7 Supplemental Methods

7.1 Characterizing the spatial distribution of population distribution in cities

We calculate measures of dispersion for population density distributions (Gini coefficient, entropy, and relative standard deviation) using standard approaches described in [10]. We calculate the "spatial dispersion index" ("spreading index" in [10]) as follows:

$$\eta_\alpha\left(P^*\right) = \frac{\frac{1}{n_\alpha(P^*)}\sum_{i,j} d(i,j)\Theta\left(P_i - P^*\right)\Theta\left(P_j - P^*\right)}{\frac{1}{n_\alpha}\sum_{i,j} d(i,j)} \tag{6}$$

where P^* is a threshold value above which cells in the area map are identified as having high population density. We calculated the dispersal index using two different values for P^*: (1) the mean population density and, following [cite], (2) the "Loubar" value derived from the Lorenz curve of population counts. To address possible bias introduced by the shape and size of the study area in each city, we calculated the above metrics over concentric circular areas of radius r centered on the map location with the highest population density in each city and report each metric across a range of r values

7.2 CDR data collection and processing

Anonymized CDR data was collected from 4415 individual network towers in the Bangkok Metropolitan Region between 1 August and 19 October 2017. We excluded two time periods from the dataset (12-13 August and 13-15 October) corresponding to two major national holidays during which total subscriber counts transiently decreased, leaving a total of 81 days in the dataset. In Dhaka Statistical Metropolitan Area, CDR data was collected from 1560 network towers. The entire data collection period in Dhaka spans 183 days from 1 April to 30 September 2017. The total number of subscribers in the Dhaka dataset increased from 22.6 million to 31.5 million during this data collection period. To minimize any potential bias introduced by this increase, and to match the duration of the data collection period of the Bangkok dataset, we used only the first 81 days of data in Dhaka dataset. Three towers in Dhaka exhibited unusual daily subscriber counts (which progressively increased during the study period) were omitted from analysis. We also identified a small number of towers ($< 1\%$ in both cities) that appear to have been activated during the data collection period. Specifically, these towers were found to have zero total subscribers until a discrete day, and from that point forward had large, non-zero subscriber numbers. We calculated the mean number of trips originating from these specific towers using only on the non-zero, post-activation days in each dataset.

As described in *Methods*, we consider two origin-destination matrices, $\mathcal{M}_{\mathrm{OD}}^{\mathrm{all}}$ and $\mathcal{M}_{\mathrm{OD}}^{\mathrm{wke}}$ specifying connectivity between nodes in the Dhaka and Bangkok city-level mobility networks. Mantel testing, performed using the untransformed origin-destination matrices as "similarility' rather than "distance' matrices, indicated statistically significant correlation between between $\mathcal{M}_{\mathrm{OD}}^{\mathrm{all}}$ and $\mathcal{M}_{\mathrm{OD}}^{\mathrm{wke}}$ for Dhaka and Bangkok ($p < 0.005$ for all comparisons with n=1000 permutations). Community structures estimated using either $\mathcal{M}_{\mathrm{OD}}^{\mathrm{all}}$ or $\mathcal{M}_{\mathrm{OD}}^{\mathrm{wke}}$ are broadly similar: similarity and dissimilarity metrics (including normalized mutual information, Rand Index, and variation of information) for community structures estimated using the two different origin-destination matrices are similar those estimated for multiple Infomap estimates using the same origin-destination matrix, i.e. the magnitude variation between community structures estimated $\mathcal{M}_{\mathrm{OD}}^{\mathrm{all}}$ or $\mathcal{M}_{\mathrm{OD}}^{\mathrm{wke}}$ is similar to the variation inherent to all Infomap-based estimates of community structure (Figures S2 & S3). The geospatial and statistical distribution of eigenvector centrality values estimated from $\mathcal{M}_{\mathrm{OD}}^{\mathrm{all}}$ and $\mathcal{M}_{\mathrm{OD}}^{\mathrm{wke}}$ are broadly similar (Figures S4 & S5).

Lastly, we evaluate whether the observed differences in estimated community structures between Dhaka and Bangkok could be attributable to unmeasured connectivity between nodes within city-level commuting networks and unobserved nodes outside each data catchment area (i.e. the BMR and DMSA shown in Figure 2 A). Specifically, to examine whether the geographically-unconstrained communities observed in Dhaka result from connectivity with unobserved nodes outside the DMSA, we test whether the geographic contiguity of the Bangkok community structure is lost if increasing numbers of nodes are excluded from the origin-destination matrix, effectively creating an increasing number of unobserved nodes. Estimated community structures in Bangkok retain strong geographic contiguity (measured as the proportion of a node's immediately adjacent neighbors that belong to the same community and reported as the average of this value for all nodes) even when the network is restricted to nodes within 15 km of the city center (Figure S7).

8 Supplemental Figures

Figure S1: Data collection areas for the Bangkok Metropolitan Region (BMR, left) and the Dhaka Metropolitan Statistical Area (DMSA, right). Mobile phone-associated movement data was collected for mobile network within the BMR and DMSA (demarcated in black).

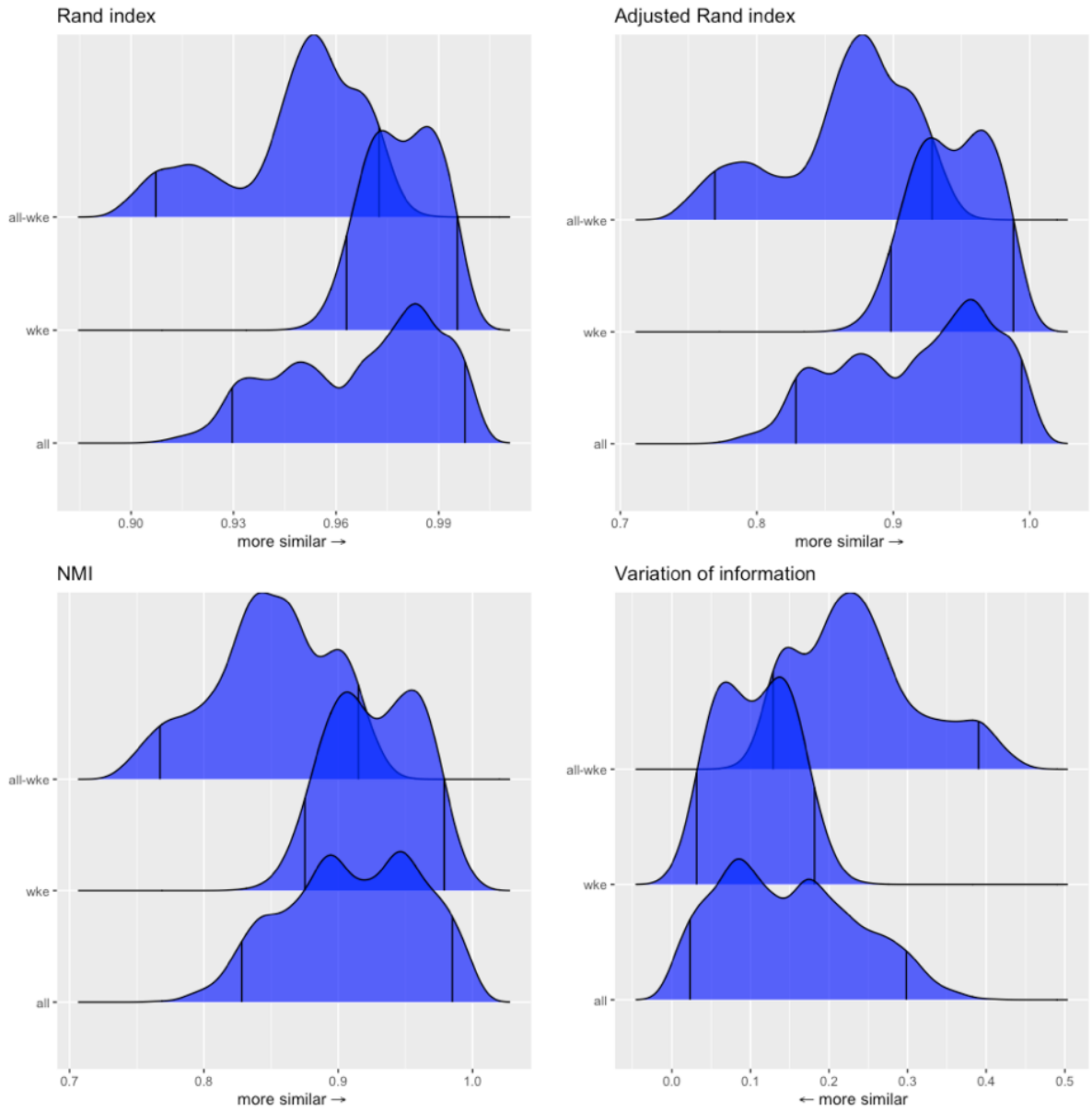

Figure S2: Infomap-generated community structures estimated using the same or different data sources in Dhaka. Distributions labeled "all" in each panel show the normalized mutual information (NMI), Rand index, or variation of infomation values for 1225 unique pairwise comparisons between n=50 unique Infomap-generated community structures estimated using $\mathcal{M}_{\mathrm{OD}}^{\mathrm{all}}$. Likewise, distributions labeled "wke" show the distribution of these metrics for pairwise comparisons between estimates based on $\mathcal{M}_{\mathrm{OD}}^{\mathrm{wke}}$. Distributions labeled "all-wke" show NMI, Rand index, and variation of information values for 2500 unique pairwise comparisons between n=50 Infomap community structure estimates using $\mathcal{M}_{\mathrm{OD}}^{\mathrm{all}}$ and n=50 estimates for $\mathcal{M}_{\mathrm{OD}}^{\mathrm{wke}}$.

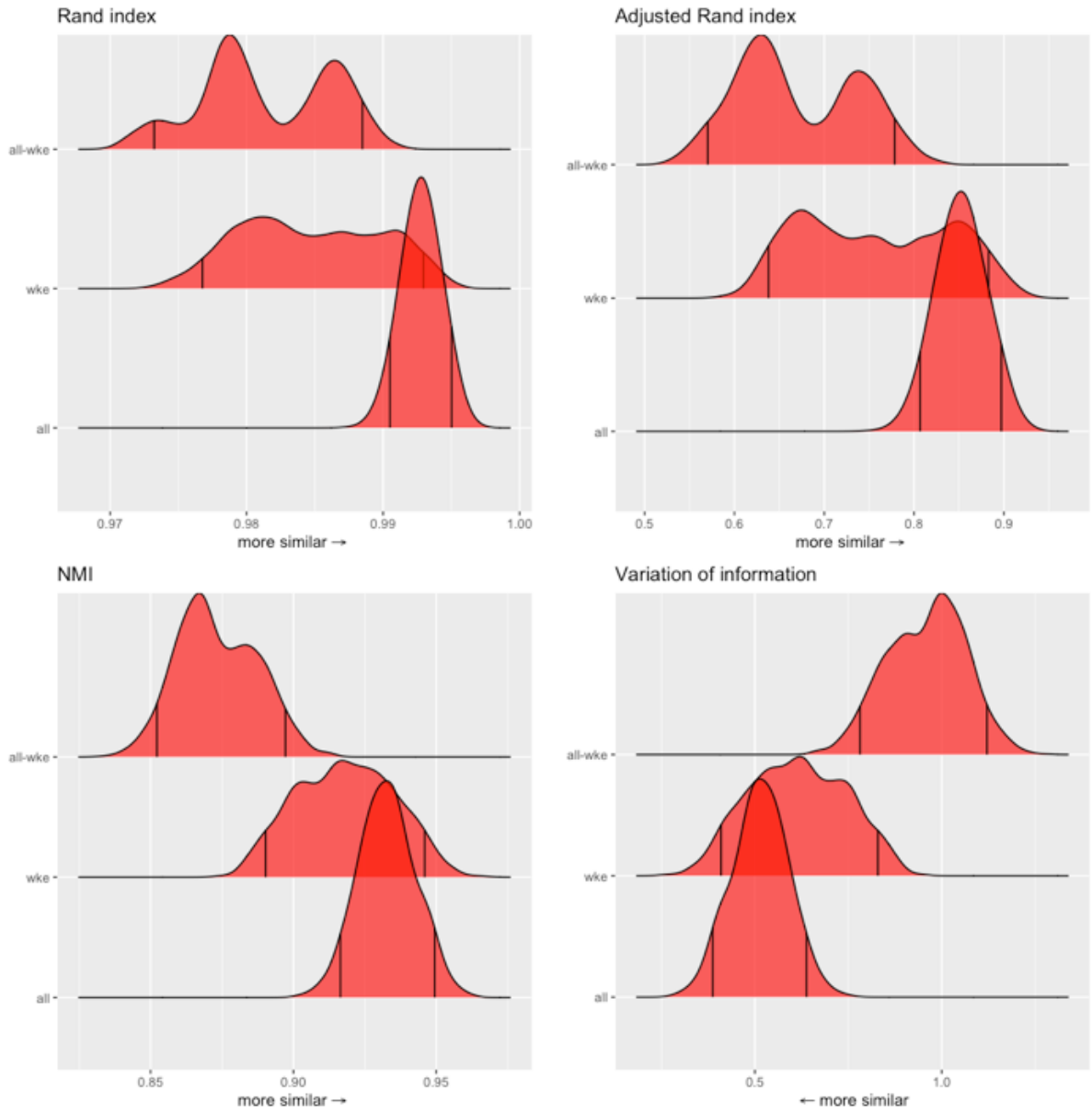

Figure S3: Infomap-generated community structures estimated using the same or different data sources in Dhaka. Distributions labeled "all" in each panel show the normalized mutual information (NMI), Rand index, or variation of infomation values for 1225 unique pairwise comparisons between n=50 unique Infomap-generated community structures estimated using $\mathcal{M}_{\mathrm{OD}}^{\mathrm{all}}$. Likewise, distributions labeled "wke" show the distribution of these metrics for pairwise comparisons between estimates based on $\mathcal{M}_{\mathrm{OD}}^{\mathrm{wke}}$. Distributions labeled "all-wke" show NMI, Rand index, and variation of information values for 2500 unique pairwise comparisons between n=50 Infomap community structure estimates using $\mathcal{M}_{\mathrm{OD}}^{\mathrm{all}}$ and n=50 estimates for $\mathcal{M}_{\mathrm{OD}}^{\mathrm{wke}}$.

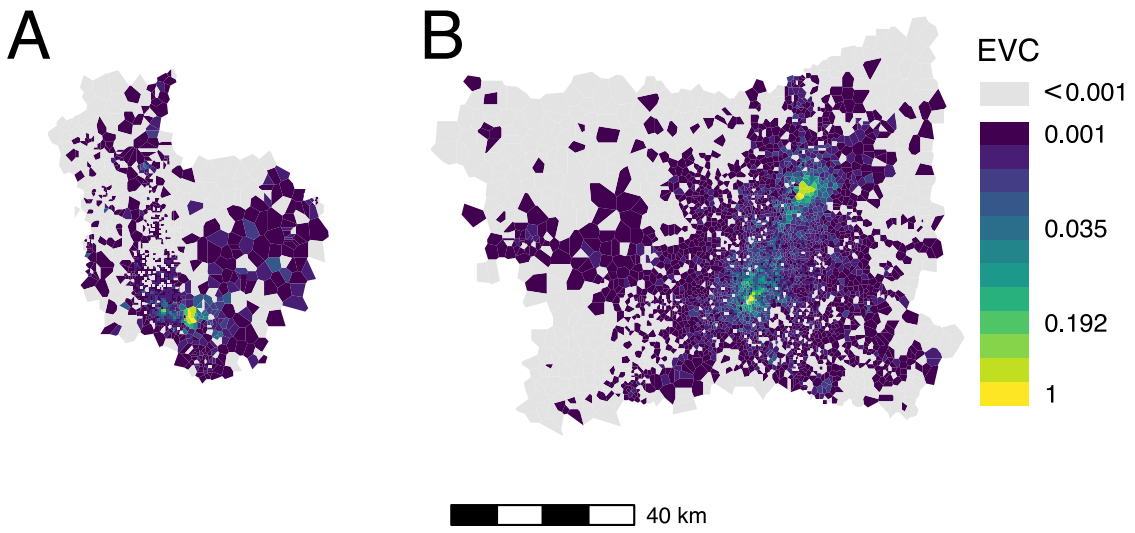

Figure S4: Eigenvector centrality by node for (A) Dhaka and (B) Bangkok using $\mathcal{M}_{\text{OD}}^{\text{wke}}$ to specify the commuter network.

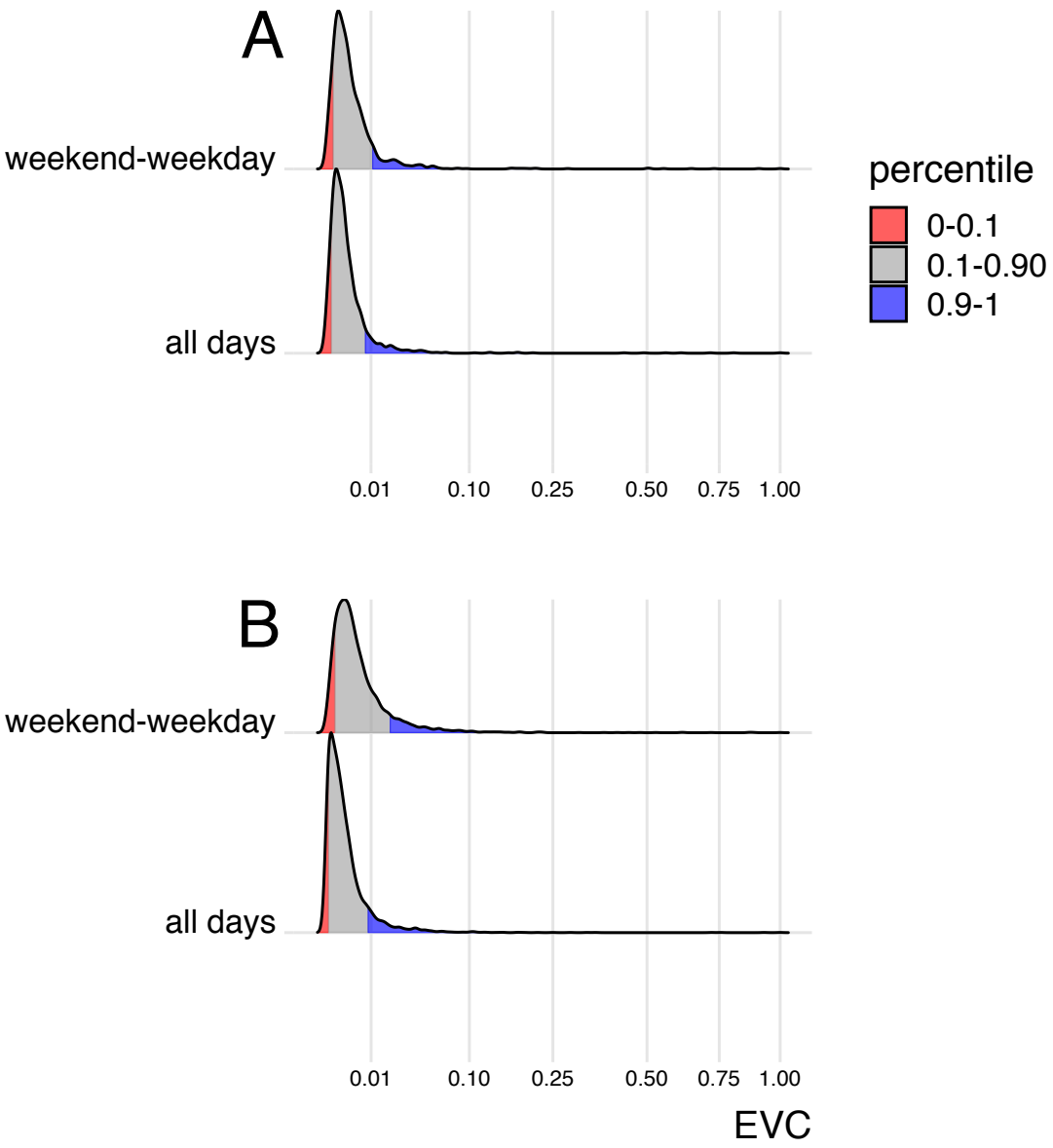

Figure S5: Distribution of eigenvector centrality values for Dhaka (A) and Bangkok (B), for commuter mobility networks specified by $\mathcal{M}_{\mathrm{OD}}^{\mathrm{all}}$ ("all days") and $\mathcal{M}_{\mathrm{OD}}^{\mathrm{wke}}$ ("weekend-weekday")

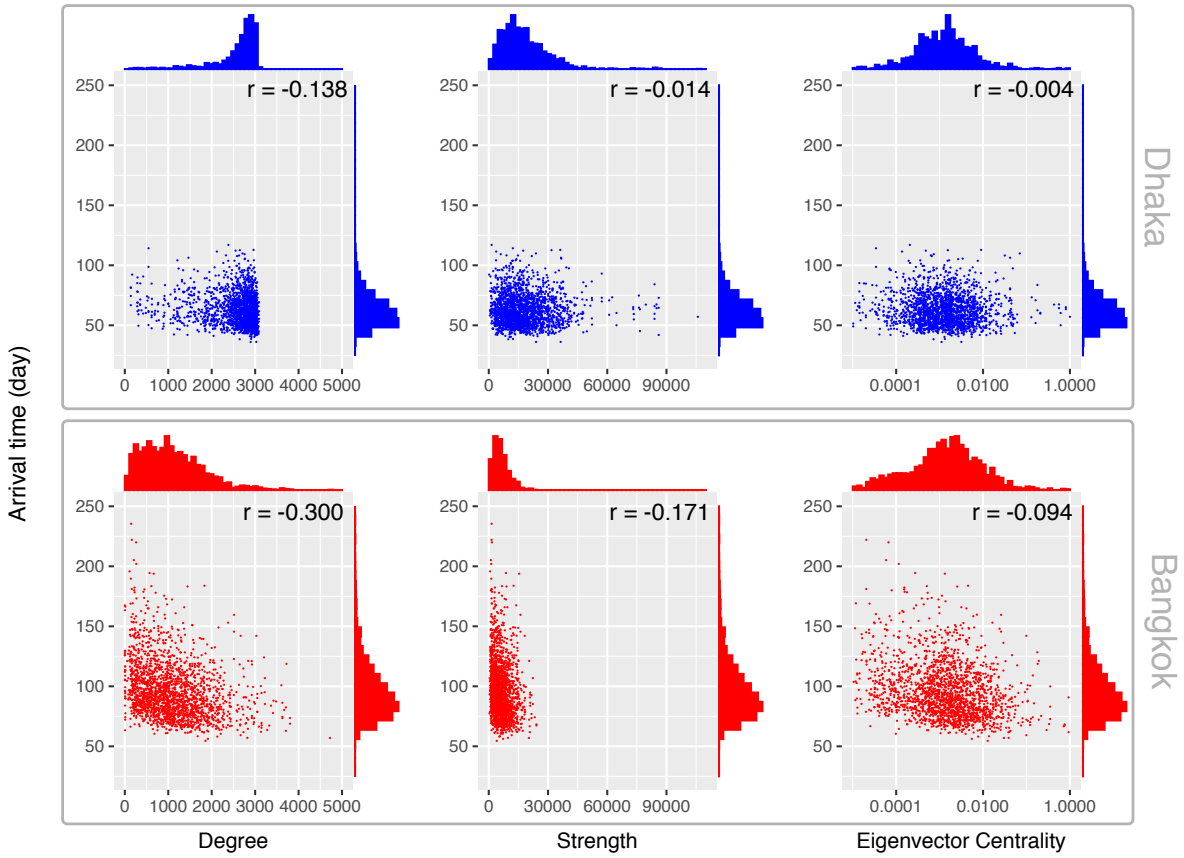

Figure S6: Origin node properties versus mean epidemic arrival time for Dhaka (blue, top panel) and Bangkok (red, bottom panel). Arrival time is calculated as the mean epidemic arrival time across all nodes for a given simulated epidemic. r: Pearson's correlation coefficient.

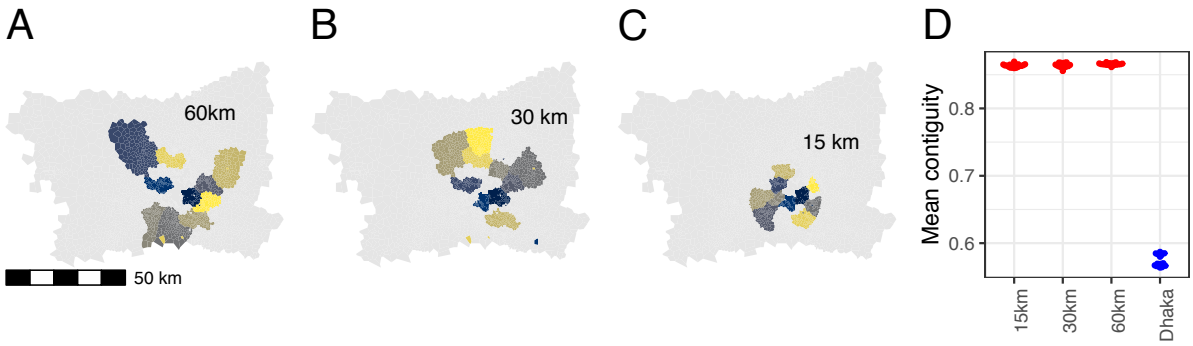

Figure S7: Example Infomap results, showing ten largest communities (by number of nodes), for Bangkok commuting network restricted to nodes within 60 km (A), 30 km (B), and 15 km (C) of the city center. Mean node-wise community contiguity (measured as the proportion of a node's nearest neighboring nodes that belong to the same community) is shown for 20 independent Infomap solutions each for nodes within 60 km (A), 30 km (B), and 15 km (C) of the city center, compared to the node-wise mean contiguity for the entire Dhaka network.

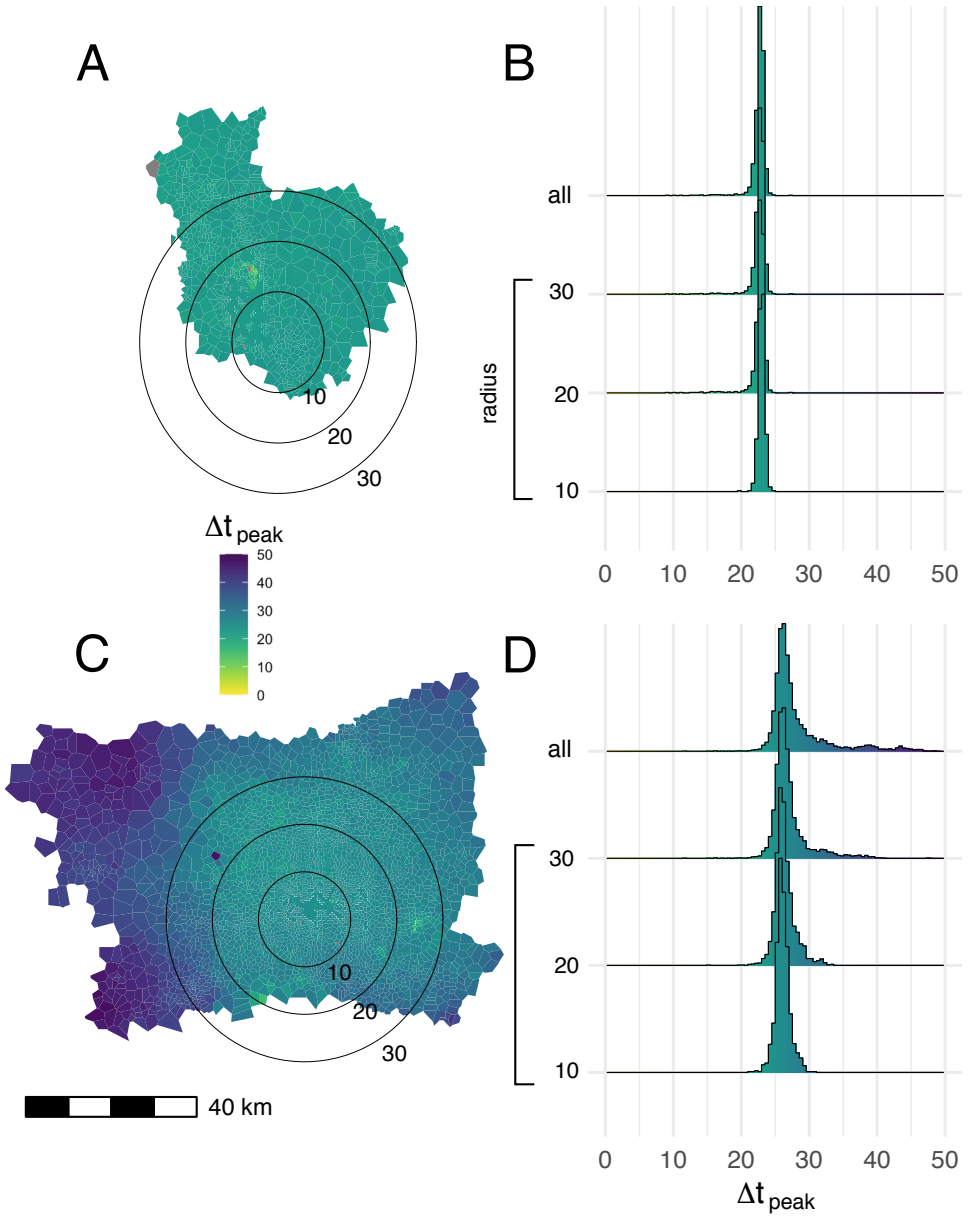

Figure S8: Mean time to epidemic peak for each node in the commuting network. Map polygons are colored according to the time lag in days between the time to epidemic peak in a given node and the earliest peaking node on the map (A: Dhaka, C: Bangkok). Results from 1,000 independent SEIR simulations, using parameters drawn via Latin hypercube sampling (as described *Methods*), are shown. The seed node for each simulation is drawn from a uniform distribution, where the probability that a given node is chosen as the seed is equal across all nodes. 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)

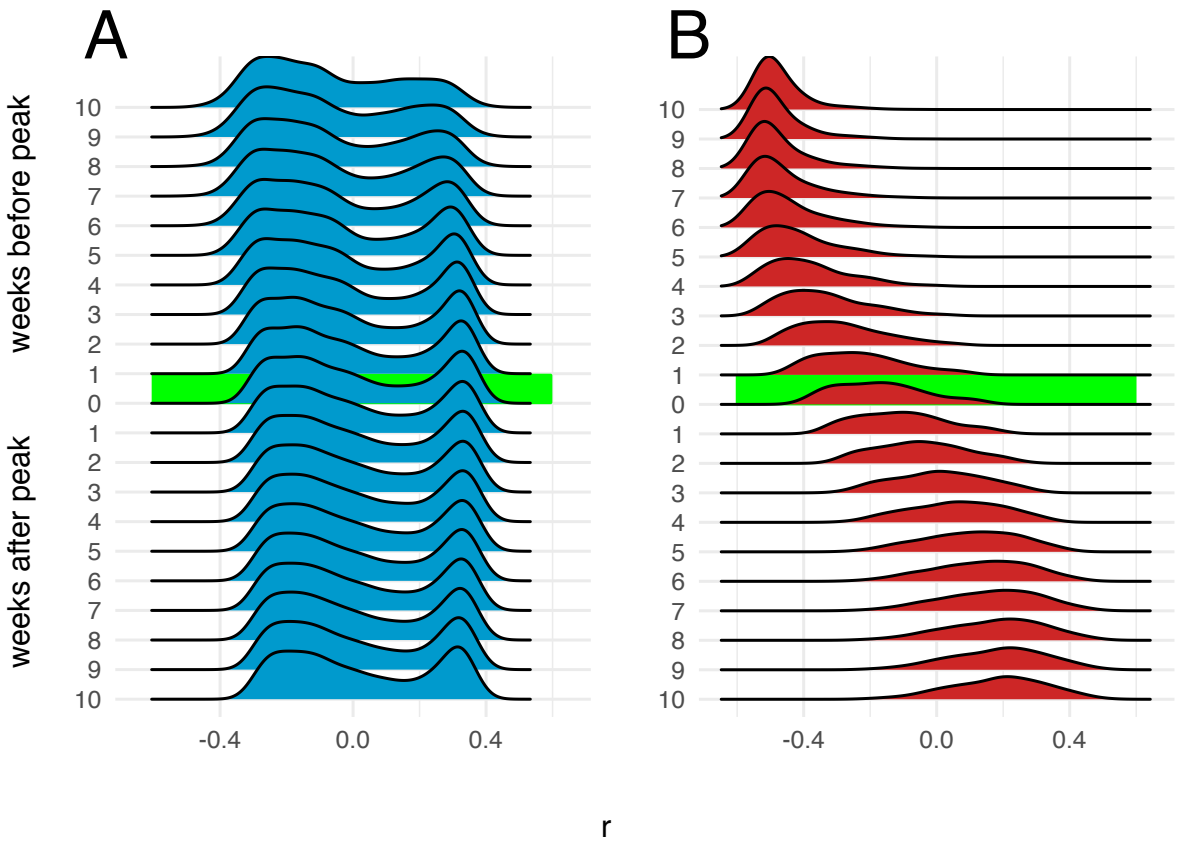

Figure S9: Correlation between number of infections and distance from epidemic origin node over time. Distributions of Pearson's coefficients for correlation between number of infections in each node and nodes' respective distances in km from the origin node, obtained from 1,000 independent SEIR simulations, are shown at different time points during simulated epidemics. The origin node for each simulation is chosen from a uniform distribution. A: Dhaka, B: Bangkok. Distribution for Pearson's correlation coefficient at the epidemic peak are highlighted.

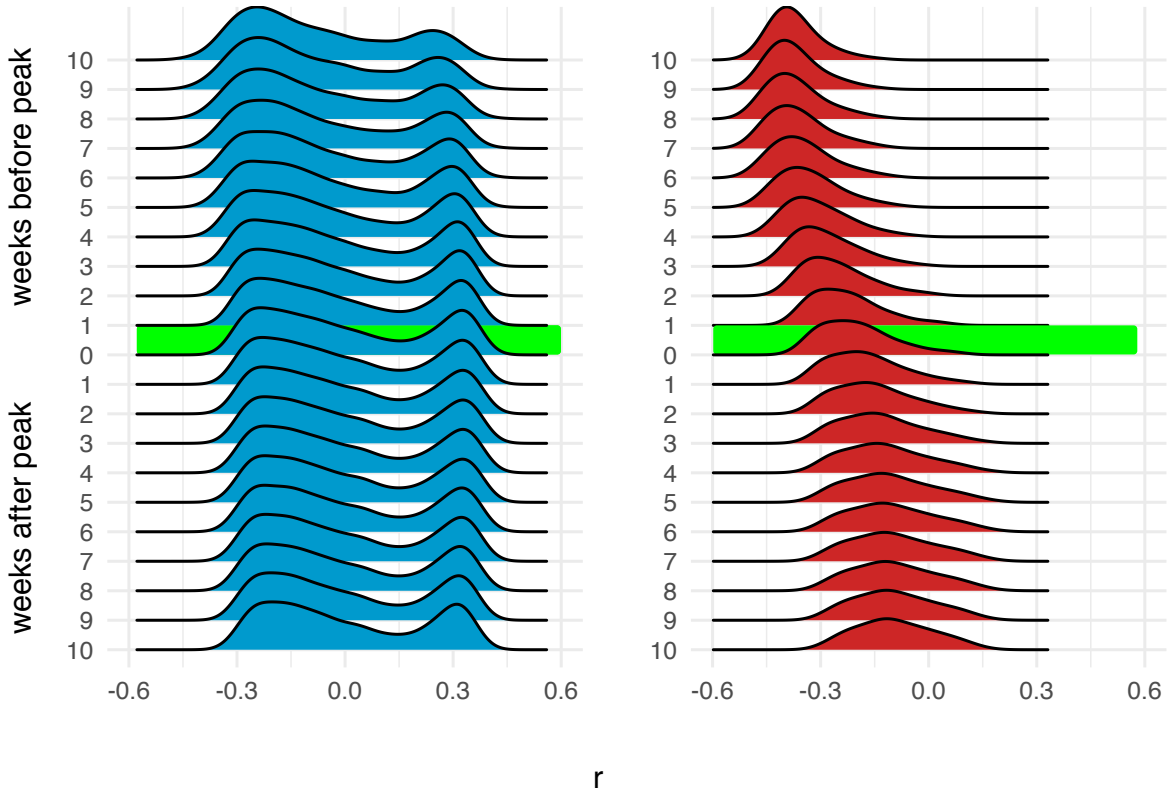

Figure S10: Correlation between number of infections and distance from epidemic origin node over time, for simulated epidemics propogated on $\mathcal{M}_{\mathrm{OD}}^{\mathrm{wke}}$. Distributions of Pearson's coefficients for correlation between number of infections in each node and nodes' respective distances in km from the origin node, obtained from 1,000 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.

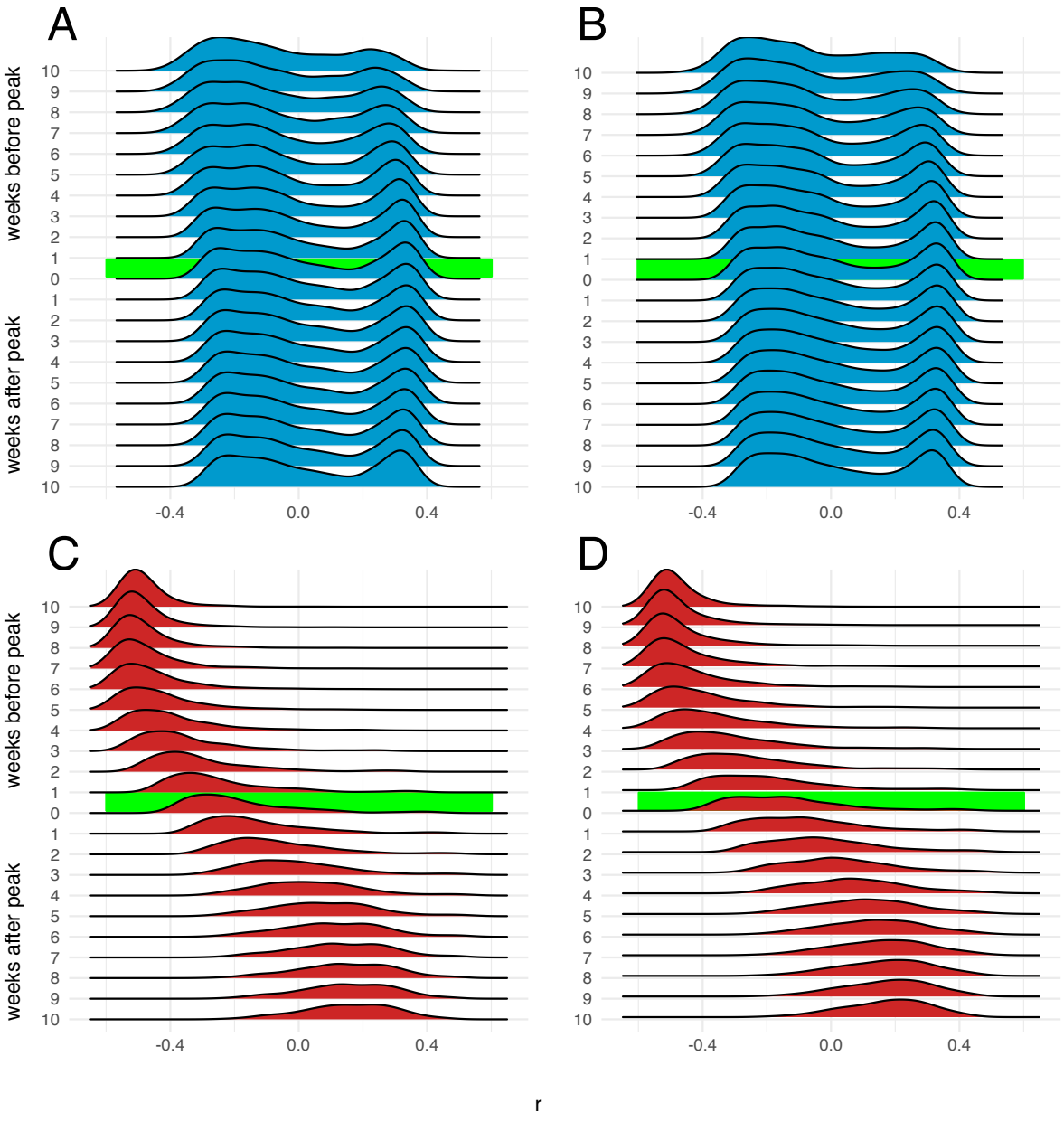

Figure S11: Correlation between number of infections and distance from epidemic origin node over time, restricted to nodes 60 km of the origin node. Distributions of Pearson's coefficients for correlation between number of infections in each node and nodes' respective distances in km from the origin node, obtained from 1,000 independent SEIR simulations, are shown at different time points during simulated epidemics. Results for Dhaka are show in panels A (origin node randomly selected from population-weighted distribution) and B (origin node randomly selected from uniform distribution), and results for Bangkok are show in panels C and D (with origin nodes selected from population-weighted and uniform distributions, respectively). Distribution for Pearson's correlation coefficient at the epidemic peak are highlighted.

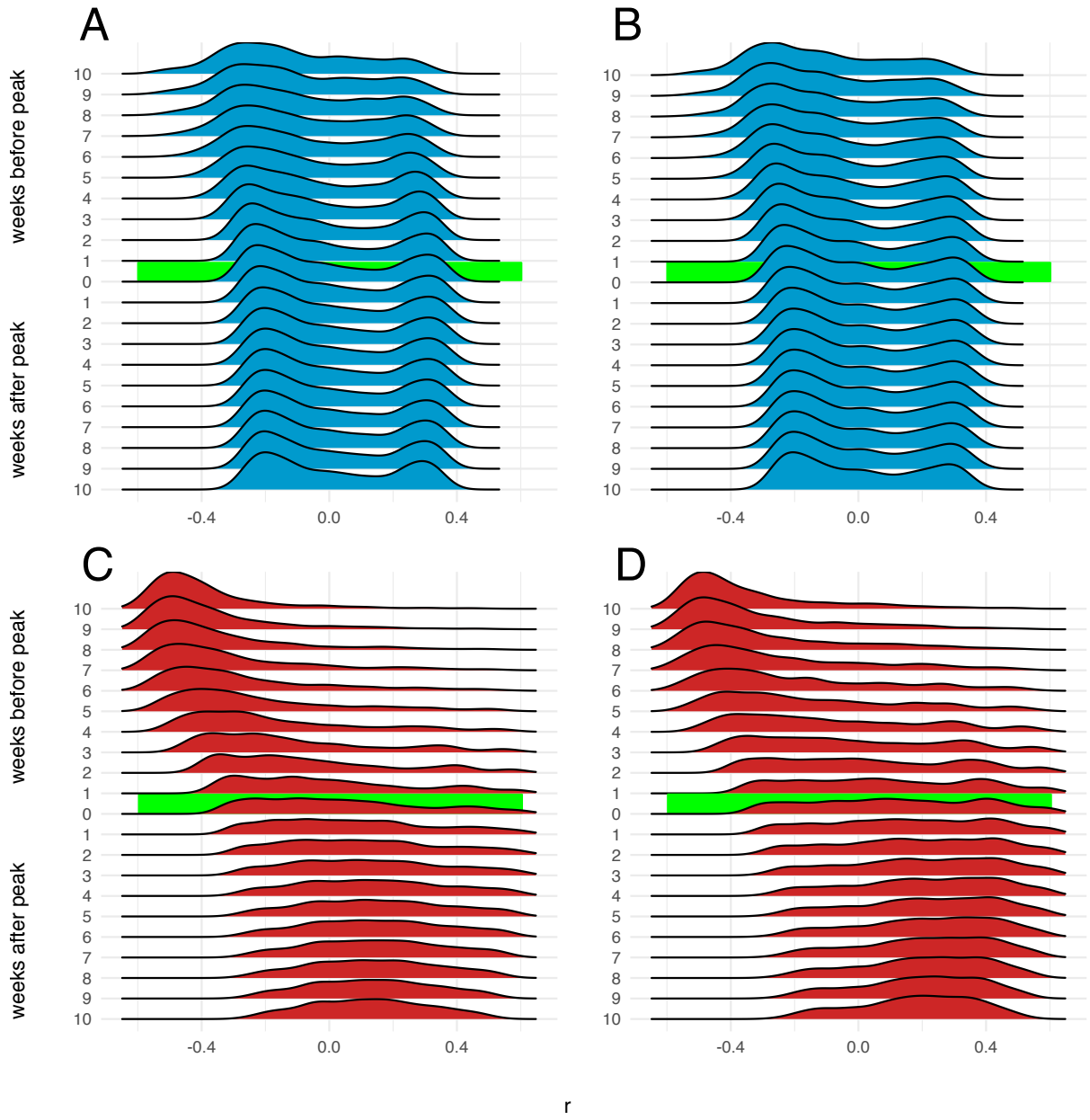

Figure S12: Correlation between number of infections and distance from epidemic origin node over time, restricted to nodes 30 km of the origin node. Distributions of Pearson's coefficients for correlation between number of infections in each node and nodes' respective distances in km from the origin node, obtained from 1,000 independent SEIR simulations, are shown at different time points during simulated epidemics. Results for Bangkok are show in panels A (origin node randomly selected from population-weighted distribution) and B (origin node randomly selected from uniform distribution), and results for Dhaka are show in panels C and D (with origin nodes selected from population-weighted and uniform distributions, respectively).

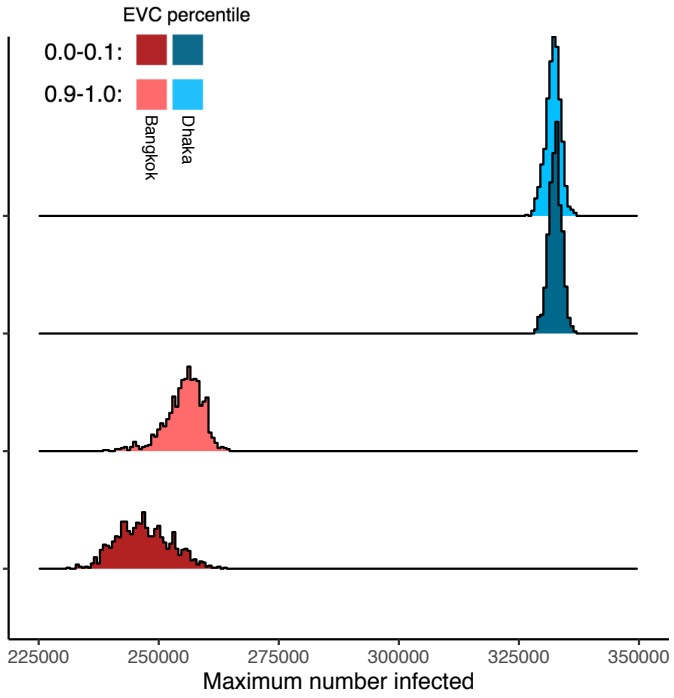

Figure S13: Maximum epidemic size by city and origin node eigenvector centrality. Distributions show the number of infected individuals at the epidemic peak (i.e. at t_{peak}) for 1,000 independent simulations originated at nodes with high (tenth decile) or low (first decile) eigenvector centrality.

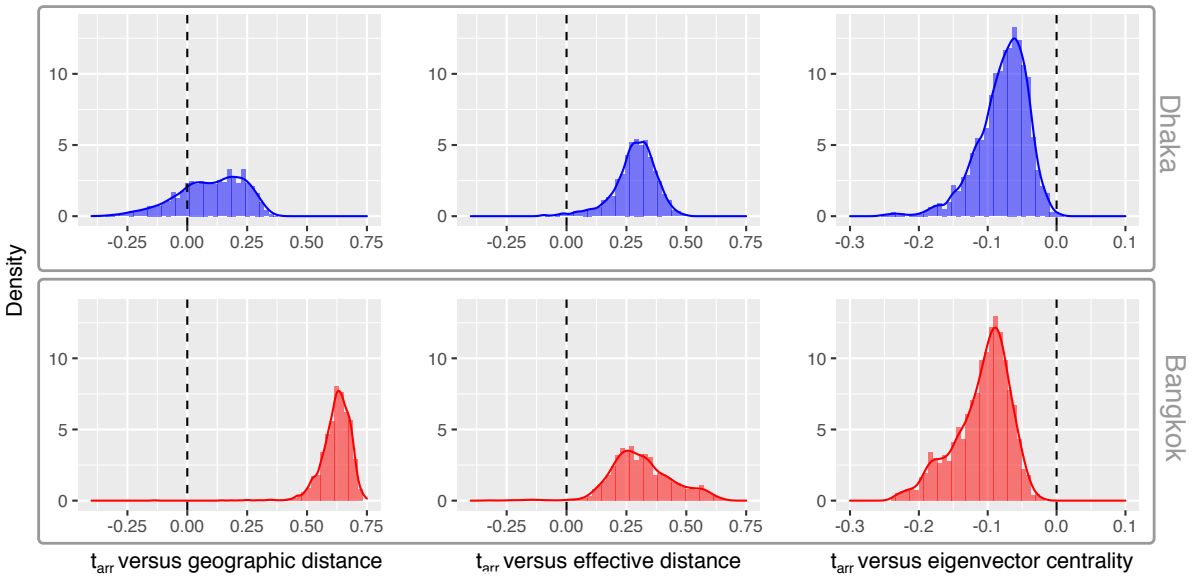

Figure S14: Distance from origin node and eigenvector centrality versus epidemic arrival time. Distributions of Pearson's correlation coefficients (r) for each variable (geographic distance from origin node, network effective distance from the origin node ([13], and eigenvector centrality) versus the epidemic arrival time for each node are shown for Bangkok and Dhaka. These distributions are generated as follows: for each of 1000 independent simulation replicates, each seeded at a single randomly-selected node, we calculate the geographic and effective network distance between the origin node and each node in the network, and then a calculate r for the correlation between each of these variables and the time of the first infection in each node (arrival time, t_{arr}) given by the simulation results. Similarly, we calculate r for the correlation between each non-origin node's eigenvector centrality and its epidemic arrival time. Each distribution is generated from 1000 r values.