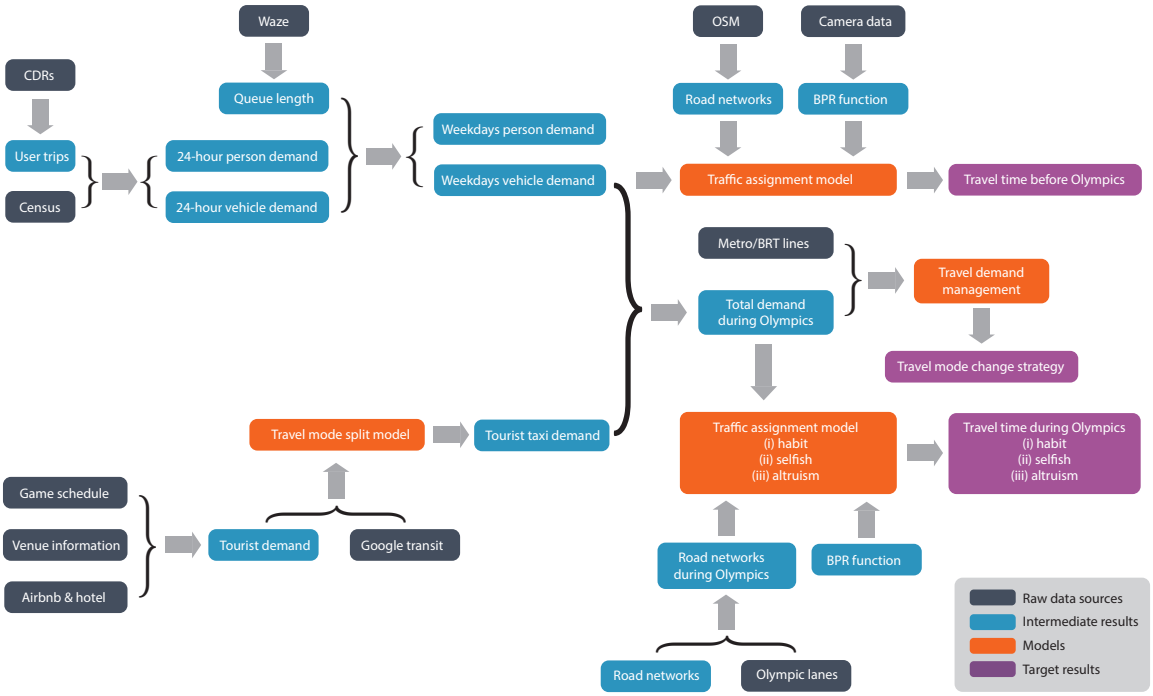
JOURNAL OF THE ROYAL SOCIETY INTERFACE

Collective benefits in traffic during mega events via the use of information technologies

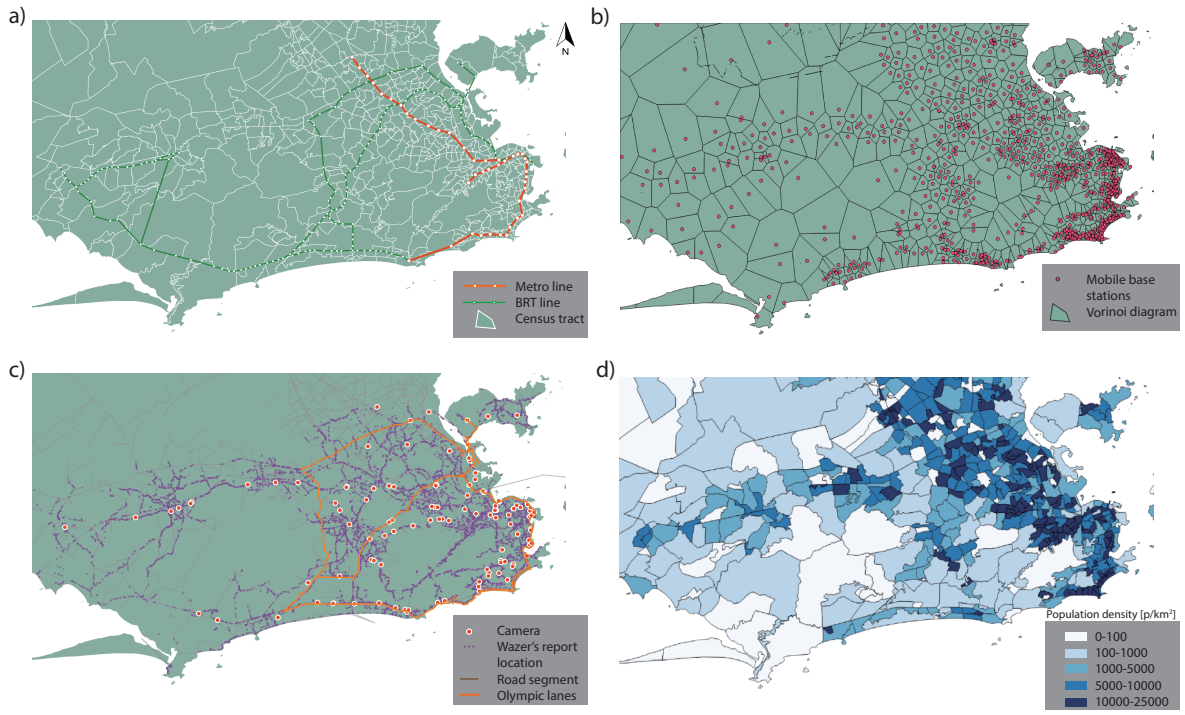
Supplementary Information

Yanyan Xu, Marta González
Department of Civil and Environmental Engineering
Massachusetts Institute of Technology

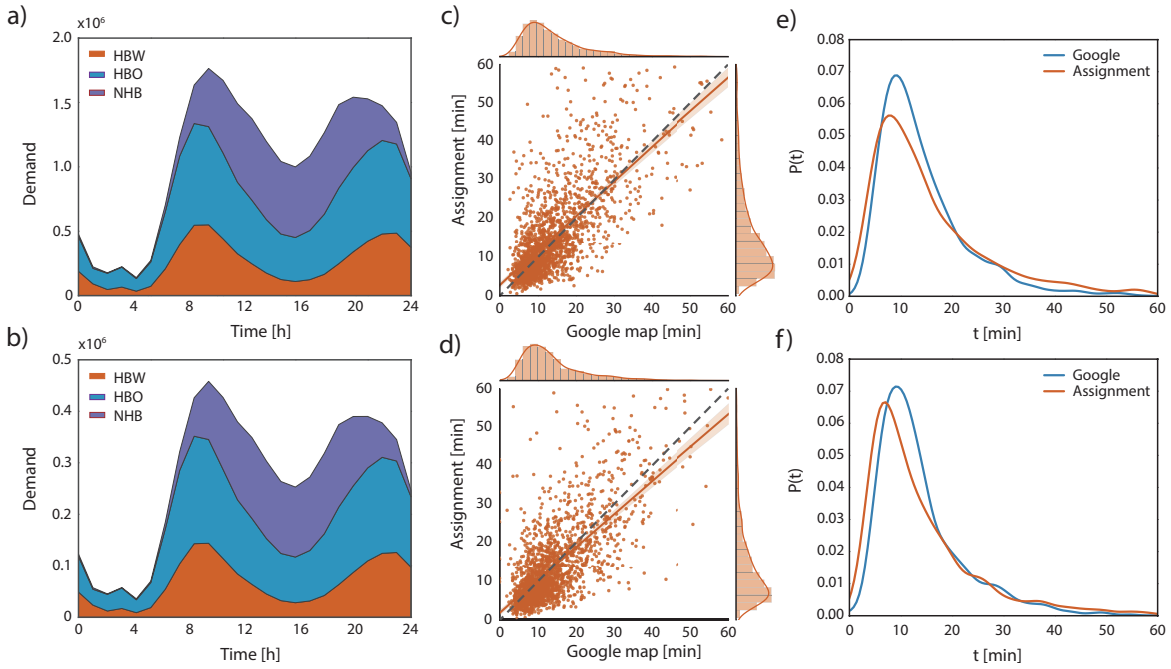
March 11, 2017



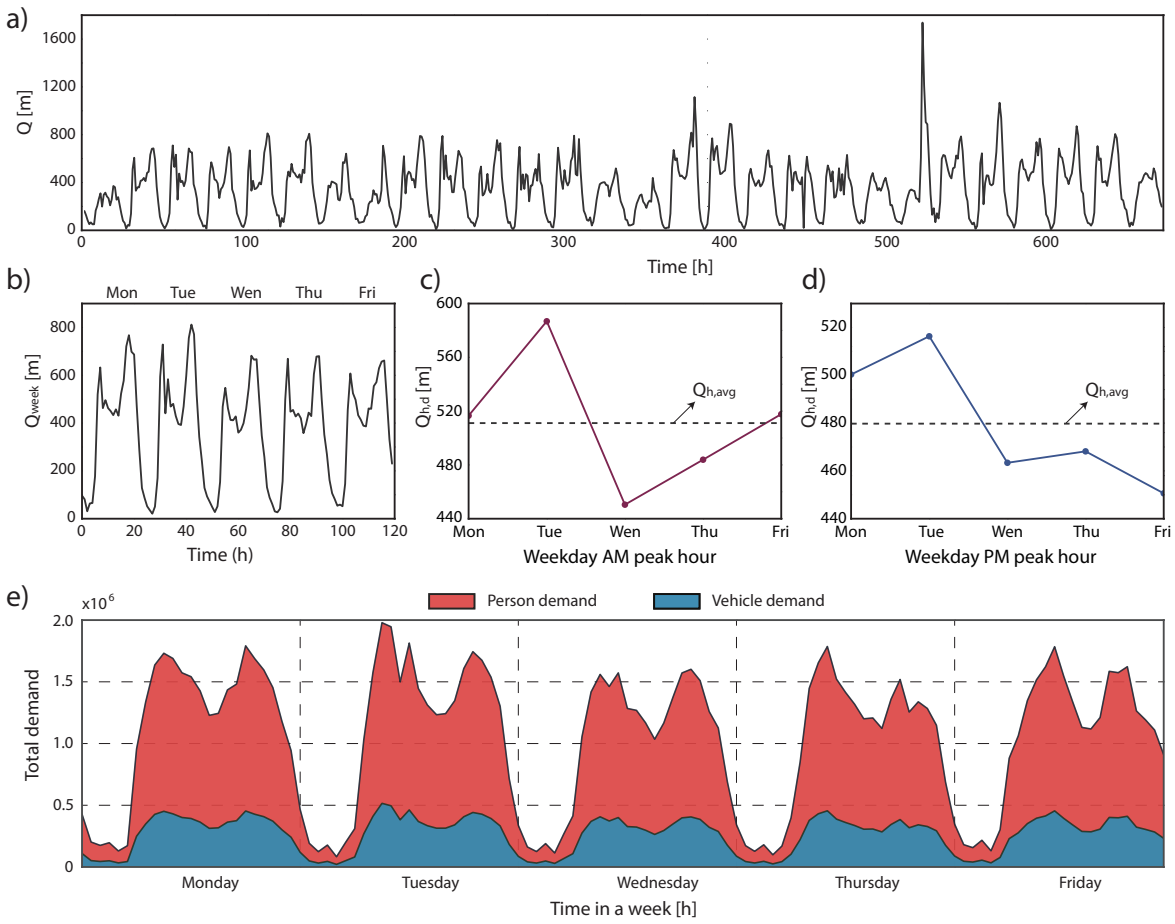
Supplementary Figure 1: The diagram of data and analysis pipeline. It shows the data resources and how they were used, the models used or developed in our work, the intermediate results, and the target results we obtained.



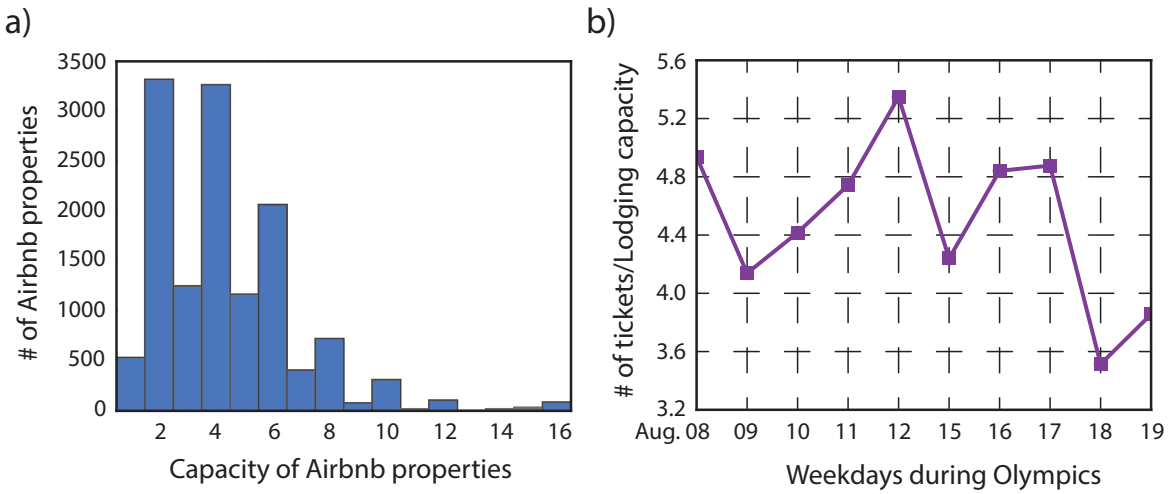
Supplementary Figure 2: Census tract, mobile base stations, traffic road network, the distribution of camera data, Wazers' report locations, and population density in the municipality area of Rio. (a) The census tracts, metro lines, and BRT lines in the municipality area of Rio. There are 514 tract, 5 BRT lines, and 4 metro lines in this area. (b) The distribution of mobile base stations and the Voronoi diagram in the municipality area of Rio. The number of towers in the whole area of Rio is 1422, and 787 of them are in the municipality area. (c) The road network extracted from OpenStreetMap, the locations of traffic surveillance cameras in our datasets, and the samples of traffic jam locations reported by the Wazers. (d) Population density (person/km^2) in spatial in the municipality area of Rio. The central area and the Governador island are two main slums in Rio, where there are few Airbnb properties. People from these areas probably do not go the Olympic games.



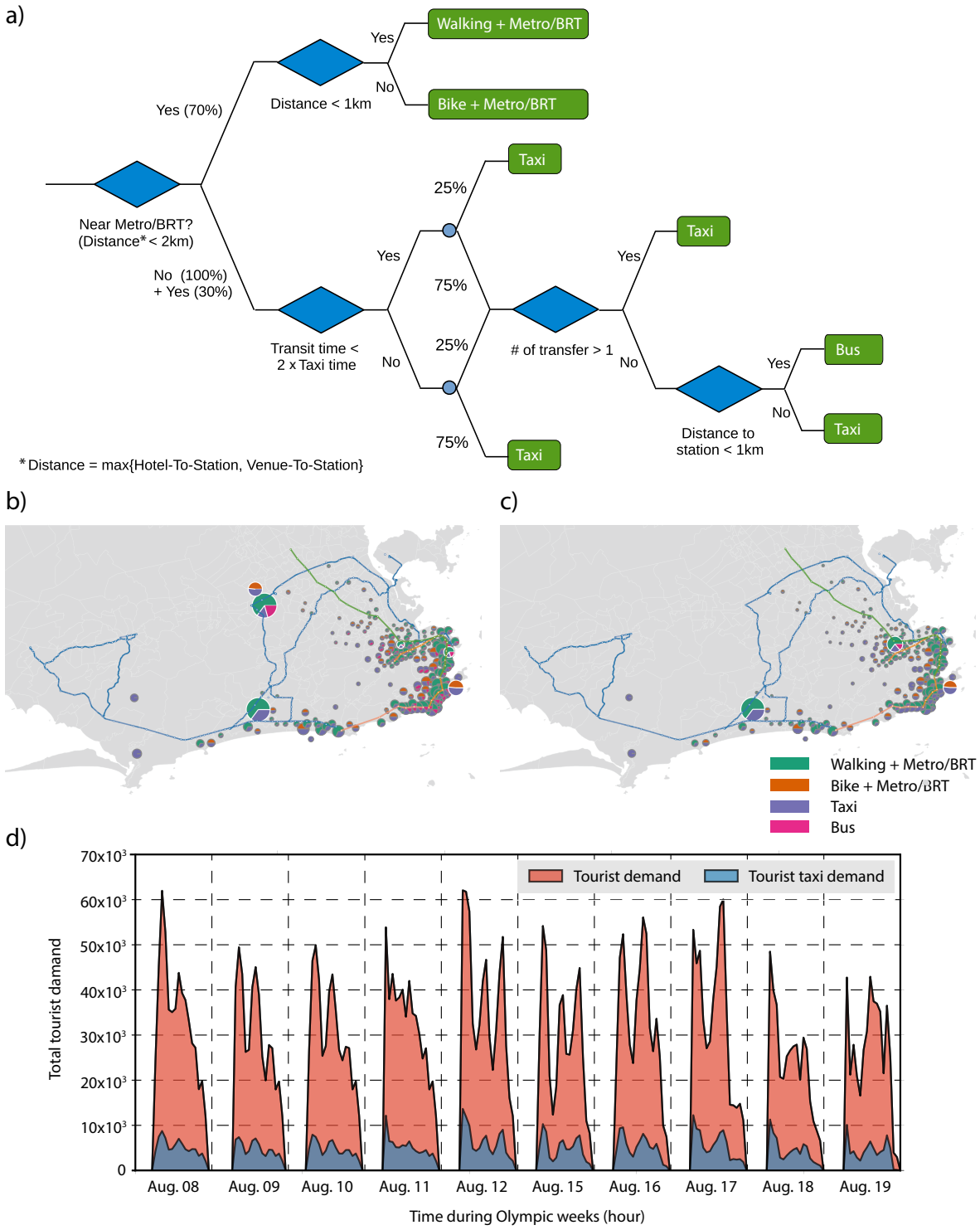
Supplementary Figure 3: Estimated travel demand and travel time validation with Google map before Olympics. (a) The stacked plot of the estimated person travel demand with different trip properties (e.g. home-based-work, home-based-other, and non-home-based) using CDRs and census data during 24 hours. (b) The estimated vehicle travel demand with different trip properties during 24 hours, which is the seed information to the Waze data to fluctuate the travel demand each of the different simulation days. (c) The scatter plot of the estimated travel time of top 5000 commuting trips versus Google map during the morning peak hour. Most scatters are distributed near the gray diagonal. Some commuting times are underestimated or overestimated compared with Google map. (d) The comparison of the distribution of estimated travel time versus Google map. The estimated commuting time has a similar distribution to the travel time from Google map. (e-f) The scatter and distribution comparison of estimated travel time and Google map during the evening peak hour.



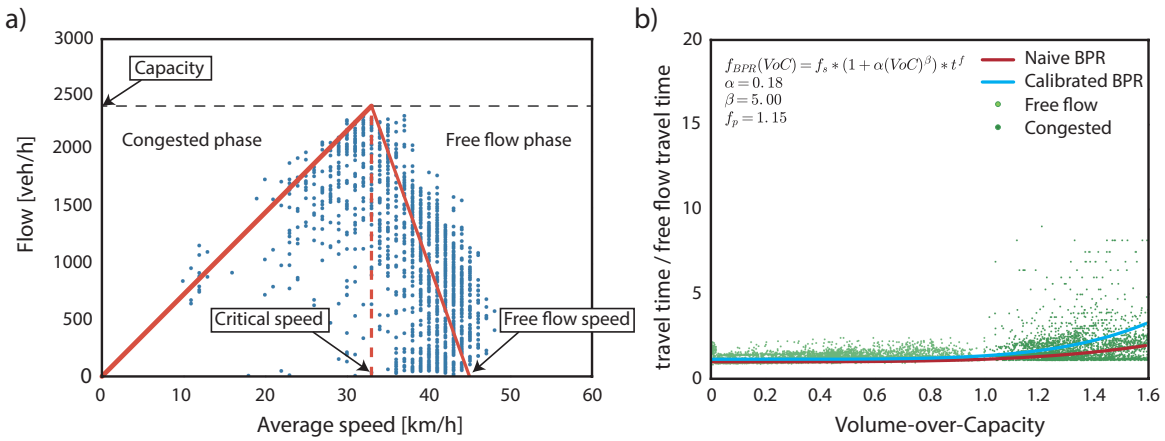
Supplementary Figure 4: Extending the average travel demand in one day to fluctuating demand on 5 weekdays using Waze data. (a) The average queue length in meters Q[m] per hour is reported by Wazers during one month. The time series plot starts on a Sunday. The queue length shows evident daily and weekly periodicity, especially on the weekdays. (b) The hourly average queue length on 5 weekdays, Q_{week}, is obtained by averaging the data in the same hour on same weekday. (c) Illustration of the hourly queue length $Q_{h,d}$ in hour h and day d for different weekdays, from Monday to Friday. The dashed line indicates the average value of $Q_{h,d}$ in that hour. (d) Illustration of $Q_{h,d}$ in the evening peak hour in 5 weekdays. (e) The fluctuating person and vehicle travel demands on 5 weekdays are obtained by expanding the hourly demand in one day using the average queue length in the same hour on different weekdays via Equation S1.



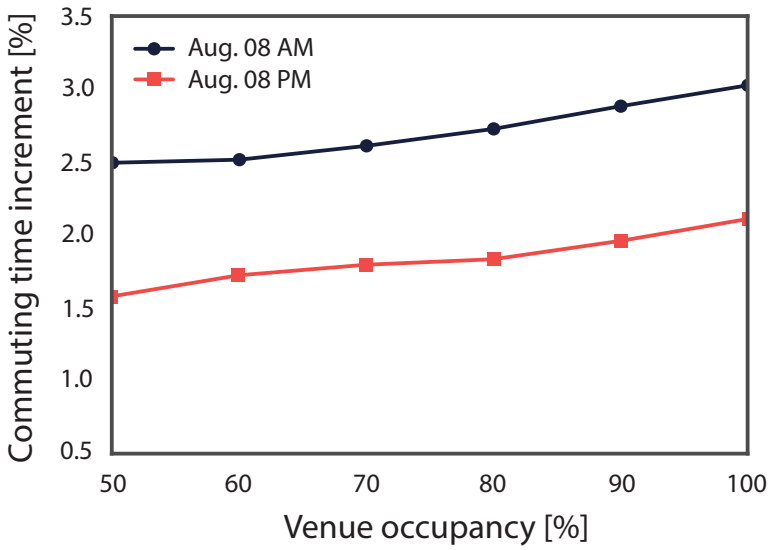
Supplementary Figure 5: Capacity of Airbnb properties and their occupancy on each weekday during the Olympics. (a) The histogram of the capacities of Airbnb properties. The maximum capacity is 16. Most properties can accommodate 2 ~ 6 persons. (b) The ratio between the number of tickets and the lodging capacity of Airbnb properties and hotels per day during the Olympics. The ratios are larger than 1 mainly because we are using the Airbnb data in Jan. 2015 and 500 large hotels in Rio. The number of Airbnb properties during the Olympics is certainly much larger than 2015 to meet the needs of tourists.



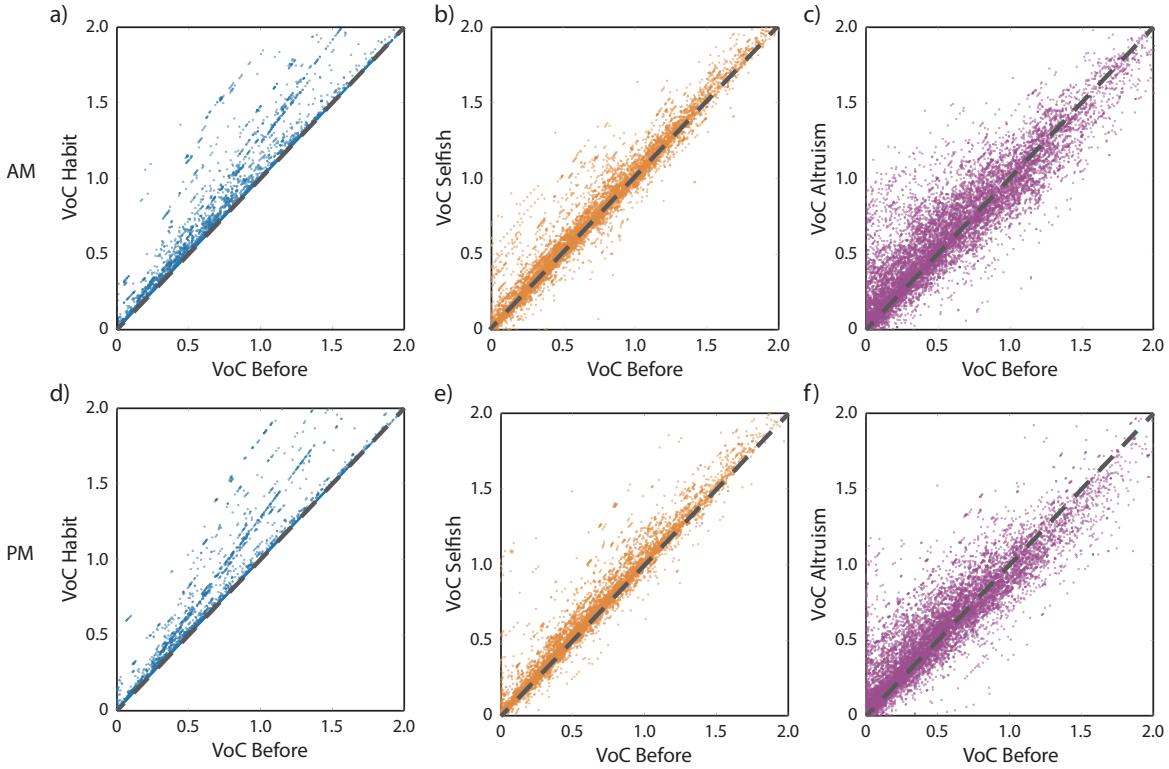
Supplementary Figure 6: Travel mode split model and the split results. (a) Tourists travel mode split model. (b) The travel mode split results during the morning peak hour on August 8. There are 6 target venues in this hour. The size of the pie indicates the flow of tourists from residential zones or to venues. The pie with white edge represents the tourists' arrival modes to the venue. The pie with gray edge represents the tourists' departure modes from their residences. (c) The travel mode split results during the evening peak hour on August 8. There are 3 target venues in this hour. (d) The estimated tourists' person and taxi demand in each hour during the Olympic weeks.



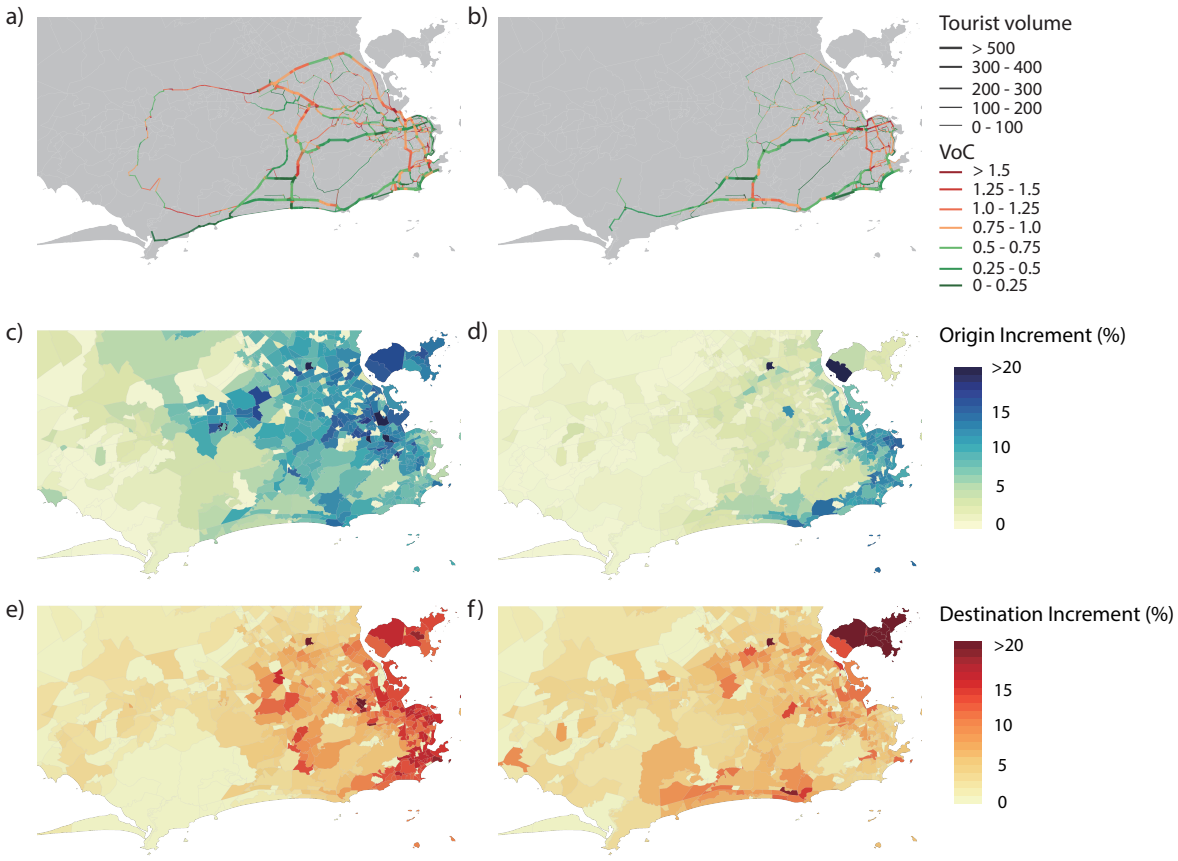
Supplementary Figure 7: Calibration of BPR function with camera data. (a) Scatter plot of traffic flow and average speed and the calibrated triangular speed-flow fundamental diagram (FD). The camera data includes traffic flow (veh/h) and hourly average speed (km/h) on the monitored road. The blue points are the field data. The red triangle is the calibrated FD. (b) The calibrated BPR function. The camera data samples in free flow phase are represented with light green points. The samples in congested phase are represented with dark green points. The data samples are used to fit the BPR function with the formation $t = f_s \cdot (1 + \alpha(VoC)^{\beta}) \cdot t^f$. To facilitate the fitting, we write BPR function as $t/t_f = f_s \cdot (1 + \alpha(VoC)^{\beta})$, which represents the relationship between t/t_f and VoC. The calibrated BPR is $t = 1.15 \cdot (1 + 0.18(VoC)^5) \cdot t^f$, which is higher than the naive BPR under the same VoC and free flow speed.



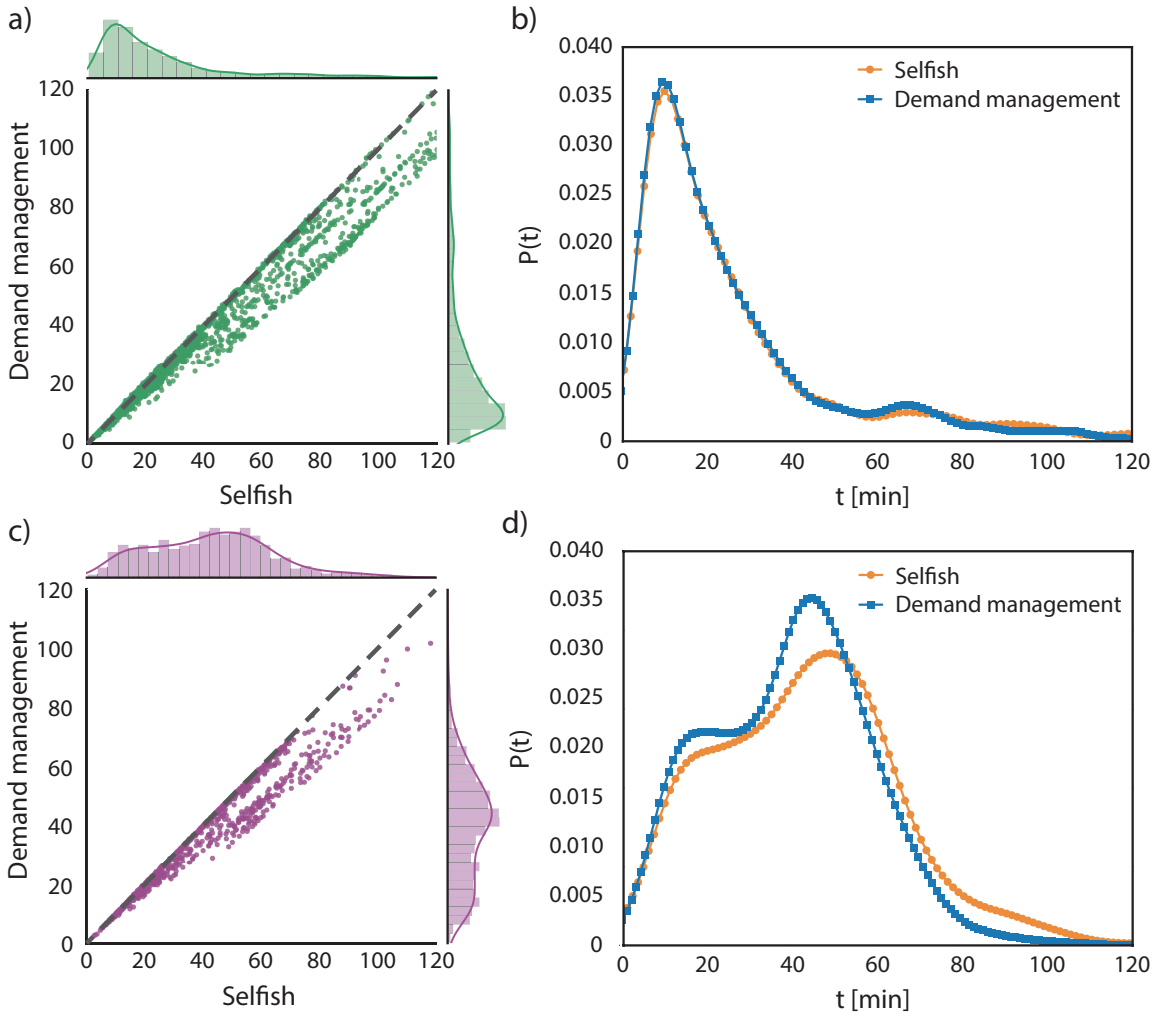
Supplementary Figure 8: The change of average percentage increment of commuters versus the venue occupancy rate O_v during the morning and evening peak hours on Aug. 8 under the *selfish* scenario. When we estimate the tourists travel demand per hour with the expected audience of each venue, the tickets of some games may not be sold out. In other words, the venues are not filled up with tourists. In this case, to investigate the increment of commuting time, we define an occupancy rate O_v for all venues. The O_v ranges from 50% to 100%. 50% implies only half tickets are sold for a game. 100% implies the upper bound case, which means all of the tickets are sold out in the given hour. Therefore, the tourists' taxi demand in the fulfilled case should be scaled with O_v. Then, we reassign the total demand to the road network under the *selfish* scenario. It's obviously that with the increase of O_v, the impact of the Olympics to local commuters becomes worse. Note that an important increment in the travel time is not only due to tourists using vehicle trips but also due to the reduced capacity in the Olympic lanes that transport athletes.



Supplementary Figure 9: Change of VoC before and during Olympics and the OD pairs suffering delay most. (a-c) Scatter plot of VoC before versus during Olympic during the morning peak hour under three scenarios. VoC > 1 reflects the demand on the road is larger than the capacity, which is over-saturated and the average speed is much lower than the free flow speed. For *habit*, the VoC of all roads are larger than before Olympics as the demand on all of the roads will increase; For *selfish*, the VoC of some roads increase and some decrease as the road resources are redistributed; For *altruism*, more under-saturated roads are getting congested and more over-saturated roads are relieved. (d-f) Scatter plot of VoC during the evening peak hour under three scenarios.



Supplementary Figure 10: Estimation of tourists' routes and impact of Olympics to the commuters in different zones on August 8 under *selfish* scenario. (a-b) The estimated tourists' travel routes during the morning and evening peak hours. The color reflects the traffic situation on the road. The width of road reflects the volume of tourists. (c) Percentage travel time increment of origins during the morning peak. (d) Percentage travel time increment of origins during the evening peak. (e) Percentage travel time increment of destinations during the morning peak. (f) Percentage travel time increment of destinations during the evening peak. People living on Governador Island suffer more heavy travel delay on their way home from work than from home to work.



Supplementary Figure 11: Effect of demand management strategy. (a) Scatter plot of the travel time of top-ranked commuting trips with versus without strategy. A fraction of travelers can benefit from the demand management strategy. Some of them save travel time almost by 20 minutes. (b) Distribution of commuting time with and without strategy. (c) Scatter plot of tourists' travel time from their residences to venues with versus without strategy. (d) Distribution of tourists' travel time with and without strategy.

SUPPLEMENTARY NOTE 1

Previous studies have estimated multiple aspects of travel demand using the mobile phone data, aka call detail records (CDRs), in conjunction with census data [1–4]. Here we briefly introduce the CDRs and the generation of travel demand before Olympics in our work.

CDRs contain the mobile phone activity logs of 2.19 million anonymous users of a specific carrier during 5 months in the metropolitan area of Rio. The activity logs includes the occurrence time of calls and SMS, and cell where the activity happens. The location of base station in a cell is approximately regarded as the location of user, and there are 1421 stations across the metropolitan area of Rio.

As the first step of trips estimation, we separate the locations of a user from CDRs into two categories: *stays* and *pass-by-points*. *Stays* are defined as the locations which the user is known to stay in for at least 10 minutes. The others are *pass-by-points* and regarded as noisy data. The *stays* are further labeled as $home$, $work$, and $other$. A user's $home$ is defined as the stay location where the user visits most frequently between 8 pm and 7 am on weeknights; $work$ is defined as the stay location other than $home$ where the user visits most between 7 am and 8 pm on weekdays; the remaining stay locations are labeled as $other$. Consequently, for one user, only one stay is labeled as $home$, and one is labeled as $work$. Once each stay is labeled by purpose, we define two consecutive $stays$ constitute a *raw trip* if they are at different places and on the same weekday. However, as CDRs only record the logs when users are interacting with their phones, we can not argue that users start their trips when they make the first call from the origin or end their trips when they make the first call from the destination. To address this issue, we estimate the conditional probability that user departed during an hour between the two consecutive calls at different stays. This conditional probability function can be estimated using observed call frequencies of all users during one day [2, 4]. Afterward, we can categorize the user's trips into three classes, e.g. home-based-work (HBW), home-based-other (HBO), and non-home-based (NHB), by their origins and destinations. Among them, HBW is essentially a commuting trip, which signifies the trip is made by a user between his $home$ and $work$, including going to $work$ from $home$ and back to $home$ from $work$; HBO signifies the trip is made between $home$ and $other$; Other trips are labeled as NHB, which includes trips between $work$ and $other$, $other$ and $other$.

In order to estimate the origin-destination (OD) matrix from zone to zone, we aggregate the trips at the census tract scale. As this scale, we obtain the population and the vehicle usage rate of residents in each tract. The census data is obtained from Instituto Brasileiro de Geografia e Estatística (IBGE). It contains 730 tracts in the metropolitan area of Rio and 514 of them locate in Rio municipality, as shown in Supplementary Figure 2a. The vehicle usage rate ranges from 0.05 to 0.40 for different tracts. Finally, to obtain a more realistic person demand, we scale the CDRs trips by *expansion factor*, which is defined as the ratio of the population of the origin tract to the number of mobile phone users whose $home$ are located in that tract. Furthermore, to estimate the vehicle demand, we weight the person demand with the vehicle usage rate in the origin tract. The validation of the estimated OD matrix with survey can be found in [2–4].

Supplementary Fig. 3a and 3b show the total estimated person and vehicle travel demand in Rio metropolitan area with different trip classes. In each hour, total person demand is the sum of persons need to travel in each tract; total vehicle demand is the sum of vehicles used for travel in each tract. Actually, a fraction of people travel by walk, bike, and public transportation. We do not consider these people in our work as they have minor impacts on the traffic conditions in the road network.

SUPPLEMENTARY NOTE 2

The travel demand obtained from the CDRs provides the average demand in 24 hours. In order to differ the travel demand on 5 weekdays, from Monday to Friday, we extend the 24-hour demand with Waze data. Waze data sets are collected by Waze, which is a GPS-based based traffic and navigation application program for smart phones developed by Waze Mobile [5]. Waze gathers complementary map data and traffic information reported by its users–Wazer. The collected information includes traffic acci-

dents, traffic jams, and police traps, etc. The data sets used in our work range from January to July 2015, and contain about 4.6 million items of reports in the 7 months, and each item contains timestamps, location, speed in jam, road type, length of queue in jam, delay in jam, and Wazer's confidence level.

Through aggregating the reported information together, we can investigate the degree of congestion. For instance, generally, more report items mean there are more congestion happening during the period. However, the number of traffic jams reported is greatly influenced by the number of app users in different hours, for instance, evening always has larger number of users than morning, thereby receives more reports under similar level of congestion. Therefore, we infer the degree of congestion with the average length of the queue, which is the driving distance of the user in traffic congestion.

In order to extend the 24-hour demand to 5 weekdays, we calculate the average queue length (a.k.a. driving distance in traffic jams reported by Wazers) Q in the municipality area of Rio in each hour first. After this step, we have one value of Q per hour. Supplementary Figure 4a illustrates the Q in one sample month. The plot starts with data on Sunday. As can be seen, the average queue length shows evident daily and weekly periodicity, especially for weekdays. Such periodicity indicates that we can use Q as an indicator to differ the traffic congestion on different weekdays. Moreover, for each hour, we compared the standard deviation of the queue lengths in this hour on the same weekdays (e.g. Monday) with the standard deviation of the queue lengths in this hour on all weekdays (include weekdays from Monday to Friday). With the former strategy, the average standard deviation on five weekdays is 92.42. The standard deviation of all weekdays is 109.12, which is larger than the separate standard deviation on each weekday. This comparison indicates that the queue length from Waze data can differ the traffic situations of different weekdays. Therefore, we then aggregate the hourly queue during 7 months into 5 weekday via averaging the values of Q in the same hour and same weekday. By doing this, we obtain 24×5 values, as shown in Supplementary Figure 4b. The next step is to expand the 24-hour travel demand to 5 weekdays demand, namely fluctuating demand, with the obtained queue lengths on different weekdays. For each hour, we extract the queue lengths on 5 weekdays. Supplementary Figure 4c and 4d show the queue length during morning and evening peak hours, respectively. Finally, for each hour, the fluctuating demand is calculated using the following equation:

$$D'_{h,d} = \omega_{h,d} \times D_h = \left(1 + \frac{Q_{h,d} - Q_{h,avg}}{Q_{h,avg}}\right) \times D_h \tag{S1}$$

where $D'_{h,d}$ is the fluctuating travel demand matrix during hour h on weekday d; D_h reflects the travel demand matrix during the same hour from CDRs; $Q_{h,avg}$ is the average queue length during hour h, as shown with the dashed line in Supplementary Figure 4c and 4d; $Q_{h,d}$ is the queue length during hour h on weekday d. $Q_{h,d}$ and $Q_{h,avg}$ are both calculated over the whole municipality area of Rio. With this method, the hourly travel demand of each tract are equally unfolded to five weekdays using the weight, $\omega_{h,d}$. Finally, the total person and vehicle demand in each hour of 5 weekdays are shown in Supplementary Figure 4e, respectively.

So far, we have obtained the OD matrix per hour on each weekday before Olympics. To assign the vehicle demand to the road network, we assume that all of the drivers are seeking the routes with shortest travel time, which bring the road network to be an user-equilibrium (UE) state. The implementation of UE model in our work is based on literature [6, 7] and described in [8].

SUPPLEMENTARY NOTE 3

In our work, the person travel demand during Olympics is an aggregation of local person demand and the number of tourists in each hour. The travel of tourists can be classifies into three cases: traveling to venues, departing from the venues, and others (going to malls, scenic spots, etc). Apparently, travelling from residences to venues is more predictable than departing from venues after events and others. Tourists probably go to other places after they watching a game, e.g. restaurant, scenic spots, instead of back to their residences. Therefore, we only consider the tourists flow from residences to venues in our

work. Consequently, the origins in the OD matrix of tourists include the residential tracts of tourists and the destinations only include the tracts with venues.

The residences of tourists contain around 13,400 Airbnb properties and 106 hotels in our data sets. The Airbnb data are available online and obtained using Airbnb API [9]. We have the number of accommodation and the location of each property in Rio. The locations of Airnbn properties as shown in Figure 1b of the paper. Supplymentary Figure 5a shows the histogram of the capacities of Airbnb properties. The capacities of Airbnb properties are mainly between 2 and 6. For the hotels, we have their locations and assume each hotel's capacity is 500. The total lodging capacity in Rio is about 111,408 individuals. For estimation of OD matrix, we aggregate the Airbnb properties and hotels to census tracts. During each hour, the number of travelling tourists are calculated with the following assumed rules: (i) all venues will be filled with tourists before the game starts; (ii) Tourists depart from their from 1 to 3 hours before the game starts. We assume 30% tourists departure 1 hour ahead; 40% tourists departure 2 hours ahead; the others departure 3 hours ahead; (iii) The tourists going to each venue are distributed in the Airbnb properties and hotels with a probability weighted by their capacities. To obtain the number of tourist trips from each tract to the venue tract, we first aggregate the Airbnb properties and hotels to the census tract scale by summing up their capacities in the given tract. Then, the travel demand in each hour are assigned to each tract via weighting the total tourist demand with a factor p_t, which is the ratio between the accommodation capacity of the tract to the total accommodation capacity of all tracts in Rio. Finally, we have the tourist flow from their residential tracts (origin) to the tracts with venues (destination) per hour per day. In addition, we compare the number of tickets of the games on each day with the lodging capacity of all Airbnb properties and hotels. The results are shown in the Supplementary Figure 5b. As observed from the figure, the ratio between the number of tickets and the lodging capacity is from 3.51 to 5.34 during the Olympics. The main reasons that the ratio is larger than 1 are: (i) our datasets only contain 500 large hotels in Rio, the actual number of hotels is larger than this; (ii) the Airbnb data we used are in January 2015, the number of Airbnb properties will certainly increase during the Olympics to meet the tourists' needs; (iii) we consider all of the expected audience are from the Airbnb properties and hotels, but actually a part of them are residents.

With the purpose of predicting the hourly vehicle demand during Olympics, we aggregate the local vehicle demand and the taxi demand of tourists together. Therefore, we present a travel mode split model by considering the location of tourists' residences, venues, the metro and BRT line, and the bus schedule. The model is represented in Supplementary Figure 6a. The metro and BRT network of Rio are connected at some stations, that is, people can travel between any two metro/BRT stations without exiting the station. Therefore, we consider the metro and BRT as one complete public transportation network.

We split the travel modes of tourists into four categories: walking + metro/BRT, bike + metro/BRT, bus, and taxi. Among them, walking + metro/BRT means the tourist walk both sides: from his/her residence to the station and from station to venue; biking + metro/BRT includes three cases: (i) bike to station and walk from station to venue, (ii) walk to station and bike from station to venue, (iii) bike to station and bike from station to venue.

To classify the tourists with travel mode, we consider certain factors impacting people's choice of travel mode, e.g. distance to the nearest station, travel time, number of transfers [10]. Consequently, we project the following principles: (i) if the distance to metro or BRT station (the larger one between the distance from residence to station and distance from station to venue) is smaller than $2km$, 70% of the tourists will take the metro/BRT. Among them, if the distance is smaller than $1km$, they will walk to the station, else they will go to the station from residences or go to the venue from station by bike. (ii) 30% of tourists who are near and 100% of tourists who are far from the metro/BRT station will choose bus or taxi to the venues. Among them, if the bus transit time is larger than twice of the taxi travel time, 75% of them will choose taxi, otherwise, 25% will. Furthermore, among the rest, they will consider the number of transfer and distance to bus station if choose bus. Only when the number of transfer is not larger than 1 and the distance to bus station is less than $1km$, they will choose bus. Otherwise, they will choose taxi to the venues. In addition, the distance from residences to station or from station to venues is regarded

as straight-line distance. The bus transit time, the number of transfer, the walking distance to the bus station, and the taxi travel time are achieved from Google map API with the departure time of tourists.

Supplementary Figure 6b and 6c illustrate the results of tourists travel mode split during the morning and evening peak hour on August 8, respectively. The size of the pie indicates the flow of tourists from residential zones or to venues. The pie with white edge represents the tourists' arrival modes to the venue. The pie with gray edge represents the tourists' departure modes from their residences. As can be seen, for the venue nearby the metro/BRT line, the main modes are walking + metro/BRT, bus, and taxi; For the venue are distance away from the metro/BRT line, the main travel modes are bike and metro/BRT, and taxi. These results conform with the reality.

SUPPLEMENTARY NOTE 4

In the camera data sets, we have 85 cameras set up on different types of roads in Rio municipality, including motorway, primary, etc. Each camera provides the field traffic volume and average speed per hour of the road. First, we match the location of surveillance camera with the right road segment. Afterward, we estimate the capacity and free flow speed on each road based on the traffic volume and average speed. The capacity signifies the maximum sustainable flow rate at which vehicles can be expected to traverse a road segment during one hour [11]. Here, the capacity is estimated as the average value of the top 5% traffic volumes collected with camera. The free flow speed is used to describe the average speed that a vehicle would travel if there were no congestion or other adverse conditions. Here the free flow speed is estimated as the average value of the top 5% speeds. The critical speed is the traffic speed when the traffic is transferring between free flow and congestion. Here we estimate the critical speed with the average value of the speeds which corresponding volumes are ranked in top 5%. Afterward, we plot the triangular speed-flow fundamental diagram (FD) of the traffic flow on each road, as shown in Supplementary Figure 7a. By doing this, the traffic data samples are separated into two categories: free flow phase and congested phase. The samples which speeds are larger than critical speed are in the free flow phase. The others which speeds are smaller than critical speed are in the congested phase [12, 13].

In our work on traffic assignment, we adopt the Bureau of Public Roads (BPR) function, $t = f_s \cdot (1 + \alpha(VoC)^{\beta}) \cdot t^f$, to estimate the travel time under a given demand volume on the road [14]. The following naive BPR function is used for highway:

$$t = (1 + 0.15(VoC)^4) \cdot t^f \tag{S2}$$

in which $\alpha = 0.15$, $\beta = 4$, and $f_s = 1.0$. However, in the urban road network, the roads contain several different levels: motorway, primary, secondary, and tertiary. Therefore, we calibrate the BPR function with the field traffic data. To this end, we transform the BPR function as

$$t/t_f = f_s \cdot (1 + \alpha(VoC)^{\beta}) \tag{S3}$$

which represents the relationship between t/t_f and VoC.

After splitting the camera data samples into free flow and congested phases, we calculate the VoC and the fraction of travel time over the free flow travel time, t/t_f, for each sample. For the free flow phase samples, the demand volume equals to the actual volume as all of the vehicles could pass through the road in a specific duration (e.g. one hour). Consequently, VoC equals to the fraction of measured volume over the capacity of the road. t/t_f equals to the fraction of free flow speed over the measured average speed. For the congested phase samples, we assume that the excess demand volume is the only reason of congestion. Therefore, the VoC is estimated by flipping the measured volume over capacity at $VoC = 1$. For instance, if the measured volume is half of the capacity in congested phase, the demand would be 1.5 times of capacity. Finally, we plot the t/t_f versus VoC in Supplementary Figure 7b. The samples in free flow phase are represented with light green points. The samples in congested phase are represented with dark green points. Then, the coefficients in Equation S3 can be calibrated using the ordinary least squares (OLS) algorithm. Finally, the calibrated BPR function has the following formation:

$$t = 1.15 \times (1 + 0.18(VoC)^5) \cdot t^f \tag{S4}$$

which generates higher travel time than the naive BPR under the same travel demand. The calibrated BPR function is used in the assignment model both before and during Olympics.

SUPPLEMENTARY NODE 5

To assign the vehicle travel demand to the road network, we need to know some properties of the roads. To this end, the free flow speed and capacity of each road are by standard convention estimated with the category of road (motorway, first, second, third, residential, etc.) The free flow travel time is calculated with the free flow speed and length of the road segment.

Before Olympics, we assign the travel demand to the road network $\mathcal{G}(\mathcal{N},\mathcal{E})$ with a User Equilibrium (UE) model, where $\mathcal{N}$ is the set of nodes, and $\mathcal{E}$ is the set of directed edges. In the UE model, all of the travelers are routing with their shortest travel times and referred as *selfish* in our paper. *Selfish* follows the first principle proposed by Wardrop [15]: *"Each user noncooperatively seeks to minimize his/her cost of transportation. Specifically, a user equilibrium is reached when no user may lower his/her transportation cost through unilateral action"* [16]. Many methods have been proposed to solve such user equilibrium problem [17]. The general objective of this optimal problem is:

$$\text{minimize} \quad Z(v) = \sum_{e\in\mathcal{E}} c_e(v_e)\, v_e \tag{S5}$$

subject to

$$\begin{cases} v_e = \sum_{p\in\mathcal{P}} \delta_{ep} f_p & \forall e \in \mathcal{E} \\ D_{i,j} = \sum_{p\in\mathcal{P}_{ij}} f_p & \forall (i,j) \in \mathcal{W} \\ f_p \geq 0 & \forall p \in \mathcal{P} \end{cases} \tag{S6}$$

where v_e is the traffic volume on edge e; $c_e(v_e)$ is the cost function depend on the flow on edge e. Under the *selfish* scenario, the travel cost equals to the travel time, that is $c_e(v_e) = t_e$. t_e is calculated using the BPR function in Supplementary Equation S4. $\mathcal{P}$ is the collection of paths of all trips in demand and p is a path of $\mathcal{P}$. f_p is the flow of path p. δ_{ep} is 1 if edge e is on path p, and 0 otherwise. $D_{i,j}$ is the demand for flow between OD pair (i,j) in the OD matrix. In our work, we solve this problem using the algorithm proposed in [6] and refined by [7]. In detail, we minimize the distance between the optimal solution and the current solution in an iterative process. In our work, the distance is measured using the following equation:

$$r_g = 1 - \frac{\sum_{o\in\mathcal{O},d\in\mathcal{D}} t'_{od} f_{od}}{\sum_{e\in\mathcal{E}} t_e v_e} \tag{S7}$$

where $\mathcal{O}$ and $\mathcal{D}$ are the set of origin and destination nodes in the road network; f_{od} is the demand of flow from o to d; t'_{od} is the shortest travel time of trip (o,d) in the current iteration. In each iteration, the shortest paths of all OD pairs will be re-calculated, and a fraction of trips will be assigned to the new shortest path. In this way, the r_g can converge to zero as the shortest travel time of each OD pair in the previous step is more and more close to the shortest one in current step. In our implementation, we repeat the iteration until r_g is smaller than a threshold, 0.01.

Moreover, as the travel demand in our work is at census tract scale, we need to convert the demand from tracts to tracts to nodes to nodes. To this end, we first find the nodes within each census tract. Second, we add a node at the geometric centre of each tract to the road network and generate links (namely dummy links) from the centroid to the nearest 3 nodes. The capacities of the dummy links are set to infinity, which means the travel time on the dummy links are zero. Then, all trips from the origin tracts are assigned to the centroid. At the destination end, the trips are assigned to the nodes uniformly. In this way, the travel demands at the tract scale are assigned to the node scale. Further details of the implementation of assignment are described in [8].

During Olympics, understanding the routing behaviour of travelers, including the local population and taxi of tourists, is premise for the collective benefits estimation. Before Olympics, the travel routes of local population are estimated using the UE model, which is a selfish routing strategy. After

the Olympics opens, we design three typical routing scenarios: *habit*, *selfish*, and *altruism*. The following content presents the demand assignment method for the three scenarios.

Habit signifies that all of the travelers in the road network follow their routine routes before Olympics. Although the travelers know that the traffic situation probably changes, they will not change their routes toward saving the travel time. In fact, the travel time of a part of travelers increase because of the additional flow in their paths and Olympic lanes. To explore this, we calculate the routes of tourists and map the tourists' taxi flow on the road network. The taxi routes from tourists' residences to venues are regarded as the shortest paths on the network before Olympics (routing toward the shortest travel time before Olympics). Afterward, the tourists flow are assigned on their routes. Thus, VoC on each road during Olympics is updated using the following equation:

$$VoC'_e = \frac{v_e^l + v_e^t}{C'_e} \tag{S8}$$

where C'_e is the capacity of road e during Olympics. Most of the roads keep their capacities, but the capacities of a small fraction of roads are reduced because of the establishment of Olympic lanes. v_e^l is the volume of local population, and is the same as before Olympics. v_e^t is the tourist volume on road e, and is calculated by aggregating the tourists trips using this road. After updating the VoC on each road, the travel time of a part of trips increase and no one could save time. Supplementary Figure 9a and 9d show the VoC of each road under *habit* versus before Olympics. As can be seen, the VoC of each road is higher than or equals to before.

Selfish signifies that all of the travelers can find and choose the routes with shortest travel time during the Olympics. Under this scenario, all of the travelers are fairly know the traffic situation in the road network and can find their shortest paths. Therefore, the road network reaches an user-equilibrium status ultimately and we re-run the UE model with the new travel demand (residential private vehicles and taxis taken by tourists) and the same BPR function with before the Olympics. In contrast with before Olympics, the demand of some OD pairs increase and the capacities of Olympic lanes reduce. As shown in Supplementary Figure 9b and 9e, it is evident that a part of roads are more busy than before, for both under-saturated ($VoC < 1$) and over-saturated ($VoC \geq 1$) roads. Besides, it is notable that some roads even become more unimpeded. That's because travelers keep away from some roads for a shorter path.

Altruism signifies that all of the travelers are selfless enough and aiming at maximizing the collective benefits of all travelers. In this sense, the optimized objective of the traffic assignment is to minimize the total travel time, which is a system optimized strategy. Different from *selfish*, which cost equals to the travel time, the cost of each edge for *altruism* equals to the marginal cost MC_e:

$$MC_e = \frac{\partial(v_e t_e)}{\partial v_e} = t_e + f_s \alpha \beta \left(\frac{v_e}{C_e}\right)^{\beta} \times t_e^f \tag{S9}$$

The essence of MC_e is the decrease of collective travel time of vehicles on this edge if one vehicle is taken out. After assigning using the same method as *selfish*, we represent the change of VoC in Supplementary Figure 9c and 9f during the morning and evening peak hour, respectively. Compared with *selfish*, *altruism* generated more VoC decreased roads, as well as increased. More importantly, most of the VoC increased roads are under-saturated before Olympics and most decreased roads are over-saturated before.

REFERENCES

[1] Shan Jiang, Gaston A Fiore, Yingxiang Yang, Joseph Ferreira Jr, Emilio Frazzoli, and Marta C González. A review of urban computing for mobile phone traces: current methods, challenges and opportunities. In *Proceedings of the 2nd ACM SIGKDD international workshop on Urban Computing*, page 2. ACM, 2013.

[2] Lauren Alexander, Shan Jiang, Mikel Murga, and Marta C González. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58:240–250, 2015.

[3] Serdar Çolak, Lauren P Alexander, Bernardo G Alvim, Shomik R Mehndiratta, and Marta C González. Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transportation Research Record: Journal of the Transportation Research Board*, (2526):126–135, 2015.

[4] Jameson L Toole, Serdar Colak, Bradley Sturt, Lauren P Alexander, Alexandre Evsukoff, and Marta C González. The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58:162–177, 2015.

[5] Waze Mobile, 2015. URL `http://www.waze.com/`.

[6] Robert B Dial. A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. *Transportation Research Part B: Methodological*, 40(10):917–936, 2006.

[7] Yu Marco Nie. A class of bush-based algorithms for the traffic assignment problem. *Transportation Research Part B: Methodological*, 44(1):73–89, 2010.

[8] Serdar Çolak, Antonio Lima, and Marta C González. Understanding congested travel in urban areas. *Nature communications*, 7, 2016.

[9] Airbnb. Rio de janeiro state of rio de janeiro brazil-airbnb, 2015. URL `https://www.airbnb.com/s/Rio-de-Janeiro--State-of-Rio-de-Janeiro--Brazil`.

[10] Avishai Ceder. *Public Transit Planning and Operation: Modeling, Practice and Behavior*. CRC Press, 2015.

[11] Highway Capacity Manual et al. Transportation research board. *National Research Council, Washington, DC*, 113, 2000.

[12] Mauro Garavello and Benedetto Piccoli. *Traffic flow on networks*, volume 1. American institute of mathematical sciences Springfield, 2006.

[13] Dirk Helbing. Derivation of a fundamental diagram for urban traffic flow. *The European Physical Journal B*, 70(2):229–241, 2009.

[14] Stella C Dafermos and Frederick T Sparrow. The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards B*, 73(2):91–118, 1969.

[15] John Glen Wardrop. Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(3):325–362, 1952.

[16] Terry L Friesz and David Bernstein. *Foundations of Network Optimization and Games*, volume 3. Springer, 2015.

[17] Michael Patriksson. *The traffic assignment problem: models and methods*. Courier Dover Publications, 2015.