Understanding Road Usage Patterns in Urban Areas

Pu Wang[1,2], Timothy Hunter[4], Alexandre M. Bayen[4,5], Katja Schechtner[6,7] & Marta C. González[2,3]

[1]School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, 410000, P.R. China, [2]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA, [3]Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA, [4]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, 94720, USA, [5]Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, CA, 94720, USA, [6]Austrian Institute of Technology, Vienna, Vienna, 1210, Austria, [7]MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA.

In this paper, we combine the most complete record of daily mobility, based on large-scale mobile phone data, with detailed Geographic Information System (GIS) data, uncovering previously hidden patterns in urban road usage. We find that the major usage of each road segment can be traced to its own - surprisingly few - driver sources. Based on this finding we propose a *network of road usage* by defining a bipartite network framework, demonstrating that in contrast to traditional approaches, which define road importance solely by topological measures, the role of a road segment depends on both: its betweeness and its degree in the road usage network. Moreover, our ability to pinpoint the few driver sources contributing to the major traffic flow allows us to create a strategy that achieves a significant reduction of the travel time across the entire road system, compared to a benchmark approach.

I n an era of unprecedented global urbanization, society faces a rapidly accelerating demand for mobility, placing immense pressure on urban road networks[1,2]. This demand manifests in the form of severe traffic congestion[3,4], which decreases the roads' level of service, while at the same time increasing both fuel consumption[5] and traffic-related air pollution[6]. In 2007 alone, congestion forced Americans living in urban areas to travel 4.2 billion hours more, purchase an additional 2.8 billion gallons of fuel, at a total cost of $87.2 billion[3]. To mitigate congestion in urban roads, urban planning[1], traffic prediction[7–9] and the study of complex networks[10–15] have been widely investigated potential influencing factors. However, without comprehensive knowledge of how roads are used dynamically, these studies are conventionally based on expensive and quickly outdated travel surveys or segmented information on traffic flow and travel time[7–9], which fail to support the researchers with the information needed to cope with modern mobility demand. Up to now our understanding of the origins of the drivers in each road remained limited and not quantitatively solved.

In this work we validate for the first time a methodology, which employs comprehensive mobile phone data to detect patterns of road usage and the origins of the drivers. Thus, providing a basis for better informed transportation planning, including targeted strategies to mitigate congestion[3,4]. We formalize the problem by counting the observed number of individuals moving from one location to another, which we put forward as the transient origin destination (*t*-OD) matrix (Fig. S5, Fig. S11 and Supplementary Information (SI) section II.A). Traditionally, ODs are costly and difficult to obtain, because they are at best based on travel diaries made every few years, which quickly become obsolete and strongly rely on provided reports[7,8]. In contrast, the rapidly increasing penetration rates and massive usage of mobile phones, with towers densely located in urban areas, can provide the most detailed information on daily human mobility[16–20] across a large segment of the population[19–25]. Thus we use three-week-long mobile phone billing records generated by 360,000 San Francisco Bay Area users (6.56% of the population, from one carrier) and 680,000 Boston Area users (19.35% of the population, from several carriers) respectively. This data set is two orders of magnitude larger in terms of population and time of observation than the most recent surveys (Table S1), providing us with a source at an unprecedented scale to generate the distribution of travel demands.

To study the distribution of travel demands over a day we divide it into four periods (Morning: 6 am–10 am, Noon & Afternoon: 10 am–4 pm, Evening: 4 pm–8 pm, Night: 8 pm–6 am) and cumulate trips over the total observational period. A trip is defined when the same mobile phone user is observed in two distinct zones within one hour (zones are defined by 892 towers' service areas in the San Francisco Bay Area and by 750 census tracts in the Boston Area). In the mobile phone data, a user's location information is lost when he/she does not use his/her

phone, but by defining the transient origin and destination with movements within one hour, we can capture the distribution of travel demands. Specifically we calculate the t-OD as:

$$t-\mathrm{OD}_{ij} = W \times \frac{F^{vehicle}_{ij}}{\sum^{A}_{ij} F^{all}_{ij}} \qquad (1)$$

where A is the number of zones. W is the one-hour total trip production in the studied urban area, a number readily available for most cities. However this number gives no information about the trip distribution between zones, which we can enhance by the information gained via mobile phones. Directly from the mobile phone data we calculate $T_{ij}(n)$, which is the total number of trips that user n made between zone i and zone j during the three weeks of study. Via calibrating $T_{ij}(n)$ for the total population we obtain: $F^{all}_{ij} = \sum^{N_k}_{n=1} T_{ij}(n) \times M(k)$, where N_k is the number of users in zone k. The ratio M scales the trips generated by mobile phone users in each zone to the trips generated by the total population living there: $M(k) = N_{pop}(k)/N_{user}(k)$, where $N_{pop}(k)$ and $N_{user}(k)$ are the population and the number of mobile phone users in zone k. Furthermore to assign only the fraction of the trips attributed to vehicles, we correct F^{all}_{ij} by the vehicle usage rate, which is a given constant for each zone and therefore obtain $F^{vehicle}_{ij}$ (see SI section II.A).

For each mobile phone user that generated the t-OD, we can additionally locate the zone where he or she lives, which we define as the *driver source*. Connecting t-ODs with driver sources allows us for the first time to take advantage of mobile phone data sets in order to understand urban road usage. In the following, we present the analysis of the road usage characterization in the morning period as a case study. Results for other time periods are presented in SI (Fig. S19).

Results

A road network is defined by the links representing the road segments and the nodes representing the intersections. Using incremental traffic assignment, each trip in the t-OD matrix is assigned to the road network[26], providing us with estimated traffic flows (Fig. 1a). The road network in the Bay Area serves a considerable larger number of vehicles per hour (0.73 million) than the one in the Boston (0.54 million). The traffic flow distribution $P(V)$ in each area can be

well approximated as the sum of two exponential functions corresponding to two different characteristic volumes of vehicles (Fig. 1a); one is the average traffic flow in their arterial roads (v_A) and the other is the average traffic flow in their highways (v_H). We measure $P(V) = p_A v_A e^{-V/v_A} + p_H v_H e^{-V/v_H}$ ($R^2 > 0.99$) with $v_A = 373$ (236) vehicles/hour for arterials and $v_H = 1,493$ (689) vehicles/hour for highways in the Bay Area (Boston numbers within parenthesis, p_A and p_H are the fraction of arterial roads and the fraction of highways). Both road networks have similar number of arterials (~20,000), but the Bay Area with more than double the number of highways than Boston (3,141 highways vs 1,267 in Boston) still receives the double of the average flow in the highways (v_H) and a larger average flow in the arterial roads.

The volume of vehicles served by a road depends on two aspects: the first is the functionality of the road according to its ability to be a connector based on its location in the road network (i.e. betweenness centrality) and the second is the inherent travel demand of the travelers in the city. The betweenness centrality b_c of a road segment[27–30] is proportional to the number of shortest paths between all pairs of nodes passing through it: we measured b_c by averaging over each pair of nodes, and following the shortest time to destination. The two road networks, analyzed here, have completely different shapes: the Bay Area is more elongated and connects two sides of a bay, while the Boston Area follows a circular shape (see Fig. 2a). But both have a similar function in the distribution of b_c: with a broad term corresponding to the arterial roads and an exponential term to the highways, which is at the tail of larger b_c. As Fig. 1b shows, we measure: $P(b_c) = p_A P_A(b_c) + p_H P_H(b_c)$ ($R^2 > 0.99$), with $P_A(b_c) \sim b_c^{-\alpha_A}$ for arterial roads and $P_H(b_c) \sim e^{-b_c/\beta_H}$ for highways. The highways in the Bay Area have an average b_c of $\beta_H = 2.6 \times 10^{-4}$, whereas a larger $\beta_H = 4.6 \times 10^{-4}$ is found for the Boston Area highways, indicating their different topological structures. Interestingly, despite, the different topologies of the two road networks, the similar shapes of their distribution of traffic flows indicate an inherent mechanism in how people are selecting their routes.

Notice that only when the traffic flow is greater than a road's available capacity, the road is congested; the ratio of these two quantities is called Volume over Capacity (VOC) and defines the level of service of a road. Surprisingly, despite the different values in average flows v and average betweeness centrality β, we find the same
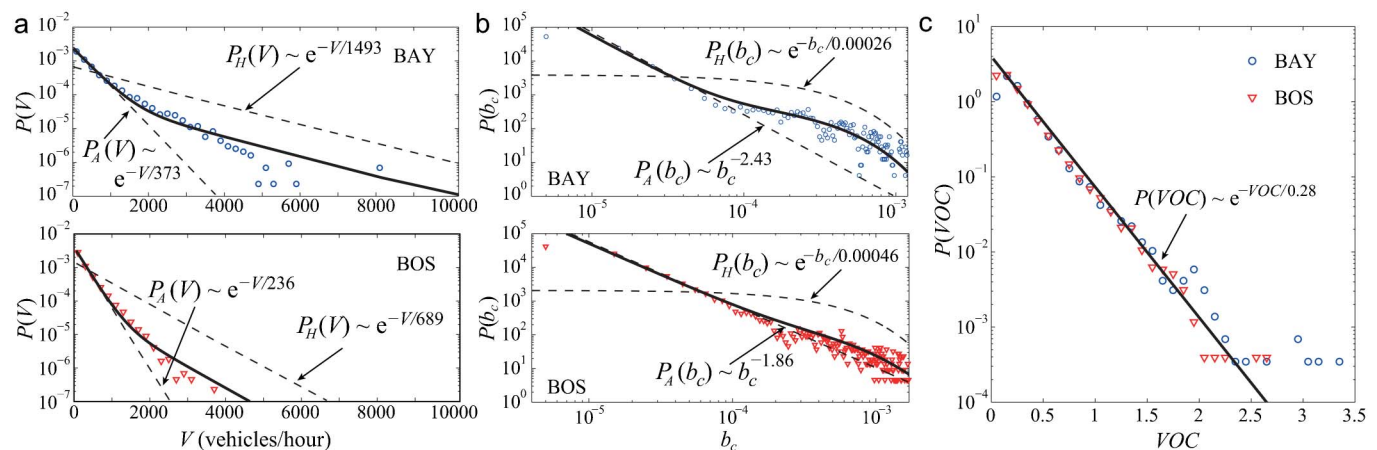

Figure 1 | Distributions of traffic flow, betweenness centrality and *VOC* in the two urban areas. (a) The one-hour traffic flow V follows a mixed exponential distribution $P(V) = P_A v_A e^{-V/v_A} + p_H v_H e^{-V/v_H}$ for both Bay Area and Boston Area, where constants p_A and p_H are the fraction of arterial roads and the fraction of highways, v_A and v_H is the average traffic flow for arterial roads and highways respectively. (b) The distribution of road segment's betweenness centrality b_c is well approximated by $p(b_c) = p_H \beta_H e^{-b_c/\beta_H} + p_A c_A b_c^{-\alpha_A}$, where the power-law distribution approximates arterial roads' b_c distribution and the exponential distribution approximates highways' b_c distribution. β_H denotes the average b_c of highways and α_A is the scaling exponent for the power-law. (c) The volume over capacity VOC follows an exponential distribution $P(VOC) = \gamma e^{-VOC/\gamma}$ with an average $VOC = 0.28$ for the two areas. Traffic flows in most road segments are well under their designed capacities, whereas a small number of congested segments are detected. For more statistical analysis of the fits, see the detailed discussion in SI section III.B.
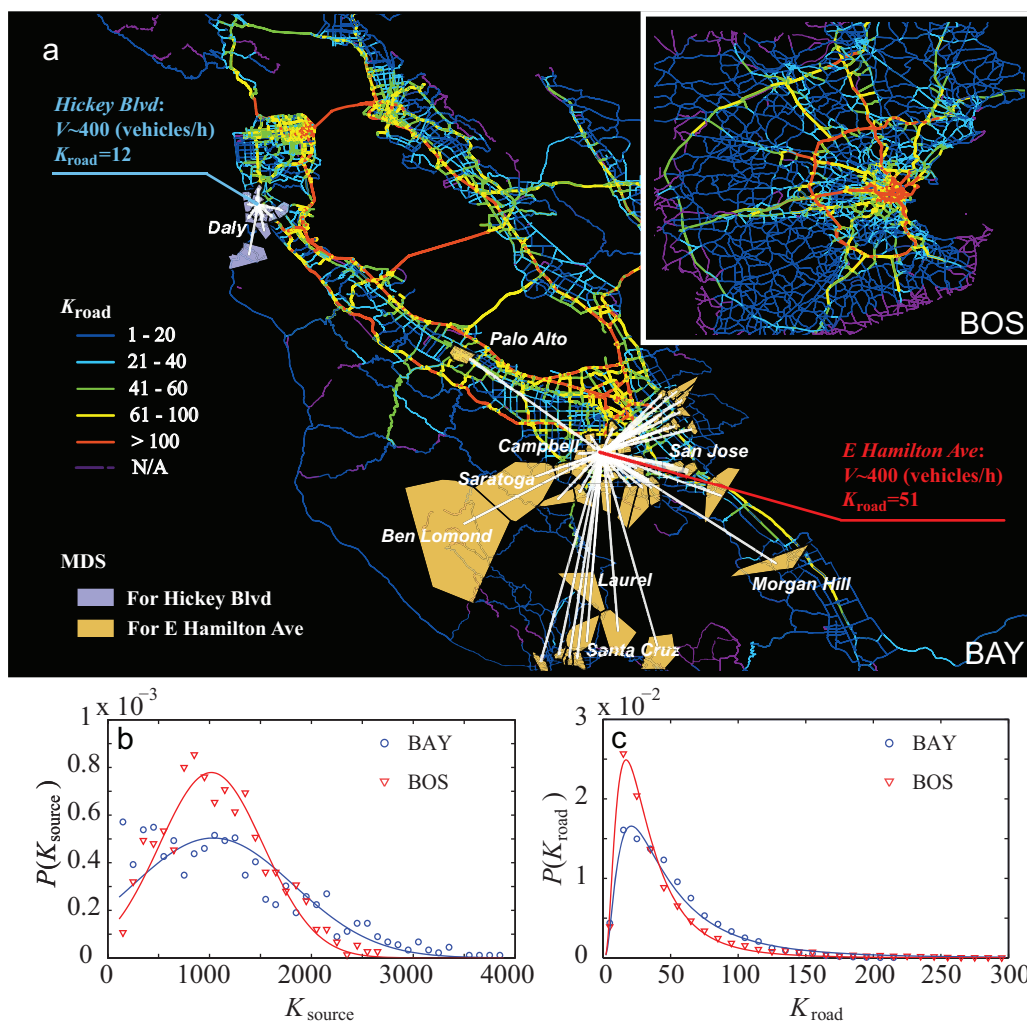
Figure 2 | Tracing driver sources via the road usage network. (a) The colour of a road segment represents its degree K_{road}. Most residential roads are found to have small K_{road}, whereas the backbone highways and the downtown arterial roads are shown to have large K_{road}. The light blue polygons and the light orange polygons pinpoint the MDS for *Hickey Blvd* and *E Hamilton Ave* respectively. The white lines show the links that connect the selected road segment and its MDS. The two road segments have a similar traffic flow $V \sim 400$ (vehicles/hour), however, *Hickey Blvd* only has 12 MDS located nearby, whereas *E Hamilton Ave* has 51 MDS, not only located in the vicinity of *Campbell City*, but also located in a few distant regions pinpointed by our methodology. (b) The degree distribution of driver sources can be approximated by a normal distribution $P(K_{source}) = e^{-(K_{source} - \mu_{source})^2 / 2\sigma_{source}^2} / (\sqrt{2\pi}\sigma_{source})$ with $\mu_{source} = 1{,}035.9$ (1,017.7), $\sigma_{source} = 792.2$ (512.3), $R^2 = 0.78$ (0.91) for Bay Area (Boston Area). (c) The degree distribution of road segments is approximated by a log-normal distribution $P(K_{road}) = e^{-(\ln(K_{road}) - \mu_{road})^2 / 2\sigma_{road}^2} / (\sqrt{2\pi}\sigma_{road}K_{road})$ with $\mu_{road} = 3.71$ (3.36), $\sigma_{road} = 0.82$ (0.72), $R^2 = 0.98$ (0.99) for Bay Area (Boston Area). For more statistical analysis of the fits, see SI section III.B.

distribution of *VOC* (Fig. 1c) in the two metropolitan areas, which follows an exponential distribution with an average *VOC* given by $\gamma = 0.28$ ($R^2 > 0.98$):

$$P(VOC) = \gamma e^{-VOC/\gamma} \qquad (2)$$

The exponential decay of *VOC* indicates that for both road networks traffic flows on 98% of the road segments are well below their designed road capacities, whereas a few road segments suffer from congestion, having a $VOC > 1$. The similarity between the two *VOC* distributions shows that in both urbanities drivers experience the same level of service, due to utilizing the existing capacities in the similar way.

The traditional difficulty in gathering ODs at large scales has until now limited the comparison of roads in regard to their attractiveness for different driver sources. To capture the massive sources of daily road usage, for each road segment with $V > 0$, we calculate the fraction of traffic flow generated by each driver source, and rank these sources by their contribution to the traffic flow. Consequently, we define a road segment's major driver sources (MDS) as the top

ranked sources that produce 80% of its traffic flow. We next define a bipartite network, which we call the *network of road usage*, formed by the edges connecting each road segment to their MDS. Hence, the degree of a driver source K_{source} is the number of road segments for which the driver source is a MDS, and the degree of a road segment K_{road} is the number of MDS that produce the vehicle flow in this road segment. As Fig. 2b shows, the driver source's degree K_{source} is normally distributed, centered in $< K_{source} > \sim 1000$ in both Bay Area and Boston Area, implying that drivers from each driver source use a similar number of road segments. In contrast, the road segment's degree K_{road} follows a log-normal distribution (Fig. 2c), where most of the road segments have a degree centered in $< K_{road} > \sim 20$. This indicates that the major usage of a road segment can be linked to surprisingly few driver sources. Indeed, only 6–7% of road segments are in the tail of the log-normal linked to a larger number of MDS, ranging from 100 to 300.

In Fig. 2a we show a road segment's degree K_{road} in the road network maps of the Bay Area and the Boston Area. Since census tracts and mobile phone towers are designed to serve similar number

of population (Fig. S2), a road segment's degree K_{road} quantifies the diversity of the drivers using it. We find that K_{road} is lowly correlated with traditional measures, such as traffic flow, VOC and betweenness centrality b_c (Fig. S15). For example, in Fig. 2a, *Hickey Blvd* in Daly City and *E Hamilton Ave* in Campbell City have a similar traffic flow $V\sim400$ (vehicles/hour), however, their degrees in the network of road usage are rather different. Hickey Blvd, only has $K_{road} = 12$, with MDS distributed nearby, whereas E Hamilton Ave, has $K_{road} = 51$, with MDS distributed not only in its vicinity, but also in some distant areas as Palo Alto, Santa Cruz, Ben Lomond and Morgan Hill.

As Fig. 2a shows, the road segments in the tail of the log-normal ($K_{road} > 100$) highlight both the highways and the major business districts in both regions. This again implies that K_{road} can characterize a road segment's role in a transportation network associated with the usage diversity. To better characterize a road's functionality, we classify roads in four groups according to their b_c and K_{road} in the transportation network (see Fig. 3 and Fig. S16). We define the *connectors*, as the road segments with the largest 25% of b_c and the *attractors* as the road segments with the largest 25% of K_{road}. The other two groups define the highways in the periphery, or *peripheral connectors*, and the majority of the roads are called *local*, which have both small b_c and K_{road} (Fig. 3). By combining b_c and K_{road}, a new quality in the understanding of urban road usage patterns can be achieved. Future models of distributed flows in urban road networks will benefit by incorporating those ubiquitous usage patterns.

Discussion

This novel framework of defining the roads by their connections to their MDS can trigger numerous applications. As a proof of concept, we present here how these findings can be applied to mitigate congestion. For a road segment, its level of congestion can be measured by the additional travel time t_e, defined as the difference between the actual travel time t_a and the free flow travel time t_f. The drivers who travel through congested roads experience a significant amount of t_e. To pinpoint these drivers, the total T_e per driver source is calculated. In contrast to the similar number of population served by each driver source (Fig. S2), the extra travel time T_e generated by driver sources can be very different, following an exponential distribution $P(T_e) = \tau e^{-T_e/\tau}$ (Fig. 4a). Some driver sources present a T_e 16 times larger than the average. This finding indicates that the major traffic flows in congested roads are generated by very few driver sources, which enables us to target the small number of driver sources affected

by this significantly larger T_e. For the Bay Area, the top 1.5% driver sources (12 sources) with the largest T_e are selected, for the Boston Area we select the top 2% driver sources (15 sources) (Fig. S17). We then reduce the number of trips from these driver sources by a fraction f, ranging from 2.7% to 27% in the Bay Area and from 2.5% to 25% in the Boston Area. The reduced numbers of trips correspond to the m total percentage of trips (m ranging from 0.1% to 1% for both areas). A benchmark strategy, in which trips are randomly reduced without identifying the driver sources with large T_e, is used as reference. Our results indicate that the selective strategy is much more effective in reducing the total additional travel time than the random strategy. In the Bay Area, the total travel time reduction δT increases linearly with m as $\delta T = k(m - b)$ ($R^2>0.90$). We find that when $m = 1\%$, δT reaches 26,210 minutes, corresponding to a 14% reduction of the total Bay Area additional travel time during a one hour morning commute (triangles in Fig. 4b). However, when a random strategy is used, the corresponding δT is only 9,582 minutes, which is almost three times less reduction than that achieved by the selective strategy (squares in Fig. 4b). Even better results are found in the Boston Area: using the selective strategy, when $m = 1\%$, δT reaches 11,762 minutes, corresponding to 18% reduction of the total Boston Area additional travel time during a one hour morning commute (diamonds in Fig. 4b), while the random strategy results only in $\delta T = 1,999$ minutes, which is six times less the reduction of that achieved by the selective strategy (circles in Fig. 4b). The underlying reason for the high efficiency of the selective strategy is intrinsically rooted in the two discoveries described above: first that only few road segments are congested and second that most of those road segments can be associated with few MDS.

Today, as cities are growing at an unparalleled pace, particularly in Asia, South America and Africa, the power of our modeling framework is its ability to dynamically capture the massive sources of daily road usage based solely on mobile phone data and road network data, both of which are readily available in most cities. Thus we validate for the first time an efficient method to estimate road usage patterns at a large scale that has a low cost repeatability compared to conventional travel surveys, allowing us to make new discoveries in road usage patterns. We find that two urban road networks with very different demand in the flows of vehicles and topological structures have the same distribution of volume over capacity (VOC) in their roads. This indicates common features in the organization of urban trips, which are well captured by the proposed bipartite network of road usage.
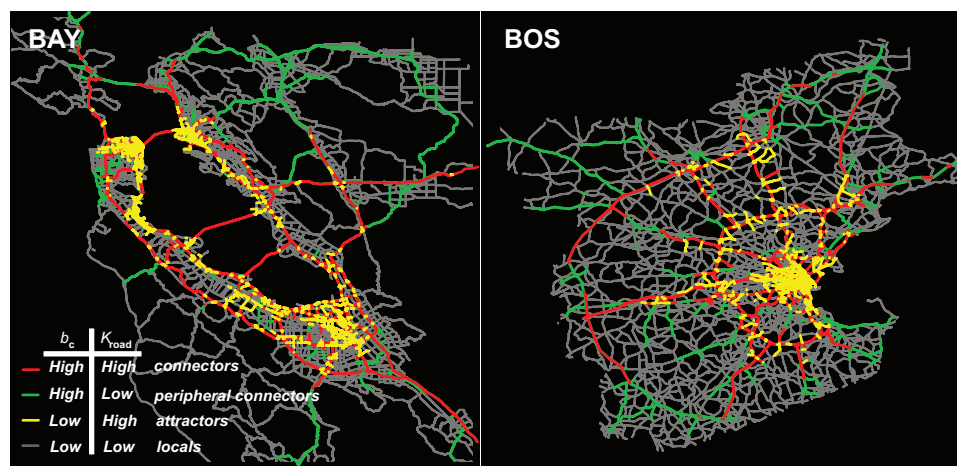

Figure 3 | Types of roads defined by b_c and K_{road}. The road segments are grouped by their betweenness centrality b_c and degree K_{road}. The red lines (connectors) represent the road segments with the top 25% of b_c and K_{road}; they are topologically important and diversely used by drivers. The green lines (peripheral connectors) represent the road segments in the top 25% of b_c, but with low values of K_{road}; they are topologically important, but less diversely used. The road segments in yellow are those with low values of b_c, but within the top 25% K_{road}; they behave as attractors to drivers from many sources (attractors). The road segments in grey have the low values of b_c and K_{road}, they are not topologically important and locally used (locals).
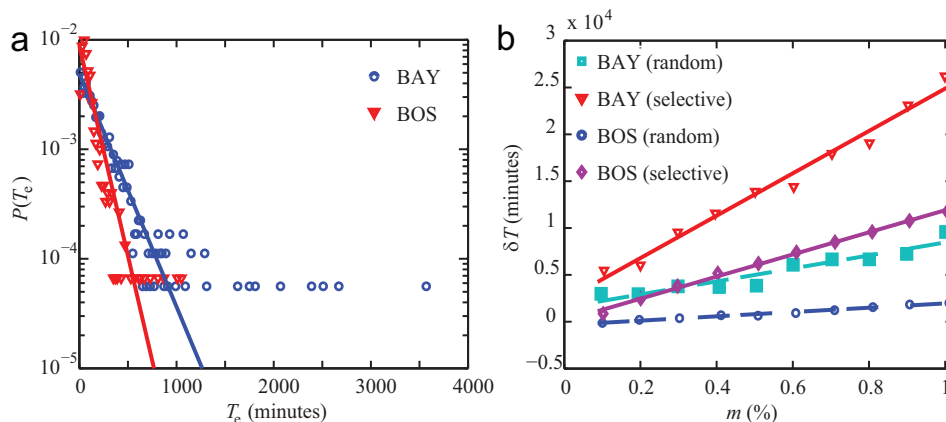
4

Figure 4 | A selective strategy based on the framework of road usage network shows better efficiency in mitigating traffic congestion. (a) The distribution of the additional time T_e for each driver source is well approximated with an exponential distribution $P(T_e) = \tau e^{-T_e/\tau}$. T_e is unevenly distributed in the two urban areas (also see Fig. S17). The maximum T_e (Bay Area 3,562 minutes and Boston Area 1,037 minutes) is significantly larger than the average T_e (204 minutes and 113 minutes respectively). (b) The total additional travel time reduction δT according to the trip reduction percentage m for the selective vs. the random strategy. The fits correspond to $\delta T = k(m-b)$, where k is the slope of the linear fit, b is close to zero for all fits. For a detailed statistical analysis of the fits, see the discussion in SI section III.B.

Based on our findings, a new quality in the understanding of urban road usage patterns can be achieved by combining the traditional classification method of assessing a road's topological importance in the road network, defined by b_c, with the novel parameter of a road's degree in the network of road usage, defined by K_{road}. The values of K_{road} and b_c together determine a road's functionality. We find that the major traffic flows in congested roads are created by very few driver sources, which can be addressed by our finding that the major usage of most road segments can be linked to their own surprisingly few driver sources. We show the representation provided by the network of road usage is very powerful to create new applications, enabling cities to tailor targeted strategies to reduce the average daily travel time compared to a benchmark strategy.

Methods

Incremental traffic assignment. The most fundamental method to assign trips to road network is provided by the classic Dijkstra algorithm[31]. Dijkstra's algorithm is a graph search algorithm that solves the shortest path problem for a graph with nonnegative edge path costs (travel time in our case). However, the Dijkstra algorithm ignores the dynamical change of travel time in a road segment. Thus to incorporate the change of travel time, we apply the incremental traffic assignment (ITA) method[26] to assign the t-OD pairs to the road networks. In the ITA method, the original t-OD is first split into four sub t-ODs, which contain 40%, 30%, 20% and 10% of the original t-OD pairs respectively. These fractions are the commonly used values[32]. The trips in the first sub t-OD are assigned using the free travel time t_f along the routes computed by Dijkstra's algorithm. After the first assignment, the actual travel time t_a in a road segment is assumed to follow the Bureau of Public Roads (BPR) function that widely used in civil engineering $t_a = t_f(1 + \alpha(VOC)^\beta)$, where commonly used values $\alpha = 0.15$ and $\beta = 4$ are selected[32]. Next, the trips in the second sub t-OD are assigned using the updated travel time t_a along the routes computed by Dijkstra's algorithm. Iteratively, we assign all of the trips in the four sub t-ODs. In the process of finding the path to minimize the travel time, we record the route for each pair of transient origin and transient destination (see SI section II.B for more detail).

Validating the predicted traffic flow by probe vehicle GPS data. Due to the lack of reliable traffic flow data at a global scale, we compare for each road segment the predicted travel time with the average travel time calculated from probe vehicle GPS data. According to the BPR function, the travel time of a road segment is decided by its traffic flow: a road segment's travel time increases with the increase of its traffic flow. Hence, obtaining the travel time from GPS probe data is an independent way to validate our result on the distribution of traffic flow. We find a very good linear relation $T_{prediction} = kT_{probe\ vehicle}$ with both travel times obtained independently (the coefficient of determination $R^2 > 0.9$ for all time periods, see SI section II.C for more detail).

Calculation of driver sources. A driver source is calculated from the mobile phone data based on the regularity of visits of mobile phone users at each time of the day[20]. This regularity is time dependent, and peaks at night when most people tend to be reliably at a home base with an average probability of 90% (Fig. S6B). Thus, we make a

reasonable assumption that a driver source is the zone where the user is mostly found from 9 pm to 6 am in the entire observational period.

1. Batty, M. The size, scale, and shape of cities. *Science* **319**, 769–771 (2008).
2. Barthélemy, M. Spatial networks. *Physics Reports* **499**, 1–101 (2011).
3. Schrank, D. & Lomax, T. *Annual urban mobility report* (Texas Transportation Institute, 2009).
4. Helbing, D. A section-based queueing-theoretical traffic model for congestion and travel time analysis in networks. *J. Phys. A: Math. Gen.* **36**, 593–598 (2003).
5. Chin, A. T. H. Containing air pollution and traffic congestion: transport policy and the environment in Singapore. *Atmospheric Environment* **30**(5), 787–801 (1996).
6. Rosenlund, M. *et al.* Comparison of regression models with land-use and emissions data to predict the spatial distribution of traffic-related air pollution in Rome. *Journal of Exposure Science and Environmental Epidemiology* **18**, 192–199 (2008).
7. Herrera, J. C. *et al.* Dynamic estimation of OD matrices for freeways and arterials. *Technical Report* (Institute for Transportation Studies, UC Berkeley, 2007).
8. Herrera, J. C. *et al.* Evaluation of traffic data obtained via GPS-enabled mobile phones: the mobile century field experiment. *Transportation Research C* **18**, 568–583 (2010).
9. Shen, W. & Wynter, L. Real-time traffic prediction using GPS data with low sampling rates: a hybrid approach. *IBM Research Report* RC25230 (2011).
10. Barthélemy, M. & Flammini, A. Modeling urban streets patterns. *Phys Rev Lett* **100**, 138702 (2008).
11. Youn, H., Gastner, M. T. & Jeong, H. Price of anarchy in transportation networks: efficiency and optimality control. *Phys Rev Lett* **101**, 128701 (2008).
12. Li, G. *et al.* Towards design principles for optimal transport networks. *Phys Rev Lett* **104**, 018701 (2010).
13. Wu, Z., Braunstein, L. A., Havlin, S. & Stanley, H. E. Transport in weighted networks: partition into superhighways and roads. *Phys Rev Lett* **96**, 148702 (2006).
14. Rosvall, M., Trusina, A., Minnhagen, P. & Sneppen, K. Networks and cities: an information perspective. *Phys. Rev. Lett.* **94**, 028701 (2005).
15. Jiang, B. Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science* **23**(8), 1033–1048 (2009).
16. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
17. Belik, V., Geisel, T. & Brockmann, D. Natural human mobility patterns and spatial spread of infectious diseases. *Phys. Rev. X* **1**, 011001 (2011).
18. Kölbl, R. & Helbing, D. Energy laws in human travel behavior. *New Journal of Physics* **5**, 48.1–48.12 (2003).
19. González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **435**, 779–782 (2008).
20. Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).
21. Eagle, N., Macy, M. & Claxton, R. Network diversity and economic development. *Science* **328**, 1029–1031 (2010).
22. Lu, X., Bengtsson, L. & Holme, P. Predictability of population displacement after the 2010 Haiti earthquake. *Proc. Natl. Acad. Sci. USA* **109**, 11576–11581 (2012).

23. Lambiotte, R. *et al.* Geographical dispersal of mobile communication networks. *Physica A* **387**, 5317–5325 (2008).

24. Sevtsuk, A. & Ratti, C. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *J. Urban Technol.* **17**, 41–60 (2010).

25. Batty, M. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals* (The MIT Press, 2007).

26. Chen, M. & Alfa, A. S. A Network design algorithm using a stochastic incremental traffic assignment approach. *Transportation Science* **25**, 215–224 (1991).

27. Newman, M. E. J. A measure of betweenness centrality based on random walks. *Social Networks* **27**, 39–54 (2005).

28. Porta, S. *et al.* Street centrality and the location of economic activities in Barcelona. *Urban Studies* 1–18 (2011).

29. Strano, E. *et al.* Street centrality vs. commerce and service locations in cities: a kernel density correlation case study in Bologna, Italy. *Environ. Plan. B: Plan. Design* **36**, 450–465 (2009).

30. Kurant, M. & Thiran, P. Layered complex networks. *Phys. Rev. Lett.* **96**, 138701 (2006).

31. Dijkstra, E. W. A note on two problems in connexion with graphs. *NumerischeMathematik* **1**, 269–271 (1959).

32. *Travel Demand Modeling with TransCAD 5.0, User's Guide* (Caliper, 2008).

Acknowledgements

Author contributions

PW and MG designed research; PW, TH, AB and MG performed research; PW, KS and MG wrote the paper.

Additional information