



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



The impact of mobility network properties on predicted epidemic dynamics in Dhaka and Bangkok

Tyler S. Brown^{a,b}, Kenth Engø-Monsen^c, Mathew V. Kiang^d, Ayesha S. Mahmud^e,
Richard J. Maude^{b,f,g}, Caroline O. Buckee^{b,*}

^a Massachusetts General Hospital, Infectious Diseases Division, United States of America

^b Harvard T.H. Chan School of Public Health, United States of America

^c Telenor Research, Norway

^d Stanford University School of Medicine, Department of Epidemiology and Population Health, United States of America

^e University of California, Berkeley, Demography Department, United States of America

^f Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Thailand

^g Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, United Kingdom

ARTICLE INFO

Keywords:

Cities
Human mobility
Mobile phone data
SARS-CoV-2 dynamics

ABSTRACT

Properties of city-level commuting networks are expected to influence epidemic potential of cities and modify the speed and spatial trajectory of epidemics when they occur. In this study, we use aggregated mobile phone user data to reconstruct commuter mobility networks for Bangkok (Thailand) and Dhaka (Bangladesh), two megacities in Asia with populations of 16 and 21 million people, respectively. We model the dynamics of directly-transmitted infections (such as SARS-CoV-2) propagating on these commuting networks, and find that differences in network structure between the two cities drive divergent predicted epidemic trajectories: the commuting network in Bangkok is composed of geographically-contiguous modular communities and epidemic dispersal is correlated with geographic distance between locations, whereas the network in Dhaka has less distinct geographic structure and epidemic dispersal is less constrained by geographic distance. We also find that the predicted dynamics of epidemics vary depending on the local topology of the network around the origin of the outbreak. Measuring commuter mobility, and understanding how commuting networks shape epidemic dynamics at the city level, can support surveillance and preparedness efforts in large cities at risk for emerging or imported epidemics.

1. Introduction

Densely populated cities are uniquely vulnerable to infectious disease epidemics (Jowell et al., 2017). The distribution and connectivity of human populations within cities plays a key role in spreading outbreaks when they occur, but the factors determining how these vulnerabilities translate into epidemic dynamics are still unclear, and this has important implications for surveillance and preparedness. Local SARS-CoV-2 epidemics have thus far exhibited substantial diversity in their scale and trajectory between cities (Bialek et al., 2020), for example, and the factors behind differential city-level risk of epidemic propagation (for both SARS-CoV-2 and other emerging or endemic pathogens) are still incompletely understood (Heroy, 2020; Stier et al., 2020). Likewise, our ability to predict the spatial distribution of disease activity within cities during epidemics (for example, by neighborhood or hospital catchment area) is constrained by lack of informative data, including geolocated or spatially-resolved epidemiological data and

reliable data on human movement in cities. Improved strategies are needed for *a priori* stratification of epidemic risk in large cities, including strategies that can estimate epidemic potential at the city level and identify local determinants of epidemic risk within cities.

Commuter mobility is an important factor in the local dispersal of directly transmissible pathogens (for example, influenza (Charaudeau et al., 2014) and for SARS-CoV-2 (Kissler et al., 2020) and differences in commuter mobility patterns predict divergent epidemic dynamics between cities (Dalziel et al., 2013). Mobile phone call detail records (CDRs) have become an important tool for estimating human mobility, and the utility of CDR data for modeling infectious disease dynamics has been demonstrated in multiple contexts (Wesolowski et al., 2015b,a). We propose that CDR-derived approximations of commuter mobility networks in cities can inform city-level predictions of epidemic

* Corresponding author.

E-mail addresses: tsbrown@mgh.harvard.edu (T.S. Brown), cbuckee@hsph.harvard.edu (C.O. Buckee).

<https://doi.org/10.1016/j.epidem.2021.100441>

Received 28 July 2020; Received in revised form 20 January 2021; Accepted 1 February 2021

Available online 22 February 2021

1755-4365/© 2021 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

risk, and we hypothesize that both node-level and higher-order properties of these networks influence the temporal and spatial trajectory of city-wide epidemics.

In this study, we employed aggregated CDR data to estimate commuter mobility networks in Bangkok, Thailand and Dhaka, Bangladesh and applied a stochastic model for propagation of a directly transmitted infection (similar to SARS-CoV-2 or influenza) on these networks. These two cities capture some of the substantial diversity in population density, spatial organization, transportation networks, and socioeconomic conditions observed across “megacities” (typically defined as cities and associated urban agglomerations with > 10 million total residents), and are thus well-suited for studying differences in epidemic dynamics between large cities. We find important differences between commuter mobility networks and predicted spatiotemporal trajectories of epidemics in each city. We report metrics on spatial distribution of population density in megacities, including Dhaka and Bangkok, to contextualize these findings. Our results support the use of mobile phone user data for evaluating city-level susceptibility to emerging or imported epidemics and for identifying locations within cities where local topology of the commuter network is expected to facilitate epidemic propagation.

2. Methods

2.1. Spatial heterogeneity in population density across megacities

To contextualize our analysis of city-level mobility and epidemic dynamics in Dhaka and Bangkok, we calculate multiple metrics summarizing the spatial distribution of population density across Dhaka and Bangkok (Volpati and Barthelemy, 2018), using data from the WorldPop project (Lloyd et al., 2019). These include both spatially-naïve metrics (Gini coefficient, entropy, and relative standard deviation) and the “spatial dispersion index”, a spatially-explicit metric characterizing heterogeneity in population density across a given area (the “spreading index” in Volpati and Barthelemy (2018)). Additional information on these metrics is provided in the Supplementary Information. To control for the different sizes and shapes of the study areas in each city, we calculate each metric over concentric circles centered on the most densely populated grid square in each city, and report the metrics by the radius r of each circle. For context, we report these metrics for 19 additional megacities located in low- and middle-income countries (Bogota, Columbia; Cairo, Egypt; Guadalajara and Mexico City, Mexico; Ho Chi Minh City, Vietnam; Hyderabad, Jaipur, Kolkata, Mumbai, and Pune, India; Istanbul, Turkey; Johannesburg, South Africa; Karachi and Lahore, Pakistan; Brazzaville-Kinshasa, Congo-Democratic Republic of Congo; Lagos, Nigeria; Manila, Philippines; Sao Paulo, Brazil; and Tehran, Iran).

2.2. Commuting networks estimated from mobile phone user data

Daily commuter flux between locations in Bangkok and Dhaka was estimated using aggregated, anonymized CDR data for 4.3 and 18.9 million average daily mobile phone subscribers in the Bangkok Metropolitan Region (BMR) and Dhaka Statistical Metropolitan Area (DSMA), respectively (Supplementary Figure S1). The BMR is an administratively-defined region that includes Bangkok and five surrounding provinces, covering 7762 km² with an estimated population of approximately 15.9 million (2048 persons/km²). The DSMA is composed of Dhaka and several surrounding administrative units covering 1353 km² with an estimated population of approximately 21 million (15,521 persons/km²). We collected CDR data over 81 consecutive days (1 August to 19 October, 2017 in Bangkok, and 1 April to 21 June 2017 in Dhaka), restricted to the mobile network towers within each administrative region (BMR or DSMA), and excluded major national and religious holidays. We aggregated data over 500 m x 500 m grid squares and used squares with at least one mobile phone tower in service during

the data collection period to define the set of nodes $\mathcal{L} = \{L_1, L_2, \dots, L_i\}$ in the commuting network. We used Voronoi polygons to define the catchment area around each node and used estimated population maps from WorldPop (Lloyd et al., 2019) to assign the population in each catchment area N_i for each node $L_i \in \mathcal{L}$.

To estimate daily trip counts between nodes, we identify the mobile network tower used for the majority of each user’s calls (the “most-visited” node) for each of two consecutive 24-hour periods. We assume that the most-visited node during the first 24-hour period and the most visited node during the following 24-hour period represent the origin and destination, respectively, of a single daily trip between the areas serviced by the tower at each node. In brief, the mean number of raw, unweighted trips originating from L_j and terminating at L_k , averaged over each daily observation in the data collection period, is denoted $T_{j,k}^{\text{raw}}$. We estimate the population-weighted number of trips originating from node L_j and terminating at node L_k by apportioning the total population in N_j by the relative proportion of all raw trips counts that originate at L_j and terminate at node L_k .

$$T_{j,k}^{\text{weighted}} = N_j \times \frac{T_{j,k}^{\text{raw}}}{\sum_{i=1}^n T_{j,i}^{\text{raw}}} \quad (1)$$

where n is the total number of nodes in \mathcal{L} . $T_{j,k}^{\text{weighted}}$ thus represents the proportion of the population at node L_j that regularly travels between L_j and L_k . We used the resulting trip counts to construct an $n \times n$ origin–destination matrix \mathcal{M}_{OD} where each entry $T_{j,k}^{\text{weighted}}$ equals the weighted mean number of daily trips between nodes L_j and L_k during the data collection period (and rows and columns are indexed by j and k , respectively).

$$\mathcal{M}_{\text{OD}} = \begin{bmatrix} T_{1,1}^{\text{weighted}} & \dots & \dots & T_{1,n}^{\text{weighted}} \\ \dots & T_{j,k}^{\text{weighted}} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ T_{n,1}^{\text{weighted}} & \dots & \dots & T_{n,n}^{\text{weighted}} \end{bmatrix} \quad (2)$$

We consider two separate versions of \mathcal{M}_{OD} : $\mathcal{M}_{\text{OD}}^{\text{wke}}$, which includes only trips originating on the last day of the weekend and terminating on the first day of the conventionally observed work week (Monday in Bangkok and Sunday in Dhaka); and $\mathcal{M}_{\text{OD}}^{\text{all}}$, which includes trips originating on all days in the data collection period. Analysis using $\mathcal{M}_{\text{OD}}^{\text{wke}}$ is motivated by the assumption that the most visited location on weekend days is likely to represent a user’s home location, and the most visited location during the first day of the work week is likely to be a user’s work location, such that $\mathcal{M}_{\text{OD}}^{\text{wke}}$ is expected to capture commuter mobility between home and work locations. $\mathcal{M}_{\text{OD}}^{\text{all}}$ captures movement between frequently-visited locations and, although there is no expectation that these are either home or work locations, provides a proxy measurement for aggregate daily human movements between locations within each city. Entries $T_{j,k}^{\text{weighted}}$ in $\mathcal{M}_{\text{OD}}^{\text{all}}$ are calculated using all 81 available origin days in the data collection period and 8 observed weekend–weekday transitions in the data collection period for $\mathcal{M}_{\text{OD}}^{\text{all}}$. $\mathcal{M}_{\text{OD}}^{\text{all}}$ and $\mathcal{M}_{\text{OD}}^{\text{wke}}$ are highly similar to one another (Mantel test $p < 0.005$) and the mobility networks estimated from the two matrices share are very similar with respect to network community structure and centrality measures (Supplementary Figures S2-S5), i.e. we observe the same overall mobility patterns between the two methods of estimating mobility flux. We use $\mathcal{M}_{\text{OD}}^{\text{all}}$ for the primary analyses that follow, given the large number of days in this data collection period, which is expected to provide more informative data compared to $\mathcal{M}_{\text{OD}}^{\text{wke}}$, which is derived from a much smaller number of observations. Additional data processing procedures and analyses using $\mathcal{M}_{\text{OD}}^{\text{wke}}$ are reported in the Supplementary Information.

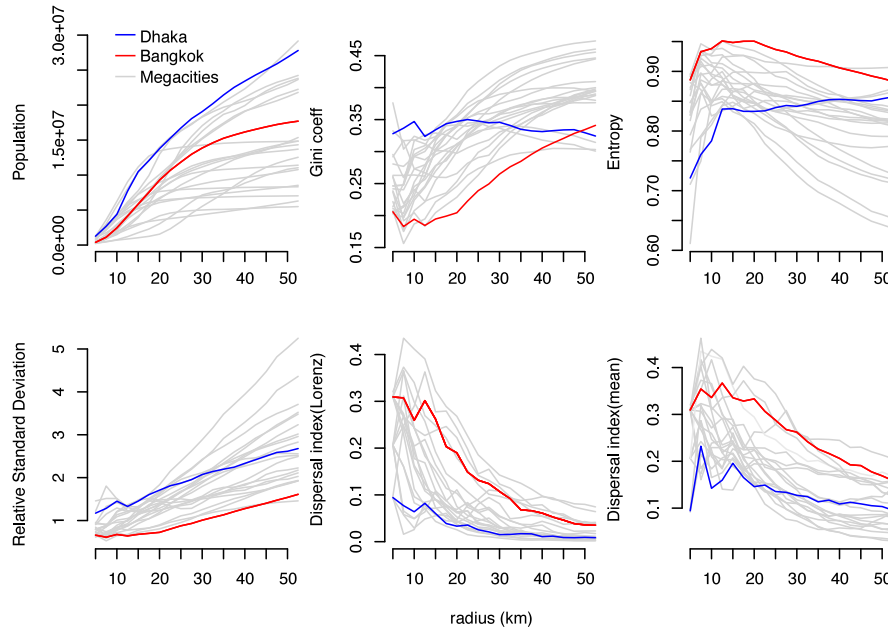


Fig. 1. Spatial distribution of population density in Bangkok and Dhaka. Total population, three spatially-naïve metrics of population density (Gini coefficient, entropy, and relative standard deviation), and one spatially-explicit metric of population density (spatial dispersion index) are shown for Bangkok (red lines), Dhaka (blue lines), and 19 additional megacities (gray lines). Each metric is estimated over concentric circles centered on the WorldPop cell with highest population density in each city (see Fig. 2). The dispersal index is estimated using two different threshold values, the mean of all population density values and the “Loubar” threshold value derived from the Lorenz curve of population density values, as described in Volpati and Barthelemy (2018). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3. Characterization of commuter mobility networks

Using the edge-weighted, directed network specified by the final origin–destination matrices (\mathcal{M}_{OD}^{all} and \mathcal{M}_{OD}^{wke}), we calculate multiple node-level metrics describing local network topology, including degree (total number of incoming and outgoing edges for each node) and strength (the sum of all incoming or outgoing edge weights given by $\sum_{j=1}^{j=n} T_{j,i}^{weighted} + \sum_{k=1}^{k=n} T_{i,k}^{weighted}$ for node L_i). We also calculate eigenvector centrality, defined as the largest positive eigenvector for the network adjacency matrix, which prior work has identified as an important indicator for spreading power in networked epidemics (Canright and Engø-Monsen, 2006). Following Brockmann and Helbing (2013), we calculate the effective distances between nodes, based on estimated connectivity rather than geographic distances, as $d_{j,k} = 1 - \log(T_{j,k}^{weighted})$ over the shortest estimated path between nodes L_j and L_k . We inferred the community structure of the commuting network, i.e., the size and membership of distinct modular subnetworks within the larger network, using InfoMap (Rosvall et al., 2009; Rosvall and Bergstrom, 2011; Csardi and Nepusz, 2006). Briefly, InfoMap estimates community structure by minimizing the length of the Huffman code (Huffman, 1952) descriptor for the path of a random walk across a given network (Rosvall et al., 2009).

2.4. Stochastic modeling of epidemic propagation on commuter mobility networks

We use a stochastic metapopulation model to estimate the propagation of a directly-transmissible immunizing infection with susceptible–exposed–infected–recovered (*SEIR*) dynamics over the city-level commuting networks in Dhaka and Bangkok. Although the duration of protective immunity acquired during SARS-CoV2 infection is still uncertain (Long et al., 2020), we assume that recovered individuals are not susceptible to infection, at least for the relatively short periods of time considered in our simulations. The framework of the model follows Tizzoni et al. (2014) and Li et al. (2020), with the

initial population of susceptible individuals distributed into origin–destination compartments specified by \mathcal{M}_{OD}^{all} and \mathcal{M}_{OD}^{wke} . We assume that all individuals are susceptible to infection at time $t = 0$.

$$S_{t=0} = \begin{bmatrix} S(0)_{1,1} & \cdot & \cdot & \cdot & S(0)_{1,n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & S(0)_{j,k} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ S(0)_{n,1} & \cdot & \cdot & \cdot & S(0)_{n,n} \end{bmatrix} = \mathcal{M}_{OD} \quad (3)$$

The number of exposed, infected, and recovered individuals in each corresponding compartment at time t are given by the $n \times n$ matrices \mathcal{E} , \mathcal{I} , and \mathcal{R} , with entries $E_{j,k}$, $I_{j,k}$, and $R_{j,k}$, respectively. Time-steps are days, and new exposures ($S \rightarrow E$ transitions) result from interaction with infected individuals at either the origin or destination, with force of infection λ^{origin} and $\lambda^{\text{destination}}$, respectively.

$$\lambda_{j,k}^{\text{origin}}(t) = \beta \frac{\sum_{k=1}^{k=n} I_{j,k}(t)}{\sum_{k=1}^{k=n} (S_{j,k}(t) + E_{j,k}(t) + I_{j,k}(t) + R_{j,k}(t))} \quad (4)$$

$$\lambda_{j,k}^{\text{destination}}(t) = \beta \frac{\sum_{j=1}^{j=n} I_{j,k}(t)}{\sum_{j=1}^{j=n} (S_{j,k}(t) + E_{j,k}(t) + I_{j,k}(t) + R_{j,k}(t))} \quad (5)$$

where β is a constant specifying the risk of infection (converted from the daily rate of infection estimated in Li et al. (2020)). $\sum_{k=1}^{k=n} I_{j,k}(t)$ and $\sum_{j=1}^{j=n} I_{j,k}(t)$ are the number of infected individuals to which the $S_{j,k}(t)$ susceptible individuals are exposed to in the origin and destination nodes j and k , respectively. For each time-step, the number of susceptible individuals $S_{j,k}(t)$ who are exposed is the sum of two draws from two different binomial distributions with probabilities λ^{origin} and $\lambda^{\text{destination}}$. Infected individuals are parsed into reported and unreported cases, with u denoting the relative infectiousness of reported versus unreported infections. The model does not consider births or deaths, such that the total populations across compartments are invariant over time and, for all times t , the denominators in Eqs. (4) and (5) equal $\sum_{k=1}^{k=n} S_{j,k}(0)$ and $\sum_{j=1}^{j=n} S_{j,k}(0)$, respectively. Progression to infectiousness ($E \rightarrow I$ transitions) and recoveries ($I \rightarrow R$ transitions) occur with fixed probabilities η and γ .

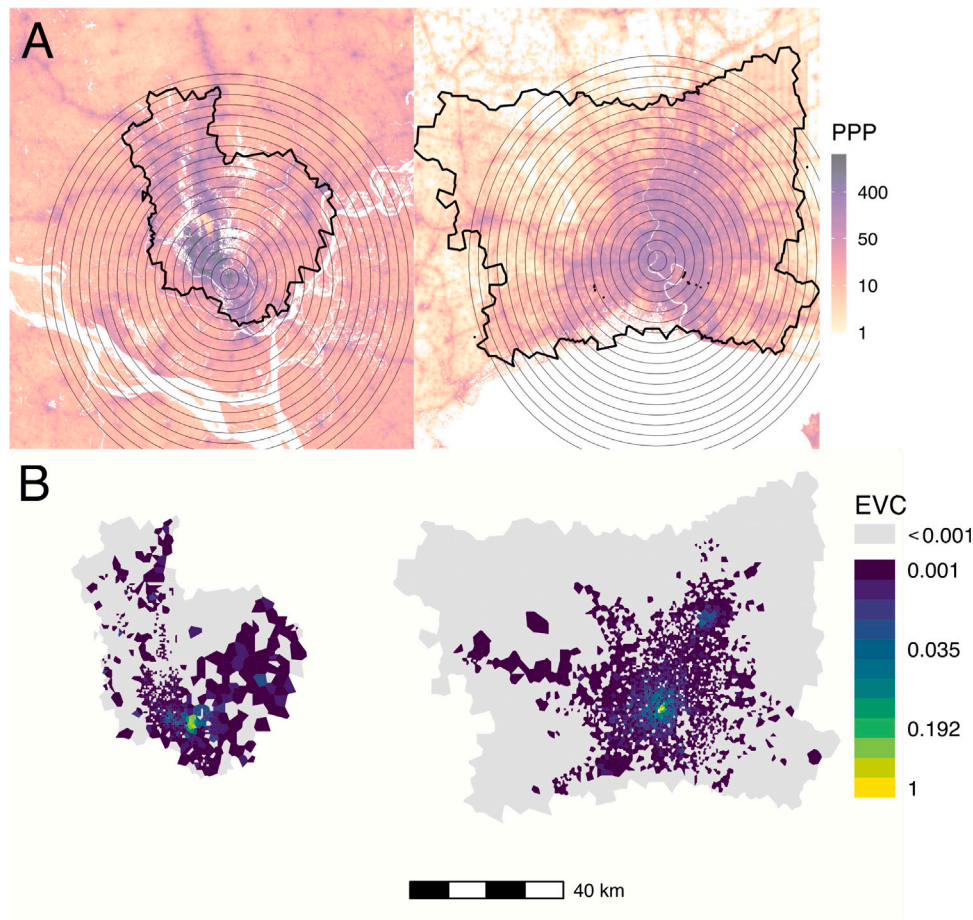


Fig. 2. Population density, mobile data catchment areas, and eigenvector centrality in Dhaka (left) and Bangkok (right). (A) Population density in persons per pixel (PPP) from WorldPop database, overlaid with concentric circles used for spatially-restricted measurements of population density in Fig. 1 (gray circles) and the outline of the mobile service tower-associated catchment areas used to estimate commuter mobility in each city. (B) Eigenvector centrality by node for Dhaka and Bangkok for the mobility network specified by \mathcal{M}_{OD}^{all} .

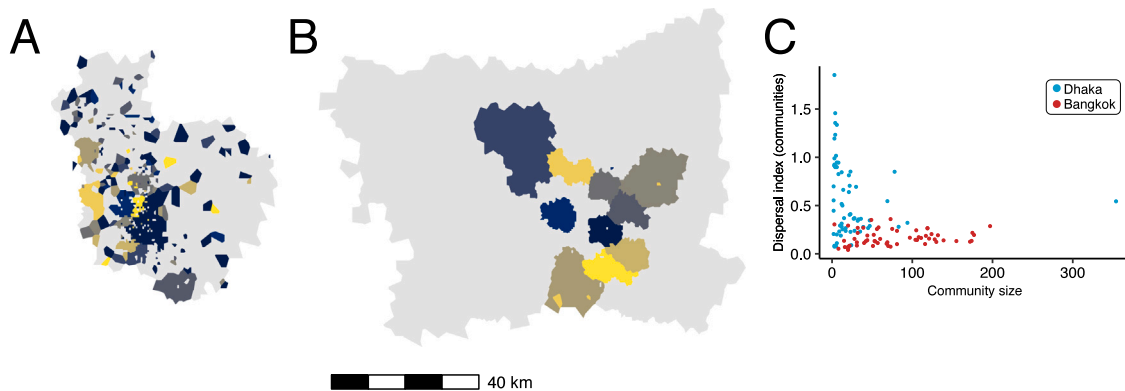


Fig. 3. Commuter mobility network structure in Dhaka and Bangladesh. The ten communities with largest membership (by number of nodes) are mapped for (A) Dhaka and (B) Bangkok. Panel (C) plots community membership size (by number of nodes) versus the spatial dispersion index calculated for the nodes in each community (blue: Dhaka, red: Bangkok). Colors for polygons in (A) and (B) are arbitrary. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.5. Modeling parameter selection and sampling

Base parameters for the stochastic SEIR model were chosen from parameters inferred for a similar model using COVID-19 disease report data from China (Li et al., 2020). We sample parameters from across

the 95% credible interval for each parameter β , η , γ , and u as estimated by Li et al. (2020) (specifically, estimates derived from data collected during early, pre-intervention stages of the epidemic) using Latin hypercube sampling. For some analyses β , η , γ , and u are set to fixed values of interest.

3. Results

3.1. City structure differences between Bangkok and Dhaka

Calculated values for the Gini coefficient and relative standard deviation, which both increase with increasing statistical dispersion, are higher across Dhaka versus Bangkok (Fig. 1), consistent with greater heterogeneity in population density in Dhaka. Likewise, calculated entropy values, which decrease with increasing statistical dispersion, are lower across all spatial scales (i.e. concentric circles of radius r) in Dhaka. The spatial dispersion index is lower in Dhaka for all values of r , indicating that cells with higher population density are distributed over smaller areas in Dhaka versus Bangkok. The spatial distributions of population density in Dhaka and Bangkok span the range of observations for other large cities in low- and middle-income countries (Fig. 1, gray lines). The spatial distribution of population density in Dhaka is generally more heterogeneous and concentrated over smaller spatial scales versus other cities, whereas the spatial distribution of population density in Bangkok is “flatter” and distributed over larger areas versus other cities (Fig. 2A).

3.2. Network properties and community structure in Dhaka and Bangkok

Node-level properties demonstrate important differences in overall connectivity and network density between Dhaka and Bangkok. The node degree distribution in Dhaka is concentrated around higher values, with most nodes sharing edges with all or almost all other nodes, with a corresponding network edge density (the ratio of the number of edges and the number of possible edges in a network) of 0.82; the degree distribution in Bangkok is less concentrated at higher values, with a corresponding network edge density of 0.11 (Supplementary Figure S6). The distribution of node strengths is more widely dispersed and has higher maximum values in Dhaka compared to Bangkok.

Network centrality in both cities is concentrated within more highly populated areas. Eigenvector centrality values are highest in centrally-located areas with high population density in both Bangkok and Dhaka; in Bangkok, we observe a second focus of nodes with high eigenvector centrality values near a known transportation hub in the northeast area of the BMR (Fig. 2B).

Network community structures are distinctly different in Dhaka and Bangkok. Fig. 3 maps the ten largest network communities in either Dhaka or Bangkok, inferred from the commuter network specified in \mathcal{M}_{OD}^{all} . In this context, network communities represent subnetworks of nodes in the commuting network that are more strongly connected to each other compared to other nodes. Network communities in Bangkok (Fig. 3B) are geographically contiguous and constrained by geographic barriers (for example, the Chao Praya River); with some exceptions, communities in Dhaka are mostly discontinuous and relatively unconstrained by geographic barriers (Fig. 3A) or distance (Fig. 3C). The geographic contiguity of network communities in Bangkok is observed even after downsampling to a smaller set of nodes within a restricted geographic area (Figure S7), indicating that differences in community structure observed between Dhaka and Bangkok are likely not related to differences in connectivity to unobserved nodes not included in the data catchment area.

3.3. Propagation of simulated epidemics: synchrony and predictability

In simulated epidemics of a directly-transmitted pathogen with SEIR dynamics, initialized using parameters inferred for SARS-CV2 transmission, we observe highly synchronized epidemic dynamics in Dhaka (Figs. 4A & 4B), with all nodes across the city reaching their highest number of predicted infections within a short, synchronized period of time. Epidemic propagation is less synchronized in Bangkok and nodes located in the peripheral areas of the BMR, specifically in the west and northeast, exhibit delayed epidemic peaks compared to

nodes in central Bangkok or the western BMR (Figs. 4C & 4D). These results are consistent across a wide range of modeling parameters and are observed regardless of how the origin node for each simulation is chosen (either from a uniform distribution or population-weighted multinomial distribution, Figure S9).

Simulated epidemics initialized on the Bangkok commuting network exhibit distinct wave-like dispersal away from an epidemic origin (Fig. 5B). Early in the simulated epidemics, the number of infected individuals in a given node is negatively correlated with geographic distance from the origin node. This correlation becomes less distinct and later reverses and becomes positive as the wave of epidemic dispersal moves outward from the epidemic origin. Simulated epidemics on the Dhaka commuter network do not exhibit any wave-like features and, unlike Bangkok, propagation of the epidemic is largely unconstrained by distance (Fig. 5A) at all time points. These features are observed regardless of how origin nodes are selected (either from a population-weighted multinomial distribution or a uniform distribution, Figure S9), and are also observed if the mobility network is specified by \mathcal{M}_{OD}^{wke} rather than \mathcal{M}_{OD}^{all} (Figure S10). Similar findings are observed when the mobility network is restricted to nodes within a specified distance from the epidemic origin (60 km in Figure S11 and 30 km in Figure S12).

3.4. Local network topology influences epidemic trajectory

We next examined dynamics of simulated epidemics originating from nodes with high and low local connectivity, as measured by their eigenvector centrality values. In simulated epidemics propagating on the Bangkok commuter network, epidemic speed (measured as the time to peak number of infections in the entire network) differs depending on the eigenvector centrality of the origin node (Fig. 6): specifically, simulated epidemics originating at nodes with lower eigenvector centrality values are slower, with later epidemic peak times, compared to those originating at nodes with high eigenvector centrality (mean time to epidemic peak, $\bar{t}_{peak} = 198$ and 222 days for nodes with eigenvector centrality values in the tenth and first deciles, respectively). For simulated epidemics initialized on the Dhaka commuter network, we observe a smaller difference in time to epidemic peak between simulations seeded at nodes with low versus high eigenvector centrality values ($\bar{t}_{peak} = 197$ and 203 days). Maximum epidemic size, i.e. the estimated number of infected individuals at the epidemic peak, also varies by eigenvector centrality of the origin node with smaller differences observed in Dhaka compared to Bangkok (Figure S13). The degree, strength, and eigenvector centrality of the origin node are more strongly correlated with mean epidemic arrival time (time to first infection in a given node, averaged over all nodes in the network) in Bangkok compared to Dhaka (Supplementary Figure S6). Geographic distance and effective network distance (a measure of mobility-based connectivity between nodes) from the origin node are positively correlated with epidemic arrival time in Bangkok (Supplementary Figure S14); in Dhaka, epidemic arrival time is positively correlated with effective network distance, but is less consistently correlated with geographic distance. In both Dhaka and Bangkok, for epidemics simulated with a wide range of model parameters and seeded at random nodes across each network, epidemic arrival time is negatively correlated with eigenvector centrality of non-origin nodes (Supplementary Figure S14).

4. Discussion

Epidemic dynamics in cities will depend on a range of factors including the connectedness and distribution of its populations. The timing, appropriateness, and efficacy of disease control interventions undertaken in response to ongoing or potential epidemics are clearly central determinants in how epidemics unfold in different cities, and may supersede the impact of many other city-level properties that might modify epidemic risk. Nevertheless, understanding how observable

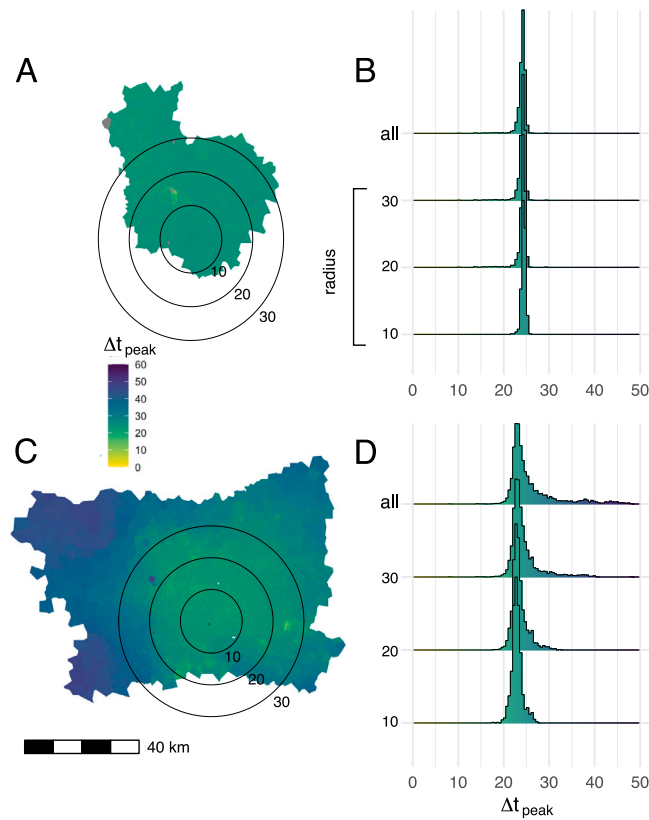


Fig. 4. Mean time to epidemic peak for each node in the commuting network. Map polygons are colored according to the time lag in days between the time to epidemic peak in a given node and the earliest peaking node on the map, Δt_{peak} (A: Dhaka, C: Bangkok). Results from 1000 independent SEIR simulations, using parameters drawn via Latin hypercube sampling (as described in *Methods*), are shown. The seed node for each simulation is drawn from a multinomial distribution, where the probability that a given node is chosen as the seed is proportional to the total population in the catchment area (Voronoi polygon) around that node. Histograms show the distribution of epidemic peak times over different geographic areas in Dhaka (B) and Bangkok (D), including nodes within concentric circles centered on most densely populated node in each city (radii: 10, 20, and 30 km) and the entirety of both study areas (“all” in panels B and D).

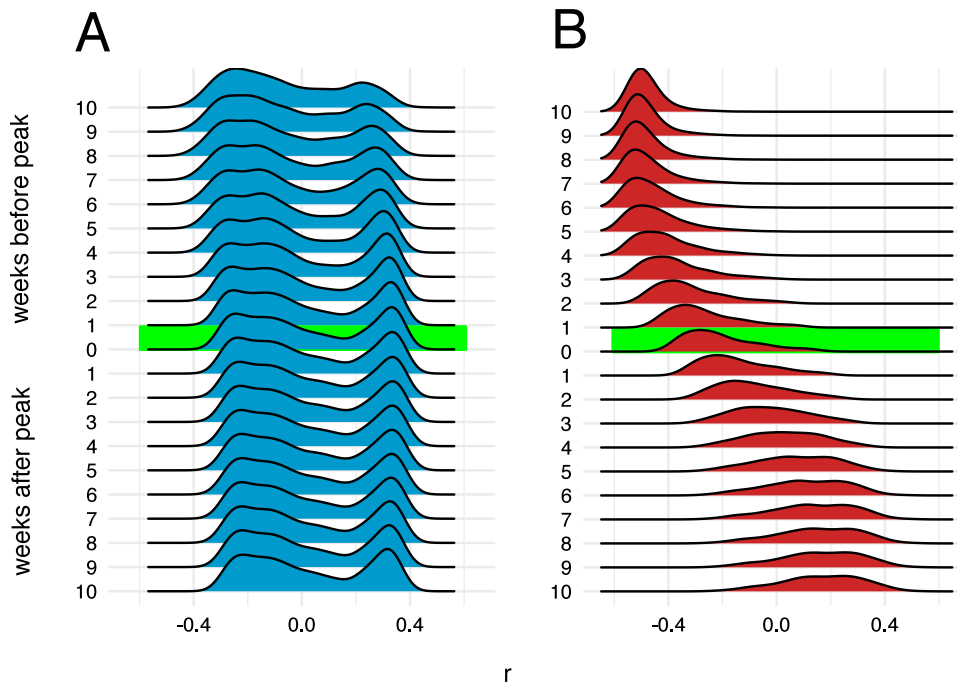


Fig. 5. Correlation between number of infections and distance from epidemic origin node over time. Distributions of Pearson’s coefficients for correlation between number of infections in each node and nodes’ respective distances in km from the origin node, obtained from 1000 independent SEIR simulations, are shown at different time points during simulated epidemics. The origin node for each simulation is chosen from a multinomial distribution weighted by the total population in the catchment area (Voronoi polygon) around each node. A: Dhaka, B: Bangkok.

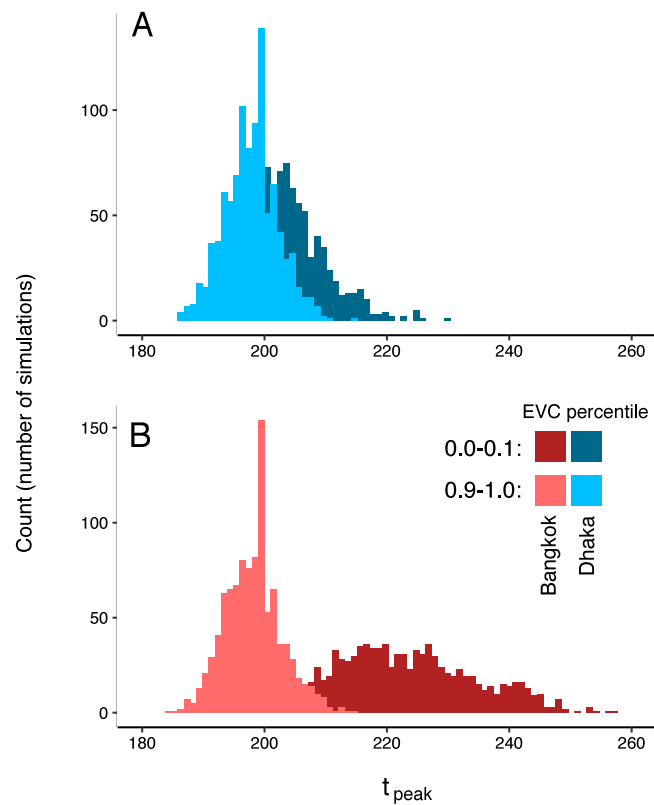


Fig. 6. Epidemic dynamics and eigenvector centrality at the origin node. Distributions for time (in days) to epidemic peak for are shown for simulated epidemics seeded at nodes with high and low eigenvector centrality values (in the first and tenth deciles, respectively). A: Dhaka; B: Bangkok. Distributions are for 1000 simulated epidemics with $\beta = 1.12$ days⁻¹.

properties of cities shape their intrinsic vulnerability to epidemic propagation remains an important scientific and public health priority. Here, we focus on mobility, acknowledging that mobility itself is closely intertwined with other important factors such as population density, economic conditions, and public infrastructure for transportation, public health, and social support.

Megacities as a group exhibit wide variation across these potential determinants of epidemic dynamics. Building on prior, foundational work on city-level mobility and epidemic risk (Dalziel et al., 2013), we have focused our analysis on Bangkok and Dhaka, two cities with distinctly different economic and social conditions that also capture the wide variation in spatial population structure across megacities (Fig. 1). The resulting analysis provides informative comparative findings, delineating key differences in underlying mobility networks and predicted epidemic dynamics in these two example cases. We also find important associations between characteristics of city-level mobility networks and multiple important epidemic features, including the synchrony, spatial dispersal, and size of simulated epidemics.

CDR-estimated mobility networks in Dhaka and Bangkok exhibit distinctly different network community structures. Network communities (i.e. modular subnetworks of nodes with relatively higher shared connectivity) in Bangkok are largely contiguous and geographically constrained, suggesting that mobility in the BMR is highly local and that daily movements of many individuals are limited to or organized around specific geographic areas (for example, neighborhoods or neighboring cities such as Nonthaburi). Communities in Dhaka are distinctly non-contiguous, suggesting movement within the DMSA is less local and not constrained by geography. Strong geographic contiguity within communities is observed in the BMR data even after restricting our analysis to a small area in central Bangkok, indicating that the geographically disorganized network communities observed in Dhaka are likely not an artifact of unmeasured connectivity with unobserved nodes outside the DMSA study area (Supplementary Information).

Simulated epidemics in Dhaka are tightly synchronized over time and space, unlike Bangkok, where simulated epidemics propagate as distinct spatial waves that arrive and peak later in outlying parts of the BMR, similar to prior observations on influenza epidemics in the United States (Viboud et al., 2006). These differences are robust over different spatial scales in each city (Fig. 4B & 4D), indicating that the synchronized epidemics in Dhaka are not simply an artifact of the relatively smaller geographic size of the DMSA. This is an important alternative explanation to consider. Given a sufficiently large catchment area around the DMSA, we would expect to see delayed epidemic peaks at outlying or weakly connected locations; indeed, using our modeling approach, this would be an expected result for simulated epidemics propagated on any mobility network in which certain nodes have weaker connectivity to the network as whole (due to distance or other factors). Still, the observations in this study have potentially important implications for public health decision-making in the BMR and DMSA. Specifically, for Dhaka, these results may suggest that resources for testing and case detection (for example, PCR-based testing for SARS-CoV2) should be deployed as early and widely as possible during an epidemic, and that geographically-restricted testing (for example, at assumed transmission “hot spots”) could result in large numbers of infections going undetected.

Several other considerations are important for contextualizing our findings. Estimating mobility from mobile phone user data requires the use of multiple simplifying assumptions, some that are inherent to all estimates made from this kind of data, some that are essential to protect the privacy of mobile data users, and some that are specific to our study. Most important, we assume that two locations, (1) the node where a user places the majority of calls during one 24-hour period and (2) the node where the same user places the majority of calls during the subsequent 24-hour period, represent the origin and terminus of a single trip by one user. This assumption has potential shortcomings and the resulting origin–destination matrix may not fully

capture daily movement. Numerous alternative approaches for estimating daily commuter movement from CDR data have been proposed and examined (for example, Alexander et al. (2015), Kung et al. (2014), Tongsinoot and Muangsin (2017)) but as yet there is no consensus on which of these approaches provides the most appropriate informative data for modeling epidemic dynamics. How origin–destination matrices derived from CDR data compare with census-derived estimates (Mamei et al., 2019), and how the use of these data sources influences models of epidemic dynamics (Tizzoni et al., 2014; Panigutti et al., 2017), are important research questions. We also note that the entries in \mathcal{M}_{OD}^{wke} and \mathcal{M}_{OD}^{all} are scaled using estimated population data from the WorldPop project, and that these estimates have important sources of bias and variance as well (Lloyd et al., 2019). We also acknowledge that mobility patterns in cities are dynamic and that analyses based on data aggregated over discrete time periods may not fully capture variability in mobility patterns over time. Likewise, population distributions and city structure may change rapidly in megacities and mobility data collected in these environments may become outdated relatively quickly. Lastly, by considering only trips between locations within the BMR and DSMA, we do not account for connectivity between these areas and more distant or outlying locations; connectivity with areas outside the BMR and DSMA is expected to not only drive epidemic dispersal into surrounding locations but also influence spatiotemporal epidemic dynamics within the central urban areas.

However, for multiple reasons, this method for estimating movement patterns is expected to capture important features of the mobility networks in each city, including features that are informative for modeling infectious disease dynamics. By using this approximation, and assuming the frequency of mobile phone use correlates to time spent in a given location, we capture the two most likely locations for each user, each day. Aggregated over millions of individual users, and averaged over an extended data collection period, we are able to obtain estimates for average connectivity between locations in a city based on large numbers of total observations. In addition, several intuitive findings from our analysis of the mobility networks specified by \mathcal{M}_{OD}^{wke} and \mathcal{M}_{OD}^{all} , including higher eigenvector centrality values at geographic points with higher expected connectivity (for example, city centers and transportation hubs), support our approach to estimating city-level mobility patterns from aggregated CDR data.

Multiple limitations of our modeling approach are important to consider. The stochastic model used in this study assumes that mixing is homogeneous in each compartment and assumes that co-location in the same node is an adequate proxy for person-person interaction. Also, this model does not account for potential differences in the average time spent in each location (“dwell time”), which is likely to vary across locations resulting in differences in average daily force of infection by location, nor does it account for geographic heterogeneity in age or household structure. Additional studies, comparing directly observed person-person interactions (for example, via Bluetooth handshake (Stopczynski et al., 2013) or RFID sensors (Cattuto et al., 2010)) to mobility trajectories estimated from CDR and other passively-collected mobile data sources, are needed to better understand the limits of CDR-based mobility estimates in this context. Lastly, we initialize our model with a homogeneously (and entirely) susceptible population, and as such this approach is poorly suited for understanding endemic infections where there is pre-existing immunity that may be heterogeneous between demographic groups and across geographic space.

In conclusion, this study describes previously unreported characteristics of mobility networks in Dhaka and Bangkok, and reports important differences in the trajectories of simulated epidemics propagated over these networks. Our findings support the continued development of passively-collected mobile user data as an important tool for understanding and predicting the *a priori* risk of epidemic propagation in different cities, and for planning disease control interventions that are predicated on understanding the spatial and temporal dynamics of epidemics at the city (including, for example, the distribution of public health resources for case detection and testing).

CRediT authorship contribution statement

Tyler S. Brown: Conceptualization, Data curation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Kenth Engø-Monsen:** Conceptualization, Methodology, Data curation, Formal analysis, Writing - original draft, Writing - review & editing, Supervision. **Mathew V. Kiang:** Conceptualization, Data curation, Writing - review & editing. **Ayesha S. Mahmud:** Conceptualization, Data curation, Writing - review & editing. **Richard J. Maude:** Conceptualization, Data curation, Writing - review & editing. **Caroline O. Buckee:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the United States National Institutes of Health (T32AI007061 to TSB and R35GM124715 to COB). This research was funded in whole or in part by the Wellcome Trust (Grant number 220211 to RJM).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.epidem.2021.100441>.

References

- Alexander, Lauren, Jiang, Shan, Murga, Mikel, González, Marta C., 2015. Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transp. Res. C* 58, 240–250. <http://dx.doi.org/10.1016/j.trc.2015.02.018>, Big Data in Transportation and Traffic Engineering.
- Bialek, S., Bowen, V., Chow, N., Curns, A., Gierke, R., Hall, A., Hughes, M., Pilishvili, T., Ritchey, M., et al., 2020. Geographic differences in COVID-19 cases, deaths, and incidence — United States, February 12– April 7, 2020. *MMWR Morb. Mortal. Wkly. Rep.* 69 (15), 465–471. <http://dx.doi.org/10.15585/mmwr.mm6915e4>.
- Brockmann, D., Helbing, D., 2013. The hidden geometry of complex, network-driven contagion phenomena. *Science* 342 (6164), 1337–1342.
- Canright, G.S., Engø-Monsen, K., 2006. Spreading on networks: A topographic view. *Complexus* 3 (1–3), 131–146.
- Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J., Vespignani, A., 2010. Dynamics of person-to-person interactions from distributed RFID sensor networks. In: Neylon, Cameron (Ed.), *PLoS ONE* 5 (7), e11596.
- Charaudeau, S., Pakdaman, K., Boëlle, P., 2014. Commuter mobility and the spread of infectious diseases: application to influenza in France. *PLoS One* 9 (1), e83002. <http://dx.doi.org/10.1371/journal.pone.0083002>.
- Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. *InterJournal, Complex Syst.* 1695 (5), 1–9. <http://igraph.sf.net>.
- Dalziel, B.D., Pourbohloul, B., Ellner, S.P., 2013. Human mobility patterns predict divergent epidemic dynamics among cities. *Proc. Biol. Sci.* 280 (1766), 20130763. <http://dx.doi.org/10.1098/rspb.2013.0763>.
- Heroy, S., 2020. Metropolitan-scale COVID-19 outbreaks: how similar are they? <https://arxiv.org/pdf/2004.01248.pdf>.
- Huffman, D.A., 1952. A method for the construction of minimum-redundancy codes. *Proc. IRE* 40 (9), 1098–1101. <http://dx.doi.org/10.1109/JRPROC.1952.273898>.
- Jowell, A., Zhou, B., Barry, M., 2017. The impact of megacities on health: preparing for a resilient future. *Lancet Planet. Health* 1 (5), e176–e178. [http://dx.doi.org/10.1016/s2542-5196\(17\)30080-3](http://dx.doi.org/10.1016/s2542-5196(17)30080-3).
- Kissler, S., Kishore, N., Prabhu, M., Goffman, D., Beilin, Y., Landau, R., Gyamfi-Bannerman, C., Bateman, B.T., Katz, D., Gal, J., Bianco, A., Stone, J., Larremore, D., Buckee, C.O., Grad, Y.H., 2020. Reductions in commuting mobility predict geographic differences in SARS-CoV-2 prevalence in New York city. <https://dash.harvard.edu/handle/1/42665370>.
- Kung, K.S., Greco, K., Sobolevsky, S., Ratti, C., 2014. Exploring universal patterns in human home-work commuting from mobile phone data. *PLoS One* 9 (6), e96180.
- Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J., 2020. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* 368 (6490), 489–493.

- Lloyd, C.T., Chamberlain, H., Kerr, D., Yetman, G., Pistoiesi, L., Stevens, F.R., Gaughan, A.E., Nieves, J.J., Hornby, G., MacManus, K., et al., 2019. Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data* 3 (2), 108–139.
- Long, Q., Tang, X., Shi, Q., Li, Q., Deng, H., Yuan, J., Hu, J., Xu, W., Zhang, Y., Lu, F., et al., 2020. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature Med.*
- Mamei, Marco, Bicochi, Nicola, Lippi, Marco, Mariani, Stefano, Zambonelli, Franco, 2019. Evaluating origin–destination matrices obtained from CDR data. *Sensors* 19 (20), 4470. <http://dx.doi.org/10.3390/s19204470>.
- Panigutti, Cecilia, Tizzoni, Michele, Bajardi, Paolo, Smoreda, Zbigniew, Colizza, Vittoria, 2017. Assessing the use of mobile phone data to describe recurrent mobility patterns in spatial epidemic models. *R. Soc. Open Sci.* 4 (5), 160950. <http://dx.doi.org/10.1098/rsos.160950>.
- Rosvall, M., Axelsson, D., Bergstrom, C.T., 2009. The map equation. *Eur. Phys. J. Spec. Top.* 178 (1), 13–23. <http://dx.doi.org/10.1140/epjst/e2010-01179-1>.
- Rosvall, M., Bergstrom, C.T., 2011. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. In: Rappallo, Fabio (Ed.), *PLoS ONE* 6 (4), e18209. <http://dx.doi.org/10.1371/journal.pone.0018209>.
- Stier, A.J., Berman, M.G., Bettencourt, L.M.A., 2020. COVID-19 attack rate increases with city size. <https://arxiv.org/pdf/2003.10376.pdf>.
- Stopczynski, A., Larsen, J.E., Lehmann, S., Dynowski, L., Fuentes, M., 2013. Participatory bluetooth sensing: A method for acquiring spatio-temporal data about participant mobility and interactions at large scale events. *IEEE*, Tizzoni, M., Bajardi, P., Decuyper, A., Kon Kam King, G., Schneider, C.M., Blondel, V., Smoreda, Z., González, M.C., Colizza, V., 2014. On the use of human mobility proxies for modeling epidemics. *PLoS Comput. Biol.* 10 (7), e1003716. <http://dx.doi.org/10.1371/journal.pcbi.1003716>.
- Tongsinoot, L., Muangsin, V., 2017. Exploring home and work locations in a city from mobile phone data. In: 2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems (HPCC/SmartCity/DSS). pp. 123–129. <http://dx.doi.org/10.1109/HPCC-SmartCity-DSS.2017.16>.
- Viboud, C., Bjornstad, O.N., Smith, D.L., Simonsen, L., Miller, M.A., Grenfell, B.T., 2006. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* 312 (5772), 447–451.
- Volpati, V., Barthelemy, M., 2018. The spatial organization of the population density in cities. <https://arxiv.org/pdf/1804.00855.pdf>.
- Wesolowski, A., Metcalf, C.J.E., Eagle, N., Kombich, Janeth, Grenfell, Bryan T., Bjornstad, Ottar N., Lessler, Justin, Tatem, Andrew J., Buckee, Caroline O., 2015a. Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *PNAS; Proc. Natl. Acad. Sci.* 112 (35), 11114–11119.
- Wesolowski, Amy, Qureshi, Taimur, Boni, Maciej F., Sundsoy, Pål Roe, Johansson, Michael A., Rasheed, Syed Basit, Engø-Monsen, Kenth, Buckee, Caroline O., 2015b. Impact of human mobility on the emergence of dengue epidemics in Pakistan. *PNAS; Proc. Natl. Acad. Sci.* 112 (38), 11887–11892.