OPEN

Encapsulating Urban Traffic Rhythms into Road Networks

Junjie Wang*, Dong Wei*, Kun He, Hang Gong & Pu Wang

School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, 410000, P.R. China.

Correspondence and requests for materials should be addressed to P.W. (wangpu@csu.edu.cn)

* These authors contributed equally to this work.

Using road GIS (geographical information systems) data and travel demand data for two U.S. urban areas, the dynamical driver sources of each road segment were located. A method to target road clusters closely related to urban traffic congestion was then developed to improve road network efficiency. The targeted road clusters show different spatial distributions at different times of a day, indicating that our method can encapsulate dynamical travel demand information into the road networks. As a proof of concept, when we lowered the speed limit or increased the capacity of road segments in the targeted road clusters, we found that both the number of congested roads and extra travel time were effectively reduced. In addition, the proposed modeling framework provided new insights on the optimization of transport efficiency in any infrastructure network with a specific supply and demand distribution.

U rban road network, one of the crucial infrastructures in cities, facilitates people's daily commutes and maintains modern society's ability to function properly[1,2]. While each city has its own constraints in geography, history and socio-economic mechanisms, road networks from very diverse cities were all developed under the co-evolutions of urban travel demands and road network supplies[2]. Faced with rapidly accelerating travel demands, transportation agencies had tried to alleviate traffic congestion through various methods, which can be classified into two groups depending upon whether they increase road supply or decrease travel demand[3–6]. For example, building new roads can increase the road network supply, whereas reducing the usage of private vehicles is capable of decreasing demand of road usage.

Given the importance of urban road networks, their robustness and efficiency have drawn widespread attention in various scientific and engineering fields. Connectedness[1], spatial accessibility[7], betweenness centrality[8,9], the minimum spanning tree[10], the price of anarchy[11] and optimal navigation strategies[12–14] were all investigated to understand road network topology and improve transport efficiency[15,16]. Additionally, laws and models of human mobility were studied using large-scale mobility data to estimate urban travel demand[17–21]. These investigations, whether employing new statistical measures developed in complex network theory[22,23], or uncovering abundant information hidden in large-scale data, tremendously revolutionized the tools for understanding urban rhythms[24] and solving urban problems[25]. However, a modeling framework was still missing, which would encapsulate travel demand information into urban roads and dynamically adapt road networks to the time-variant and rapidly growing mobility demands in this era of unprecedented global urbanization.

Here, based on large-scale road GIS data and mobile phone data in the San Francisco Bay area and the Boston area (see Methods), a method was developed to target the road clusters most heavily used by drivers from time-variant sources of traffic congestion, thus providing a basis for better informed urban traffic control, which includes the mitigation of traffic congestion. To capture the dynamical urban travel demand, transient ODs (t-ODs) for the morning period (6:00–10:00 a.m.), the noon/afternoon period (10:00 a.m. to 4:00 p.m.) and the evening period (4:00–8:00 p.m.) were estimated (Fig. S2b). First, we defined a *trip* as a displacement occurring within one hour in a specific time period (Fig. S2a, see Methods for definition of displacement) and counted the number of trips F_{ij} between each pair of zones i and j. In this study, zones were defined by towers' servicing areas in the Bay area and census tracts in the Boston area. The different zone definitions were resulted from the different features of location records in the two mobile phone datasets (see Methods). The defined zones were only used in the process of generating the t-ODs. All measurements regarding the dynamical driver sources (Fig. 1c–f) were based on the census tracts for both Bay area and Boston area.

In order to avoid the data sampling bias caused by an unevenly distributed user penetration rate, we defined the down-scale ratio ($M(i) < 1$) or the up-scale ratio ($M(i) \geq 1$) as follows (Fig. S2c):

$$M(i) = N_{pop}(i) / N_{user}(i) \qquad (1)$$
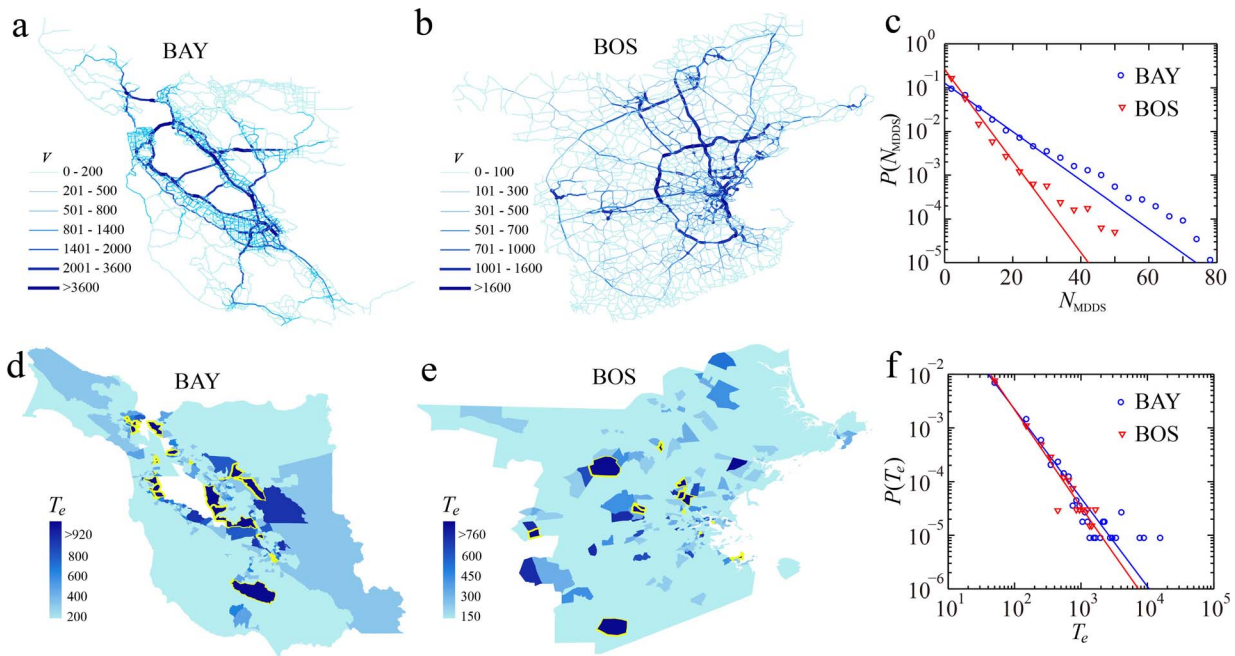
Figure 1 | Sources of traffic congestion in the Bay area and Boston area. (a) Traffic flow in the Bay area road network. (b) Traffic flow in the Boston area road network. (c) The number of major dynamical driver sources N_{MDDS} of a road segment is well approximated by the exponential distribution $P(N_{\text{MDDS}}) = 0.13 e^{-0.13 N_{\text{MDDS}}}$ ($R^2 > 0.99$) in the Bay area and the exponential distribution $P(N_{\text{MDDS}}) = 0.29 e^{-0.28 N_{\text{MDDS}}}$ in the Boston area ($R^2 > 0.99$). (d) The color of a Bay area census tract (1,398 in total) represents the total extra travel time T_e experienced by drivers whose trips originated from that census tract during one of the peak morning hours. We defined the top 2% census tracts with the largest T_e as *congested driver sources* and highlighted those using yellow polygons. (e) The extra travel time T_e of Boston area census tracts (750 in total). Congested driver sources have been pinpointed using yellow polygons. (f) Extra travel time T_e follows a power-law distribution $P(T_e) = 4.46 T_e^{-1.65}$ ($R^2 > 0.99$) in the Bay area and a power-law distribution $P(T_e) = 8.94 T_e^{-1.81}$ ($R^2 > 0.99$) in the Boston area. The maps in (a), (b), (d), (e) were generated using TransCAD 5.0 and ArcGIS.

where $N_{pop}(i)$ and $N_{user}(i)$ are the population and the number of mobile phone users in zone i. We adjusted F_{ij} using Eq. (2), thus the number of trips generated by residents in a zone is proportional with its population:

$$F_{ij}^{all} = \sum_{n=1}^{N_k} T_{ij}(n) \times M(k) \qquad (2)$$

where N_k is the number of phone users in the kth zone and $T_{ij}(n)$ is the number of trips that user n made between zone i and zone j during the three weeks' observational period. Using the car usage rate data[26], we estimated the number of trips produced by vehicles $F^{vehicle}_{ij}$ (Fig. S2d). Finally, we upscaled the estimated distribution of travel demands with the hourly trip production W for the entire population, thus defining the estimated t-OD (Fig. S2e):

$$t\text{–OD}_{ij} = W \times \frac{F^{vehicle}_{ij}}{\sum_{ij}^{A} F_{ij}^{all}} \qquad (3)$$

where A is the number of zones. To assign trips to the road networks, we map each t-OD pair from zone based t-OD to intersection-based t-OD. We find the road intersections within a zone and randomly select one intersection to be the origin or destination in the intersection-based t-OD (Fig. S2b). With the intersection-based t-ODs, the traffic flow from the two road networks was predicted using an incremental traffic assignment method[27] (see Methods). In the following, we present the results from the morning period as a case study. For more detailed information on generating t-ODs, please see the Supplementary Information (SI).

Results

Locating the dynamical driver sources of road segments. The driver source was defined by a mobile phone user's home location where the user could be found from 9:00 p.m. to 6:00 a.m[16]. It

indicates each road segment where its users (drivers) live and offers useful information for urban traffic planning[16]. Unfortunately, the information is not applicable for real-time traffic control because it is unable to capture the time-variant vehicle origins for a road at different times of a day. In this work, we have defined the *dynamical driver sources* of a road segment as the census tracts where its drivers' trips start (not necessarily driver's home) during a particular time window, thus packaging the dynamical travel demand information into a road segment. The intersection-based t-ODs record the starting road intersection for each trip; thus, the census tract where a trip starts can be located by finding the census tract for the starting road intersection.

For each road segment, we calculated the fraction of traffic flow contributed from each dynamical driver source and then ranked these sources according to their contributions to traffic flow. Consequently, we defined a road segment's major dynamical driver sources (MDDS) as the top ranked sources that produced 80% of the traffic flow (similar to the definition of major driver source (MDS)[16]). As Fig. 1c shows, the number of major dynamical driver sources N_{MDDS} follows an exponential distribution $P(N_{\text{MDDS}}) = 0.13 e^{-0.13 N_{\text{MDDS}}}$ in the Bay area and an exponential distribution $P(N_{\text{MDDS}}) = 0.29 e^{-0.28 N_{\text{MDDS}}}$ in the Boston area. For 90% of the Bay area (Boston area) road segments, the N_{MDDS} were smaller than 18 (9), implying that the time-variant traffic flow in most roads could be attributed to the trips starting from a low number of census tracts (Fig. 1d & e). Surprisingly, the largest N_{MDDS} only represents 5.6% (7.0%) of the Bay area (Boston area) census tracts, while the largest N_{MDS} was found to occupy 19.0% (38.9%) of the Bay area (Boston area) census tracts. This indicates that over a specific time period, a road's vehicle origins are confined to a much smaller scale than drivers' home locations. This is inspiring news for applying dynamical driver source predictions in practical urban traffic control, since fewer sources need to be paid attention when controlling a road segment's traffic (Fig. S3).
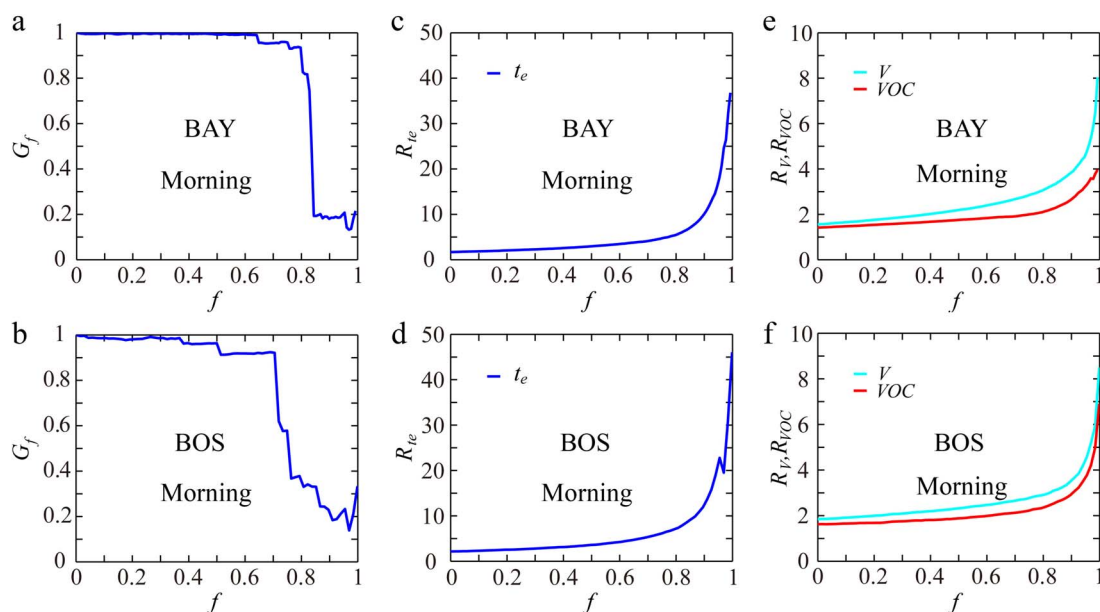
Figure 2 | Locating the road segments used extensively by drivers from congested driver sources. (a) The size of the largest road cluster G_f with the fraction of removed road segments f. We observed a sharp decrease of G_f at a threshold fraction $f_c \sim 0.82$ in the Bay area. Below the threshold f_c, the largest cluster size G_f was larger than 0.9, above the threshold f_c, the largest cluster size G_f was smaller than 0.2. (b) The size of the largest road cluster G_f with f in the Boston area. (c), (d) The ratio of average extra travel time of the remaining road segments and that of all road segments was denoted by $R_{te} = <t_e>_{\text{remain}}/<t_e>_{\text{all}}$. The ratio R_{te} increased slowly with f when $f < f_c$, but increased sharply when $f > f_c$. (e), (f) The ratio of the average traffic flow V and the average VOC of the remaining road segments and those of all road segments were denoted by R_V and R_{VOC}. Similar to R_{te}, the ratios R_V, R_{VOC} increased slowly with f when $f < f_c$, but increased sharply when $f > f_c$. Results from the noon/afternoon period and evening period, please refer to Figure S5.

Pinpointing road clusters heavily used by drivers from congested driver sources. Defining dynamical driver sources, we next built a connection between a road network and its sources of traffic congestion. For a road segment, the level of congestion can be quantified by the extra travel time t_e, which is defined as the difference between the actual travel time t_a and the free flow travel time t_f ($t_e = t_a - t_f$). The free flow travel time t_f equals to the length of the road segment divided by the speed limit of the road segment. The actual travel time t_a was calculated using the Bureau of Public Roads (BPR) function that widely used in civil engineering:

$$t_a = t_f\left(1 + \alpha(VOC)^{\beta}\right) \qquad (4)$$

Commonly used values $\alpha = 0.15$ and $\beta = 4$ were selected[26]. The volume over capacity $VOC = V/C$, where V is the hourly traffic flow and C is the hourly road capacity.

The drivers who travel through congested roads experience a significant amount of extra travel time. To pinpoint these drivers, the total extra travel time T_e generated by drivers from each census tract was calculated. Despite that the populations in the census tracts are similar (Fig. S1c & d), the extra travel time T_e generated by the census tract can be very different. It is observed that T_e follows a power-law distribution $P(T_e) = 4.46T_e^{-1.65}$ in the Bay area and a power-law distribution $P(T_e) = 8.94T_e^{-1.81}$ in the Boston area (Fig. 1f). Some census tracts present a T_e 116 times larger than the average; during a specific time window the major traffic flow in congested roads is attributed to few dynamical driver sources.

Defining the top 2% dynamical driver sources with the largest T_e as *congested driver sources* (28 in the Bay area, 15 in the Boston area), we pinpoint the major sources of congestion during the morning peak (yellow polygons in Fig. 1d & e). Collecting road segments used at least once by drivers from congested driver sources, we obtained two giant road clusters in the Bay area and the Boston area (with 12,288 road segments and 6,809 road segments respectively). Drivers from the congested driver sources heterogeneously used road segments

within the giant road clusters (Fig. S4). For each road segment, traffic flow generated by the congested driver sources was denoted as F. In the Bay area (Boston area), more than 90% of the road segments have an F smaller than 292 (106), while the largest F reaches 4,589 (1,509).

In order to pinpoint the road clusters used most heavily by drivers from the congested driver sources, we ranked road segments by their traffic flow F; each time, we removed a road segment with the lowest F from the giant road cluster. The fraction of removed road segments was denoted as f. We found that in a wide range of f, the largest road cluster covered almost all road segments left; however, when f passed a threshold $f_c \sim 0.82$ ($f_c \sim 0.70$) in the Bay area (Boston area), the size of the largest road cluster G_f dropped sharply, meaning that the giant road cluster broke down (Fig. 2a & b).

Interestingly, in the phase $f > f_c$, where giant road clusters fell apart, dramatic changes were also observed for the properties closely related to traffic congestion (Fig. 2c–f). Taking the Bay area as an example, the ratio R_{te} is the average extra travel time of the remaining road segments over the average extra travel time of all road segments. We found that R_{te} increased slowly with f when $f < f_c$, but sharply when $f > f_c$. In the case of the 500 road segments remaining, the ratio R_{te} was as large as 20, implying that the road segments left had much greater extra travel times than average (Fig. 2c). Similar results were found for average traffic flow V and average volume over capacity (VOC), which represented roughly 5 times the overall average when there were 500 road segments remaining. In conclusion, road segments with extremely high t_e, V, VOC were targeted in the phase $f > f_c$ using our methodology. For more statistics relating to t_e, V, VOC, please see Fig. S4 in SI. For results from the noon/afternoon period and evening period, please see Fig. S5 in SI.

Of the two observed phases, the phase $f > f_c$ is of great interest. In this phase, a few road clusters that identified the roads most frequently used by drivers from congested driver sources, are pinpointed in particular regions of the two urban areas. We measured the sizes of the road clusters after removing a fraction of $f = 0.96$ road segments from the giant road cluster in the Bay area, finding that
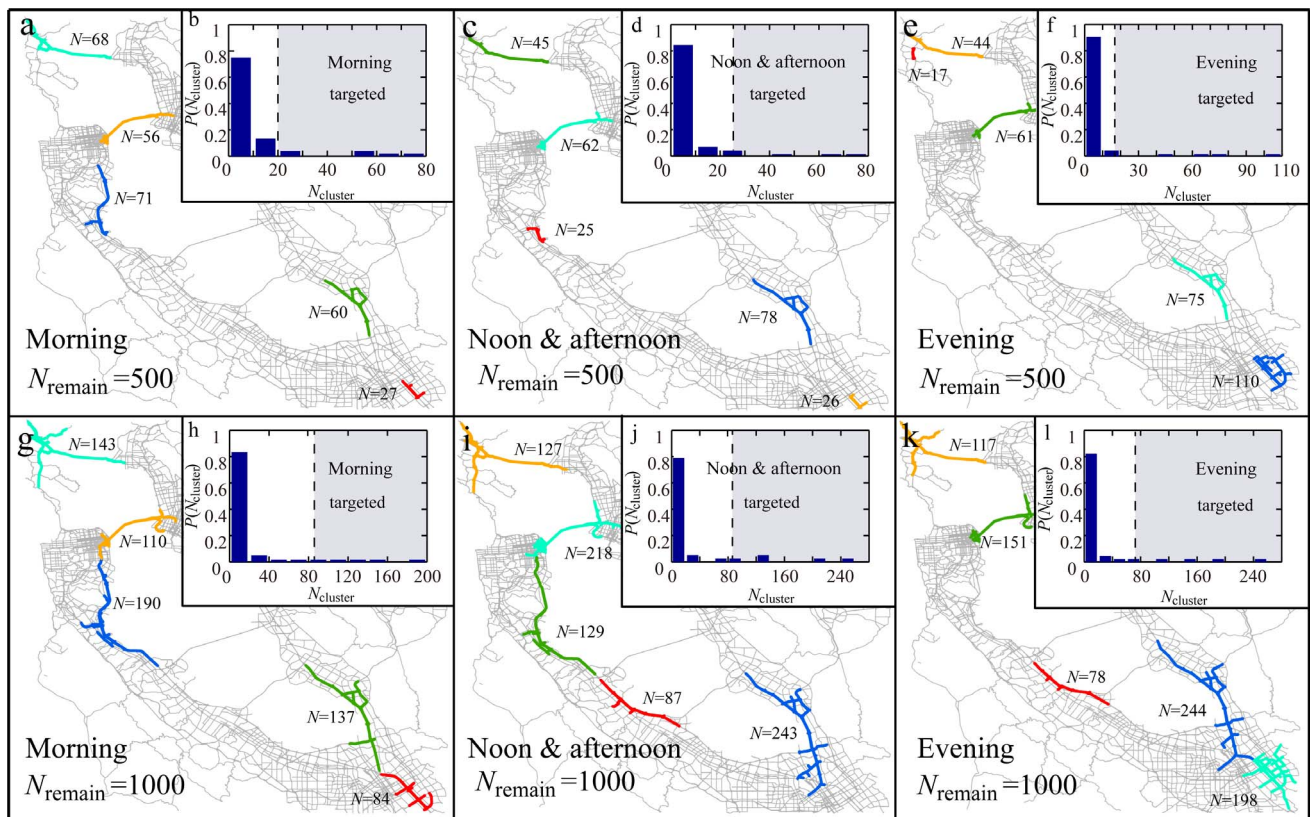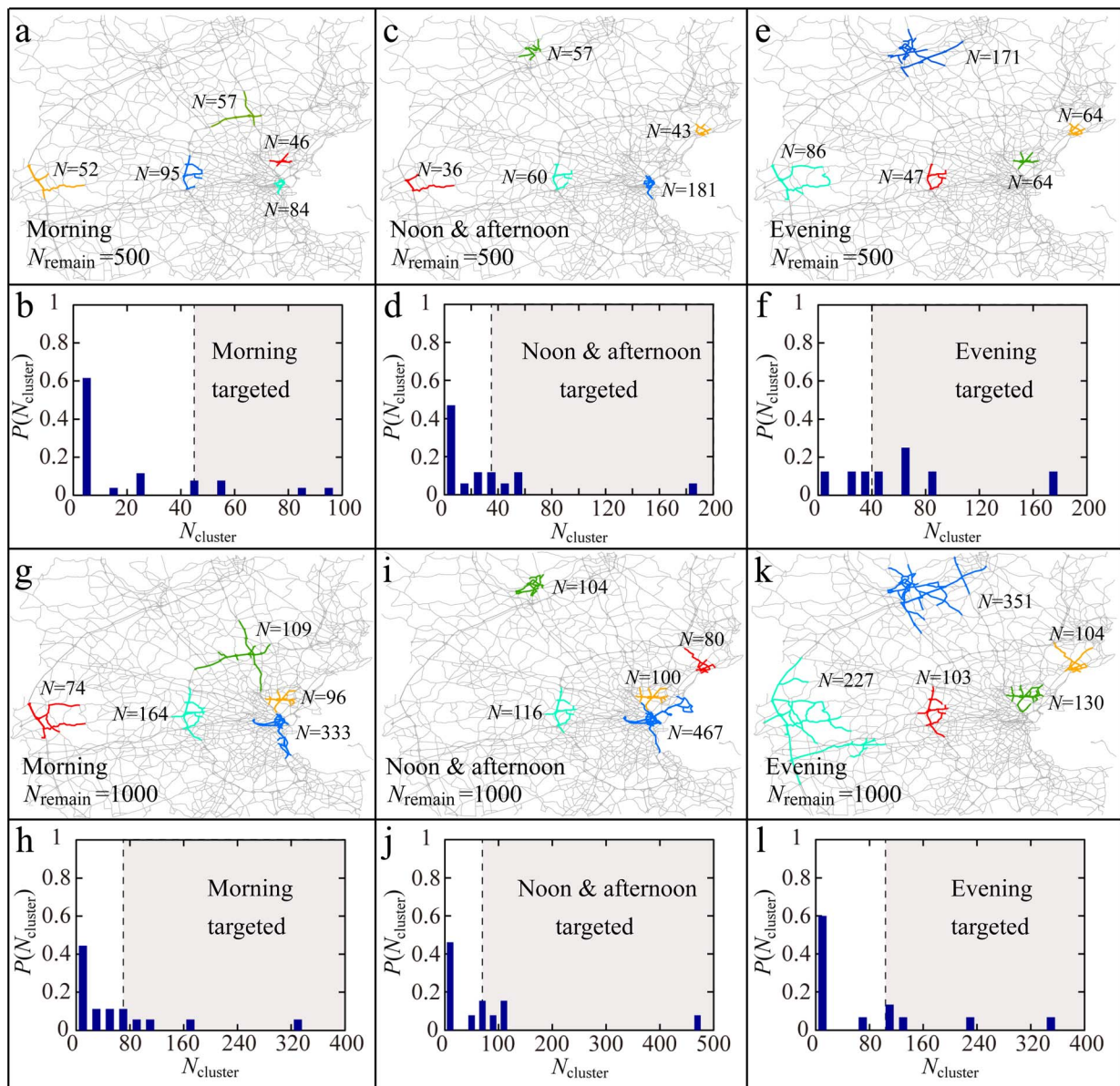
Figure 3 | Spatial distribution of targeted road clusters in the Bay area. The targeted road clusters are depicted in different colors (dark blue: the 1st largest, light blue: the 2nd largest, green: the 3rd largest, orange: the 4th largest, red: the 5th largest). (a), (c), (e), (g), (i), (k) Targeted road clusters show different spatial distributions during morning period, noon/afternoon period and evening period, which can be addressed by the different distributions of travel demand over the three periods. (b), (d), (f), (h), (j), (l) The distribution of the size of the targeted road clusters for morning period, noon/afternoon period and evening period when 500 road segments remain ($f = 0.96$) and 1,000 road segments remain ($f = 0.92$). The maps in (a), (c), (e), (g), (i), (k) were generated using TransCAD 5.0 and ArcGIS.

most road clusters were small ($N < 10$); a low number of clusters constituted a large fraction of road segments left (Fig. 3a & b). During the morning period, the top five largest road clusters represented 56% of the road segments left (Fig. 3b). The top road clusters captured the minority group of roads closely related to congestion; at the same time it avoided scattering tiny road clusters, making this method feasible in practical implementations. The top five road clusters located at different time periods for the Boston area are shown in Figure 4. Interestingly, at different times of a day, the targeted road clusters had different spatial distributions. For instance, during the morning peak hours in the Bay area, the largest targeted road cluster was located near the San Francisco airport; however, during the evening period, no targeted road cluster was found near the airport (Fig. 3). In the Boston area, the largest targeted road cluster was observed in downtown Boston during the morning period and noon/afternoon period, while during the evening period no targeted road cluster was observed downtown (Fig. 4). The proposed modeling framework endowed road networks with the capability of sensing urban traffic rhythms, which we believe can provide useful urban travel demand information for new generations of urban traffic control technologies.

Improving road network efficiency based on targeted road clusters. The method of targeting road clusters closely related to traffic congestion can trigger numerous applications. As proof of this concept, we will present how these findings can be applied to mitigate traffic congestion. There are two ways to mitigate traffic congestion. The first is to increase the capacity of congested roads, thus reducing their extra travel time. However, civil engineers

discourage this method because it may disrupt urban textures or attract more vehicles to the road. The second is to reduce the number of vehicles on congested roads by increasing the related travel costs (i.e., congestion pricing), thus forcing drivers to avoid using congested roads by either changing their routes or abandoning their trips. As a proof of concept here, the capacity of road segments in the targeted road clusters was increased to reduce extra travel time, whereas the speed limits on road segments in the targeted road clusters was lowered in order to direct drivers to avoid congested roads. As a reference, we used a benchmark strategy in which five road clusters were randomly generated that were the same sizes as the targeted road clusters (Fig. 5a, see Methods).

We first increased the capacity of road segments in the targeted road clusters (targeted strategy) by a fraction of ΔC ($\Delta C = 0.1, 0.2, 0.3$). The total travel time T slightly decreased, while the total extra travel time Σt_e significantly decreased in both urban areas (Fig. 6a & b). As ΔC equals to 0.3, the total extra travel time decreased 26% (17%) in the Bay area (Boston area), which equaled 48,065 (11,373) minutes of extra travel time saved in one hour during peak morning hours and 4 times (16.7 times) of that achieved by the benchmark strategy. The number of congested road segments also decreased 7% (14%) in the Bay area (Boston area), which is 7 times (8.8 times) that achieved by the benchmark strategy (Fig. 6c). We next decreased the speed limit of road segments on the targeted road clusters by a fraction of ΔS ($\Delta S = 0.1, 0.2, 0.3$), observing that the total travel time T increased slightly in the Bay area (Boston area) (Fig. 6d). This is not surprising because lowering the speed limit reduced road network capacities. However, the interesting finding was that the total extra travel time Δt_e and the number of congested roads $N(VOC > 1)$

Figure 4 | Spatial distribution of targeted road clusters in the Boston area. Measurements in Figure 3 were conducted for the Boston area. The maps in (a), (c), (e), (g), (i), (k) were generated by TransCAD 5.0 and ArcGIS.

decreased with almost no construction costs. As ΔS equals to 0.3, the total extra travel time Σt_e decreased 8% (5%) in the Bay area (Boston area), which equaled 14,259 (3,354) minutes of extra travel time saved during one hour of the peak morning hours and 12.6 times (3.1 times) of that achieved by the benchmark strategy (Fig. 6e). The number of congested road segments decreased 3% (5%) in the Bay area (Boston area), while the number of congested road segments slightly increased when using the benchmark strategy (Fig. 6f).

We next explored the effects of congestion mitigation on different scales of targeted road clusters. Increasing the scale of the targeted road clusters, more prominent reductions of total travel time T, total extra travel time Σt_e and number of congested road segments $N(VOC > 1)$ were observed when increasing the road capacity (Fig. S6, Fig. S7). However, this method is usually subject to high construction costs as well as other side effects, such as urban texture disruption; therefore, careful evaluations need to be taken before practical implementations. Increasing the scale of targeted road clusters upon which the speed limit is reduced, both total travel time T and total extra travel time Σt_e increased, while the number of congested road segments $N(VOC > 1)$ decreased (Fig. S6, Fig. S7). This suggests that a

balance needs to be made between reducing extra travel time and reducing the number of congested roads. To avoid a large increase of travel time, lowering the speed limit is appropriate to a small group of carefully selected road segments. The advantages of this method rely on its low operation/construction cost, flexibility and adaptability to dynamical travel demands.

Discussion

More detailed travel demand information can be added into the current modeling framework. For example, to capture the different human mobility patterns between weekdays and weekends, we divided the mobile phone data according to the types of days (weekday or weekend), and targeted the dynamical driver sources. We observed slightly different spatial distributions of the targeted road clusters in weekdays and weekends, indicating that more detailed travel demand information would lead to more accurate estimations of the dynamical sources of traffic congestions (Fig. S8–Fig. S12 in SI). The empirical investigations of the time-variant travel demand and congested driver sources provided us a way to introduce more
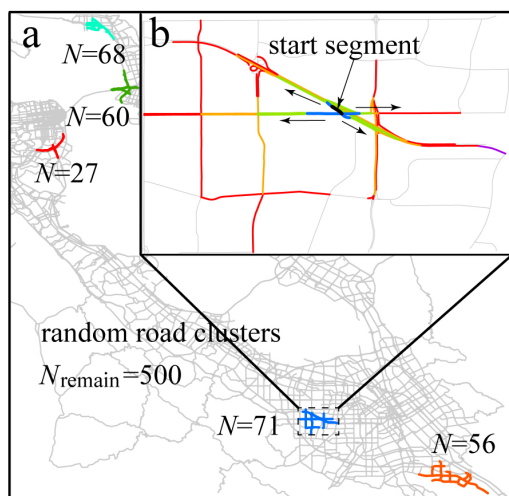
Figure 5 | Illustration of the random road clusters (a) A random road cluster has the same size with its corresponding targeted road cluster. Five random road clusters were generated in the benchmark strategy. (b) A random road cluster starts from one randomly chosen road segment and grows by finding its neighboring road segments. Different colored road segments represent the growing layers. The growing process stops when the size of random road cluster reaches the size of its targeted counterpart. The maps in the figure were generated using TransCAD 5.0 and ArcGIS.

realistic modeling properties and a guidance to generate practical models dedicated to transport networks' improvements.

We firmly believe that the proposed methodology can be applied to a wider range of transport networks and potentially appeal the interest of the general audience from diverse research fields. The generality of our results relies on the ubiquitous existence of the sources of basic transporting elements in many networks. The proposed modeling framework can be extended to other supply-demand related networks, from transportation networks[9–16] to the Internet[28,29], in which the basic transporting elements have "sources." It can also be applied to epidemic spreading networks[30,31], in which the

basic diffusing elements have "sources." In this study, using road network as an example, we developed a method to target the link clusters that were heavily correlated with the congestion of the transporting elements, thus, offering new insight to understand similar scenarios in other related studies of network flows.

Taken together, we located the dynamical driver sources for each road segment in the Bay area and the Boston area. For most road segments the major dynamical driver sources are limited, thus congested driver sources can be pinpointed. Giant road clusters involving all roads used by drivers from the congested driver sources were generated in the two urban areas. As we gradually removed the less-used roads from the giant road clusters, we observed phase transition phenomena for the size of the largest road cluster and the properties closely related to traffic congestion. In the phase $f > f_c$, extra travel time, traffic flow and VOC increased much faster with f than they did when $f < f_c$, enabling us to target road clusters highly correlated with traffic congestion. As a proof of concept, we lowered the speed limit and increased the capacity of road segments in the targeted road clusters to improve transport efficiency. While increasing the road capacity in targeted road clusters have a more prominent effect in reducing extra travel time, its implementation is usually expensive, inflexible and may induce more travel demands. Lowering the speed limit in targeted road clusters, which can be incorporated with almost no implementation cost, can adapt to the dynamic urban travel demands and mitigate traffic congestion at the expense of slightly increased travel time.

Methods

Road GIS data and mobile phone data. This study employed the Bay area and Boston area road networks, which consist of highways and arterial roads. For each road segment, the speed limit, the number of lanes and the direction of traffic were extracted from the database. The capacity of a road segment was estimated according to the 2000 Highway Capacity Manual[32]. Different mobile phone operators collected the mobile phone data in the Bay area and the Boston area. When a person used a phone in the Bay area, the mobile phone tower providing the service was recorded; in the Boston area, the user's location coordinates were estimated using a standard triangulation algorithm (Fig. S1a & b). Due to the different features of location records, user displacement was defined as the change of servicing towers in the Bay area and the change of census tracts in the Boston area.
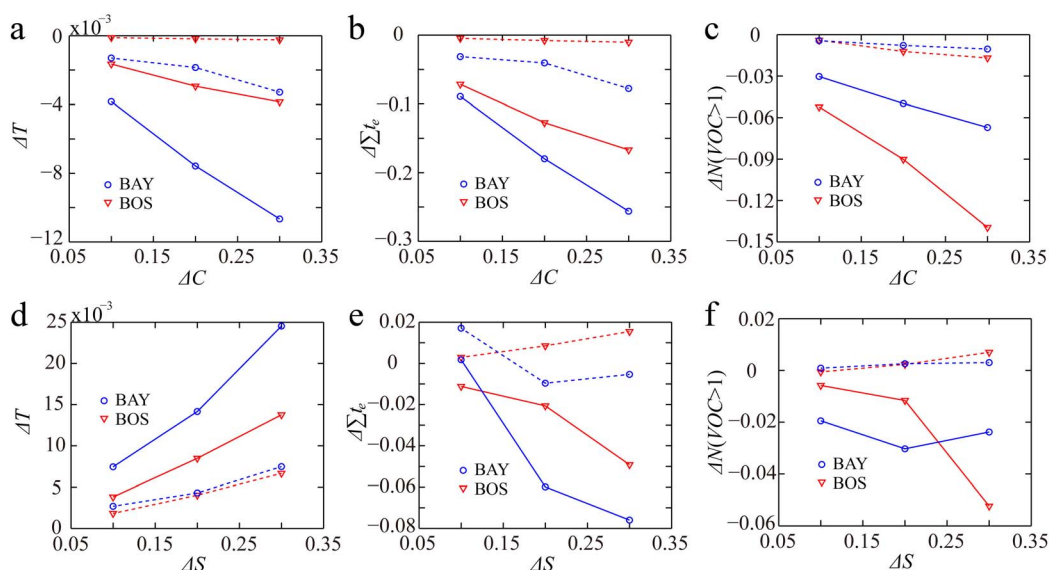
Figure 6 | **Mitigating traffic congestion by lowering speed limit or increasing capacity.** Blue and red solid (dashed) lines show the effects of congestion mitigation using the Bay area and Boston area targeted (random) road clusters. (a) The decrease of total travel time ΔT with the increase of capacity ΔC. (b) The decrease of total extra travel time $\Delta\Sigma t_e$ with the increase of ΔC. (c) The decrease of number of congested road segments $\Delta N(VOC > 1)$ with the increase of ΔC. (d) The change in total travel time ΔT with the reduction of speed limit ΔS. (e) The decrease of total extra travel time $\Delta\Sigma t_e$ with the reduction of ΔS. (f) The decrease of number of congested road segments $\Delta N(VOC > 1)$ with the reduction of ΔS.

Incremental traffic assignment. With the intersection based t-ODs calculated, we assigned the trips to the two road networks. The most fundamental method is provided by the Dijkstra algorithm, commonly used for routing in transportation networks[33]. With the Dijkstra algorithm, we could find the shortest path with minimum travel time between the origin and destination in a road network. However, the Dijkstra algorithm ignores the dynamical change of travel time with traffic flow. Therefore, we apply the incremental traffic assignment (ITA) method[27] to assign the t-OD pairs to the road networks. In this method, we first split the original t-OD of a specific time period into four sub t-ODs, containing 40%, 30%, 20% and 10% (commonly used values[26]) of the OD pairs. The OD-pairs (trips) in the four sub t-ODs were randomly selected from the original t-OD. Trips in the first sub t-OD were assigned using free travel time t_f along the routes computed by Dijkstra's algorithm. After that, the actual travel time t_a in a road segment was updated using the BPR function (Eq. 4). Next, the trips in the second sub t-OD were assigned using the updated t_a, consequently the actual travel time t_a in a road segment was updated again. Iteratively, we assigned all trips in the four sub t-ODs. In the process of finding the shortest path, we measured traffic flow for each road segment.

Method to generate random road clusters. A random road cluster starts from a randomly chosen road segment and diffuses to its neighboring segments that share the same intersections. The random road cluster grows until its size reaches that of the targeted road cluster (Fig. 5). Road segments that were visited by one random road cluster were not used in generating other random road clusters. As Figure 5a illustrates, five random road clusters were generated in the Bay area road network; their sizes were 71, 68, 60, 56 and 27, the same sizes of the targeted road clusters depicted in Fig. 3a.

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, College Station, 2009).
4. Hin, L. T. W. & Subramaniam, R. Smart-card system keeps Singapore in the fast lane. *Nature* **411**, 737–737 (2001).
5. Rhoads, T. A. & Shogren, J. F. Why do cities use supply side strategies to mitigate traffic congestion externalities? *Economics Letters* **92**, 214–219 (2006).
6. Hoffman, K., Berardino, F. & Hunter, G. Congestion pricing applications to manage high temporal demand for public services and their relevance to air space management. *Transport Policy* **28**, 28–41 (2013).
7. Li, G. *et al.* Towards design principles for optimal transport networks. *Phys. Rev. Lett.* **104**, 018701 (2010).
8. Newman, M. E. J. A measure of betweenness centrality based on random walks. *Social Networks* **27**, 39–54 (2005).
9. Crucitti, P., Latora, V. & Porta, S. Centrality measures in spatial networks of urban streets. *Phys. Rev. E* **73**, 036125 (2006).
10. Tero, A. *et al.* Rules for biologically inspired adaptive network design. *Science* **327**, 439–442 (2010).
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. Kleinberg, J. M. Navigation in a small world. *Nature* **406**, 845 (2000).
13. Caretta Cartozo, C. & De Los Rios, P. Extended navigability of small world networks: exact results and new insights. *Phys. Rev. Lett.* **102**, 238703 (2009).
14. Guimera, R. *et al.* Optimal network topologies for local search with congestion. *Phys. Rev. Lett.* **89**, 248701 (2002).
15. 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).
16. Wang, P. *et al.* Understanding road usage patterns in urban areas. *Sci. Rep.* **2**, 01001 (2012).
17. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
18. González, M. C., Hidalgo, C. A. & Barabási, A. L. Understanding individual human mobility patterns. *Nature* **435**, 779–782 (2008).
19. Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).
20. 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).
21. Expert, P., Evans, T. S., Blondel, V. D. & Lambiotte, R. Uncovering space-independent communities in spatial networks. *PNAS* **99**, 7663–7668 (2011).
22. Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
23. Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–96 (2002).
24. Zhang, F., Wilkie, D., Zheng, Y. & Xie, X. Sensing the pulse of urban refueling behavior. *15th ACM International Conference on Ubiquitous Computing*, ACM, Zurich (2013).
25. Zheng, Y., Liu, F. & Hsie, H. U-Air: When urban air quality inference meets big data. *19th SIGKDD conference on Knowledge Discovery and Data Mining*, KDD, Chicago (2013).
26. Caliper Corporation. *Travel demand modeling with TransCAD 5.0, user's guide* (Caliper, Newton, 2008).
27. Chen, M. & Alfa, A. S. A network design algorithm using a stochastic incremental traffic assignment approach. *Transportation Science* **25**, 215–224 (1991).
28. Pastor-Satorras, R. & Vespignani, A. *Evolution and structure of the Internet* (Cambridge University Press, Cambridge, 2004).
29. Yook, S. H., Jeong, H. & Barabási, A. L. Modeling the internet's large-scale topology. *PNAS* **99**, 13382–13386 (2002).
30. Wang, P., González, M. C., Hidalgo, C. A. & Barabási, A. L. Understanding the spreading patterns of mobile phones viruses. *Science* **324**, 1071–1076 (2009).
31. Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *PNAS* **106**, 21484–21489 (2009).
32. Transportation Research Board. *Highway capacity manual: 2000* (Transportation Research Board, Washington, D.C., 2000).
33. Dijkstra, E. W. A note on two problems in connexion with graphs. *NumerischeMathematik* **1**, 269–271 (1959).

Acknowledgments

Author contributions

Conceived and designed the experiments: J.W., D.W. and P.W. Wrote the paper: K.H. and P.W. Analyzed the empirical data and did the analytical and numerical calculations: J.W., D.W. and H.G.

Additional information