transmission patterns of COVID-19 outbreak? An age-specific social contact characterization. EClinicalMedicine 2020;22:100354.

[7] Walker PGT, Whittaker C, Watson OJ, Baguelin M, Winskill P, Hamlet A, Djafaara BA, Cucunubá Z, Mesa DO, Green W, Thompson H, Nayagam S, Ainslie KEC, Bhatia S, Bhatt S, Boonyasiri A, Boyd O, Brazeau NF, Cattarino L, Cuomo-Dannenburg G, Dighe A, Donnelly CA, Dorigatti I, van Elsland SL, FitzJohn R, Fu H, Gaythorpe KAM, Geidelberg L, Grassly N, Haw D, Hayes S, Hinsley W, Imai N, Jorgensen D, Knock E, Laydon D, Mishra S, Nedjati-Gilani G, Okell LC, Unwin HJ, Verity R, Vollmer M, Walters CE, Wang H, Wang Y, Xi X, Lalloo DG, Ferguson NM, Ghani AC. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. Science 2020;369:413–22.

[8] Laxminarayan R, Wahl B, Dudala SR, Gopal K, Mohan B. C, Neelima S, Reddy KSJ, Radhakrishnan J, Lewnard JA. Epidemiology and transmission dynamics of COVID-19 in two Indian states. Science 2020;370:691–7.

[9] Dickens BL, Koo JR, Lim JT, Park M, Quaye S, Sun H, Sun Y, Pung R, Wilder-Smith A, Chai LYA, Lee VJ, Cook AR. Modelling lockdown and exit strategies for COVID-19 in Singapore. Lancet Reg Health West Pac 2020;1:100004.

[10] Wong NS, Lee SS, Kwan TH, Yeoh E-K. Settings of virus exposure and their implications in the propagation of transmission networks in a COVID-19 outbreak. Lancet Reg Health West Pac 2020;4:100052.

[11] Thu TPB, Ngoc PNH, Hai NM, Tuan LA. Effect of the social distancing measures on the spread of COVID-19 in 10 highly infected countries. Sci Total Environ 2020;742:140430.

[12] Kucharski AJ, Klepac P, Conlan AJK, Kissler SM, Tang ML, Fry H, Gog JR, Edmunds WJ. on behalf of the CMMID COVID-19 working group, effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. Lancet Infect Dis 2020;20:1151–60.

[13] Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, Liao Q, Kwan MYW, Lee SL, Chiu SS, Wu JT, Wu P, Leung GM. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health 2020;5:E279–88.

[14] Dickens BL, Koo JR, Lim JT, Park M, Quaye S, Sun H, Sun Y, Pung R, Wilder-Smith A, Chai LYA, Lee VJ, Cook AR. Modelling lockdown and exit strategies for COVID-19 in Singapore. Lancet Reg Health West Pac 2020;1:100004.

[15] Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. Nat Commun 2020;11:5710.

[16] Sun K, Wang W, Gao L, Wang Y, Luo K, Ren L, Zhan Z, Chen X, Zhao S, Huang Y, Sun Q, Liu Z, Litvinova M, Vespignani A, Ajelli M, Viboud C, Yu H. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. Science 2021;371:eabe2424.

[17] Adam DC, Wu P, Wong JY, Lau EHY, Tsang TK, Cauchemez S, Leung GM, Cowling BJ. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nat Med 2020;26:1714–9.

[18] Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Leskovec J. Mobility network models of COVID-19 explain inequities and inform reopening. Nature 2021;589:82–7.

[19] DATA.GOV.HK, Data in coronavirus disease (COVID-19), https://data.gov.hk/en-data/dataset/hk-dh-chpsebcddr-novel-infectious-agent.

[20] Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the international AAAI conference on weblogs and social media; 2009 https://gephi.org/users/publications/.

[21] The government of Hong Kong special administrative region, COVID-19 thematic website, https://www.coronavirus.gov.hk/eng/index.html.

[22] O.T.&.P. Healthcare, A timeline of COVID-19 and ot&p updates, https://www.otandp.com/covid-19-timeline.

[23] Westbrook L. Dance off: the niche Hong Kong social scene behind city's biggest Covid-19 cluster. South China Morning Post 2020 https://ph.news.yahoo.com/dance-off-niche-hong-kong-101207169.html.

[24] Duhalde M, Choy G, Wong D. Visualising Hong Kong's biggest Covid-19 superspreader event. South China Morning Post 2021 https://multimedia.scmp.com/infographics/news/hong-kong/article/3115971/hk-covid-cluster/index.html.

[25] News.gov.hk, Restricted area in Jordan set up (2021), https://www.news.gov.hk/eng/2021/01/20210123/20210123_111902_776.html?type=category&name=covid19.

[26] R. Pyne, More areas come under stricter COVID testing rules. rthk.hk (2021), https://news.rthk.hk/rthk/en/component/k2/1572465-20210126.htm.

[27] https://data.gov.hk/en-data/dataset/hk-dh-chpsebcddr-novel-infectious-agent/resource/f3ec2c1a-2c8e-4b10-acd4-91aeca19f694

[28] https://www.chp.gov.hk/en/healthtopics/content/24/102466.html