Intentional Travel Groups and Social Networks during the COVID-19 Pandemic

Mingzhi Zhou
University of Hong Kong
Shuyu Lei
University of Hong Kong
Jiangyu Wu
University of Hong Kong
Hanxi Ma
University of Hong Kong
David Levinson
University of Sydney
Jiangping Zhou (zhoujp@hku.hk)
University of Hong Kong

Article

Keywords:

Posted Date: May 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1548702/v1

Abstract

Although face-to-face social contacts decreased during the COVID-19 pandemic, some people remain physically traveling and meeting as a group to gain benefits like sustaining intimacy and increasing productivity. Using multiday continuous smartcard data, we identify intentional group travel patterns in Hong Kong Metro system. Those patterns serve as our proxies for physical (visible) interactions and social (invisible) contact networks among people who intentionally travel as a group (ITGs). We measure the spatial centrality of ITGs and persistent group riders (PGRs), a subset of ITGs remaining active amid the pandemic, to infer different locales' social interactions in the city. By examining the social network formed by ITGs across time, we found that its size and interconnections varied during the pandemic and PGRs might be the most influential vertices in maintaining the networks' topological properties. The findings could facilitate transit-usage-and-virus-spread modelling and the formulation of more effective pandemic countermeasures in transit-reliant cities.

Background

The COVID-19 virus often spreads within social interactive networks[1-2]. In response, face-to-face meetings decreased, as a consequence of both voluntary reductions by people and anti-pandemic measures implemented by governments to restrict group activities, e.g., bans on gatherings and closure of non-essential businesses[2-4]. Hence, COVID-19 greatly suppressed physical social interactions, part of which were increasingly integrated into the digital interactive network[5-6] and others were probably forgone. However, there remain in-person social activities that encapsulate those resilient, meaningful and enduring social bonds such as friends, relatives and family in the city. Studies hitherto have examined behavioral changes and the susceptible-infected-recovered process with various non-traditional data sources, e.g., cellular signal data, online maps, socio-economic data, and time-use diaries[6-9]. Yet, to our knowledge, little research has been conducted on how particular physical social interactions reacted to the pandemic. This is largely due to the lack of empirical data concerning those interactions.

The Hong Kong Metro, including its supporting facilities like platforms, comprises important public spaces where different social interactions such as encounters, familiar strangers, and group travel behaviors frequently occur[10-12]. In this article, we probe intentional travel groups (ITGs) within a Metro system based on transit smartcard ("Octopus") data from Hong Kong amid the pandemic. The data spans the year 2020 and represents a notable percentage of physical social interactions in the city, where Metro accounts for 41% of the local passenger trips[13]. We assume that two riders in one ITG know each other and intentionally travel together (Figure 1c). They are grouped through consistency in timing, location, and contact frequency (see Methods and Supplementary Information). ITGs are a subset of inter-relationship among riders and they form an invisible but meaningful "social network" ("ITG network"). Only weekend data is used since there are fewer concurrent trips for commuting and more social activities for leisure and maintenance. In this network, individual ITG riders can be regarded as nodes and the group travel interactions among them, i.e., invisible connections (lines in a social network

graph) between individual ITG riders, can be regarded as links. Based on Octopus records ITG riders' information, we can identify their spatial and temporal distribution across the city.

After operationalizing the ITG network using Octopus data, we examine whether and to what degree the resulting ITG network follows commonly-reported typological properties of other social interactive networks that have been previously studied[14]. These property characteristics often include: 1) the degree of the network is scale-free and has power-law-like distribution; 2) the interconnections of the network, measured by the average degree and the relative size of the largest cluster, often increase as time progresses.

Results

Big picture. Through the year, there were 1,337,774 ITGs and 1,542,050 distinct ITG riders identified in the network. Daily ITG riders account for 10–16% of all daily riders. About half of all ITGs consist of two adult riders, and the proportion climbs to over 60% when we particularly examine ITGs with different home stations. We observe 23.5% of all ITGs include one adult and one child (Fig. 1a). In particular, we define riders' "home" stations as those riders start their first metro trip of a day most frequently within a year. It could be found that 74% of all ITGs include two riders with the same "home" stations. These ITGs have an average of 10.96 ITG trips throughout the year, which is more than that of other ITGs with 8.24 trips on average. Considering "family" members are more likely to live together and have higher contact frequency, the ITGs including riders with the same home stations could at least partially reflect the share of "family" or "neighbors" in the dataset.

Figure 1b presents monthly ITG riders. Hong Kong suffered from four surges of locally confirmed COVID-19 caseloads during 2020 since the first such case emerged on Jan 23[15]. As expected, the numbers of both ITGs and ITG riders were significantly suppressed after COVID-19 hit the city. Specifically, the quantity of ITGs decreased from 313,441 in January 2020 to 129,858 in February 2020. The size of ITG riders in February dropped to only 44% of that in January. In terms of persistent group riders (PGRs), who remain travelling as ITGs, 85,809 PGRs, 45% of all ITG riders in January, retained their intentional group travel in February after the local pandemic outbreak. Despite the presence of COVID-19, 9,461 PGRs remained in ITG for all the twelve months in 2020.

Destination distribution of ITG Riders. To explore how ITG riders' destinations were impacted by the pandemic and how related potential transmission risks by destination changed across different surges, we compared the spatial distribution of those destinations before and after the outbreak of the pandemic. To quantify the spatial distribution, a percolation analysis was performed, which enabled us to efficiently single out the top destinations (Metro stations) with the most incoming ITG trips[16]. The analysis' detail is given in the Methods.

Figure2 presents the spatial patterns of the top Metro stations in terms of accommodating the most incoming ITG trips across different periods based on the percolation analysis. Not surprisingly, before the

local pandemic outbroke, the ITG riders centered around Kowloon South and Hong Kong Island North, where there is the highest concentration of local leisure, commercial, catering, and business facilities (Fig. 2(b)). Specifically, stations serving Causeway Bay (Station 28), Tsim Sha Tsui (Station 3), and Mong Kok (Station 6) stood out among all the local Metro stations.

After the local pandemic outbroke, only PGRs continued travelling as a group. The top Metro stations that these riders visited varied notably from those for ITG riders pre the pandemic. First, Tesung Kwan O (Station 50) and its neighboring stations became one of the top 3 destinations. Second, Mong Kok was no longer within the top 3 or even top 10 list. Third, Causeway Bay was still one of the most popular destinations whereas other stations such as Tai Wai emerged in the list of these destinations across the four surges of locally confirmed cases. Throughout the four surges, Tesung Kwan O was a destination that PGRs continued visiting. Given that on top of and around the station and its neighboring stations, it is one the latest "Rail + Property" new towns of Hong Kong, one can say that the corresponding new town was quite self-contained for PGRs before and amid COVID-19, i.e., local facilities and opportunities were able to satisfy most need of these riders [(c.f., 17)]. Most if not all riders/residents in this new town only need a short Metro ride to have their needs met, which exemplifies both the well-discussed neighborhood planning[18] and the emerging 15-minute city concept[19–20].

A Social Interactive Network among Metro Riders during the Pandemic. Regarding ITG as a proxy for social interactions among Metro riders, a Metro "social network" (ITG network) is constructed. This network is facilitated by the presence and usage of the local Metro system. The network's nodes are individual ITG riders, and the links exist when any two ITG riders intentionally travel as a group.

Against the backdrop of the COVID-19 pandemic, we examine whether and how the network properties changed significantly amid the pandemic surges. Overall, the daily quantity of all riders, ITG riders, and ITGs decreased after the local pandemic outbroke, hit rock bottom in the first and second surge, and then rebounded gradually in the third and fourth surge (Fig. 3a). Particularly, Fig. 3b shows the average quantity of ITG riders by hour of a weekend day pre-pandemic and in each surge. In addition to an obvious decrease after the pandemic outbroke, the hourly ITG riders were fewer in the first and second surge than those in later two surges. During the surges, the pre-pandemic peak hour (5 pm) was flattened, which indicates some trips might shift from peak to off-peak hours. Despite the variations in quantity, we found that the average lag between any two ITG trips of all the ITG riders fluctuated amid the pandemic, with 82 hours during the pre-pandemic time and then 92, 138, 237, and 167 hours in the following four surges, respectively. It seems that the lag increased right after the local pandemic outbroke while decreased when people became aware that there could be multiple surges of locally confirmed cases.

To explore the variation among different periods of each surge, Fig. 3c depicts the COVID-19 case surges and their respective quiet (the time pre-COVID-19 or when daily confirmed cases remained low), climbing (when daily confirmed cases were increasing), and declining periods (when daily confirmed cases were decreasing) in Hong Kong in 2020. All ITGs were allocated to the different periods of each surge. Concerning ITGs across the three periods, we proposed and tested the first hypothesis:

Hypothesis 1

The size and interconnections in the ITG network varied significantly across the quiet, climbing, and declining periods of each pandemic surge.

Figure 3d shows the variation in the ratio of ITG riders to all the riders as the time approaches a surge's zenith, when the number of daily locally confirmed cases reached the highest of the given surge. The ratio of ITG riders to all the riders in each surge fluctuated. Except the fourth surge, a "smiling" curve in variation could be found for the first, second, and third surges. In the fourth surge, one half of the "smiling" curve could also be seen despite of incomplete data. This means that there was a negative correlation between the share of ITG riders and the daily confirmed cases. Quantity of ITG riders decreased disproportionately strongly. Interestingly, even though there were a much larger number of locally confirmed cases in the third and fourth surges, the share of the daily ITG riders in these two surges were no smaller and the slope of the corresponding "smiling" curves were even flatter than their counterparts in the first and second surges. This implies that ITG riders became less sensitive to the locally confirmed cases as the fight against COVID-19 had become a long-lasting task[21].

Student's t-tests were performed to examine the variation in the ITG network's interconnections, measured by the degree (k) of ITG riders, and size, measured by the average number (n) of ITG trips of each Metro rider per weekend day among the different periods of each surge. The degree (k) equals the number of other riders that a given ITG rider has ITG trips with. One hypothesis for the analysis is that k and n during the quiet period of a given surge would be lower than that within the surge.

Table 1 presents the t-test results, which were statistically significant for k or n models for the last three surges. This indicates that the pandemic situations did reduce interconnectedness and size of the ITG networks. Another hypothesis is that the two parameters would differ between two periods within a surge. Results of most n models show statistically significant lower values in declining periods compared with the climbing periods. This complies to the previous study that responses from Metro travel behaviors would lag behind the variation in COVID-19 transmission[22]. However, k of the last three surges is statistically significantly lower during the climbing periods than the declining periods, which implies lower interconnectedness of the network while newly confirmed cases are growing. Some analysis results incompatible with the above explanations might be impacted by implementation of local anti-pandemic measures during the certain periods. For example, closure of cross-border check points occurring within the climbing period of the first surge, and bans on gatherings and international travel were implemented during the climbing periods of the second surge.

Table 1
t-test Results for the Intentional Travel Group (ITG) network

	Average Degree for the ITG Riders (*k*)			Average ITG trips per rider (*n*)		
Surge	$k_s < k_q$	$k_d > k_c$	$k_d < k_c$	$n_s < n_q$	$n_d > n_c$	$n_d < n_c$
1st Surge	44.54	-	2.46***	10.81	-	37.17***
2nd Surge	14.17***	16.87***	-	11.83***	10.25***	-
3rd Surge	13.69***	9.68***	-	24.93***	-	6.89***
4th Surge	3.72	11.03***	-	6.71***	-	9.06***

t-statistic is reported. ***p < 0.01, **0.01≤ p < 0.05, *0.05≤ p < 0.1. *s* denotes data during the given surge which consists of both the climbing and declining periods; *q* denotes data in the quiet period before the given surge; *c* denotes data during the climbing period and *d* denotes data during the declining period. Data is normalized into daily average.

A Scale-free ITG network. Anti-pandemic measures and people's fear of infection could reduce physical contacts in the city. Subsequently, the ITG network could collapse. However, most social interactive networks are found to be scale-free in distribution[14] and scale-free networks are usually robust when faced with malfunctional nodes and exogenous attacks[23–24]. That is, a scale-free network hierarchically includes dominant nodes with a large degree followed by less-dominant nodes. In the network, nodes with a small degree account for vast majority of the network. Therefore, the overall structure of the network usually would not be affected if we randomly remove a small fraction of nodes. We consider that this would also be true for the ITG network amid the pandemic. Such properties would increase metro travel anxiety of many people and notably reduce outgoings of many Metro riders. Hence, the following two hypotheses are proposed:

Hypothesis 2

The structure of the ITG network remains robust amid the pandemic.

The topology of the network remains relatively robust as a whole despite of the pandemic. Figure 4a compares the complementary cumulative distribution functions (CCDF) of ITGs network before the pandemic outbroke with those in different COVID-19 surges. The ITG networks perform statistically significant in power-law distribution spanning all periods with p-value of KS-tests smaller than 0.01. The α decreased after the pandemic outbroke and then fluctuated across surges around 3 to 5. This indicates that riders with a higher degree account for a larger proportion of ITG riders of a local surge compared to the pre-pandemic time.

Hypothesis 3

The ITG network increases in time and follows a scale-free distribution.

Shown in Fig. 4b, interconnections of the network can be measured by two indicators: average degree (*k*) and the relative size of the largest cluster (*S*). Despite the *k* mentioned above, the *S* is the ratio between the total number of riders in the largest cluster and that in the whole network. Both *k* and *S* increased in an approximately linear manner. The growth of *k* has three contributors. One, new ITG riders made ITG trips together and joined the existing ITG network. Two, these new riders travelled with existing ITG riders so new ITGs trips emerged. Three, new ITG trips emerged among existing ITG riders. A cluster of the ITG network represents a subset of ITG riders who became connected because of the presence of ITG trips. The ITG network of Metro riders is embedded in the large social interactive network in the city. A larger *S* indicates that an increasingly dominant cluster in the ITG network and possibly in the large social interactive network had emerged.

As for the network structure, we used the whole year's ITG dataset to examine distribution of frequency and degree of the ITG riders. As shown in Fig. 4c, the trip frequencies of most riders in the network typically increases as the degree increases. For all ITG riders, their trip frequencies' mean and maximum are 28 and 1,132, respectively. Figure 4d presents the degree distribution in the logarithmic binning format. The distribution of all the ITG riders and that of non-commuting ITG riders remain similar. The average degree of all the ITG riders is 1.73 and the max value is 52, lower than 150, the suggested cognitive limit to the number of people with whom one could maintain stable social relationships[25]. In particular, we found that the probability *P(x)* that a rider is connected with at least x other riders in the ITG network decays following a power law, which can be expressed as $P(x) \sim x^{-\alpha}$, where α = 5.67. The Kolmogorov-Smirnov test (KS-test) was performed to examine whether the network's degree distribution statistically follows the power-law distribution. The test produces a p-value smaller than 0.01. This result demonstrates that the ITG network indeed resembles other social networks with a scale-free distribution in degree.

Who maintains the ITG network? According to what we know about the structural robustness of scale-free networks[23], a probable explanation for the ITG network's relative stability amid is the continued presence of PGRs, who remained active in ITG travel despite of the surges in locally confirmed cases. Table 2 presents the t-test results regarding the degree and number of the PGRs and non-PGRs. The results corroborate the hypothesis that the degree and number of PGRs would be significantly higher than those of non-PGRs in any given surge ($p < 0.01$). Not surprisingly, even the PGRs were impacted by the surge of the locally confirmed cases–their *k* and *n* in the four climbing periods were mostly smaller than those in the corresponding quiet periods. To further explore the role of PGRs in maintaining the ITG network's robustness, Hypothesis 4 was proposed.

Table 2
t-Test Analysis for Persistent Group Riders (PGRs) in the Intentional Travel Group ITG network

	Average Degree for PGRs (*k*)		**Average ITG trips for PGRs (*n*)**	
Surge	$k_s < k_q$	$k_{np} < k_p$	$n_s < n_q$	$n_{np} < n_p$
1st Surge	19.82***	130.54***	7.90	52.18***
2nd Surge	67.70***	115.46***	52.17***	80.02***
3rd Surge	122.69***	142.99***	97.44***	60.09***
4th Surge	116.11***	141.15***	79.67***	20.69***

t-statistic is reported. ***p < 0.01, **0.01≤ p < 0.05, *0.05≤ p < 0.1. *s* denotes data during the given surge which consists of both the climbing and declining periods; *q* denotes data in the quiet period before the given surge; *p* denotes data on PGRs occurring before the local pandemic outbroke and then remaining in the network within all four surges; *np* denotes data on non-PGRs within the given surge.

Hypothesis 4

PGRs influenced the properties of the ITG network more than non-PGRs.

To test this hypothesis, we conduct an experiment to compare how removal of PGRs, who remain active in group travel amid the pandemic, and the same number of riders randomly drawn from the ITG dataset would affect the interconnectedness in the network. The interconnectedness can be measured by k and S. We validated that this is the case in Hong Kong. More details were presented in the Methods.

Figure 5a shows k becomes significantly lower under the “Failure 1” (removing PGRs from the ITG network), while the k’s for “Failure 2” (removing the same amount of randomly drawn riders) are almost the same as those without removing any riders from the ITG network in the same period. Similarly, if we assume it is highly likely for ITGs and related interactions to transmit viruses when one ITG rider has been infected, *S* would indicate the largest number of riders potentially affected by an infected rider in the ITG network. Figure 5b shows that *S* apparently remains more stable when randomly removing ITG riders (Subset 1), compared with removing PGRs (Subset 2–5). For the PGRs in different subsets, the more frequently they presented in the four surges, the more influential they were in *S.* Removing those PGRs who presented in all the four surges (Subset 5) resulted in the smallest *S* for failures. These indicate that the PGRs played a more influential role in interconnections of the network than wandering riders who entered or left during the given period.

Discussion

Cities serving as the magnet and mecca for billions of people is not by coincidence. They prompt more impactful social relationships among people, which are foundations for meaningful coordination,

collaborations, and innovations. Over time, cities as mankind's greatest "invention", have become increasingly richer, smarter, greener, healthier, and happier because of these collective activities[26]. Group travel constitutes one special form of these activities and serves as prerequisites for others. By exploiting a year's smartcard data in Hong Kong, we constructed an ITG network of Metro riders, which generated the following insights and findings.

First, it demonstrates that the longitudinal Metro smartcard data can help us identify an ITG network of Metro riders, which serves as a proxy for the general ITG network by the local residents. Second, we systematically examined and tracked the probable physical interactions and mobility patterns of the ITG Metro riders and PGRs based on network properties like the stable power-law-like degree distribution and the variations in interconnections and size. This points to some general laws of group interactive behaviors and probable virus transmission routes before and amid pandemics. Third, we found that the interconnections of the ITG networks are determined by a subset of riders, that is, PGRs in our study. Given that virus spread often relies on a social interactive network, this argues for investigating whether PGRs are more likely to spread or succumb to the virus. Identifying PGRs thus can be useful in more targeted preventive measures and in tracing important close contacts of the infected. Fourth, the changing and stable popular destinations among ITG riders and PGRs pre- and amid COVID-19 indicates that some stations and their surroundings can have better supply of facilities and opportunities than others to meet the demand of riders/people. Our work can help identify inequities in the distribution of essential locales, facilities and opportunities amid and post COVID-19[27].

Limitations of this research should also be mentioned. Although the ITGs identified from the Metro smartcard data could partially represent the local physical social interactions in transit-reliant cities like Hong Kong, they overlook interactions among people who use other modes of travel such as cycling, buses, private cars or walking. The smartcard data only allows us to identify probable ITGs, and we still know little about how many ITGs there were in total, who the ITG riders are, and how much they actually benefited from group travel (or lost because of forgoing ITG trips). In the longer term, we need to design surveys or interviews to better address these questions, which anonymous smartcard data alone cannot handle well.

Methods

Metro Smartcard Data. The metro riders' smartcard swipe records used in this article were obtained from the Hong Kong MTR Corporation Limited (MTR) via an MOU between MTR and one of the authors' employers. MTR operates 96 stations on its metro networks as of 2020 (Supplementary Fig. 1). The Hong Kong Metro network had a total length of 263 km in 2020 and carried 96 stations (3.15 million trips per day). The records include metro usage information of smartcard holders, that is, a metro trip between two stations with the swipe-in and swipe-out data in sequence. All records were anonymized to protect privacy. A dataset covering all weekends in each month of 2020, with a total number of 104 Saturdays/Sundays, is used to examine group travel behaviors in different periods of the pandemic. The data consist of the following variables: hashed smartcard ID, date of the day, entry and exit time, station

ID of each transaction, and card type. In this article, data of all types are used to identify intentional travel groups (ITGs). However, four card types are particularly used to analyze the potential relationship between two riders in an ITG, including adult, child (3–11 years old), student (full-time day course students of 12–25 years old) and senior (65 years old and above). The article excluded the trips those entry stations and exit stations are the same (where trips might not truly exist or were abandoned).

Identify intentional travel groups (ITGs). The premise of methods to identify the group travel in this article is that riders in ITGs have intentional and repeated social contacts (interactions) rather than those meeting coincidently. We filter ITG trips from encounters in the metro system based on their consistency in travel time, locations, and contact frequency. Firstly, an initial potential group travel dataset is constructed by setting time thresholds and fixing locations for the trips. Only when two riders, i and j, entering and existing the same stations within one minute are considered finishing a potential group trip (g_{ij}). Specifically, g_{ij} is required to satisfy the following conditions:

o_i = o_j

d_i = d_j

|ent_i - ent_j| ≤ 1 min

|ext_i - ext_j| ≤ 1 min

where, o_i and o_j are the origin station ID of the rider i and j, respectively. d_i and d_j are the destination station ID of the rider *i* and *j*, respectively. ent_i represents the time when rider i swiped in the origin station o_i and ext_j represents the time when rider *i* swiped out the destination station d_i. A similar explanation could be applied to ent_j and ext_j.

Then, since the above strategy might include a volume of riders traveling together by chance, further work needs to be done to identify ITGs, that is, two riders in a group indeed know each other and conduct group trips together. We reason that two riders are more likely to be “friends” if they have more frequent group trips observed in the system.

Assuming that the exit time difference of group trips between two riders in an intentional travel group would generally be smaller than that between those travelling together by coincidence[12], we did KS-tests of the distribution of the exit time difference between group trips adopted for the dataset. The total number of group trips of rider *i* and *j* in a month is defined as $\sum g_{ij}$. When $\sum g_{ij} \geq \beta$, where β is a given threshold of group trip times, we consider the group with rider *i* and *j* as an ITG and it appears at least β times in the potential group travel pool in the certain month. The KS-tests of the distribution of the exit time difference between group trips were adopted for various potential β. For an ideal β, intentional travel groups occurring at least β times and those following a higher threshold should follow the same distribution in exit time difference of trips. Accordingly, the hypothesis for the tests is:

H_0 : *exit time difference of group trips in potential travel groups with group times more than* β *and more than* β *+ 1 in a month follow the same distribution.*

H_1 : *exit time difference of group trips in potential travel groups with group times more than* β *and more than* β *+ 1 in a month follow the different distribution.*

Considering impacts of the pandemic, we did KS-tests for both weekend group trips in January, as a sample month before the local pandemic broke out, and February, as a sample month after. Supplementary Fig. 2 shows that for either January or February, the null hypothesis could be rejected before β = 4 with a *p-value* smaller than 0.05. Hence, the ITGs in this article would be defined as those with group trips at least 4 times within a month.

Non-commuting ITGs. Considering commuters who work on both weekdays and weekends might lead to coincident group trips, we tried identifying these groups of people and removed them from the ITG datasets when examining degree distribution. To be specific, we first select out regular daily commuters in Hong Kong. Commuters are defined as who: (1) are adult riders and have at least one consecutive round trip in a day, with at least six hours between trips of the round trip; (2) have trips like (1) between the same pair of stations at least eight times per month. Accordingly, the “home” stations here are those where commuters start and end their round trips of a day, and the “work” stations are those where commuters stay at least six hours between trips of their round trip of a day. Hence, a dataset for local regular commuters is acquired by examining metro smartcard data for the first eight months of the year 2020.

Then, we identify group commuting trips during the weekends as those conducted by two metro commuters with the same “home” and “work” stations and travel together between these two stations during the weekends. These potential group commuting trips are removed from the dataset of potential group travel during the weekends. In this way, the non-commuting ITGs in this article would be defined as those with non-commuting group trips at least 4 times within a month. Throughout the year 2020, 1,316,073 ITGs and 1,524,000 distinct non-commuting ITG riders were identified.

A Percolation Analysis. The spatial centrality of how ITG riders maintain their group travel is measured within the local metro system through a percolation analysis, which define a station’s central status based on its connectivity to other stations through the ITG riders’ group travel flows. This method hierarchically organizes the measured locations and avoid the bias by deciding an arbitrary threshold to define centers[16].

Firstly, we construct an O-D matrix by counting the incoming ITG trip flows of each station, that is, the number of ITG trips from other metro stations to the given station. Then, the unique trip flows, which represents the unique values of a list of ITG trips between any two stations, are extracted. On the basis, the unique flows will be arranged in descending order and each of them will be used as a threshold for analysis. Only when the number of group trips is equal to or larger than the given threshold that these trips will be permitted.

$$T_{p,(nm)} = \begin{cases} T_{nm}, if T_{nm} \geq u_p \\ 0, if T_{nm} < u_p \end{cases}$$

where $T_{nm}$is the number of incoming ITG trips from Station *n* to Station *m*. T_p represents the O-D matrix that consists of all permitted ITG trip flows $T_{p,(nm)}$ given at the unique trip flow u_p. Accordingly, the sum (T_{nm}), the total number of incoming ITG trips from other stations to the given Station *m*, is expected to decrease when the threshold value increase. Station *m* is considered disconnected with the system when sum (T_{nm}) becomes zero, which means all incoming group flows of the station is smaller than the given threshold. Hence, the stations those are connected longer or with higher sum (T_{nm}) at more thresholds could be considered as more central ones in the metro system.

An experiment for failures in the ITG network. As an experiment for failures in the ITG network, we compared how removal of PGRs, who remain active in group travel amid the pandemic, and the same number of riders randomly drawn from the ITG dataset would affect the interconnectedness in the network. The interconnectedness can be measured by k, the average degree, and S, the relative size of the largest cluster. As for the experimental failures for measurement on k, we called removing PGRs who remains in the ITG network up to a certain period “Failure 1” and removing the same amount of randomly drawn riders from the network “Failure 2”. *S* would indicate the largest number of riders potentially affected by an infected rider in the ITG network. Taking the “social network” of Metro riders in the whole year as the subject, we separately remove the same fraction (number) of different subsets of riders, 50 times in total and up to 19% of the network size, to investigate how that would influence S. The value of 19% is acquired since it is the proportion of the largest cluster in the whole-year network. Our subsets of ITG riders include those randomly selected throughout the year (Subset 1) and PGRs presented in at least one (Subset 2), two (Subset 3), three (Subset 4), and four (Subset 5) surges throughout the year. The subsets allow us to examine different failures of the network.

Declarations

Acknowledgments

The research underlying this paper was financially supported by the Platform Technology Funding, University of Hong Kong (URC012530226) and General Research Fund of Hong Kong (Grant 17603220).

Author contributions

J. Z. and M. Z. designed research; M. Z, J. Z., S. L., J. W., and H. M. performed research and analyzed the data and contributed new reagents and analytic tools; J. Z., M. Z. and D. L. wrote and reviewed the paper.

Data availability statement

The data that support the findings of this study are available from the Hong Kong MTR Corporation Limited (MTR) but not publicly available. Restrictions apply to the availability of these data, which were

used under license for the current study. Data are however available from corresponding author upon reasonable request and with permission of the Hong Kong MTR.

Competing interests

The authors declare no competing interests.

References

1. WHO. Coronavirus disease (COVID-19): How is it transmitted? Available at https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted.
2. Block, P. Hoffman, M. Raabe, I. J. et al. Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. Nat Hum Behav. 4, 588–596 (2020).
3. Yabe, T., Tsubouchi, K., Fujiwara, N et al. Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic. Sci Rep. 10, 18053 (2020).
4. Benzell, S. Collis, Nicolaides, A. C. Rationing Social Contact during the COVID-19 Pandemic: Transmission Risk and Social Benefits of US Locations. *Proc. Natl. Acad. Sci. U.S.A.* 117(26), 14642–14644 (2020).
5. McQuire, S. Geomedia: Networked Cities and the Future of Urban Space. Chapter 1 (Cambridge, UK; Malden, MA: Polity Press, 2016).
6. Carlsen, H. B. Toubol, J. Brincker, B. On Solidarity and Volunteering during the COVID-19 Crisis in Denmark: The Impact of Social Networks and Social Media Groups on the Distribution of Support. European Societies. 23(1), 122–140 (2021).
7. Chang, S. Pierson, E. Koh, P. W. et al. Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening. Nature. 589: 82–87 (2020).
8. Feng, S., Kirkley, A. Integrating online and offline data for crisis management: Online geolocalized emotion, policy response, and local mobility during the COVID crisis. Sci Rep. 11, 8514 (2021).
9. Sullivan, O. Gershuny, J. Sevilla, A et al. Using Time-use Diaries to Track Changing Behavior across Successive Stages of COVID-19 Social Restrictions. *Proc. Natl. Acad. Sci. U.S.A.* 118(35), e2101724118 (2021).
10. Sun, L. J. Axhausen, K. W. Lee, D. Huang, X. F. Understanding Metropolitan patterns of daily encounters. *Proc. Natl. Acad. Sci. U.S.A.* 110(34), 13774–13779 (2013).
11. Zhou, J. Yang, Y. Ma, H. Li, Y. “Familiar Strangers” in the Big Data Era: An Exploratory Study of Beijing Metro Encounters. Cities. 97, 102495 (2020).
12. Zhu, K. Yin, H. D. Qu, Y. C. Wu, J. J. Group Travel Behaviour in Metro System and its Relationship with House Price. Physica A. 573, 125957 (2021).
13. Transport Department, The Government of the Hong Kong Special Administrative Region. Available at: https://www.td.gov.hk/en/transport_in_hong_kong/public_transport/railways/index.html.

14. Barabasi, A. L. Jeong, H. Neda, Z. et al. Vicsek, Evolution of the Social Network of Scientific Collaborations. Physica A, 311, 590–614 (2002).
15. HK Gov. Coronavirus Disease (COVID-19) in HK (Geodatabase). Available at: https://chp-dashboard.geodata.gov.hk/covid-19/zh.html.
16. Sarkar, S. Wu, H. Levinson, D. M. Measuring Polycentricity via Network Flows, Spatial Interaction and Percolation. Urban Studies, 57(12), 2402–2422 (2020)
17. Cervero, R. Murakami, J. Rail and Property Development in Hong Kong: Experiences and Extensions. Urban Studies. 46(10), 2019–2043 (2003).
18. Silver, C. Neighborhood Planning in Historical Perspective. Journal of the American Planning Association. 51(2), 161–174 (1985)
19. Kissfazekas, K. Circle of paradigms? Or ’15-minute’ neighborhoods from the 1950s. Cities. 123, 103587 (2022).
20. Leung, P. How '15-minute cities' will change the way we socialize. Available at: https://www.bbc.com/worklife/article/20201214-how-15-minute-cities-will-change-the-way-we-socialise (September 15, 2021).
21. Petherick, A. Goldszmidt, R. Andrade, E. B. et al., A worldwide assessment of changes in adherence to COVID-19 protective behaviours and hypothesized pandemic fatigue. Nat Hum Behav, 5, 1145–1160 (2021).
22. Zhang, N. Jia, W. Wang, P. et al., Changes in local travel behavior before and during the COVID-19 pandemic in Hong Kong. Cities. 112, 103139 (2021).
23. Albert, R. Jeong, H. Barabasi, A. L. Error and Attack Tolerance of Complex Networks. Nature. 406, 378–381 (2000).
24. Dou, B. L. Wang, X. G. Zhang, S. Y. Robustness of Networks Against Cascading Failures. Physica A. 389, 2310–2317 (2010).
25. Dunbar, R. I. M. Coevolution of neocortical size, group size and language in humans. Behavioral and Brain Sciences. 16(4), 681–735 (1993).
26. Glaeser, E. Triumph of the City: How Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier. (New York: Penguin Books, 2012).
27. Jay, J. Bor, J. Nsoesie, E. O. et al., Neighbourhood Income and Physical Distancing during the COVID-19 Pandemic in the United States. Nat Hum Behav. 4,1294–1302 (2020).

Figures

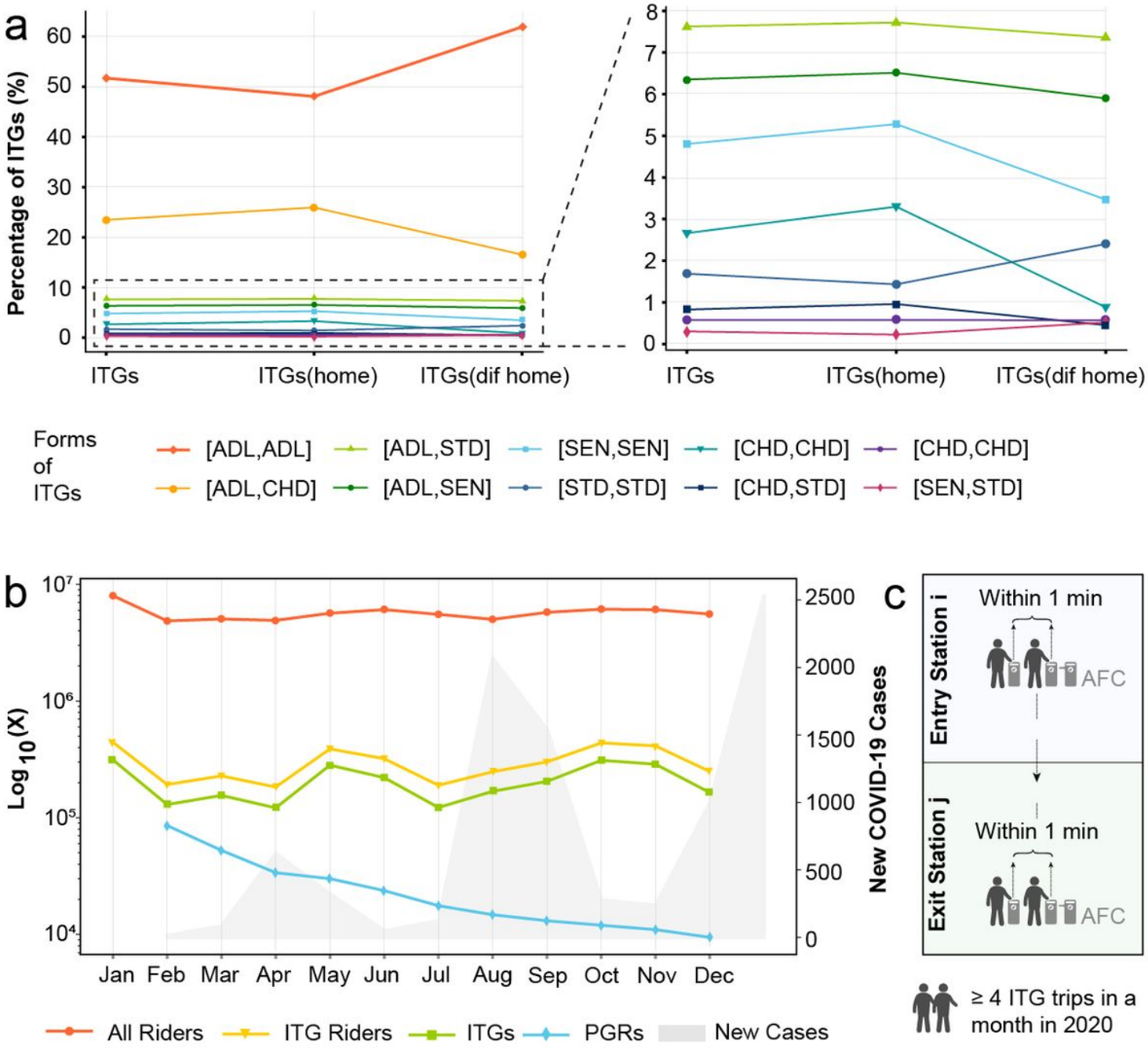

Figure 1

Intentional Travel Groups (ITGs) during the COVID-19 pandemic. a. Percentage of ITGs consisting of different card types, including the adult (ADL), senior (SEN), student (STD), and child (CHD). "ITGs" represent identified ITG with these four card types. Among them, "ITGs (home)" include two riders with the same home stations, while "ITGs (dif home)" include two riders with different home stations. **b.** Quantity of all riders, ITG riders, PGRs, ITGS, and new COVID-19 confirmed cases by month in 2020. Here, PGRs represent those who occurred as ITG riders in the certain month and all previous months. X could

be all riders, ITG riders, PGRs, and ITGS. **c.** Illustration of how an ITG is identified. An ITG is counted when its two riders enter the same station i and exit the same station j within 1 min, and have such pattern at least 4 times in a month of 2020.

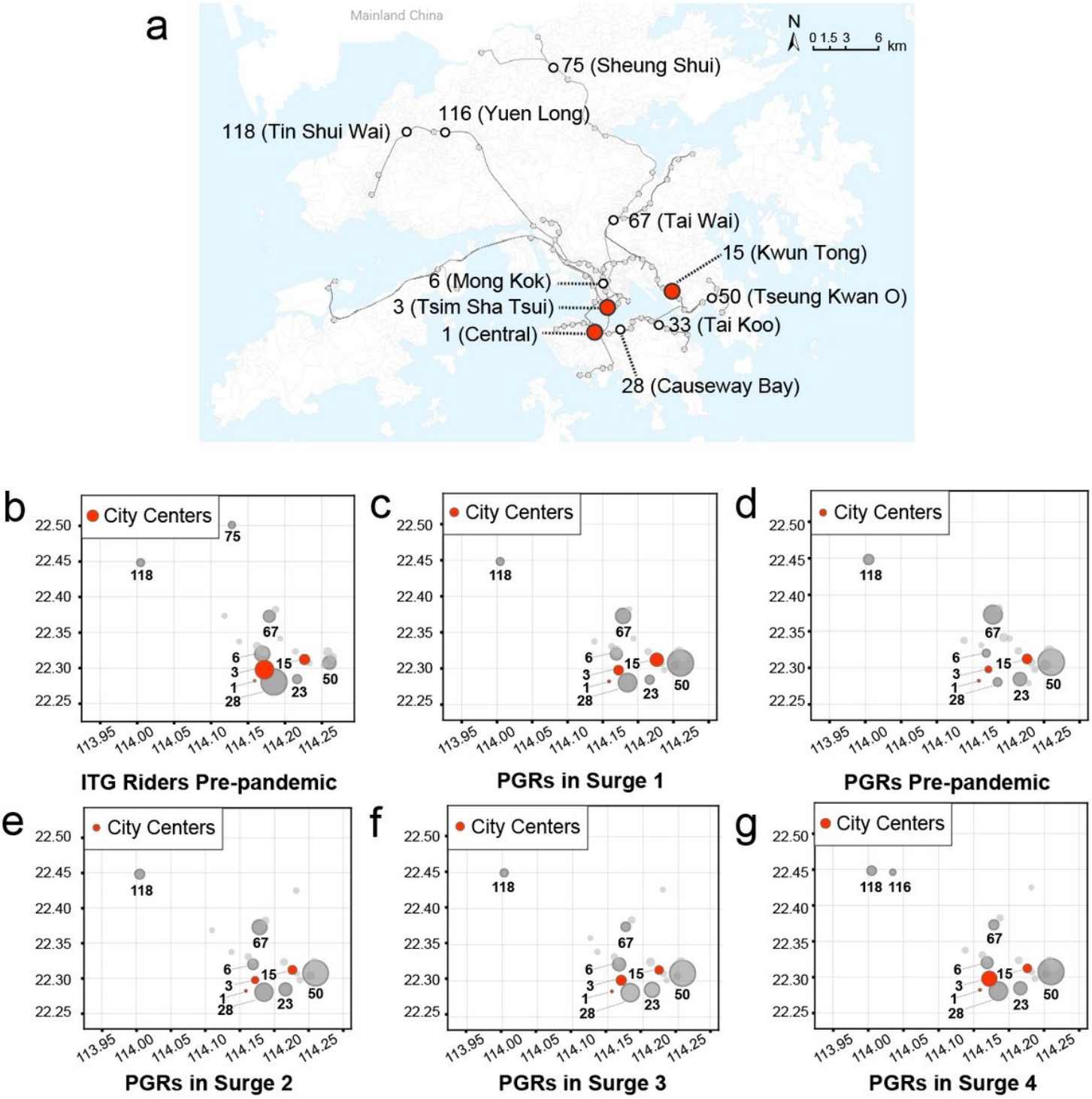

Figure 2

Spatial centrality of Intentional Travel Groups (ITGs) and persistent group riders (PGRs). a. City centers and selected Metro stations with the most incoming ITG trips. **b-g.** Spatial pattern of incoming ITG trips

based on the percolation analysis. Top 20 stations of (b) all ITGs riders before the pandemic, and PGRs occurring before the pandemic and in all four surges. ITGs trips of these PGRs during (c) the pre-pandemic period, (d) the first surge, (e) the second surge, (f) the third surge and (g) the fourth surge. The bubble size reflects the ranking of stations. The red point represents the three city centers. Central Station is not within the Top 20 stations.

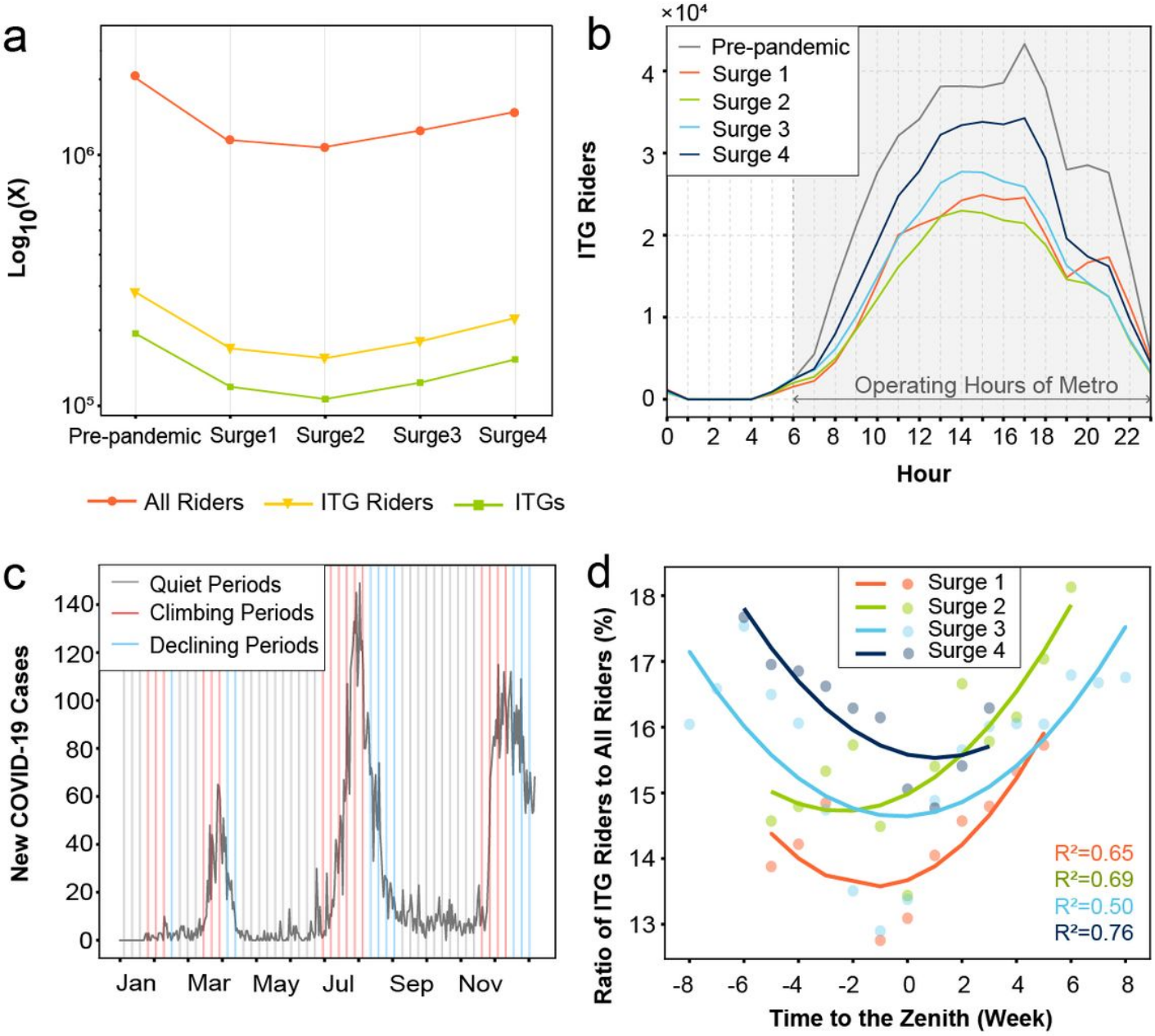

Figure 3

Variations in size of ITG network during the pandemic. a. Daily quantity pre-pandemic and in each surge. X could be all riders, Intentional Travel Group (ITG) riders, and ITGs. **b.** The average quantity of ITG riders by hour of a weekend day pre-pandemic and in each surge. **c.** Periods of pandemic surges. Each vertical line represents a weekend and its color differentiates different surges of the pandemic. **d.** the ratio of ITG riders to All riders before and after a surge's zenith. (Data are aggregated at daily level by calculating the average value of Saturday and Sunday in each week. Here, each surge's ITG riders or All riders include those in climbing and declining periods as well as three weekends before and after. The negative value on the X-axis denotes the time before the zenith and the positive is that after.

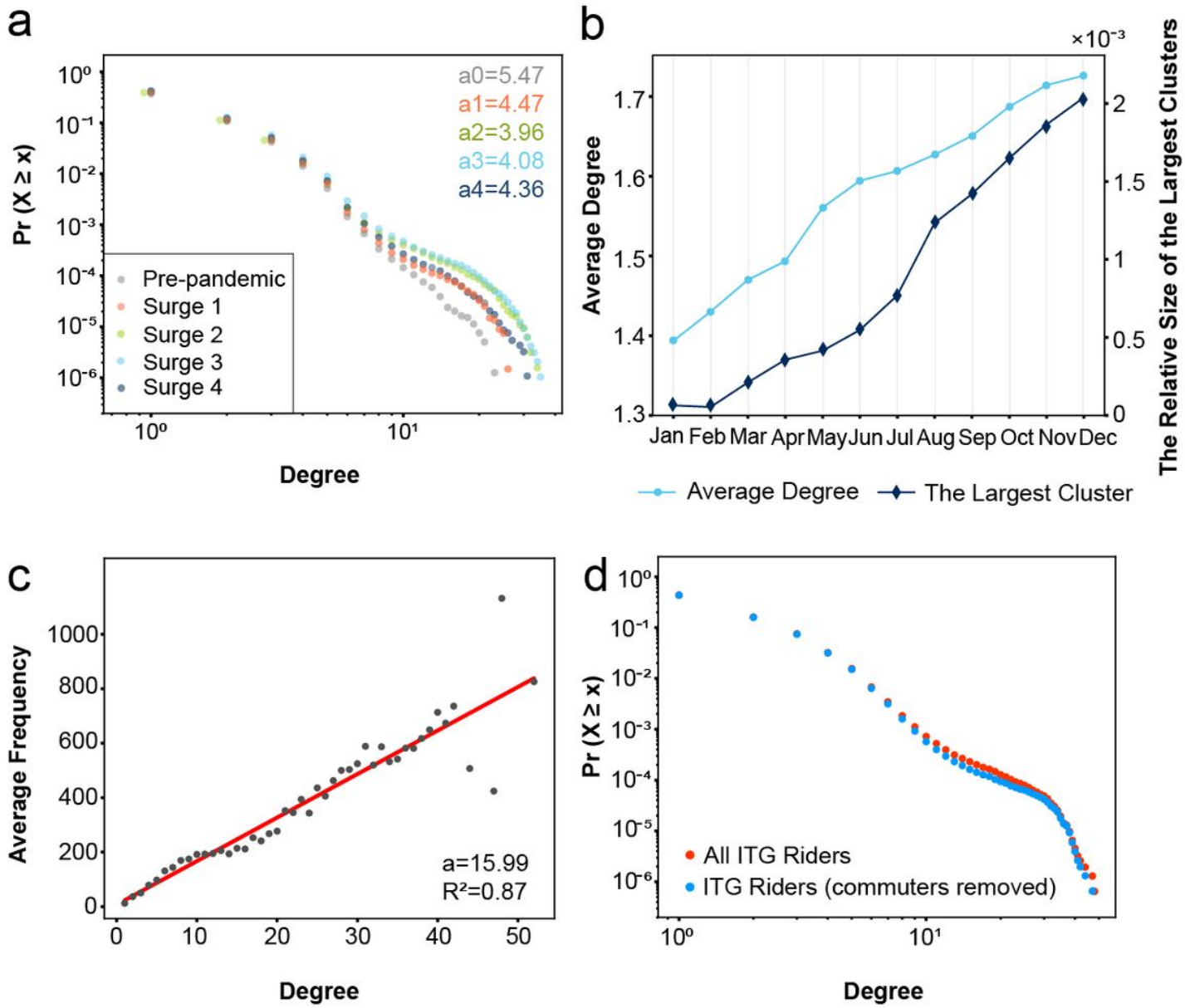

Figure 4

Typological properties of the ITG network. a. Comparison of the complementary cumulative distribution functions (CCDF) of group riders' degree during the pre-pandemic time and each pandemic surge. **b.** The average degree *k* and relative size of the largest cluster *S* in the network. Results are computed on the cumulative Intentional Travel Groups (ITGs) and their members from Jan up to the given month; **c.** The average frequency per rider by degree for the year 2020 (the average frequency represents the average number of ITG trips of all ITG riders with a given degree); **d.** The CCDF power-law scaling of ITG riders' degree in the ITG network for the year 2020. The red points denote all ITG riders identified (N=1,337,774; $\alpha = 5.67$; $p < 0.01$), and the blue points denote non-commuting ITG riders.

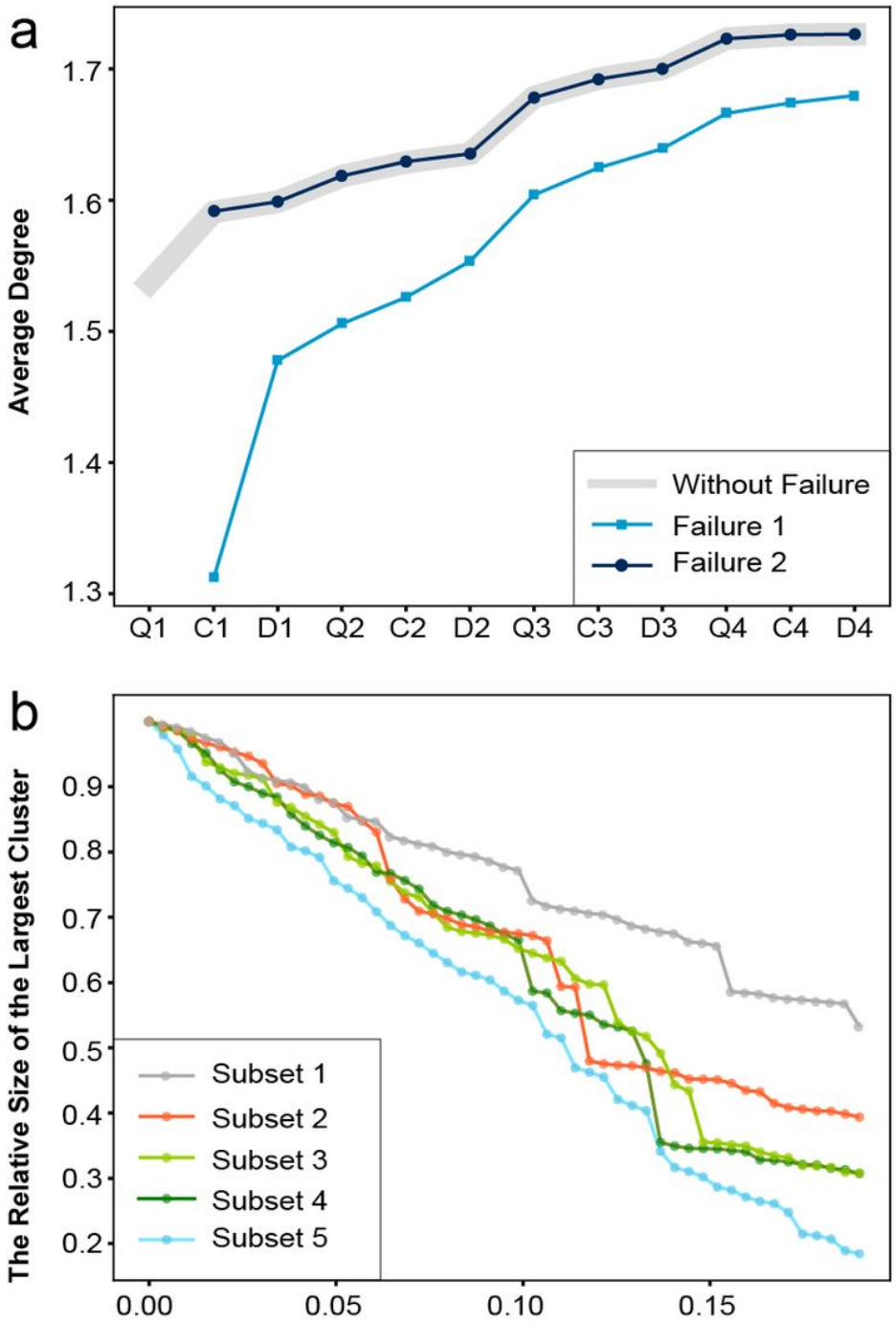

Figure 5

Experimental failures in the ITG network. a. The average degree of the cumulative network in each pandemic period when without failures, under Failure 1 (removing Persistent Group Riders (PGRs) who occurred in a certain period and all previous periods), and under Failure 2 (randomly removing Intentional Travel Group (ITG) riders as the PGRs removed in the same period). Q denotes the quiet periods; C denotes the climbing periods and D denotes the declining periods. **b.** The relative size of the largest cluster in the network by randomly removing ITG riders or removing PGRs. The X-axis represents the

fraction of removed ITG riders for the whole system, and the Y-axis represents the fraction of riders contained in the largest cluster. The subsets of ITG riders include those randomly selected throughout the year (Subset 1) and PGRs presented in at least one (Subset 2), two (Subset 3), three (Subset 4), and four (Subset 5) surges throughout the year.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ScientificReportsSupplementaryInformation.docx