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

Yanyan Xu
Department of Civil and Environmental Engineering, MIT, Cambridge, MA 02139, USA

Marta C. González*
Department of Civil and Environmental Engineering, MIT, Cambridge, MA 02139, USA and Engineering Systems Division, MIT, Cambridge, MA 02139, USA
(Dated: January 8, 2017)

Information technologies today can inform each of us about the route with the shortest time, but they do not contain incentives to manage travelers such that we all get collective benefits in travel times. To that end we need travel demand estimates and target strategies to reduce vehicles from target roads in a feasible way. During large events the traffic inconveniences in large cities are unusually high, yet temporary, and the entire population may be more willing to adopt collective recommendations for collective benefits in traffic. In this paper, we integrate, for the first time, big data resources to estimate the impact of events in traffic and propose target strategies for collective good at the urban scale. In the context of the Olympic Games in Rio de Janeiro, we first predict the expected increase in traffic. To that end, we integrate data from mobile phones, Airbnb, Waze, and transit information, with game schedules and expected attendance in each venue. Next, we evaluate different route choice scenarios for drives during the peak hours. Finally, we gather information on the trips that contribute the most to the global congestion which could be redirected from vehicles to transit. Interestingly, we show that (i) following new route alternatives during the event with individual shortest times can save more collective travel time than keeping the routine routes used before the event, uncovering the positive value of information technologies during events; (ii) with only a small proportion of people selected from specific areas switching from driving to public transport, the collective travel time can be reduced to a great extent. Results are presented on-line for the evaluation of the public and policy makers *.

I. INTRODUCTION

Daily traffic has important implications for the functioning of our cities [1–4]. It affect total energy use, equity, air pollution, and urban sprawling. Given this impact, master plans of urban transportation need to be technically sound, economically affordable and publicly acceptable [5–11]. This becomes a more pressing need when preparing for large events, which unusually stress the use of the available infrastructures and put at risk the overall success of the event.

In their best attempts, goals of an urban transportation plan seek to: (a) avoid long and unnecessary motorized travel, (b) shift the movement of people to socially efficient modes, such as walking, biking, and public transit, and (c) improve the technology and operational management of transportation services. To reach these goals, plans today try to promote the use of bus rapid transit (BRT), congestion charging, or bike-sharing. But much less is done to develop real time information platforms that provide the value of choices for the social good. Nowadays, the most popular information platforms, such as Waze or Google Transit Feeds, give us individual information about travel times but do not take into account global information, e.g. providing incentives to reduce global costs regarding our route choices. One limitation may be that the main set of infrastructures in urban transportation planning of mature cities were developed in the seventies, before the information age, and relied on the results of travel diaries, limiting the communication with the majority of the actual travelers. Second, demand management faces the “tragedy of the commons”. Space in streets are a shared-resource system where individual users act independently according to their own self-interest, behaving contrary to the common good by depleting that resource. The population, however, may be more prone to take actions for collective benefits while hosting a big event.

We propose demand management strategies during mega events. Large-scale events happen every year around the world, such as Olympic games, world expositions, concerts, pilgrimages, etc. They attract large number of participants and tourists traveling to one destination, thereby producing increased pressure on transportation, especially for cities with already large population [12, 13]. Past research has tried to estimate the impact of events on the economy and air quality of the host city [14–16]. Moreover, in the face of changing conditions in cities, a new topic – city resilience – has drawn attention from academics and decision makers in recent years [17–20]. In the case of transportation networks, researchers mainly study the network resilience to cope with the unexpected damage or perturbations of transportation facilities [21–24] or guide the long-term transportation construction [25, 26]. For instance, Donovan and Work quantified the resilience of transportation system to extreme events using GPS data from taxis [23]. Their model detects the event from his-
historical data, thereby can not forewarn the impact of the forthcoming events. In the context of traffic management during large-scale events, previous efforts focus on ensuring the efficient travel of participants. However, the disruptions to the travel of the local population are not taken into account. The most used policy today to reduce motorized travels is to limit the number of vehicles with specific ending plate number, but without efficient strategies to target congested bottlenecks [27, 28]. Consequently, the new paradigm is to achieve the collective benefits of all travelers during events by integrating multiple data resources using information technologies to calculate the costs and communicate the benefits of various strategies [29, 30].

Specifically, we evaluate the impact of large-scale events on the traffic in the host city and evaluate the impact of strategies to overcome it. We aim at understanding the change of travel demand when during large-scale events, and to address reasonable demand management strategies to mitigate the traffic congestion during the event. We take the Summer Olympics 2016 in Rio de Janeiro as a case of study to estimate the impact of large-scale events on the travel of the local population. Rio de Janeiro is one of the most congested cities in the world according to the TomTom’s report on global traffic congestion [31]. A study released by the Industry Federation of the State of Rio de Janeiro (FIRJAN) confirms that traffic congestion has tremendous economic costs as well. The study found that congestion costs of the cities of Rio and São Paulo added roughly USD 43 billion in 2013 alone. The loss amounts nearly 8% of each metropolitan area’s Gross Domestic Product (GDP). This is greater than the estimated budget for transport capital investment in Brazil, Mexico, and Argentina combined. Traffic congestion originates from the imbalanced development of travel demand of vehicles and the road network supply [32, 33]. For a booming city, the traffic congestion can be mitigated through constructing more roadways and transit infrastructure. But for mature urban areas like Rio, opportunities for further investments in transportation infrastructure are often limited [34].

The International Olympic Committee (IOC) estimates 0.48 million tourists in Rio for Olympics, which is about 7.5% of the Rio population. To understand the impact of Olympics, we estimate the travel demand of the local population and their fraction in private vehicles using mobile phone data, also known as call detail records (CDRs) combined with Waze data. The travel times of commuters taking private cars are estimated during the morning and evening peak hours and compared with Google maps in the same hour. During the Olympics, we estimate the origin and destination of tourists using the Olympic Games’ schedule, information on expected audience in each venues, and Airbnb properties [35] and hotels. To estimate the increase in vehicular traffic, we estimate the taxi demand of tourists going to the events each hour and also the reduced capacity in the dedicated Olympic lanes. Both the tourists’ taxi demand and the local vehicle demand are assigned to the road network under three routing scenarios: habit, selfish, and altruism. The goal is to assess how if certain routing recommendations are followed we can gain collective benefits in vehicular traffic. To evaluate the results, we estimate the travel time of tourists and travel time increment of local commuters’ under the three scenarios in the commuting peak hours. In addition, we also propose a mode change strategy, that targets a selected fraction of travelers to change from driving to metro and BRT. To this end, we uncover the origin destination pairs with the most contribution to the collective travel time and consider the overall benefit of taking one vehicle out of that pair. Finally, we demonstrate the effectiveness of the proposed demand management strategy by comparing it with a benchmark program that reduces the same number of vehicles randomly distributed (which is similar to car reductions by plate numbers). A detailed diagram of the data and modeling pipeline can be found in Supplementary Figure 1.

II. RESULTS

A. Travel Demand Estimation

Travel demand estimation before the Olympics.

Previous studies generated the average hourly travel demand at the census tract scale using CDRs from mobile phones (include the timestamp and location for every phone call or SMS of anonymous users), census records, and surveys data in Rio de Janeiro [11, 36–38]. In the travel demand estimation framework, the stay locations of each user are recognized and labeled as home, work, or other. The most visited place during weekday nights and weekends is labeled as home, the most visited place during weekday working hours is labeled as work, and the rest are labeled as other. Consequently, we classify the trips of each person as: home − based − work (or commuting, includes travel from home to work and from work to home), home − based − other (trips between home and other), and non − home − based (trips between non-home places, e.g. work, and other). After aggregating the trips to census tract scale with the geographical locations of their origins and destinations, we get the number of mobile phone users traveling from tract to tract at hourly basis. Then, the travel demand of all residents is estimated by scaling the user demand with an expansion factor, which is defined as the ratio between the actual population of the origin tract from the census and the number of users whose home are located in that tract. In this way, we could get a reasonable person origin-destination (OD) matrix with different trip purposes. To assess the traffic in the road network we need to estimate the vehicles demand. Namely the vehicles OD matrix, counting the number of private vehicles used by residents from their origin to their destination tracts.
In this work we only consider the motor vehicles used by residents and thereby simply scale the person demand of each OD pair with the vehicle usage rate in its origin tract. The estimated vehicle demand is 0.44 million from 1.69 million trips of residents during the morning peak hours and 0.44 million vehicles and 1.61 during the evening peak hours in the Rio de Janeiro municipality. The 24-hour trip demand with different trip classes are given in the Supplementary Figure 3a and 3b, in Supplementary Note 1.

Next, we extend the vehicle demand to small fluctuations in five weekdays, using the records of Waze Mobile [39]. Waze provided the records of Wazers for 7 months in 2015. The data sets include the location of user, timestamp, level and duration of jam, average speed, and length of the queue. We relate the fluctuations in the average length of the queue of traffic jams in the entire road network as proportional to the fluctuations of the total vehicle demand in this hour (previously estimated with mobile phone data). In other words, we calculate the average queue length in the whole municipality area of Rio in each hour each weekday, and use that value as a global congestion index. Using it to uniformly extend the travel demand of all OD pairs in 5 weekdays (see details in the Supplementary Figure 2c, Figure 4, and Note 2).

Travel demand estimation during the Olympics.
To build the OD matrices during the Olympics, the two pieces of information to be added locations of the origins and destinations and the flow between them. During the Olympics, the trips of tourists mainly contain the following three categories: traveling from residences to venues, departing from the venues, and others (going to restaurants, malls, scenic regions, etc). Among them, the trips from residences to venues are more predictable and the most important piece to increase traffic in a particular time. Therefore, we only consider the flow from spectators’ residences to the game venues. Figure 1a represents the location of 12 Olympic venues, the distribution of Airbnb properties and hotels, the metro and the BRT lines in Rio. Most of the tourists’ residences are distributed in the southeast coastal area. As planned by the municipal government, most venues are located around the metro or BRT stations, which makes public transportation quite convenient for most of the spectators.

The person travel demand equals to the sum of local demand before the Olympics and the number of people going to stadiums from their residences in the same time interval. To estimate this increase, the number of spectators arriving to each venue is estimated each hour based on the Olympic game schedule and the expected audience in each venue. For each hour, we add the expected audience of the venue if there are games that start in the venue in the given hour. This information was provided by the city together with the games schedule. Figure 1b shows the results on weekdays during the Olympics. The maximum number of spectators is nearly 0.1 million, which is a considerable fraction of the number of commuters in the peak hour. To determine the departure from hotels/Airbnb places to venues, we make the following assumptions: (i) 30% spectators depart 1 hour ahead; 40% spectators depart 2 hours ahead; the others depart 3 hours ahead. (ii) We use the distribution of Airbnb properties also to capture the distribution of origins of the local population that can afford the tickets. Namely, all the spectactors are distributed from the Airbnb properties and hotels, and are named tourists in the rest of the paper (see Supplementary Figure 2d for the distribution of the population density in Rio de Janeiro). As the Airbnb properties and hotels distribute across tracts, we first aggregate them to tracts with their geographical locations and assign a accommodation capacity to each tract. Then, for each tract, we define a factor $p_t$ as the ratio between the accommodation capacity of the tract to the total accommodation capacity in Rio. While the 12 venues are each located in a different tract. Finally, the number of trips from each tract to the venues are estimated by scaling the total demand to the venue with the factor $p_t$ in the origin tract (see Supplementary Figure 5 and Note 3).

To estimate the additional vehicle demand during the Olympics, we estimate the travel mode of tourists in each hour. Based on their required travel distances, a considerably fraction of them may use public transportation or taxi whilec will not affect our vehicle traffic and the subsequent strategies. We allow travel mode of tourists in 4 categories: walking and Metro/BRT, bike and Metro/BRT, taxi, and bus. To that end we simply take into account the distance to metro/BRT stations, the total travel time, and the number of mode transitions. Figure 2a shows the results of travels by mode on August 8 (Monday). As expected, most tourists choose Metro/BRT because both of their hotels and venue are near to Metro/BRT stations. Nonetheless, during the daytime, we estimate that about 10,000 tourists choosing taxis to the venues per hour, which produces a considerable vehicle demand added in the streets to only 12 destinations. As the tourists do not likely move alone, we assume taxi occupancy as 2, that is, 2 tourists per taxi per trip (see Supplementary Figure 6).

Figure 2b shows the total trips and the individual car trips on 10 weekdays from August 8 to 19. Car trips add the local vehicle demand in private cars estimated from CDRs and the taxi trips estimated for tourists. The morning peak is around 9:00 and the evening peak is around 18:00. During the peak hour, about 27% of the total trips occur. The number increase in about 60,000 trips during the Olympics. Consequently, traffic in the city will be specially congested for the paths from tourists’ residences to venues.
FIG. 1. Locations of venues, tourists’ residences and the number of spectators per venue per hour. (a) The locations of 12 Olympic venues, the Metro and BRT lines in Rio, the locations of hotels and distribution of Airbnb properties. Most venues are near to Metro/BRT stations, as well as hotels and most Airbnb properties. We distribute tourists around the 13,400 Airbnb properties and 106 hotels. Metro and BRT will likely be the first choice for most spectators. (b) The number of spectators arriving at each venue per hour on 9 weekdays during the Olympic Games. The largest indoor stadium, Carioca arena is also the busiest one.

FIG. 2. Estimated travel mode of tourists and total travel demand during the Olympics. (a) Tourist travel modes on August 8. A large proportion of tourists are using public transportation. About 1/3 of them may use taxis. (b) Estimates of total trips and vehicle trips per hour on 10 weekdays during the Olympics from August 8 to 19. The total trip estimates add local travelers and tourists going to venues in the given hour. The vehicle trips add the local population in their private cars vehicle and the estimated number of taxis used by tourists.

B. Travel time estimates and analysis of impacts in vehicular traffic

Before the Olympics, we assign the drivers to the routes distributing them via their shortest travel times and taking into account the resulting congestion as streets fill up. This is a common approximation to model the complex problem of route selection. Namely, the user equilibrium (UE) model, which implies no driver can unilaterally reduce his/her travel time by changing routes. In our implementation of UE model, the travel times of links depend on the volume-over-capacity ratio (VoC), calculated with the Bureau of Public Roads (BPR) function:

\[ t_e(v_e) = f_s \left[ 1 + \alpha \left( \frac{v_e}{C_e} \right)^\beta \right] \times t_f^e \]  

(1)

where \( t_e(v_e) \) is the average travel time on link \( e \); \( t_f^e \) is the free flow travel time on this link; \( f_s \) is a scale factor and not less than 1. The coefficients in BPR are calibrated using field data collected by surveillance cameras as \( f_s = 1.15, \alpha = 0.18, \beta = 5.0 \). Finally, we compare our estimated travel times of top commuter OD pairs with the results of the Google maps API in the same hour, finding very good agreement (see Supplementary Figure 3, Figure 7, and Note 4).
The Olympics will disrupt the routes of a fraction of travelers, especially those with routine routes being in the path to the games or trough the reduced capacity of the lanes dedicated to the Olympics. These are lanes where only buses with athletes and stuff can travel. In the trips assignment during the Olympics, this reduced capacity also generates traffic.

Our goal is to evaluate the impact in travel times under three types of vehicular route choices: (i) habit: All travelers will follow their routine travel routes even if this route is having more congestion during the Olympics; (ii) selfish: travelers have good knowledge of the traffic situation and each of them will choose the route with shortest travel time, which follows the UE model; (iii) altruism: travelers follow the travel routes for the best case scenario for the collective travel time. In this case, the travel route of each traveler is assigned taking into account their effects in the total travel time. We evaluate the results of routing strategy both on taxis and residential vehicles. The traffic states on the roads are diverse under the three scenarios (see Supplementary Figure 9).

Figure 3a and 3b illustrate the box plot of the distribution of tourists’ travel times during the morning and evening peak hour on 10 weekdays, respectively. The habit scenario always perform worse than selfish and altruism as local travelers will not give way to tourists and get considerable increase in their journeys. Selfish and altruism scenarios, by contrast, allow travelers to choose their routes toward their own or others’ benefit. Interestingly, in the morning peak hour, tourists’ travel times under altruism are globally similar with selfish, while they are much worse than selfish and habit in the evening. The reason is that in the morning, the flow direction of tourists (mainly from urban to suburban) is opposite to most of the commuting trips (mainly from suburb to the urban core). While in the evening peak hour, more commuters have similar direction with tourists (mainly from the urban core to the suburbs). In this case, under altruism some taxis would detour, taking a longer travel time than selfish and habit.

Further, we evaluate the impact of the Olympics on local commuters, calculating the average percentage increment of commuter’s travel times as

\[
I_{\text{comm}} = \frac{\sum_{o,d} (t_{o,d}^{\text{Olym}} - t_{o,d}^{\text{before}}) f_{o,d}^{c} \cdot 100\%}{\sum_{o,d} t_{o,d}^{\text{before}} f_{o,d}^{c}}
\]

where \(o,d\) is one of all the OD pairs; \(t_{o,d}^{\text{Olym}}\) refers to the number of commuters; \(t_{o,d}^{\text{Olym}}\) and \(t_{o,d}^{\text{before}}\) refer to the travel time in the \(o,d\) route before and during the Olympics, respectively. \(I_{\text{comm}}\) can be negative as selfish or altruism allows that some commuters find shorter paths than before. Figure 3c and 3d depict the distribution of commuters travel time in a log scale on weekdays. More people have larger travel times \((I_{\text{comm}} > 20\%)\) under the habit scenario than under selfish or altruism scenarios. Moreover, in contrast with selfish, altruism raises the number of commuters suering longer travel time but earns overall benets for the majority of commuters. Figure 3e and 3f illustrate the average percentage increment per day. The increment with the habit scenario is always larger, then followed by selfish and altruism. Furthermore, certain peak hours are subject to most serious delays, e.g., morning peaks on August 09 and 16, evening peaks on August 12, 15, and 19. This is the essence of the altruism strategy: while a small fraction of people sacrifices with longer travel times via detours to less popular routes [11], the overall saving in travel time is larger than in the selfish strategy. While in previous work it was already observed that altruism vs. selfish strategies do not produce large differences [11], here we see that both strategies have considerable differences with the habit scenario. This shows the benefits of information technologies to help decrease congestion during the events when people can select alternative routes different from their routine routes.

To further evaluate this effect we see the effects of the interplay among habit and selfish, meaning a fraction of people changing routes toward their shortest travel times, and others keeping their routine routes. To examine such intermediate states, we define a selfish parameter \(\Lambda\) to account for the fraction of selfish travelers. \(\Lambda\) ranges from 0 to 1, where 0 implies the habit scenario, and 1 implies the selfish scenario. Specifically, the travelers in each OD pair seek their shortest travel time with a percentage of \(\Lambda\) and their routes need to be reassigned, others are following their habit routes. For each link, it can be occupied by habit flow and selfish flow. The habit flow is calculated as \(v_e^{\text{habit}} \cdot (1 - \Lambda)\), where \(v_e^{\text{habit}}\) is the link volume under habit scenario. The selfish flow \(v_e^{\text{selfish}}\) is obtained by assigning the selfish demand using UE model. Therefore, the VoC is calculated by:

\[
\text{VoC}_e = \frac{v_e^{\text{habit}} (1 - \Lambda) + v_e^{\text{selfish}}}{C_e}
\]

and the BPR function in Equation 1 is used to estimate the travel time on each link. For each OD pair, the total commuting time also contains two parts: \((1 - \Lambda) \cdot f_{o,d}^{c} \cdot t_{o,d}^{\text{habit}}\) and \(\Lambda \cdot f_{o,d}^{c} \cdot t_{o,d}^{\text{selfish}}\), where \(t_{o,d}^{\text{habit}}\) is the travel time under the habit scenario and \(t_{o,d}^{\text{selfish}}\) is the shortest travel time under the selfish parameter \(\Lambda\). Figure 3g and 3h indicate the average increment for commuters on each weekday with different values of \(\Lambda\). The increment percentage decreases with the increase of \(\Lambda\), indicating that the impact of the Olympics recedes if more travelers are selfishly looking for their best routes as opposed to using their routine routes.

Most of the transportation planning strategies designed to reduce motorized vehicles are applied independently of origin and destination of the travelers, in consequence they are very costly in terms of the percentage of car reduction (usually 20% of the cars selected by the ending digit in the plates). They achieve very modest benefits in travel times, usually of the order of 2% [40]. Based on the estimation of travel delays of commuters under the selfish scenario, we explore the spatial impact...
of the Olympics on commuters. To achieve this, we average the percentage increment of commuter trips to origin and destination zones. Results indicate that commuters who live in the northeast of Rio suffer serious impact in the morning peak hour (see Supplementary Figure 10). In addition, people working in the eastern coastal area suffer travel delays the most in both the morning and the evening peak hours. We also find that the densely populated Governador Island suffers critical delays as one of the two bridges between the island and mainland are set as Olympic lanes.

To facilitate the policy-making, we visualize the travel time before and during the Olympics all over the metropolitan area of Rio, as shown in Figure 4. From the visualization, travelers can explore their travel time increment during the peak hours due to the Olympics. Besides, the platform provides the travel time under different scenarios, which helps the travelers and policy makers realize the collective benefits generated by the travel demand management strategy. This is just a proof of concept that needs to be managed with a more smooth access to the information for the users to be scalable.

C. Informed mode change

Aiming to mitigate the traffic congestion during the Olympics, the government of Rio de Janeiro has made important investments, such as enhancing the capacity of the traffic network, extensions of the public transporta-

tion infrastructures, including construction of new metro and BRT lines. As a complement to those efforts, in this work we propose an efficient strategy to manage the travel demand with the present transport infrastructure, concretely, reducing a fraction of vehicle demand toward relieving congestion during the peak period to the most extent.

With the purpose of selecting the critical trips to reduce, we evaluate the contribution of each OD trip to the collective travel time. Namely, we consider the following question: how much time will we save collectively if we take one vehicle out from a given OD route? We represent the road network as a directed acyclic graph \( G(\mathcal{N}, \mathcal{E}) \), where \( \mathcal{N} \) is the set of nodes, and \( \mathcal{E} \) is the set of directed edges. After assigning the travel demand to the road network, each road segment \( e \in \mathcal{E} \) associated with volume \( v_e \) and travel time during traffic \( t_e \).

First, for a road segment, we estimate the travel time saving of others if we reduce one vehicle using the marginal edge cost, which is the partial gradient of total travel time over the current volume. For each edge, we have:

\[
MC_e = \frac{\partial(t_{e}(v_e))}{\partial v_e}
= t_e(v_e) + f_s \alpha \beta \left( \frac{v_e}{C_e} \right)^{\beta} \times t_e^{f}
\]

(4)

where the edge travel time \( t_e \) is calculated using the calibrated BPR function in Equation 1. The marginal edge cost \( MC_e \) consists of two terms: the first one \( t_e \) reflects
FIG. 4. Interactive visualization of travel times before the Olympics and during the Olympics via various strategies of mobility. The purple hexagon reflects the origin of trips. The white hexagons are associated with the Google travel times for comparison. The colors of other hexagons reflect the travel time from the chosen origin to them. Results are presented on-line at www.flows-rio2016.com.

The travel time of one vehicle and the second would be the saved travel time by other vehicles in the same edge. The travel route $p_{i,j}$ of each OD trip $(i,j)$ is the set of edges on the path. Consequently, we calculate the marginal path cost of OD pair $(i,j)$ as the sum of $MC_e$ for the edges is traversed by the path:

$$MC_p = \sum_{e \in E} \delta_{ep} MC_e$$

(5)

where $\delta_{ep}$ is the delta function, which is 1 if edge $e$ is traversed by path $p$, 0 otherwise.

Larger values of $MC_p$ indicate more collective travel time would be saved if we take the trip out. Consequently, a sensible strategy is to reduce the demand from top-ranked OD pairs. To formulate a feasible strategy, we only consider the trips which origins and destinations both located nearby the metro or BRT stations, which means these people could switch to public transport other than driving. In our experiments, we define the maximum distance from the centroid of a zone to the nearest station. We evaluate the effects of this distance defined as $1km$, $2km$, or $3km$. First, we select the trips with origins and destinations within distance to the nearest station and destination within the metro or BRT station. Then, calculate the $MC_p$ for each trip and reduce 60% demand from the top-ranked trips. The number of top-ranked trips ranges from 1,000 to 10,000. Finally, we reassign the remainder demand to the road network and check the reduction of the collective travel time. As a benchmark, we keep the same number of total reduced trips but uniformly distribute them to all OD pairs near Metro and BRT stations.

Figure 5a illustrates the reduction of collective travel times as a percentage of the travel time before the strategy, which approximately follows a linear relationship with the number of reduced OD pairs. Interestingly, in contrast to the uniform benchmark, the strategy based on marginal costs is more effective by a factor of five. For example, reducing 60% of the flow from the selected 5000 OD pairs at the range of $2km$, represents 1.14% of the total vehicle flows. In that case, the reduction in the percentage of collective travel time is more than 10% with the marginal cost strategy and only 2% with the uniform benchmark case. In addition, different distance thresholds produce similar results for the same number of OD pairs. However, as shown in Figure 5b, greater distances indicate a lower percentage in the total flow. This is because larger distances provide more options to the collective travel time saving and does affect the car ridership in a less concentrated fraction.

Figure 5c presents the spatial distribution of the reduced car demand for $\{2km, 6000 OD pairs\}$. This case reduces the collective travel time by 10.6% at the expense of 1.4% decrease of the total car demand, and improve the average speed of all vehicles from $37.08km/h$ to $39.94km/h$. Implying that a considerable fraction of the travel times of local commuters and tourists decrease, especially for the travelers with long trips (see Supplementary Figure 11). Interestingly, the distribution of destinations concentrates a very small area in the Centro of Rio (Downtown). Meanwhile, the distribution of origins concentrates in two areas, the west end of the BRT line and the west end of the metro line. This suggests that people living in two neighborhoods in the West Zone of Rio (e.g. Santa Cruz and Paciência) and three neighborhoods in the North Zone of Rio (e.g. Guadalupe, Marechal Hermes, and Bento Ribeiro) would have to switch from driv-
FIG. 5. Informed changes from vehicles to transit during the morning peak hour. (a) Collective travel time reduction under different strategies. The reduction of collective travel time rises linearly over the number of reduced OD pairs. The slopes of the marginal strategy are $1.45 \times 10^{-3}$, $1.64 \times 10^{-3}$, and $1.67 \times 10^{-3}$ for 1km, 2km, and 3km, respectively; while the slopes of the uniform benchmark strategies are $0.47 \times 10^{-3}$, $0.39 \times 10^{-3}$, and $0.40 \times 10^{-3}$ for 1km, 2km, and 3km, respectively. The informed strategy reduces the collective travel time more than 5 times the reduction of the uniform case. (b) Reduced demand under different threshold distances. As expected larger number of OD pairs imply larger percentage of the demand reduced. (c) Added ridership to the Metro/BRT line and reduced vehicles around stations with configuration {2km, 6000 OD pairs}. The width of the metro and BRT line reflects the increased ridership by strategy. Blue and red in different areas reflect the origin and destination of reduced demand, respectively. Darker colors imply more people switching to transit from BRT or metro lines during the morning peak hour, if they work in Downtown. Moreover, Figure 5c gives the additional ridership in each segment of the metro and BRT line. As can be seen, the maximum increase is 5,000 travelers in the morning peak hour, in contrast with the capacity of metro and BRT, $\sim$ 30,000 passengers per hour per direction. If the current capacity of BRT can not meet the needs of mode change strategy, it is convenient and economical to add buses to the current BRT system. Finally, to investigate the impact in travel times of the proposed strategy on the individuals changing modes, we compared their average driving travel time with the transit time from the Google API (GTFs). Based on these estimates, average travel time would drop from 96.3 mins to 80.5 mins if taking BRT and metro during the morning peak. This may be encouraging to the individuals to cooperate with the proposed travel demand management strategy.

III. DISCUSSION

Mega events can greatly benefit the host city in many aspects, such as attracting investment and tourism and stimulating the economy. Nevertheless, it also exerts disruptions in the routine of the city. One of the most feared
costs by the population is the increase in travel times, especially for already dense cities, which are more likely to host the event. In the run-up to the Olympics, city planners need estimates on how the traffic will be affected, in order to establish proper policies to cope with the impact. However, the current impact evaluation on travelers is mostly confined to qualitative studies with anecdotal experience of events management. We lack quantitative methods to support the strategies. This is mostly due to difficulties of data availability to estimate travel demand. In this work, we present a method to estimate urban travel demand and the time increments to commuters during a large event by integrating multiple and large scale data resources. Moreover, we evaluate the effects of various routing strategies in the increase in congestion.

As a case of study, we analyzed the 2016 Summer Olympics in Rio de Janeiro. The large inflow of tourists increases the travel demand while the establishment of Olympic lanes decreases the road network supply. The first task is to estimate the rise in the demand-to-supply ratio in the streets and how this will affect travel times. First, we estimate the person and vehicle travel demands during the Olympics in Rio by estimating the number of tourists and their travel modes. In particular, we expect a greater number of tourists traveling during the morning peaks of August 8th, 12th, and 15th, as well as the evening peaks of August 12th, 15th, 16th, and 17th. By estimating the routes of vehicles under three distinct scenarios, habit, selfish, and altruism, we assess quantitatively the impact of the Olympics on commuters. We find that the habit scenario produces the greatest travel times, followed by selfish and altruism. For some peak hours, the increment in the percentage of travel times of all commuters can be up to 7% if people follow their routine routes. The selfish scenario, which is the maximum benefit possible via changing routes, still produces about 5% of the increment for the most affected peak hours. This is in agreement with the magnitude of savings reported by Çolak et al. [11] in routine conditions. They showed that the collective travel times could be decreased at most by 4.7% – 7.7% by routing strategies (altruism).

The most effective strategy to reduce traffic is informed mode change. This improves existing practices of restricting cars by the ending digit of the plate numbers. To generate this strategy, we calculate the contribution of each OD pair to the collective travel time. Namely, the drivers who are mostly involved in traffic bottlenecks are encouraged to change from driving to public transportation. Finally, by reducing 1% of the total cars, but targeting the zones near metro and BRT lines, the decrease of overall travel time reaches about 9%. Wang et al. reported that 1% target decrease in demand can achieve 14% and 18% decrease in travel times for the San Francisco Bay Area and Boston, respectively [40]. However, the proposed countermeasures did not consider the existing alternatives for travel modes. In contrast, our strategy only targets drivers within 3km of public transportation both in their origins and their destinations. For incentives, discounts for transit and ridership services in the selected communities could be tested.

Overall, we propose a methodology to give travel recommendations to the users toward their collective benefits using information technologies. Specifically, we showed that the use of information to target mode change can be the most cost-effective alternative to increasing capacity in transportation. This information-based approach is convenient not only for relieving congestion, but also potentially increasing the use of public transport, which would deliver better environmental outcomes, stronger communities, and more sustainable cities. We have estimated how the travel demand in each zone contributes differently to the overall congestion; these results can be helpful for the planning of routes of public transportation. In future studies, we can calculate the reduction in emissions associated with the improvement in travel times when taking one vehicle out from the selected OD pairs, thereby managing vehicle demand to improve air quality. The data resources used in our work are the byproducts of the use of communication technologies (CDRs, Waze) or open source repositories (event schedule, venue property, Airbnb, hotels, OpenStreetMap). Consequently, the proposed methods are portable for events in other cities. Meanwhile, as the data resources are becoming more abundant, our work represents a feasible application for demand prediction and management that improves urban well-being.

We evaluated three ideal scenarios and their impacts during the Olympics. We expect that the most likely routing behavior to be between habit and selfish, meaning that only a fraction of the population may find their shortest routes while the others will follow their habit routes before the Olympics. To have an idea of such scenarios, we have defined a selfish parameter A, and report the results for different values that go from the habit to selfish case.

An interesting avenue is to estimate empirical routing behavior [41]. Collecting data about individual route choices before and after the event is useful to understand the changes of behavior during large events. Also, comparing the actual change of traffic conditions from Waze during the Olympics vs. our estimates, as well as the comparison of other mitigation strategies, such as ridesharing, and changes in the departure times of travelers, can be key emphasis in the future work.

IV. METHOD

A. Data sets

The data resources used in this work are: mobile phone data (CDRs), Waze data, camera data, Airbnb data, hotel data, the Olympic game schedules and information of venues, as well as the OpenStreetMap. CDRs consist of 5 months of 2.19 million anonymous users and
are used to estimate the 24-hour routine ODs before the Olympics; Waze data sets contain 4.6 million reports during 7 months and are used to extend the 24-hour ODs to 5 weekdays. We argue that the larger number of cars. Also, camera data sets provided the relation between traffic volumes and average speeds in 85 main streets and are used to calibrate the relationship between volume-over-capacity and actual travel time. Airbnb data sets were crawled on the website of Airbnb in January 2015 [35], and contain 13,400 properties and each property provides its location and the number of accommodations available. We estimate the distribution of tourists’ residences using Airbnb data set together with 106 hotels information. OpenStreetMap provides the road network we used in our demand assignment. Game schedules and locations and capacities of the venues are used for the estimating the tourists’ destination and departure times. Among the data sets, CDRs, Waze data, and camera data are the byproduct of the activity. Other data sets are all publicly available (see Supplementary Figures 1-5, 7, and Notes 1-4).

B. Tourists travel mode split

In order to estimate the added vehicle demand during the Olympics, the taxi demand of tourists must be calculated from the tourists total demand in each hour. We define four mobility modes for tourists: walking and metro/BRT, bike and metro/BRT, taxi, and bus. The reason we merge the metro line and BRT lines together is that they are a closed-loop, as shown in Figure 1. Walking and metro/BRT implies the origin and destination of tourists is near enough to the stations (1km); Bike and metro/BRT implies they are near enough for biking (2km). The tourists will consider bus if its travel time and number of transfers are both acceptable (less than 3). Otherwise, they will choose taxi to the venues. We assume the occupancy of taxis by tourists is 2.0, meaning two tourists will take one taxi in average during the Olympics (see Supplementary Figure 6 and Note 3).

C. Travel time estimation

To estimate travelers’ delay during the Olympics, we represent their routing trips before and during the Olympics using a traffic assignment model. Traffic assignment aims to estimate the travel time and volume on each road segment. The estimation is implemented by appointing reasonable (usually shortest travel time) travel paths for all of the trips from their origin to destination. Before the Olympics, we assume that all travelers have found their routes with shortest travel time and assign the demand with UE model. To validate the estimated travel time, we compare the travel times of top 5000 commuter OD pairs with Google maps APIs during the morning peak hour. The results show that the estimates are reasonable (see Supplementary Figure 3). During the Olympics, both of the demand and the capacity of the road networks change. For the habit scenario, all of the travelers follow their routes before the Olympics, we update the volume and travel time on each edge only considering the additional tourists flow. Tourists’ routes are chosen according to the shortest travel path before the Olympics. For the selfish scenario, we assign the new demand with UE model as before the Olympics. For altruism scenario, we calculate the shortest path with respect to the marginal cost for each OD pair, which makes the entire road network reach system optimum. Aiming at a more realistic estimates of travelers’ routing during the Olympics, we argue that only a fraction of people can find their shortest path, which means one fraction of the drivers follow their routine routes, while the remaining fraction is assigned using the UE model to the available space (see Supplementary Figure 9, 10, and Note 5).

ACKNOWLEDGMENTS

We thank Shan Jiang and Yingxiang Yang for informative discussions and code at the beginning of this project. Armin Akhavan for helpful suggestions with the visualization platform and the collection of the Airbnb data. We specially thank all the members of the Sala Pensa of Rio de Janeiro’s Major Office, lead by Pablo Cerdeira for discussions and data provided. We thank Prof. Alexandre G. Evsukoff for data and contextual information on Rio de Janeiro. The research was partly funded by Ford, the Department of Transportation’s grant of the New England UTC Y25, the MIT Portugal Program, the MIT-Brazil seed Grants Program, the Center for Complex Engineering Systems at KACST-MIT and the MIT-QCRI research collaboration.


J. L. Peel, M. Klein, W. D. Flanders, J. A. Mulholland, and P. E. Tolbert, Research Report (Health Effects Institute) 3 (2010).


C. Black and E. Schreffler, Transportation Research Record: Journal of the Transportation Research Board, 81 (2010).


