

Research



Cite this article: He H, Deng H, Wang Q, Gao

J. 2021 Percolation of temporal hierarchical mobility networks during COVID-19. *Phil. Trans. R. Soc. A* **380**: 20210116.

Trans. R. Soc. A **380**: 20210116.

<https://doi.org/10.1098/rsta.2021.0116>

Received: 15 May 2021

Accepted: 8 July 2021

One contribution of 14 to a theme issue 'Data science approaches to infectious disease surveillance'.

Subject Areas:

statistical physics, mathematical physics, computer modelling and simulation, computational physics

Keywords:

percolation theory, COVID-19, human mobility, network science

Author for correspondence:

Jianxi Gao

e-mail: gaoj8@rpi.edu

Percolation of temporal hierarchical mobility networks during COVID-19

Haoyu He¹, Hengfang Deng², Qi Wang² and

Jianxi Gao¹

¹Department of Computer Science and Center for Network Science and Technology, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

²Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115, USA

HD, 0000-0001-8083-8441; JG, 0000-0002-3952-208X

Percolation theory is essential for understanding disease transmission patterns on the temporal mobility networks. However, the traditional approach of the percolation process can be inefficient when analysing a large-scale, dynamic network for an extended period. Not only is it time-consuming but it is also hard to identify the connected components. Recent studies demonstrate that spatial containers restrict mobility behaviour, described by a hierarchical topology of mobility networks. Here, we leverage crowd-sourced, large-scale human mobility data to construct temporal hierarchical networks composed of over 175 000 block groups in the USA. Each daily network contains mobility between block groups within a Metropolitan Statistical Area (MSA), and long-distance travels across the MSAs. We examine percolation on both levels and demonstrate the changes of network metrics and the connected components under the influence of COVID-19. The research reveals the presence of functional subunits even with high thresholds of mobility. Finally, we locate a set of recurrent critical links that divide components resulting in the separation of core MSAs. Our findings provide novel insights into understanding the dynamical community structure of mobility networks during disruptions and could contribute to more effective infectious disease control at multiple scales.

© 2021 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

1. Introduction

The unprecedented pandemic of the coronavirus disease 2019 (COVID-19) is affecting more than 200 countries and infected more than 32 million causing 578 530 deaths in the USA as of 9 May 2021 [1]. In a short one-year period, the USA has experienced two waves of transmissions [2]. In response to the fast and dynamic transmission of SARS-CoV-2, different levels of government have employed restricting population movement as a key non-pharmaceutical intervention (NPI) in limiting contacts and increasing social distancing [3]. These measures have caused significant changes in people's mobility patterns [4]. Also, the employment of stay-at-home orders has substantially reduced long-distance travel as well as local commuting [5]. Indeed, human mobility is both a key driver of and NPI to control COVID-19 [6–8].

Numerous studies have investigated the correspondence between mobility patterns and the spread of infectious disease such as SARS, influenza and malaria [9–13] as well as COVID-19 in different countries [14–20]. Studies have shown that the early spatial patterns of COVID-19 infection in China correspond to the human mobility fluxes and such correlation drops as some local control policies were executed [21,22]. Other studies support the finding that regions and places with high volumes of mobility are contributing to the spread of the deadly virus [15,23–25]. On the flip side, social distancing and mobility reduction are associated with a decreased level of transmission of COVID-19 [21,26,27].

Mobility-based networks are constructed to accurately model the transmission of COVID-19 in social networks, but they suffer from a few limitations. First, when building the networks, compromise on resolution or scale is often necessary to account for both the spatial and temporal dynamics of mobility networks. Studies have aggregated human mobility on the census block group or county levels [28–32]. The aggregation removes details and information on other levels and leads to inaccurate prediction of transmission patterns. Second, previous studies validate the effectiveness of inter-city travel restrictions in reducing the imported COVID-19 incidence rate [23,33]. Yet, there is still limited understanding of how local measures and responses could change the mobility network structure. Specifically, communities may still face challenges from intra-city mobility, which must be considered in network models. Third, a recent study has discovered the percolation process and phase transition in human mobility networks on the county level, which indicates the possibility of devising effective strategies to control mobility flows at critical bridges and contain the transmission of COVID-19 [31]. However, it is unknown if percolation processes govern the structural changes in mobility networks with a higher geographical granularity or multi-level mobility networks.

To address these limitations, here we propose a multilevel approach to capture the hierarchical and the dynamical property of mobility networks. We construct two-layer hierarchical networks of 378 Metropolitan Statistical Areas (MSAs) in the USA. Such dynamic multiplex networks consist of a macro-level and meso-level layer for daily mobility. On the macro level, the MSAs are nodes, and the travels among the metropolises are links. On the meso level, we treat census block groups as nodes and the travels between as links. This hierarchical construction allows us to analyse local and long-range trips simultaneously and examine their interactions. Thus, the multiplex network supports our effort to uncover the associations between different layers and node attributes such as COVID-19 incidence rate and collective response metrics.

To understand the structural changes of multiplex networks, we adopt an analytical approach that originated from percolation theory. As a fundamental concept in network science, percolation theories are studied extensively, contributing to understanding and conducting materials and network-based applications [34–36]. Specifically, percolation theory is an essential step for coping with complex models and dynamical processes occurring on the networks [37–44]. It is a

crucial approach to identify the hierarchical structure and determine any discontinuous phase transition within the complex systems. Previous applications in network science include critical phenomena in urban traffic planning [45–47], epidemic modelling [48,49] and cascading processes on networks [50,51]. These strengths make the approach effective in better understanding the hierarchy of dynamic mobility networks and thus unearth variations in mobility stabilization and emergent structure patterns.

2. Mobility data

We use the Social Distancing Metrics dataset from SafeGraph to quantify the daily travel changes (see Data availability). The data track the GPS pings of more than 20 million anonymous mobile devices and records the related information of their daily movements, including travelling trajectory, home-dwelling time, distance travelled from home, etc. The smallest geological unit used is the census block group; therefore, a device's location is spatially joined to the census block group containing it. Here the home block group of a device is defined as its common nighttime (18:00–7:00) census block group for 6 consecutive weeks. The mobility network is built based on the movements between different block groups. The dataset has a field *destination_cbgs* which is a dictionary that reports the destination census block groups and the number of devices that visited them from each home census block group. The dataset ranges from 1 January to 31 December 2020, allowing us to study human mobility patterns before and after the national emergency declared on 13 March 2020. The data track each device's daily trajectory among census block groups in the USA during daytime (7:00–18:00). A human mobility network is formed based on it, in which a directed link from block group i to block group j is created if a device is detected living in block group i and travelling to block group j within the same day. The weight is the number of devices detected travelling from i to j .

3. The hierarchical networks structure of the USA

We build a mobility-based, hierarchical network of 378 Metropolitan Statistical Areas (MSAs) in the USA (figure 1). Our approach aggregates and normalizes the regional travel flux based on the number of devices observed in both origin and destination regions on two levels: inter-MSA (figure 1*a,b*) and intra-MSA (figure 1*c–f*). For the intra-MSA networks, denoted as α -networks, we take each census block group as a node. A directed link from block group i to block group j is created if a device is detected living in block group i and travels to block group j within the same day. The weight is the number of travellers. One issue that needs to be addressed is that human mobility is significantly different between weekdays and weekends, and yet patterns from both periods are important and should be considered in this research. Because of this, we average the data from every 7 consecutive days (3 days before to 3 days after) to construct the networks. This approach ensures every network includes a set of weekdays and weekends. Thus, it captures a weekly periodicity of routine mobility and reduces the effects of outliers. The edge weight is the mean of the weights on the same edge for these 7 days.

For the inter-MSA networks, denoted as ω -networks, we aggregate all block groups in an MSA and treat it as a node. Therefore, there are 378 nodes in the inter-MSA networks. The weight of a link is the mean of the normalized number of total travellers between two MSAs for 7 consecutive days.

To simulate the mobility flow from node i to node j to calculate the probability of a infected person travelling from i to j , we let population of node i be m , infected population of i be n , and k people travel from i to j . Thus, we can roughly present the probability of infection for each person in i be n/m and we can calculate the probability of not spreading COVID-19 to j as a function of n/m , $f(n/m) = 1 - (1 - (n/m))^k$. Taking its binomial expansion, we have

$$f\left(\frac{n}{m}\right) = 1 - \left(1 - \frac{n}{m}\right)^k = 1 - \sum_{s=0}^k \binom{k}{s} 1^{k-s} \left(-\frac{n}{m}\right)^s = \frac{nk}{m} - \frac{k(k-1)}{2} \left(\frac{n}{m}\right)^2 + \dots$$

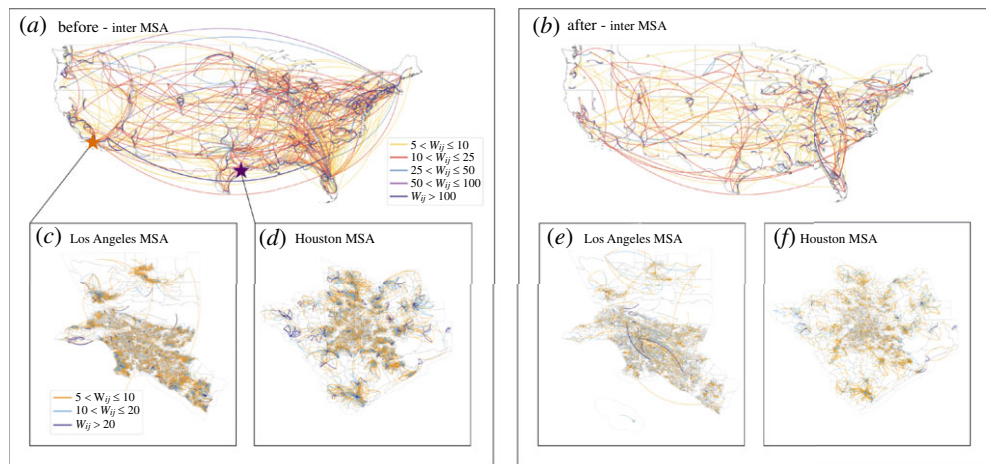


Figure 1. Inter- and intra-MSA mobility networks before and after the national emergency declaration. In (a,b), the nodes represent MSAs before and after the national emergency declaration. The colours of the links demonstrate the number of 7-day average flux normalized by the device count between MSAs. Panel (c,d) demonstrates the intra-MSA networks of Los Angeles MSA and Houston MSA before the national emergency declaration. The nodes are census block groups and the colours are based on the numbers of 7-day average normalized influx between pairs of nodes. (Online version in colour.)

Since the infected population at any particular time point is relatively small compared with total population of that area, we can consider $n/m \ll 1$ and terms in this binomial series are converging to 0 rapidly. Thus, the value from this function is largely determined by the first term nk/m and the rest of the terms can be ignored. Since the infected population, n , is an inaccurate constant for various reasons (i.e. false-negative rate, asymptomatic individuals), we use k/m to represent the weight of links. Therefore, we develop a normalized weighting approach using device count to mitigate the biases of mobility flow:

$$W_{ij} = \frac{R_{ij}}{D_i} + \frac{R_{ji}}{D_j},$$

where R_{ij} is denoted as the number of people travelling from node i to j , and D_i is the device count of node i on the same day. The constructed hierarchical networks simultaneously examine local mobility flows within MSAs and long-distance travel between MSAs. The construction also enables systematic exploration of the interactions between the two levels.

To show the computational advantage of the hierarchical network model compared with traditional ones, we run the percolation process on both the hierarchical model and the traditional model with the same day of data from our data set. The average time of analysing the percolation process of one-day data is about 30 min for the hierarchical model and about 10 h for the traditional model using the same computer. The processing time of the traditional model cannot support real-time mobility analysis, and thus the hierarchical model is necessary.

Figure 1 shows significant changes in mobility flows before and during COVID-19. Figure 1a shows the ω -network structure in the week of 1 February, and figure 1b in the week of 1 June. As the weight of the edge gets larger, the colours of the links in graphs get darker. Comparing figure 1a,b, we observe that edge weights of most links decrease substantially. Therefore, fewer people travel between MSAs and the entire network becomes sparser. Figure 1c,d shows the intra-MSA network (i.e. α -network) structure of Los Angeles and Houston in the week of 1 February and figure 1e,f the one in the week of 1 June. The ω - and α -networks display similar geographical distributions even though the networks become sparser due to a decline in mobility.

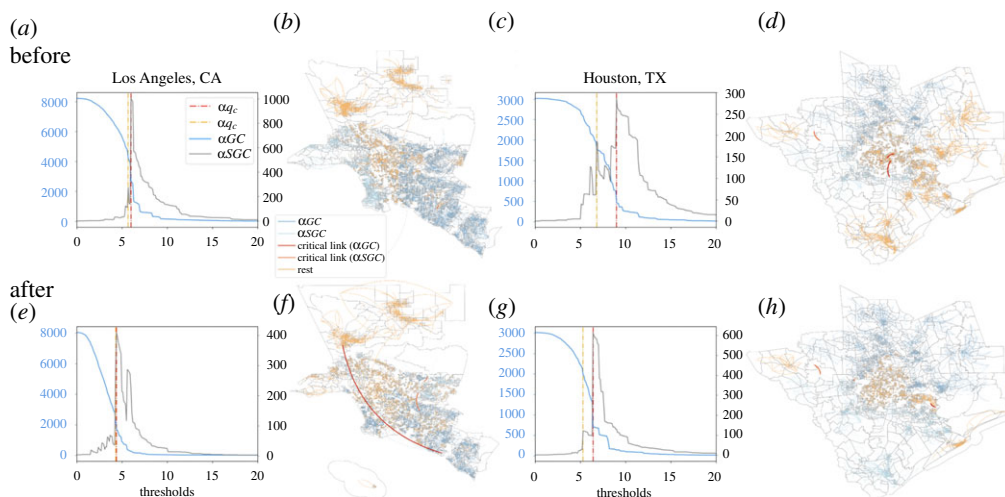


Figure 2. Percolation process of mobility networks in Los Angeles and Houston MSAs. Panel (a) shows the two largest connected components (αGC and αSGC) with the change of αq_c in the week of 29 January to 4 February 2020 in Los Angeles MSA. Panel (b) shows its network when the weight of edges equals αq_c in the week of 29 January to 4 February 2020. Panel (c) shows αGC with the change of αq_c in the week of 29 January 29 to 4 February 2020 in Houston MSA. Panel (d) shows the network of Houston MSA when the weight of edges equals αq_c in the week of 29 January to 4 February 2020. (e) shows the two largest connected components with the change of αq_c in Los Angeles MSA size in the week of 29 May to 4 June 2020. Panel (f) shows the network with the change of αq_c in Los Angeles MSA in the week of 29 May to 4 June 2020. Panel (g) shows the network with the change of αq_c in Houston MSA in the week of 29 May to 4 June 2020. Panel (h) shows the network map of Houston when the threshold of the weight of edges equals αq_c in the week of 29 May to 4 June 2020. (Online version in colour.)

4. Intra-MSA percolation

We apply a percolation-based approach to analyse the connected components of both the α - and ω -networks and study their phase transitions. We first set the threshold of the edge weight 0 and increase it by a stepsize after each iteration. With each increase, the edges with weights lower than the threshold are removed from the graph and a new graph is generated with the remaining edges and nodes. The method is adapted from other research such as a traffic study in Beijing which removes less crowded to more crowded routes [45] and a study on the regional structure of Britain by adding edges to cities sorted by distancing between them from nearest to furthest [52]. After each update, we calculate the connected components, especially the largest, i.e. the giant, component (GC), and the second-largest, i.e. the second giant, component (SGC). Figure 2a,c shows that the size of αGC of an intra-MSA network declines and αSGC increases. At the critical point of αq_c , the largest component experiences a sudden drop and αSGC simultaneously reaches its maximum size.

The observed change aligns with results reported in [31], showing that mobility networks on different scales, i.e. county and block groups, experience similar percolation processes. It also indicates that the intra-MSA α -network experiences a phase transition at αq_c . After the transition, the network becomes sparser, and components are more likely to disconnect from each other.

Besides αq_c , we also examine a second critical point αq_{c2} , at which αSGC is the largest before αq_c . Thus, αq_c and αq_{c2} indicate that the network experiences hierarchical phase transitions [31,45].

The percolation approach allows us to examine networks in different MSAs. We analyse the mobility data from 29 January to 4 February 2020 in Los Angeles MSA and find the critical point at $\alpha q_c = 6$ (red line in figure 2a). At this point, the second-largest component experiences a significant increase, and its size reaches 1001 with a total of 8211 census block groups which is at

its maximum. Simultaneously, about the same amount of nodes disconnects from αGC , indicating that αSGC separates from the largest one.

It is worth noting that the network becomes sparse at αq_c (red line in figure 2a) where the largest possible component other than αGC is disconnected from αGC . At this point, αGC experienced the most significant decrease throughout the whole process and constitutes only a small portion, 33.13% (2720 out of 8211 block groups), of the original network. Thus, the network lost most of its connectivity and cannot effectively spread COVID-19 inside the MSA. In other words, COVID-19 is more likely to spread within each subcomponent without transmitting to other ones. Another finding is that prior to αq_c , we discover a less significant phase transition at αq_{c2} (yellow line in figure 2a) where the MSA can also be partitioned into subcomponents. The same patterns are observed in Houston MSA (figure 2c).

After identifying the critical points, i.e. αq_c and αq_{c2} , we locate the critical links whose elimination causes αSGC to disconnect from the largest one. Figure 2b shows the network of Los Angeles MSA in the week of 29 January when the threshold is αq_c . The dark blue component is αGC , the light blue one is αSGC , and the orange the rest. The critical link is marked in red (figure 2b) with the weight of αq_c . The link connects αGC in the east MSA Los Angeles and αSGC in the west. Similarly, in Houston MSA (figure 2d), the critical link separates αGC in the North and αSGC in the South. The findings align with a previous study on the county level [31] and suggest that local mobility networks have bottlenecks when going through the percolation process.

Figure 2e–h demonstrates the same percolation process after the state emergency declaration (29 May 29 to 4 June 2020). Comparing figure 2a,e, we find that the sizes of both αGC and αSGC are similar. However, αq_c and αq_{c2} are smaller than those before the national emergency declaration. A similar pattern is observed in Houston MSA (see figure 2c,g). Thus MSAs are less robust after the national emergency declaration and they could be broken into isolated clusters with smaller sacrifices to effectively contain COVID-19 within a small area. At the q_{cs} , the locations of αGC and αSGC in Los Angeles MSA almost remain the same as before (figure 2b) and after (figure 2f). The similarity is also observed in Houston MSA (figure 2d,h). However, local mobility decreases substantially after the state emergency declaration, and thus the components become more sparse. A sparse network means that mobility networks could be disconnected and controlled more easily. Also, the weights of edges decline in a similar proportion and the patterns of major clusters at q_c are almost the same before and after the national emergency declaration. Thus, it is feasible to devise effective measures in controlling the spread of COVID-19 locally.

5. Intra-MSA correlation of αq_c with other attributes

We compute daily αq_c s of all MSAs from 1 January to 31 December 2020 and examine these values with MSA attributes. Figure 3a shows αq_c s in different MSAs before and after the national emergency declaration. The values are positively related. Then we explore a few attributes of the MSAs and their correlations with αq_c to better understand network structures. Three attributes are tested, and their regression trendlines are shown in figure 3b–d. Figure 3b shows the correlation between αq_c and the sizes of the largest components in the MSAs. As αq_c increases, the sizes of the largest components at critical points quickly decrease. Comparing the critical points before and after the national emergency declaration, we find that the sizes of the largest components at critical points remain consistent. Yet, αq_c s during the pandemic are larger than the ones before. Figure 3c shows the correlation between αq_c and total flux. It is found that the total flux of an MSA increases as αq_c increases. Also, the total flux after the national emergency declaration is smaller by comparing points before and after. The change is caused by the decline of local mobility due to the stay-at-home order. Also, median edge weight and αq_c are positively correlated (figure 3d). There has been no substantial change in the linear correlation magnitude between median edge weight and αq_c before and after the national emergency declaration. However, the scatter plot in figure 3d suggests a reduction of the median edge weights for αGC s after the declaration with about 50% decrease across most MSAs.

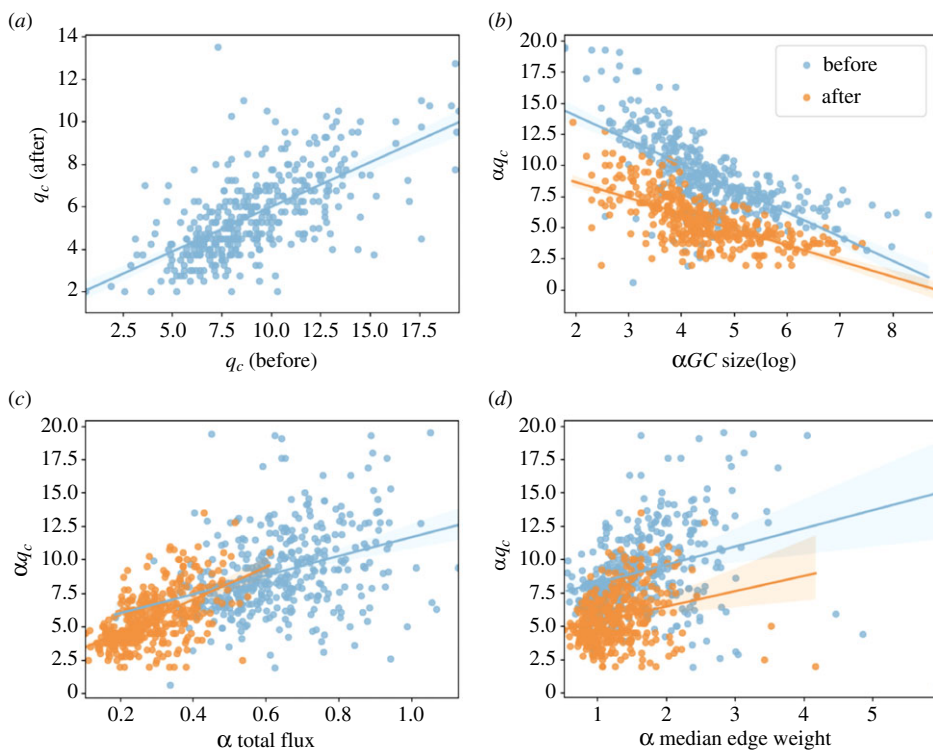


Figure 3. The relationship between αq_c and MSA attributes before (29 January to 4 February) and after (29 May to 4 June) national emergency declaration. (a) αq_c before the national emergency declaration versus αq_c after. Panel (b–d) shows αq_c versus the size of αGC at αq_c , total flux, and median edge weight of each MSA before and after national emergency declaration, respectively. (Online version in colour.)

6. Inter-MSA percolation

After analysing the intra-MSA network, we examine the percolation process of the inter-MSA network (i.e. ω -network) in the USA. In figure 4, we control two parameters: αq_c for each intra-MSA network and the edge weights between MSAs. As the threshold for weight increases (rows in figure 4), the number of edges decreased, especially for the long-distance links. As the threshold for αq_c increases (columns in figure 4), fewer nodes remain and the network becomes sparser. By comparing the corresponding graphs at the same thresholds for edge weight and αq_c before and after the national emergency declaration, i.e. figure 4a,g, we found that αq_c for each MSA is smaller in most cases. Also, most long-distance edges in the inter-MSA network before the national emergency declaration disappear in the one after. The changes indicate that while human mobility substantially decreased under the influence of COVID-19, long-distance travels are particularly excluded from people's travel behaviours. The behaviour changes divide the ω -network into isolated clusters. Also, people are more likely to travel locally under the influence of the pandemic since more edges appear in the local networks, especially in the southeastern region, after the national emergency declaration. While this could indicate strengthened social bonding during the difficult time, the increase could also be a factor contributing to the fast growth of infection cases in these regions.

By applying our percolation approach used in the intra-MSA networks, we obtain a similar phase transition pattern for inter-MSA networks. In figure 5a, we found that when thresholds reach 4, the size of ωSGC increases abruptly to 80 out of the 378 MSAs. This means nearly $\frac{1}{5}$ of MSAs disconnected from ωGC when the edge weight is 4. At the critical point, the size of

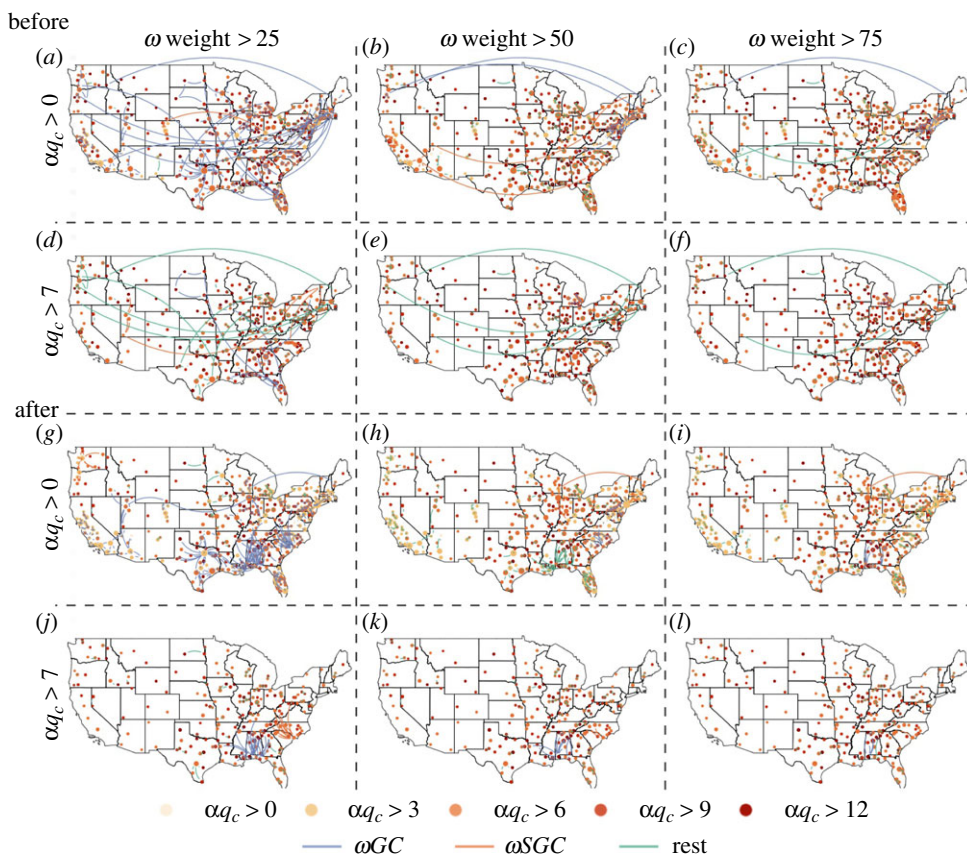


Figure 4. Inter-MSA networks before (i.e. 29 January to 4 February, (a–f)) and after (i.e. 29 May to 4 June, (g–l)) the national emergency declaration. The networks are controlled by two parameters: αq_c for each MSA, and the edge weights between MSAs. (Online version in colour.)

ωSGC decreases to less than 100 nodes (figure 5a,b), and the inter-MSA network loses its strong connectivity to spread COVID-19 effectively.

We observe a similar phase transition during the pandemic (in figure 5d). There is an abrupt decrease in the inter-MSA ωGC and an increase in the intra-MSA ωSGC . However, both ωq_c and ωq_{c2} after the national emergency declaration are smaller than the ones before. The decreases demonstrate people are less likely to travel long-distance under the influence of pandemic and travel restriction.

Figure 5b shows the inter-MSA network map at its critical point before the national emergency declaration and figure 5e shows one after. By comparing the inter-MSA maps at their critical points, we find several subcomponents in ωGC or ωSGC , including: west region, midwestern region, northeast region, southeast region and south region. The comparison shows that the critical link usually connects these regions. While internal connections within each region remain stable even during the pandemic, cutting off the critical links can effectively disconnect the regions.

The CCDF plot in figure 5c shows the difference of distribution edge weight and in figure 5f shows the difference of distribution edge distance between two nodes in these two inter-MSA networks. The total numbers of edges of the two networks are the same since there are travellers in any pair of MSAs every day. However, the distribution becomes more heavy-tailed after the national emergency declaration, suggesting more centralized low weight edges and short-distance edges in the ω -networks during the pandemic.

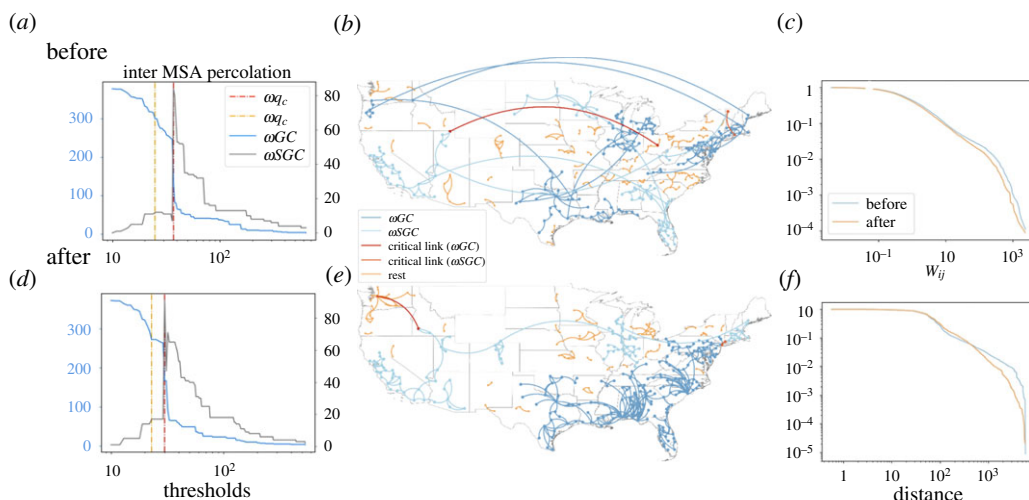


Figure 5. Percolation of inter-MSA mobility network. *a,d* show the changes of the sizes of ω_{GC} and ω_{SGC} from 29 January to 4 February 2020 and 29 May to 4 June, respectively. Panel (*b,e*) shows the geographical networks when the weight of edges equals ω_{q_c} in the same week. Panel (*c*) shows the complementary cumulative distribution function (CCDF) of normalized edge weight from 29 January to 4 February (blue line) and 29 May to 4 June (red line). (*f*) CCDF of distances of the edges from 29 January 29 to 4 February (blue line) and 29 May to 4 June (red line). (Online version in colour.)

Then we expand our percolation process to two-level thresholds, α_{q_c} as each intra-MSA network and ω weight as the normalized weight of edges between MSAs, to take the robustness of different MSAs into consideration. In figure 6, we control these two parameters and create a heat map for both ω_{GC} and ω_{SGC} before and after the national emergency declaration. We observe that before and after the national emergency declaration, both ω_{GC} and ω_{SGC} get smaller as α_{q_c} or ω weight get larger. By comparing the same components at different times (ω_{GC} in figure 6*a,b* and ω_{SGC} in figure 6*c,d*), we can see that both are more centralized in lower α_{q_c} and ω weight after the national emergency declaration than before. Thus besides ω weight as we showed, mobility is also more sensitive to α_{q_c} after the national emergency declaration.

7. Time series of percolation threshold predictability

Next, we investigate the chronological progression of α_{q_c} and measure the predictability with various features for all intra-MSA networks. Figure 7*a* captures the time series of descriptive statistics of α_{q_c} for each MSA. We find that before the national emergency declaration, the α_{q_c} remains stable, with a median value of 8.5. The Interquartile Range (IQR) spans between 7 and 11 with longer tails near the upper quantiles. Right ahead of the declaration, the median value witnessed a sudden increase and reached 10 followed by an abrupt and blunt drop to 5 by the end of March. It is notable that during this sharp decrease, the variance between different MSAs shrank remarkably to less than 1. The reduced variance indicates the universality of perturbation from the emergency declaration across different regions. After the downward trend hitting the lowest point at mid-April, the median value of intra-MSA α_{q_c} s gradually increased back to 7 around early June and remains stable again.

Figure 7*b* compares the predictive strength of α_{q_c} with a different set of features over time. The baseline for the predictors includes the 7-day mobility fluxes with both mobility influx (the total number of travellers from other MSA), outflux (the total number of travellers to other MSA), and intra-flux (the total number of travellers between each pair of census block groups within MSA). We then append different sets of features using the social distancing metrics from SafeGraph, 7-day COVID-19 incremental infection rate, and socio-demographic characteristics.

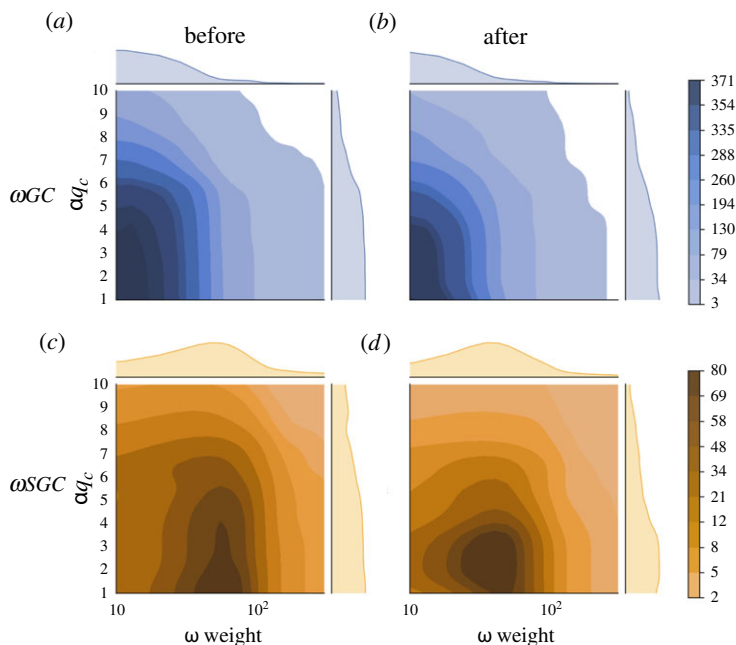


Figure 6. Heat map of ωq_c with different thresholds of αq_c and ω weight in percolation of inter-MSA mobility map before (29 January to 4 February) and after (29 May to 4 June) the national emergency declaration. Panel (a) shows the heat map of ω_{GC} with αq_c and ω weight in the week of 29 January to 4 February 2020. Panel (b) shows the heat map of ω_{GC} with αq_c and ω weight in the week of 29 May to 4 June 2020. Panel (c) shows the heat map of ω_{SGC} with αq_c and ω weight in the week of 29 January to 4 February 2020. Panel (d) shows the heat map of ω_{SGC} with αq_c and ω weight in the week of 29 May to 4 June 2020. (Online version in colour.)

All the attributes, including αq_c , are obtained as 7-day rolling average except the COVID-19 incremental infection rate, which is computed by 7-day case increase divided by the total population. To mitigate the multicollinearity effect from a high level of correlations between these features, we use principal component analysis (PCA) to ensure the orthogonality before the configuration of the regression models. Given the varying correlation between input features across different days, we apply a uniform threshold of 95% variance to determine the number of principal components being used for prediction. Such a threshold guarantees a minimum of 95% of variance explained across all MSAs on different days. We tested both R^2 and adjusted R^2 to consider the change of the number of predictors each day, and the increasing predictive capability over time can be observed for both metrics.

Figure 7b shows the R^2 values with different sets of transformed features. We can see that most combinations of these factors became more predictive of the percolation threshold during the mobility falloff except using mobility flux solely. The overall predictive strength with mobility flux is rather low, with R^2 values fluctuating around 0.2 over time. With social distancing metrics, the R^2 values jumped to between 0.3 to 0.35 before mid-March and approached 0.4 around early May. Furthermore, adding the COVID-19 infection rate improves the predictive performance by increasing the R^2 by 0.1 at most times and exceeded 0.5 around early May. Surprisingly, socio-demographic indicators do not improve the predictive strength and cause even a very minor decrease, which could be due to the lack of variations of these variables on the MSA level. Figure 7c,d is the scatter plots of prediction versus actual intra-MSA q_c s, where figure 7c shows the week of 29 January to 4 February 2020 when the R^2 is relatively low (0.32) while figure 7d is for 6 May to 12 May 2020 with an R^2 of 0.51. It is notable that for both cases, the predictions are mostly overestimating the actual outcome. There is also heteroskedasticity present: the error variance changes across different q_c s. In conclusion, the overall predictive

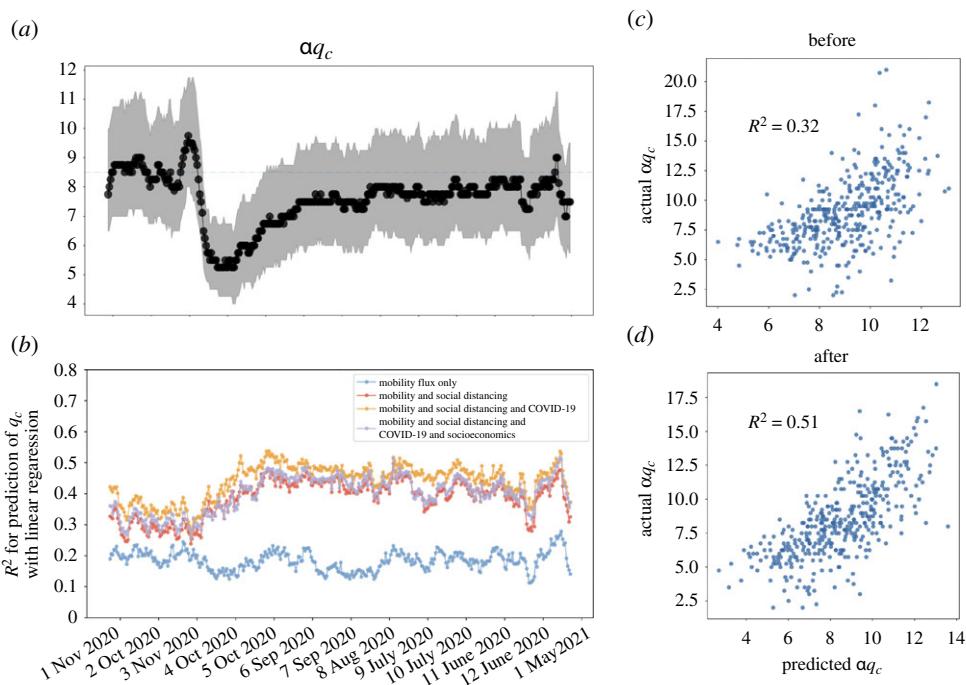


Figure 7. Time series of critical threshold value from 1 January 2020 to 31 December 2020. (a) interquartile range of αq_c and the line represents the median of αq_c in intra-MSA over time. (b) Times series of R^2 for prediction of αq_c with linear regression. Scatter plots of predicted αq_c with linear regression with respect to the actual value with weekly mean of 29 January to 4 February 2020 (before, c) and 6 May to 12 May 2020 (after, d). (Online version in colour.)

strengths of the percolation threshold with linear methods are moderate. The social distancing behaviour and mobility flow can only partially explain the percolation threshold, while social distancing has played a progressively crucial role since the beginning of the outbreak.

8. Discussion and conclusion

In this paper, we study the percolation of the human mobility network that includes over 175 000 of the census block groups in the USA. Two block groups have a connection on a day when users travel between them on that day. If more users travel between them, the weight becomes stronger. However, analysing its percolation is very time-consuming due to the large size, heterogeneous weights, and number of networks. Thus, this paper proposes a novel hierarchical structure of the mobility network to understand human dynamics during the pandemic. In our model, the inter-MSA network (i.e. ω -networks) is composed of 378 MSAs in the USA. Each MSA comprises many census block groups, which is the intra-MSA (i.e. α -networks) level. During the percolation process, we remove the q fraction of the weakest links and measure the giant component of the network because it is easier to break the weak links than the strong links. We find a critical q_c , and when $q > q_c$, the network breaks down abruptly because the large enough second-largest component is disconnected from the giant cluster. We compute daily q_c for each intra-MSA and inter-MSA. The analysis allows us to better understand the robustness of each MSA's connectivity and interactions between MSAs.

In this paper, our analysis of a network's vulnerability is based on its percolation processes and the transition points. The strength of our percolation analysis is that it breaks the network into isolated clusters with minimum cost. Thus, it can effectively reduce the propagation of COVID-19. However, the method has several limitations. The first limitation is that the percolation process

is solely based on the connectivity of a graph [42,43] and thus neglects some essential features of nodes (i.e. ratio of different races, economic status, etc.) and edges (i.e. purpose of people travelling on this route). Taking these features into account can help us better understand the mobility flow and transmission of COVID-19 and develop effective NPI strategies. The second limitation is that mobility behaviours are incredibly complex and dynamic. Thus mobility restriction based on percolation approaches should be considered part of holistic solutions with other NPIs [27]. The third limitation is that results from the percolation process can help contain COVID-19 within a small cluster at the early stage of the pandemic. However, if large-scale community transmission has already started, the method obtained by percolation analysis will be less effective.

The study has three main contributions. The first contribution is the revelation of the hierarchical structures of the temporal mobility network. Given the underlying principle of the cluster forming process in classic percolation theory [34,36], we unveil that spatially hierarchical structure is an essential characteristic of dynamical mobility networks under the impact of COVID-19. Our finding raises the significance of the interaction between different network layers. Thus, the construction of the hierarchical networks enables identifying links that could serve as the critical bridges at different scales, a break of which leads to a disconnection between giant components and second-largest components. Our numerical results provide opportunities to mitigate or prevent macro-transmission with strategic mobility controls.

Secondly, we reveal the universality of phase transition of mobility networks at critical points using percolation theory from statistical physics. Although previous research has demonstrated that the percolation process governs the mobility networks on the county level [31], here we show the percolation process controls the phase transitions in these networks on different levels (i.e. block groups and inter-MSAs) universally. It is supported by the resemblance of functions between the largest components sizes and the percolation threshold across all levels of granularity. The universality demonstrates the possibility of predicting mobility patterns regardless of geographical regions and scales.

Furthermore, the understanding of the association between critical threshold αq_c and the key features and characteristics in local MSAs improves our ability to assess and predict the vulnerability of mobility networks. Previous studies show that topological and hierarchical properties of mobility networks and heterogeneous local shocks determine the critical threshold and the emergence of vital adaptive links [31,45]. Here, we quantify the association with various data sources and find that local social distancing metrics are critical factors in predicting the percolation features for time-dependent mobility networks. Surprisingly, despite socio-economic heterogeneities observed from COVID-19's incidence rate [30], we do not find socio-economic characteristics to be deterministic of the percolation patterns. Future research is needed to help identify other key features and support public health authorities and policymakers to assess the potential local transmission risk.

There are still some limitations in our hierarchical network model, promoting new research directions in the future. First, the data are from device trajectory within a day; thus, some cases are not captured, i.e. people without smartphones, and some are overly counted, i.e. people with multiple devices. Second, we use the partition rule that treats each MSA as a node representing a 'local network.' The assumption is that mobility with MSAs is self-containing, i.e. routine travel happens mostly inside MSAs. The drawback of this partition rule is that some rural areas (i.e. vast mid-west region with few MSAs) with low population are not considered in this model, which may cause it to lose some accuracy.

Data accessibility. Data are private and protected by SafeGraph (<https://docs.safegraph.com/docs/social-distancing-metrics>). Python code used to process the data and generate the results is made publicly available on the author's github page (<https://github.com/he-h/Covid-Mobility-Network-Analysis>).

Authors' contributions. H.H., H.D., Q.W. and J.G. designed and developed the study. H.H. and H.D. developed the codes for the analyses described in the paper. H.H. and H.D. developed the figures and wrote the first draft. H.H., H.D., J.G. and Q.W. revised the manuscript.

Competing interests. We declare we have no competing interests.

Funding. We acknowledge the support of the US National Science Foundation under grant no. 2047488, the Rensselaer-IBM AI Research Collaboration, and the Health Care Cost Institute COVID-19 Research Accelerator Grant. Q.W. is supported in part by the US National Science Foundation (grant no. 2027744).

References

1. CDC. 2021 COVID data tracker.
2. Viner RM *et al.* 2021 Susceptibility to SARS-CoV-2 infection among children and adolescents compared with adults: a systematic review and meta-analysis. *JAMA Pediatr.* **175**, 143–156. (doi:10.1001/jamapediatrics.2020.4573)
3. Perra N. 2021 Non-pharmaceutical interventions during the COVID-19 pandemic: a review. *Phys. Rep.* **913**, 1–52. (doi:10.1016/j.physrep.2021.02.001)
4. Wang S, Wei K, Lin L, Li W. 2020 Spatial-temporal Analysis of COVID-19's impact on human mobility: the case of the United States. (<https://arxiv.org/abs/2010.03707>).
5. Fowler JH, Hill SJ, Levin R, Obradovich N. 2020 The effect of stay-at-home orders on COVID-19 cases and fatalities in the United States. (<https://arxiv.org/abs/2004.06098>).
6. Hsiehchen D, Espinoza M, Slovic P. 2020 Political partisanship and mobility restriction during the COVID-19 pandemic. *Public Health* **187**, 111–114. (doi:10.1016/j.puhe.2020.08.009)
7. Ugarov A. 2020 *Inclusive costs of NPI measures for COVID-19 pandemic: three approaches*. medRxiv preprint 2020.03.26.20044552.
8. Soltesz K *et al.* 2020 The effect of interventions on COVID-19. *Nature* **588**, E26–E28. (doi:10.1038/s41586-020-3025-y)
9. Merler S, Ajelli M. 2010 The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proc. R. Soc. B* **277**, 557–565. (doi:10.1098/rspb.2009.1605)
10. Belik V, Geisel T, Brockmann D. 2011 Natural human mobility patterns and spatial spread of infectious diseases. *Phys. Rev. X* **1**, 011001. (doi:10.1103/PhysRevX.1.011001)
11. Bajardi P, Poletto C, Ramasco JJ, Tizzoni M, Colizza V, Vespignani A. 2011 Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS ONE* **6**, e16591. (doi:10.1371/journal.pone.0016591)
12. Findlater A, Bogoch II. 2018 Human mobility and the global spread of infectious diseases: a focus on air travel. *Trends Parasitol.* **34**, 772–783. (doi:10.1016/j.pt.2018.07.004)
13. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO. 2012 Quantifying the impact of human mobility on malaria. *Science* **338**, 267–270. (doi:10.1126/science.1223467)
14. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. 2020 Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 389–394. (doi:10.1038/s41586-020-2284-y)
15. Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM. 2020 Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *Lancet Infect. Dis.* **20**, 1247–1254. (doi:10.1016/S1473-3099(20)30553-3)
16. Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, Rinaldo A. 2020 Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc. Natl Acad. Sci. USA* **117**, 10 484–10 491. (doi:10.1073/pnas.2004978117)
17. Hadjidemetriou GM, Sasidharan M, Kouyialis G, Parlikad AK. 2020 The impact of government measures and human mobility trend on COVID-19 related deaths in the UK. *Transp. Res. Interdiscip. Perspect.* **6**, 100167. (doi:10.1016/j.trip.2020.100167)
18. Aleta A, Hu Q, Ye J, Ji P, Moreno Y. 2020 A data-driven assessment of early travel restrictions related to the spreading of the novel COVID-19 within mainland China. *Chaos Solitons Fractals* **139**, 110068. (doi:10.1016/j.chaos.2020.110068)
19. Sirkeci I, Yucesahin MM. 2020 Coronavirus and migration: analysis of human mobility and the spread of COVID-19. *Migr. Lett.* **17**, 379–398. (doi:10.33182/ml.v17i2.935)
20. Zhong L, Diagne M, Wang W, Gao J. 2021 Country distancing increase reveals the effectiveness of travel restrictions in stopping COVID-19 transmission. *Commun. Phys.* **4**, 1–12. (doi:10.1038/s42005-021-00620-5)
21. Kraemer MUG *et al.* 2020 The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497. (doi:10.1126/science.abb4218)

22. Si H, Wang Y, Yuan J, Shan X. 2010 Mobility prediction in cellular network using hidden markov model. In *2010 7th IEEE Consumer Communications and Networking Conf.*, pp. 1–5. New York, NY: IEEE.
23. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Leskovec J. 2021 Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87. (doi:10.1038/s41586-020-2923-3)
24. Noland RB. 2021 Mobility and the effective reproduction rate of COVID-19. *J. Transp. Health* **20**, 101016. (doi:10.1016/j.jth.2021.101016)
25. Cartenì A, Di Francesco L, Martino M. 2020 How mobility habits influenced the spread of the COVID-19 pandemic: results from the Italian case study. *Sci. Total Environ.* **741**, 140489. (doi:10.1016/j.scitotenv.2020.140489)
26. Nouvellet P *et al.* 2021 Reduction in mobility and COVID-19 transmission. *Nat. Commun.* **12**, 1090. (doi:10.1038/s41467-021-21358-2)
27. Yang J, Zhang Q, Cao Z, Gao J, Pfeiffer D, Zhong L, Zeng DD. 2021 The impact of non-pharmaceutical interventions on the prevention and control of COVID-19 in New York city. *Chaos* **31**, 021101. (doi:10.1063/5.0040560)
28. Wang Q, Phillips NE, Small ML, Sampson RJ. 2018 Urban mobility and neighborhood isolation in America's 50 largest cities. *Proc. Natl Acad. Sci. USA* **115**, 7735–7740. (doi:10.1073/pnas.1802537115)
29. Prestby T, App J, Kang Y, Gao S. 2020 Understanding neighborhood isolation through spatial interaction network analysis using location big data. *Environ. Plann. A: Econ. Space* **52**, 1027–1031. (doi:10.1177/0308518X19891911)
30. Chang SY, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Leskovec J. 2020 Mobility network modeling explains higher SARS-CoV-2 infection rates among disadvantaged groups and informs reopening strategies. *medRxiv*.
31. Deng H, Gao J, Wang Q. 2021 Network percolation reveals adaptive bridges of the mobility network response to COVID-19. (<https://arxiv.org/abs/2104.11405>).
32. Phillips NE, Levy BL, Sampson RJ, Small ML, Wang RQ. 2019 The social integration of american cities: network measures of connectedness based on everyday mobility across neighborhoods. *Sociol. Methods Res.* **50**, 1110–1149.
33. Quilty BJ *et al.* 2020 The effect of inter-city travel restrictions on geographical spread of COVID-19: evidence from Wuhan, China. *BMC Med.* **18**, 259. (doi:10.1186/s12916-020-01712-9)
34. Shante VK, Kirkpatrick S. 1971 An introduction to percolation theory. *Adv. Phys.* **20**, 325–357. (doi:10.1080/00018737100101261)
35. Sahini M, Sahimi M. 2003 *Applications of percolation theory*. Boca Raton, FL: CRC Press.
36. Li M, Liu RR, Lü L, Hu MB, Xu S, Zhang YC. 2021 Percolation on complex networks: theory and application. *Phys. Rep.* **907**, 1–68. (doi:10.1016/j.physrep.2020.12.003)
37. Cohen R, Erez K, Ben-Avraham D, Havlin S. 2001 Breakdown of the internet under intentional attack. *Phys. Rev. Lett.* **86**, 3682–3685. (doi:10.1103/PhysRevLett.86.3682)
38. Albert R, Barabási AL. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97. (doi:10.1103/RevModPhys.74.47)
39. Newman ME. 2003 The structure and function of complex networks. *SIAM Rev.* **45**, 167–256. (doi:10.1137/S003614450342480)
40. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S. 2010 Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028. (doi:10.1038/nature08932)
41. Gao J, Buldyrev SV, Stanley HE, Havlin S. 2012 Networks formed from interdependent networks. *Nat. Phys.* **8**, 40–48. (doi:10.1038/nphys2180)
42. Gao J, Buldyrev SV, Havlin S, Stanley HE. 2011 Robustness of a network of networks. *Phys. Rev. Lett.* **107**, 195701. (doi:10.1103/PhysRevLett.107.195701)
43. Huang X, Gao J, Buldyrev SV, Havlin S, Stanley HE. 2011 Robustness of interdependent networks under targeted attack. *Phys. Rev. E* **83**, 065101. (doi:10.1103/PhysRevE.83.065101)
44. Gao J, Barzel B, Barabási AL. 2016 Universal resilience patterns in complex networks. *Nature* **530**, 307–312. (doi:10.1038/nature16948)
45. Li D, Fu B, Wang Y, Lu G, Berezin Y, Stanley HE, Havlin S. 2015 Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proc. Natl Acad. Sci. USA* **112**, 669–672. (doi:10.1073/pnas.1419185112)

46. Zeng G, Li D, Guo S, Gao L, Gao Z, Stanley HE, Havlin S. 2019 Switch between critical percolation modes in city traffic dynamics. *Proc. Natl Acad. Sci. USA* **116**, 23–28. (doi:10.1073/pnas.1801545116)
47. Zeng G *et al.* 2020 Multiple metastable network states in urban traffic. *Proc. Natl Acad. Sci. USA* **117**, 17 528–17 534. (doi:10.1073/pnas.1907493117)
48. Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A. 2005 Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *J. Theor. Biol.* **235**, 275–288. (doi:10.1016/j.jtbi.2005.01.011)
49. Wang W, Tang M, Stanley HE, Braunstein LA. 2017 Unification of theoretical approaches for epidemic spreading on complex networks. *Rep. Prog. Phys.* **80**, 036603. (doi:10.1088/1361-6633/aa5398)
50. Gleeson JP, Cahalane DJ. 2007 Seed size strongly affects cascades on random networks. *Phys. Rev. E* **75**, 056103. (doi:10.1103/PhysRevE.75.056103)
51. Watts DJ. 2011 A simple model of global cascades on random networks. In *The structure and dynamics of networks*, pp. 497–502. Princeton, NJ: Princeton University Press.
52. Arcaute E, Molinero C, Hatna E, Murcio R, Vargas-Ruiz C, Masucci AP, Batty M. 2016 Cities and regions in Britain through hierarchical percolation. *R. Soc. Open Sci.* **3**, 150691. (doi:10.1098/rsos.150691)