Macroscopic dynamics and the collapse of urban traffic

Luis E. Olmos[a,b,c,d], Serdar Çolak[b], Sajjad Shafiei[e], Meead Saberi[f], and Marta C. González[b,c,d,1]

[a]Department of Physics, National University of Colombia, Bogotá 111321, Colombia; [b]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; [c]Department of City and Regional Planning, University of California, Berkeley, CA 94720; [d]Energy Analysis & Environmental Impacts Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [e]Institute of Transport Studies, Civil Engineering Department, Monash University, Melbourne, VIC 3800, Australia; and [f]School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia

Edited by Paul Trunfio, Boston University, Boston, MA, and accepted by Editorial Board Member Pablo G. Debenedetti June 14, 2018 (received for review February 15, 2018)

Stories of mega-jams that last tens of hours or even days appear not only in fiction but also in reality. In this context, it is important to characterize the collapse of the network, defined as the transition from a characteristic travel time to orders of magnitude longer for the same distance traveled. In this multicity study, we unravel this complex phenomenon under various conditions of demand and translate it to the travel time of the individual drivers. First, we start with the current conditions, showing that there is a characteristic time τ that takes a representative group of commuters to arrive at their destinations once their maximum density has been reached. While this time differs from city to city, it can be explained by Γ, defined as the ratio of the vehicle miles traveled to the total vehicle distance the road network can support per hour. Modifying Γ can improve τ and directly inform planning and infrastructure interventions. In this study we focus on measuring the vulnerability of the system by increasing the volume of cars in the network, keeping the road capacity and the empirical spatial dynamics from origins to destinations unchanged. We identify three states of urban traffic, separated by two distinctive transitions. The first one describes the appearance of the first bottlenecks and the second one the collapse of the system. This collapse is marked by a given number of commuters in each city and it is formally characterized by a nonequilibrium phase transition.

urban traffic gridlock | nonequilibrium phase transition | directed percolation | human mobility | mobile phone

The steady increase of traffic congestion not only translates into overpowering travel times (1–3), but also erodes economic growth (4) and has negative environmental impacts (5–7). Urban traffic has extensively been studied through computer simulations with particular interest in the characterization of the transition from free flow to congestion (8–10). Traditional approaches in traffic engineering have described the phenomenon within the framework of a well-defined relationship between network-wide average flow and the density of cars (11–17). Considering the network density as the control parameter, a congested state at the network level emerges when car outflow starts decreasing when the vehicular demand exceeds a certain value. In parallel, critical loads R_c at which vehicles start accumulating in a road network have been studied in abstract frameworks unrelated to empirical travel demand of cities (18–20). All these approaches fall short by assuming steady-state traffic conditions and using state parameters unrelated to the individual travelers. Congestion at the urban scale is by nature unevenly distributed in space and the volume of cars varies strongly during the day.

The recent availability of data on personal tracking devices has enriched the study of traffic models. Origin–destination (OD) tables can be extracted from call-detailed records (CDRs) of mobile phones (21, 22) and GPS-equipped vehicles can act as sensors of traffic conditions (15, 23). Patterns of individual mobility have been uncovered (24, 25) and allow us to model individual daily mobility from passively collected sources (26). Comparing various cities, scaling of urban indicators emerges (27, 28). Examples are travel times and road network characteristics as a function of population and socioeconomic characteristics (27, 29, 30). For operational and planning purposes, a macroscopic description of the urban traffic dynamics and their vulnerability to collapse, measured in terms of car volumes, road network supply, and individual travels, is essential, yet still missing. In other words, In what way does the information contained in the ODs determine the travel time of target individuals and how can these dynamics be understood in terms of actionable quantities to explain when the system will collapse?

As a first step in that direction, Çolak et al. (22) used a framework of static equilibrium to compare the morning conditions of congested travel times (t_{ue}) along the routes of five cities as a function of the road supply to vehicle demand (Γ). In this work we go a step farther, fully studying the nonequilibrium dynamics of six cities for which we have information on their empirical vehicular demand. We show that the state of the system can be characterized by the time that a representative number of commuters take to depart at the beginning of the peak hour and arrive at their destinations. We formulate the measurements as the unloading time of individuals departing within the same hour and then measure after they reach their maximum number in the network. As recently proposed in simulations of the Chicago road network (31), the concept of the recovery period is an important element to diagnose urban congested states. Here we focus in the travelers departing at the beginning of the peak hour in diverse cities.

First we directly measure the reported times of individual travel diaries in Boston, San Francisco, and Bogotá, and we build their reported temporal profile of car trips vs. time in the morning peak period. We focus on the target group of vehicles that enter the network within the peak hour (7:00–8:00 AM) and find that they have an exponential unloading period. This empirical observation is corroborated by a calibrated and validated

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Modeling and Visualizing Science and Technology Developments," held December 4–5, 2017, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and video recordings of most presentations are available on the NAS website at www.nasonline.org/modeling_and_visualizing.

Author contributions: L.E.O. and M.C.G. designed research; L.E.O. and M.C.G. performed research; L.E.O., S.Ç., S.S., M.S., and M.C.G. analyzed data; and L.E.O. and M.C.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. P.T. is a guest editor invited by the Editorial Board.

Data deposition: The data of the OD matrices, software to replicate the method, and the appropriate documentation can be accessed on GitHub at https://github.com/leolmoss/CollapseUrbanTraffic.

[1]To whom correspondence should be addressed. Email: martag@berkeley.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1800474115/-/DCSupplemental.

Published online December 10, 2018.

simulation-based dynamic traffic assignment model of the entire city of Melbourne. We argue that this characteristic unloading time τ faced by commuters can be related by macroscopic characteristics that contain the overall road capacity and travel demand. To uncover the macroscopic dynamic that explains τ, we implemented a cellular automata (CA) model on the road networks of five cities around the world, namely Boston, Porto, Lisbon, Rio de Janeiro, and the San Francisco Bay area. Informed by the empirical trip demand derived from CDRs from each city, we analyze the morning peak hour by loading constantly during 1 h and letting the drivers arrive at their destination. We show that the exponential form of the unloading time is a consequence of the log-normal distribution exhibited by the commuting distances. As expected, $\tau \sim \Gamma$, each city has a different τ every morning, and the differences are explained by the ratio of the total vehicle demand to their available street capacity. We further conduct a scenario analysis for a different number of cars entering for the peak hour, keeping the empirical distribution of OD trips in the morning peak. In doing so, we uncover three different states of urban traffic and the critical demand beyond which the system collapses. As an illustration, snapshots of these states at the same hour are shown in Fig. 1*A* for Boston. We show that the dynamical response is independent of the level of detail of the traffic model and the city under consideration.

The Unloading Time of Urban Road Networks

Peak-hour traffic congestion is inherent to the circadian rhythm of modern city life. After an intense and heterogeneous loading period, there is a recovery period where traffic jams on major arterial roads dissipate, thus relieving the city traffic. We focus, in this work, on the recovery period after the morning peak hours under current traffic conditions and further see how it changes as a function of the number of cars in diverse cities, paying particular attention to the collapse of the system.

We start by gathering empirical observations on urban congestion. We analyze individual travel diaries in Boston, San Francisco, and Bogotá. This allows us to build an approximate temporal profile of car trips in the network. To have a controlled sample, we focus on the target group of vehicles that enter the network within the peak hour, defined as the hour immediately before the peak, here from 7:00 AM to 8:00 AM. Remarkably, we observe a generic behavior: After a loading period, the number of target group vehicles still in the network, $N(t)$, follows a characteristic unloading time that follows an exponential decay that is depicted in Fig. 1*B*. It is important to note that this observation is in contrast with that of all of the vehicles of commuters also shown in Fig. 1*B*. This sample has a more heterogeneous mix of times to unload, not properly characterized by a single relaxation time.

We further measure the loading and unloading with the simulation-based dynamic traffic assignment (DTA) model of the city of Melbourne (32). This model has been carefully calibrated for the entire city and simulates almost 2.1 million commuters in a 4-h morning peak period (*Materials and Methods*). Similar to the empirical results reported in surveys, we observe a recovery period that follows an exponential decay that can be written as

$$N(t)_{t \geq t_0} = N(t_0)\, e^{-\frac{t-t_0}{\tau}}, \qquad [1]$$

with $t_0 = 8{:}30$ AM and the unloading time $\tau = 0.51(4)$ h (Fig. 1*C*). This implies a proportionality between the exit function $G(N(t))$ and the number of vehicles, namely

$$G(N(t)) = \left| \frac{dN(t)}{dt} \right| \propto N(t) \quad \text{for} \quad t \geq t_0. \qquad [2]$$

This relation is confirmed by simulations (*SI Appendix*, Fig. S2). Thus, τ should indicate the network response to the congestion, including also vehicles not belonging to the target group. We further study how τ depends on network characteristics and different travel demands over diverse cities.

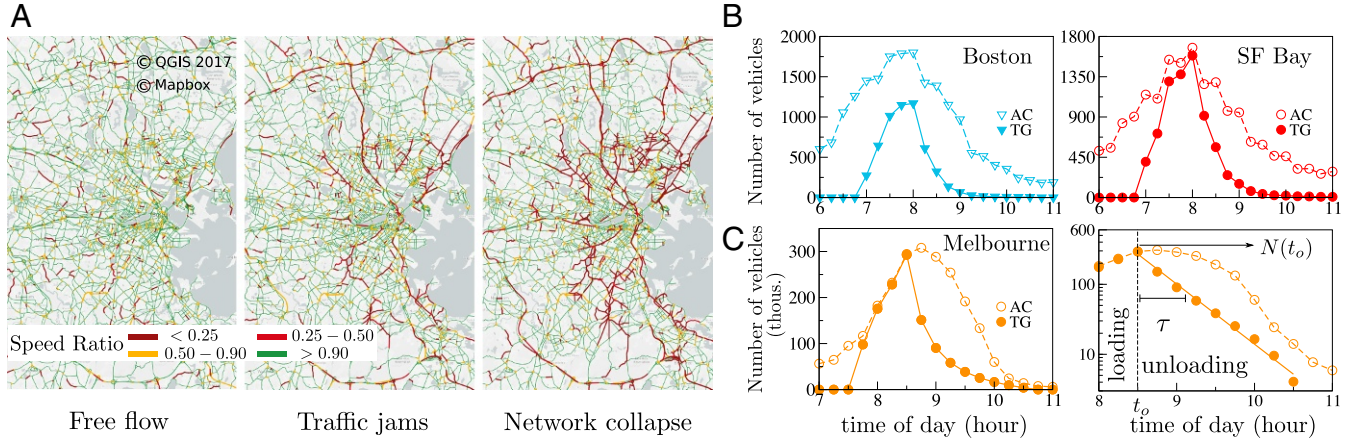

Fig. 1. Urban traffic dynamics. (*A*) Snapshots of the simulated Boston network just in the morning peak hour (8:30 AM) for three different demand levels. For very low traffic demand (free flow), interaction between vehicles can be neglected and the spatial distribution of short queues at intersections looks uncorrelated. For current traffic demands, traffic jams emerge along the roads, showing a clear heterogeneous spatial distribution, but even so vehicles still flow. For high demands, these traffic jams grow to such an extent that they merge into each other, leading to the eventual network collapse. Road segments are classified into four depicted categories according to the speed ratio, i.e., the current speed divided by the speed limit of the street (empty roads are included in the green category). (*B*) Temporal profile of the vehicle trips in the road network during the morning peak hour built from travel diaries in Boston and San Francisco. For the target group (TG) of vehicles entering from 7:00 AM to 8:00 AM, we observe a clear exponential unloading process; i.e., $N(t) = N(t_o)\exp(-\frac{t-t_o}{\tau})$. In contrast, the profile for all commuters (AC) shows a more heterogeneous behavior. Similar results for Bogotá are presented in *SI Appendix*, Fig. S1. (*C*) A simulation-based dynamics traffic assignment model of Melbourne confirms the exponential decay of the TG, here defined from 7:30 AM to 8:30 AM. In contrast, all of the vehicles of commuters have a more heterogeneous unloading, not properly characterized by a single relaxation time.

CA Model

Due to the complexities of large-scale traffic simulations, to compare cities we implement a CA model (33). As an input, we use validated travel demand models obtained by Çolak et al. (22). We focus on traffic demand from 7:30 AM to 8:30 AM for the subject cities (22) (*Materials and Methods*). To model the observed recovery, we load the road network during 1 h; i.e., every time step Δt, the network is loaded with R randomly chosen trips. After that, we stop the loading and let the system recover within a long time window. We select a sufficiently long time of observation that allows us later to observe the dynamics of long-lasting traffic jams.

The initial route in the road networks is precalculated with the congested traveled time t_{ue} as weights assuming the shortest time path. Vehicles depart from origin nodes (intersections) and then are inserted into the network according to the initial routing strategy, such that, for high insertion rates, queues of new vehicles can be formed at origin nodes. Once vehicles are inserted into the network, the trips toward their destinations are determined by two processes: the dynamics along the streets and the routing at the intersections (Fig. 2*A*).

We implement the deterministic Nagel–Schreckenberg CA model (33), where for simplicity each edge only has one lane. When a car is traveling in free flow and approaches an intersection, it decreases its speed to v_{uts}. In the implementation presented here our unit time step (uts) is $\Delta t = 1.2375[\text{s}]$ and the cell sizes are $l = 5.5[\text{m}]$. This defines the minimum distance per uts that vehicles can travel in the absence of congestion, which corresponds to $v_{\text{uts}} = 16[\text{km/h}]$. Each vehicle keeps a gap distance which creates congestion when cars accumulate in the streets, taking speeds in the range $[0, v_{\max}]$, where $v_{\max}$ is the speed limit of the road segment.

In each time step in a random sequence, the intersections transmit the first vehicles of the originating streets. The vehicles pass through a street segment (*i*) if the first cell in the next desired street is empty and (*ii*) if the road capacity of the originating street allows it. In the latter case, the vehicle is delivered with a probability p proportional to the road capacity of the originating street, C_e (*SI Appendix*, Table S1). In the case of long waiting times, we introduce a basic dynamic routing strategy, commonly known as adaptive driving: A vehicle that has been stopped at an intersection during more than $t_{\text{wait}} = 96$ time steps (approximately 2 min) can decide to reroute to a less congested destination street and recomputes its route.

In *SI Appendix*, Fig. S3, we show how this simple model offers a reasonable description of important empirical features of urban traffic reported in previous works (12, 15, 17).

Comparing Morning Traffic

Due to the one-lane representation of the streets and the simplicity of the vehicle dynamics, the networks studied via the CA model are more susceptible to congestion if using the empirical volume demand (V). We thus rescaled V by the ratio between the space demands in simulated and real networks, given by

$$V_{ca} \equiv 0.7 \times \frac{\sum_{e \in E} x_e \cdot \ell_e}{\sum_{e \in E} x_e \cdot \ell_e \cdot n_e} \cdot V, \qquad [3]$$

where ℓ_e and n_e are the length (in kilometers) and number of lanes that a road segment e contains in the shortest path from origin to destination, and x_e is the number of cars that pass through it. With the calibration factor 0.7, we obtain a reasonable correlation between the travel times obtained from our

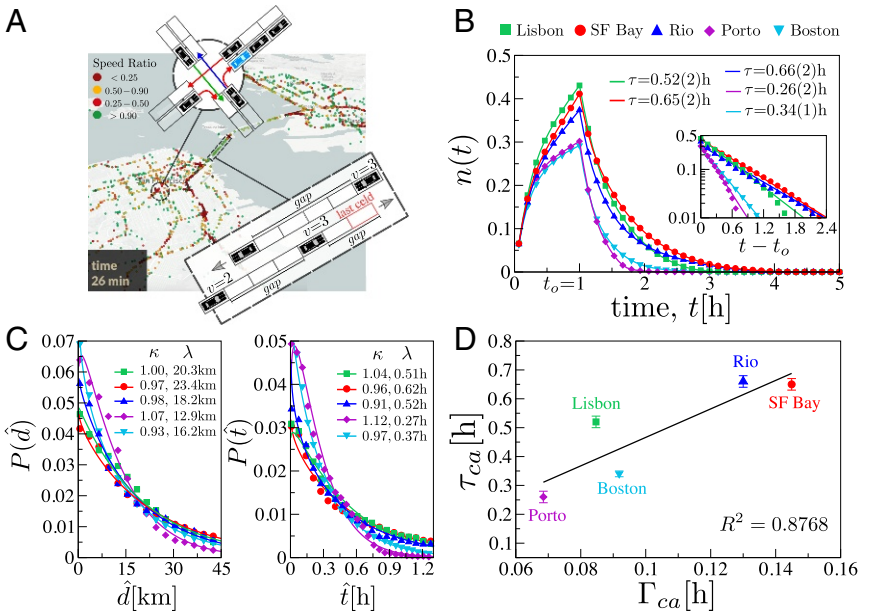

Fig. 2. Comparison of congestion levels for five studied cities. (*A*) Schematic for the proposed CA model. (*Lower Inset*) The snapshot illustrates the various cases of the NaSch CA dynamics (33) using $v_{\max} = 3$ cells/Δt; black vehicles correspond to the configuration at time t and gray ones correspond to the position at $t + 1$ after the velocity updating. (*Upper Inset*) Colored arrows illustrate several possibilities of moving in the intersections: green (successful), red (not successful), and blue (a new route chosen). A light-blue vehicle means that the cell is occupied by a vehicle that has crossed before or by one that just entered the network. (*B*) During the first hour every city is loaded homogeneously with a rescaled loading rate $R = R_{ca}$ (Table 1) from their empirical travel demand. *Inset* shows the log-lin plot of the recovery period with the unloading time τ of the fitted exponential function depicted in the key. Every curve here is an average over 30 realizations. (*C*) After the loading hour, the remaining distance $\hat{d}$ and the remaining travel time $\hat{t}$ for the cars in the network can be fitted by a Weibull distribution with parameters shown in the keys (see *SI Appendix*, Fig. S7 and Table S2, for more details). This explains the exponential behavior in *B*. (*D*) Under current traffic conditions, we observe a linear trend between τ and Γ, the actual demand to supply ratio for each city. Error bars show the SD. A comparison between τ and both the TomTom traffic index and data derived from GoogleAPI is shown in *SI Appendix*, Fig. S8.

Table 1. A comparison of the actual travel demand and road network in the subject cities

City characteristics	City				
	Boston	Porto	Lisbon	Rio de Janeiro	San Francisco Bay
Population, millions	4.5	1.7	2.8	12.6	7.15
Area, 1,000 km^2	4.6	2.0	2.9	4.6	18.1
Roads, thousand miles	12	3	7	6	30
Total volume, V, millions	0.916	0.171	0.324	0.432	1.015
V_{ca}/V	0.292	0.282	0.290	0.307	0.274
$R_{ca}[veh/\Delta t]$	92	16	33	45	94

simulations and the publicly available Google Maps Application Programming Interface (API) (*SI Appendix*, Fig. S4). Despite the differences in road infrastructure, there are small differences between these ratios for the considered cities, as shown in Table 1.

Next, we distribute the calibrated volume demand (V_{ca}) in the 1-h loading, such that $R_{ca} = \frac{V_{ca}}{1\,\mathrm{h}} \cdot \Delta t$ randomly chosen trips are inserted each time step Δt.

The observed patterns here are independent of the simulation modeling methodology. Both the simplified simulation and the more complex DTA model reveal similar outcomes. Fig. 2*B* shows the loading and unloading in terms of the fraction of vehicles in the network, $n(t) \equiv \frac{N(t)}{V_{ca}}$. The exponential decay is reproduced with unloading times τ_{ca} ranging from 0.26 h to 0.66 h (Fig. 2*B*, *Inset*). As expected, we can see the network responds differently from city to city. Note that the values of τ_{ca} of the empirical representation are higher than those of the empirical travel conditions. Because the same simplifications are done for each city, the relative differences of τ_{ca} are preserved.

The exponential recovery can be explained by commuting trip distances, d, which can be approximated by a log-normal distribution (*SI Appendix*, Fig. S5), $P(d;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma d} e^{-(\ln(d)-\mu)^2/2\sigma^2}$ with medians ranging from 10 [km] to 16 [km] and the SD ranging from 2.13 km to 2.91 km: $\mu(\sigma)_{\text{Boston}} = 2.31(1.07)$, $\mu(\sigma)_{\text{Porto}} = 2.52(0.76)$, $\mu(\sigma)_{\text{Lisbon}} = 2.77(0.91)$, $\mu(\sigma)_{\text{Rio}} = 2.48(1.04)$, and $\mu(\sigma)_{\text{San Francisco Bay}} = 2.73(0.97)$. After the first hour there is an important fraction of vehicles that have not reached their destination. The remaining distance of these vehicles follows a Weibull distribution, $P(\hat{d};\kappa,\lambda) = \frac{\kappa}{\lambda}(\frac{\hat{d}}{\lambda})^{\kappa-1} e^{-(\hat{d}/\lambda)^{\kappa}}$ (presented in Fig. 2*C*, *Left*). In a noncollapsed state, the remaining travel times, $\hat{t}$, follow the same distribution, as shown in Fig. 2*C*, *Right*. Thus, in the recovery period after the first hour, the number of vehicles as a function of time, $t_r = t - t_o$, can be estimated as

$$n(t_r) = n(t_o) \cdot \left(1 - \int_0^{t_r} P(\hat{t};\kappa,\lambda)\, d\hat{t}\right) = n(t_o) \cdot e^{-(t_r/\lambda)^{\kappa}}. \quad [4]$$

Note that κ takes values very close to 1.0 and the λ values are similar to the τ_{ca} of each city; this explains the exponential recovery time. Similar results are obtained for Melbourne (*SI Appendix*, Fig. S6).

We can further explain τ_{ca} as a function of the vehicle miles traveled and the road network capacity. To that end, we measure the demand-to-supply ratio, defined as

$$\Gamma \equiv \frac{\sum_{e\in E} \ell_e x_e}{\sum_{x_e>0, e\in E} \ell_e C_e}. \quad [5]$$

This measure is a simple ratio. The numerator is the distance that needs to be traveled from origin to destination, in the static equilibrium with the estimates of current demand, adding the trips of the target group that departs in the peak hour. The denominator is the upper bound of the total vehicle kilometers the road network can support per hour. Thus, it captures the spatial distribution of the loading on the available road infrastructure of the city. This definition differs from the one used in ref. 22; here Γ has dimensions of time. The values of this parameter under current traffic conditions, Γ_{ca}, are shown in Table 2. As shown in Fig. 2*D* there is an increasing relation between τ_{ca} and Γ_{ca}. Remarkably, the congestion levels suffered by commuters in diverse cities can be defined by a metric that synthesizes the trip distribution of a target group and the available road infrastructure. *SI Appendix*, Fig. S9 shows a sensitivity analysis of τ_{ca} with respect to the parameter t_{wait} of the CA model.

Dynamics of Urban Traffic

To have a complete understanding of τ and the macroscopic dynamics of urban traffic, we analyze different demand levels, keeping the spatial distribution of trips and the loading in the morning peak hour. In doing so, we uncover three different states of urban traffic (Fig. 3). We call them free flow, traffic jam, and network collapse, also known as gridlock. We further use the superscript i to emphasize that the values are calculated for each city i. First, for very low Γ^i, as vehicles do not affect each other, $\tau^i = \tau^i_o$ is independent of the volume of cars and depends only on the distribution of travel distances in each city (Table 3). After a certain value Γ^i_1, queues start to emerge at the intersections. As a

Table 2. Comparison of metrics associated with urban traffic in the subject cities

City metrics	City				
	Boston	Porto	Lisbon	Rio de Janeiro	San Francisco Bay
Demand to supply, Γ_{ca}[h]	0.092	0.068	0.085	0.130	0.145
Unloading time, τ_{ca}[h]	0.34(1)	0.26(2)	0.52(2)	0.66(2)	0.65(2)
Mean commuting distance d[km]	10.07	12.45	15.99	11.98	15.33
SD commuting distance [km]	2.91	2.13	2.48	2.82	2.64
Mean remaining distance $\hat{d}$[km]	16.22	11.48	18.52	18.34	19.53
Mean remaining travel time, $\hat{t}$[h]	0.321	0.258	0.501	0.534	0.629

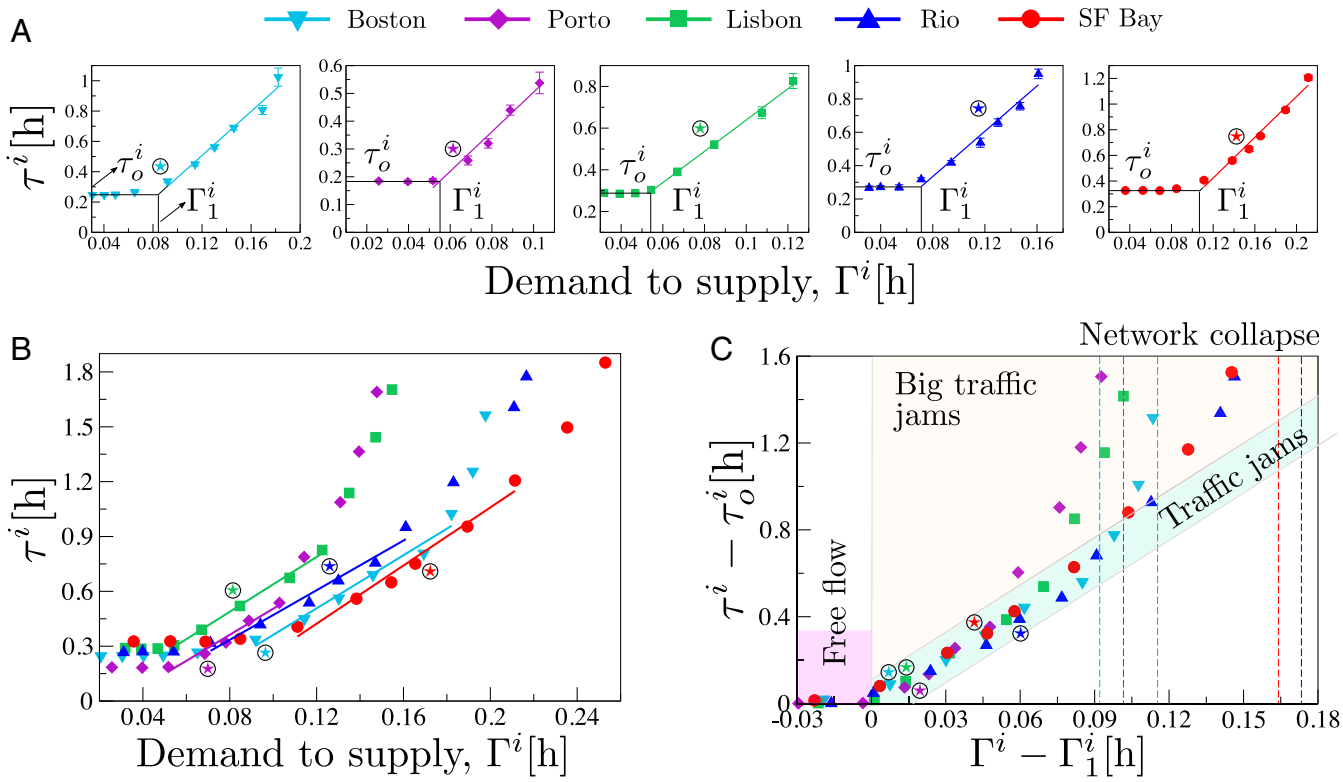

Fig. 3. Effects of Γ on the unloading time τ. (*A*) When scanning different demand levels, we can identify a transition from a free-flow state with a constant $\tau^i = \tau_o^i$ to a traffic jam phase where τ^i increases linearly with Γ^i. (*B*) Comparison between previous curves for all cities for a wider travel demand range. Beyond the linear region, τ increases dramatically as evidence of the transition to a collapsed state. (*C*) Plot of $\tau^i - \tau_o^i$ vs. $\Gamma^i - \Gamma_1^i$. The traffic jam phases for all cities collapse onto a thin band, meaning that urban traffic networks respond similarly to the congestion. Vertical lines identify the critical threshold to the network collapse. Every point is an average over 30 realizations. Stars represent the current traffic conditions for each city.

result, while preserving the exponential recovery, τ depends linearly on Γ. This is the state observed in the studied cities under the current commuting conditions. Interestingly, when we plot $\tau^i - \tau_o^i$ vs. $\Gamma^i - \Gamma_1^i$, the linear regions for all cities collapse into a band as shown in Fig. 3*C*, indicating that after their particular threshold Γ_1^i, all observed cities respond similarly to the congestion. It is worth mentioning that Γ_1^i coincides with the critical insertion rates studied in previous works with constant loadings (18–20) (*SI Appendix*, Fig. S10). Now, if the demand keeps growing, long-lasting traffic jams emerge, deforming the initially exponential unloading, and thus τ increase dramatically as shown in Fig. 3*C*. This is the onset of the transition to the network collapse, because for a large number of cars we do not observe an exponential unloading anymore. From this state, τ is measured as how long it takes to unload the network to reach the value of $n(t_o)\exp(-1)$.

Γ is a state variable that allows us to compare the congestion level for different cities. Interestingly, according to its definition (Eq. **5**), Γ is an extensive variable, and thus we can expect a dependency on the city size. *SI Appendix*, Fig. S11 suggests that Γ_{ca} may be set just by population (P) and spatial extent of the urban area (A). We find $\Gamma_{ca} \sim (PA)^{0.21(4)}$, a result aligned with the recently reported scaling laws for urban systems (28–30).

Urban Vulnerability and Transition to a Collapsed State

While the considered cities already face high traffic demand, the studied exponential decay indicates a characteristic time in the recovery without the occurrence of long-lasting traffic jams that expand to the majority of the network. We further compare the emergence of collapse induced by the number of cars.

Fig. 4*A* shows the recovery period for all five cities under high loading rates. At a certain critical value R_c^i (colored symbols), cities recover algebraically over a wide time window up to an exponential cutoff. As depicted, the power-law decay is fitted as $n(t) \sim t^{-\alpha}$, where the critical exponent α is obtained from the slopes of the black solid lines. For $R > R_c$, the unloading follows a slower recovery until an inflection point appears, where the outflow rate decreases to very low levels and remains so for a considerable time, as shown in *SI Appendix*, Fig. S12. In this transient collapsed state, most of the vehicles remain trapped in long-lasting gridlocks. Eventually, due to the rerouting possibilities, the system recovers and unloads completely.

The dynamics of $n(t)$ resemble the critical behavior of the directed percolation (DP) universality class, with the difference that traffic systems do not have irreversible absorbing states due to the rerouting rules. Prompted by this analogy, we follow the scaling approach in the DP framework (34–36) for characterizing the transition to the network collapse. As Fig. 4*B* shows, we find that curves collapse when $n(t) \cdot t^{\alpha}$ is plotted as a function of $(t - t_o)\varepsilon$, where $\varepsilon = \frac{|R^2 - R_c^2|}{R_c^2}$ is the deviation from criticality.

Table 3. Comparison of metrics associated to the phase transitions observed in the CA model of the subject cities

	City				
Critical parameters	Boston	Porto	Lisbon	Rio de Janeiro	San Francisco Bay
Γ_1^i[h]	0.0843	0.055	0.0531	0.0703	0.1077
τ_o^i[h]	0.247(2)	0.184(2)	0.287(2)	0.269(2)	0.326(1)
R_c^i $\frac{\text{vehs}}{\Delta t}$	178	37	66	86	168
α	1.54(1)	0.60(2)	0.56(1)	0.65(1)	0.75(1)
β	0.8(1)	0.21(2)	0.27(2)	0.61(3)	0.53(5)

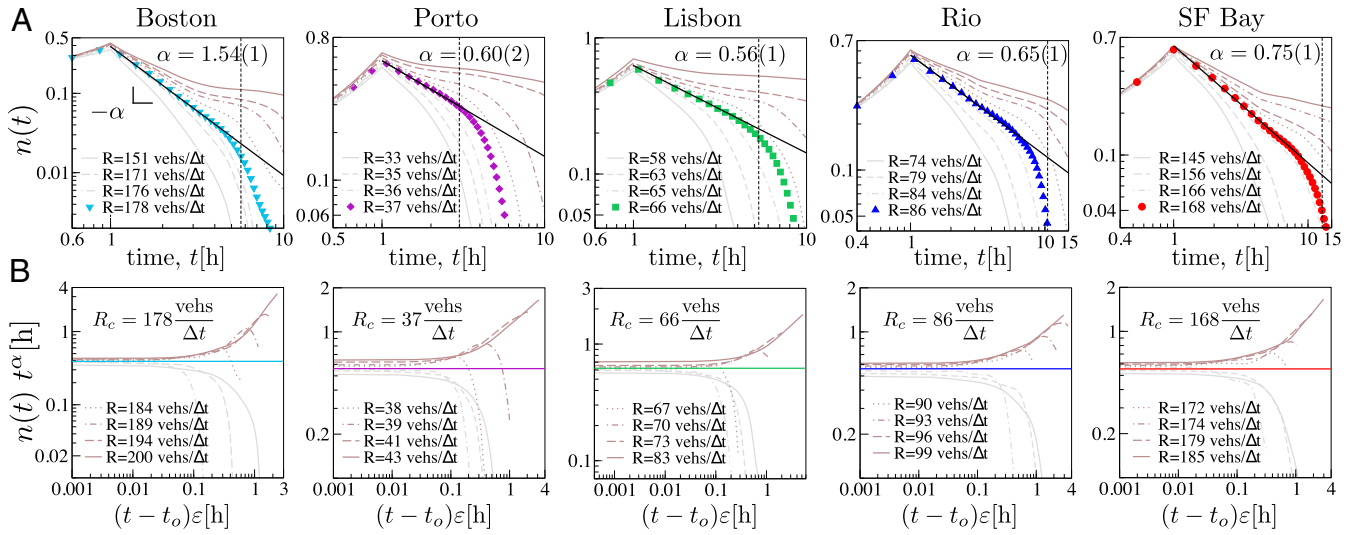

Fig. 4. Critical behavior of the fraction of vehicles in the network $n(t)$. To study the transition to urban gridlock, we systematically increase the traffic demand per time step in each city. (*A*) Number of vehicles in the network $n(t)$ after the loading, for several loading rates. The critical condition is indicated by the solid black lines where the $n(t)$, the order parameter, follows a power law $t^{-\alpha}$. The nonuniversal values of the critical exponent α are shown for each city. The critical loading rate R_c is estimated as the average loading rate of the two colored curves. Vertical lines define the time at which the order parameter $n(R)$ is measured (*SI Appendix*, Fig. S13). (*B*) Scaling plot of data in *A*, where $\varepsilon = \frac{|R^2 - R_c^2|}{R_c^2}$ is the deviation from criticality. Above the critical point $R > R_c$, the system falls into a gridlock state that eventually resolves after several hours.

Even though α is different for each city, indicating nonuniversality, this behavior evidences that the onset of the traffic gridlock is a nonequilibrium phase transition (34–36).

We further study the transition in the supercritical region, using the loading rate as a control parameter. The natural order parameter is the remaining percentage of vehicles in the

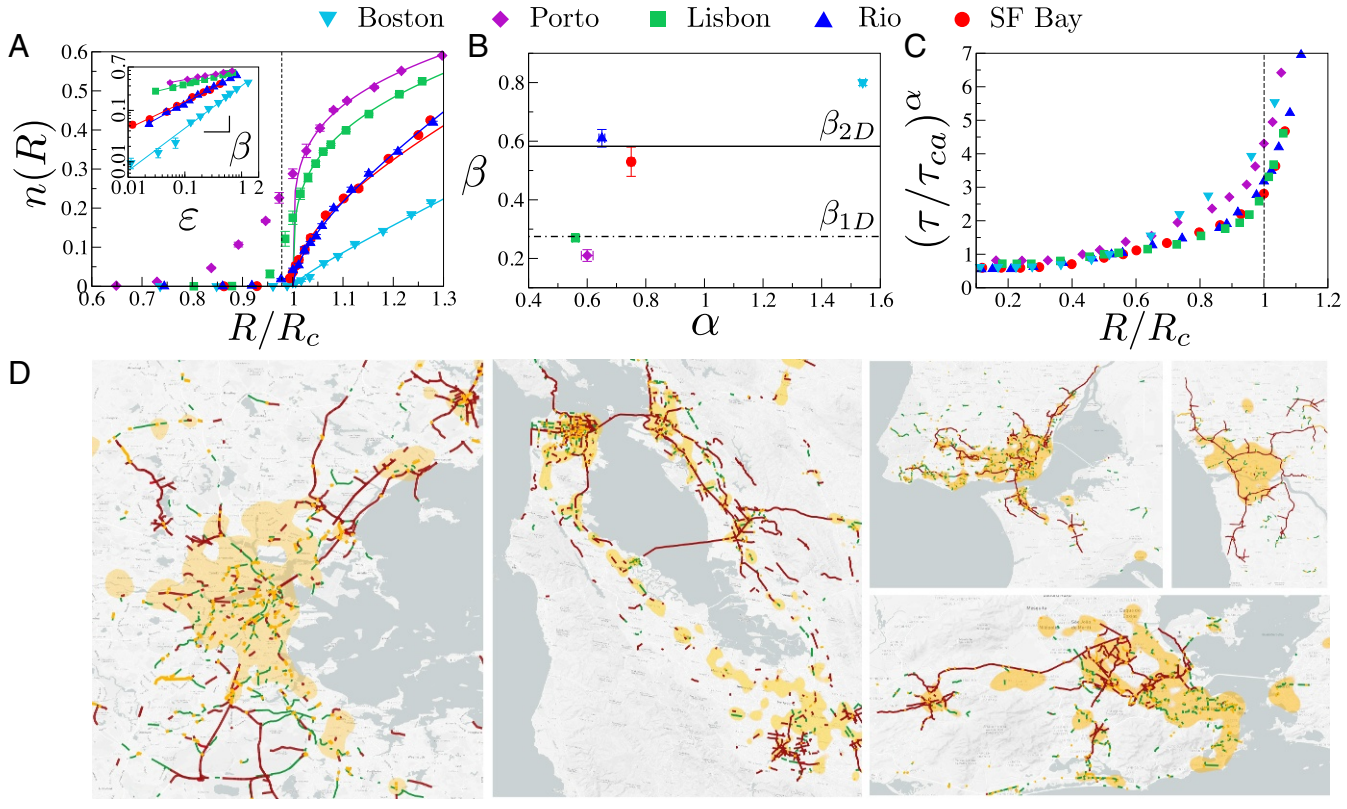

Fig. 5. Phase transition vs. the control parameter R/R_c. (*A*) Order parameter $n(R)$ vs. R/R_c for the five cities. *Inset* shows the same data in logarithmic scales, and the solid lines show $n(R) \sim \varepsilon^{\beta}$. (*B*) Relationship between the critical exponents α and β. Black lines depict the universal values of β_{1D} (dashed-dotted line) and β_{2D} (solid line) of DP. (*C*) Recovery time T diverges at the critical point. A data collapse can be obtained by plotting the quantity $(T/\tau)^{\alpha}$. (*D*) Snapshots of long-lasting gridlock configurations for all cities with $R > R_c$. The color code of the road segments is the same as in Fig. 1*A*, and the orange shading indicates zones with a high density of destinations.

network at very long times, $n(R)$, where it is expected that a scaling law $n(R) \sim \varepsilon^{\beta}$. However, since the gridlocks always recover, it becomes difficult to define the proper time t to measure $n(R)$. We define it as the moment at which the best fit of β is obtained for each city, indicated by the vertical lines in Fig. 4*A* (also *SI Appendix*, Fig. S13).

Fig. 5*A* shows $n(R)$ as a function of $\frac{R}{R_c}$ for all cities. As depicted, with increasing $\frac{R}{R_c}$, $n(R)$ increases continuously at the vicinity of $\frac{R}{R_c} = 1.0$. The power law $n(R) \sim \varepsilon^{\beta}$ in this second-order transition is shown in Fig. 5*A*, *Inset*. Fig. 5*B* shows how β increases with increasing α. For comparison, the universal exponents of (1 + 1) and (2 + 1) dimensions in DP are also sketched.

To compare the response of various cities, we define $(\tau/\tau_{ca})^{\alpha}$ as the temporal order parameter. Note that τ_{ca} are the reference values that resemble the traffic conditions in the city today. Fig. 5*C* shows that this quantity diverges at the vicinity of $\frac{R}{R_c} = 1.0$ and, once again, we obtain a data collapse onto a single curve. Thus, the transition point separates a traffic jam phase with finite recovery times and a collapsed phase in which the dynamics are almost frozen.

The dynamics of the presented results do not depend on the rules of the CA. A sensitivity analysis (*SI Appendix*, Fig. S14) indicates that, as expected, R_c increases and α decreases with increasing t_{wait}. In contrast, the critical exponent β seems to be robust, suggesting that it does not depend on details of the model such as rerouting strategy, but rather depends on the structural shape of the traffic gridlocks in the city. To have some insights about the differences in the cities' critical behavior, we show snapshots of long-lasting gridlock configurations along with the business districts for each city. At one extreme, Porto exhibits a single mega-jam spreading in its unique business district. At the other end, Boston shows several big traffic jams, one for each working zone. The differences in the spatial structure from a monocentric to a polycentric organization generate differences in the values of the exponents that characterize the collapse. Further spatial studies of traffic dynamics under actual conditions have been studied with percolation theory (23). In *SI Appendix*, Figs. S15 and S16, we study the percolation transition with various demand volumes. From the perspective of the nonequilibrium phase transitions, future work should focus on studying the spatial and temporal correlation lengths around the criticality (R_c). Further comparative studies of cities can give insights into evaluating mitigation strategies like congestion pricing (37–39) and routing policies (40). A discussion on this can be found in *SI Appendix*, section A.

Discussion

We have uncovered an exponential recovery in urban traffic, given by the unloading time of a target group in the peak hour. This value relates individual mobility with traffic congestion levels and depends on the road infrastructure and travel demand. Within this framework, we measure the vulnerability of urban networks to loading vehicles, while keeping the trip distributions and street capacities unchanged. We show that the transition to urban gridlock resembles the DP universality class, and it is studied as a nonequilibrium phase transition. In a data-rich reality, the aim of this work is to open avenues for direct measurements and description of urban traffic with methods of statistical physics.

Materials and Methods

Surveys Datasets. Data were derived from the 2010/2011 Massachusetts Travel Survey (MTS) and the 2000 Bay Area Transportation Survey (BATS).

Extraction of Validated Travel Demand Information. Travel information can be extracted from the analysis of CDRs from mobile phones. Çolak et al. (22) estimated and validated the OD tables during the morning peak hour for the same five cities discussed here. Using these OD tables, along with road networks publicly available on OpenStreetMaps (OSM), the congested travel time (t_{ue}) for each road was estimated through a static traffic assignment algorithm. In summary, we implement a four-step travel model (31, 32). For the CA simulations, we start from OD matrices aggregated at a level of census tracts. Once a trip is chosen to enter the network, both the origin and destination nodes are assigned at random from the nodes within the corresponding census tract.

Simulations of Melbourne. DynaMel is a large-scale simulation-based DTA model of Melbourne, Australia. The model has been calibrated and validated to simulate the 6:00 AM to 10:00 AM morning peak period. The model consists of 55, 719 links and 24, 502 nodes and simulates almost 2.1 million commuters in the 6:00 AM to 10:00 AM morning peak period. As an input of the DTA model, DynaMel applies a machine-learning–based technique to calibrate the traffic-flow fundamental diagram using the observed traffic data from hundreds of freeway loop detectors across the entire network.

Details of the CA Model. Vehicle dynamics along road segments are modeled with the deterministic Nagel–Schreckenberg CA model (33). For simplicity, each edge has only one lane. Therefore, every road segment is discretized in cells of equal length, $l = 5.5$[m], and no more than one vehicle can occupy a cell at every time step $\Delta t = 1.2375$[s]; hence the speed unity is fixed to $v_{\text{uts}} = 16$[km/h] corresponding to 1 cell per time step. At each time step, all vehicles update in parallel their velocities and move according to the rule $v_{t+1} = \min(v_t + 1, gap, v_{\max})$, where *gap* is the distance from the car ahead or to the next intersection and $v_{\max}$ taken from the speed limit of the road network data (Fig. 1*A*). When a car is traveling in free flow and approaches an intersection, it decreases its speed to v_{uts}. The dynamic at the intersection has three steps. First, the first new vehicle in the queue is chosen to pass. Second, incoming streets are checked in a random sequence asking for the destination of the first vehicle in the street. If the intersection is just the destination node, then the vehicle is removed from the network. Otherwise, if the first cell of the desired destination street is free, the vehicle is delivered with a probability p proportional to the empirical flow capacity of the originating street, C_e (see Table S1 for the values used). Finally, in the case of long waiting times, there is the rerouting strategy we have already explained.

For contractual and privacy reasons, we cannot make the raw data available. We are pleased to make available the data of the OD matrices, software to replicate the method, and the appropriate documentation. This information may be accessed at the GitHub repository https://github.com/leolmoss/CollapseUrbanTraffic. This repository is sufficient to reproduce the results of this paper.

ACKNOWLEDGMENTS. This work was supported by the Massachusetts Institute of Technology Energy Initiative (MITEI), and the MIT Environmental Solutions Initiative as well as by Lawrence Berkeley National Laboratory's Laboratory Directed Research and Development (LDRD) funds. Views and conclusions in this document are those of the authors and should not be interpreted as representing the policies, either expressed or implied, of the sponsors.

1. Hanna R, Kreindler G, Olken BA (2017) Citywide effects of high-occupancy vehicle restrictions: Evidence from "three-in-one" in Jakarta. *Science* 357:89–93.
2. Molloy M (2015) World's worst traffic jam? Thousands of cars left stranded on motorway in China. *The Telegraph*, Section World News - China. Available at https://www.telegraph.co.uk/news/worldnews/asia/china/11919370/Worlds-worst-traffic-jam-Thousands-of-cars-left-stranded-on-motorway-in-China.html. Accessed July 5, 2018.
3. TomTom (2016) Traffic congestion index. Available at tomtom.com/. Accessed November 2016.
4. Sweet M (2014) Traffic congestion's economic impacts: Evidence from US metropolitan regions. *Urban Stud* 51:2088–2110.
5. McDonald RI, et al. (2011) Urban growth, climate change, and freshwater availability. *Proc Natl Acad Sci USA* 108:6312–6317.
6. Shindell D, et al. (2011) Climate, health, agricultural and economic impacts of tighter vehicle-emission standards. *Nat Clim Change* 1:59–66.
7. Lelieveld J, Evans JS, Fnais M,Giannadaki D, Pozzer A (2015) The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* 525:367–371.
8. Helbing D (2001) Traffic and related self-driven many-particle systems. *Rev Mod Phys* 73:1067–1141.
9. Chowdhury D, Santen L, Schadschneider A (2000) Statistical physics of vehicular traffic and some related systems. *Phys Rep* 329:199–329.

10. Treiber M, Kesting A (2013) *Traffic Flow Dynamics* (Springer, Berlin).
11. Godfrey J (1969) The mechanism of a road network. *Traffic Eng Control* 11:323–327.
12. Herman R, Prigogine I (1979) A two-fluid approach to town traffic. *Science* 204:148–151.
13. Mahmassani H, Williams J, Herman R (1984) Investigation of network-level traffic flow relationships: Some simulation results. *Transportation Research Record: Journal of the Transportation Research Board*, ed Kaplan EW (Transportation Research Board, Washington, D.C.), pp 121–130.
14. Daganzo CF, Geroliminis N (2008) An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B* 42:771–781.
15. Daganzo CF, Geroliminis N (2008) Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Res B* 42:759–770.
16. Daganzo CF (2007) Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Res B* 41:49–62.
17. Mazloumian A, Geroliminis N, Helbing D (2010) The spatial variability of vehicle densities as determinant of urban network capacity. *Philos Trans R Soc A* 368:4627–4647.
18. Arenas A, Díaz-Guilera A, Guimerà R (2001) Communication in networks with hierarchical branching. *Phys Rev Lett* 86:3196–3199.
19. Echenique P, Gómez-Gardeñes J, Moreno Y (2005) Dynamics of jamming transitions in complex networks. *Europhys Lett* 71:325–331.
20. Çolak S, Schneider CM, González MC (2013) On the role of spatial dynamics and topology on network flows. *New J Phys* 15:113037.
21. Toole JL, et al. (2015) The path most traveled: Travel demand estimation using big data resources. *Transport Res C Emerg Technol* 58:162–167.
22. Çolak S, Lima A, González MC (2016) Understanding congested travel in urban areas. *Nat Commun* 7:10793.
23. Li D, et al. (2015) Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proc Natl Acad Sci USA* 112:669–672.
24. Alexander L, Jiang S, Murga M, González MC (2015) Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transport Res C Emerg Technol* 58:240–250.
25. González M, Hidalgo C, Barabási A (2008) Understanding individual human mobility patterns. *Nature* 453:779–782.
26. Jiang S, et al. (2016) The TimeGeo modeling framework for urban mobility without travel surveys. *Proc Natl Acad Sci USA* 113:E5370–E5378.
27. Bettencourt LM, Lobo J, Helbing, DKC, West GB (2007) Growth, innovation, scaling, and the pace of life in cities. *Proc Natl Acad Sci USA* 104:7301–7306.
28. Bettencourt LMA (2013) The origins of scaling in cities. *Science* 340: 1438–1441.
29. Samaniego H, Moses ME (2008) Cities as organisms: Allometric scaling of urban road networks. *J Transport Land Use* 1:21–39.
30. Louf R, Barthelemy M (2014) How congestion shapes cities: From mobility patterns to scaling. *Sci Rep* 4:5561.
31. Mahmassani HS, Saberi M, Zockaie A (2013) Urban network gridlock: Theory, characteristics, and dynamics. *Transp Res C Emerg Technol* 36:480–497.
32. Shafiei S, Gu Z, Sarvi M, Saberi M (2018) Calibration and validation of a simulation-based dynamic traffic assignment model for a large-scale congested network. *Simul Model Pract Th* 86:169–186.
33. Nagel K, Schreckenberg M (1992) A cellular automaton model for freeway traffic. *J Phys France* 2:2221–2229.
34. Henkel M, Hinrichsen H, Lübfcbeck S (2008) *Non-Equilibrium Phase Transitions* (Springer, Dordrecht, The Netherlands), Vol 1.
35. Marro J, Dickman R (1999) *Nonequilibrium Phase Transitions in Lattice Models* (Cambridge Univ Press, New York).
36. Grassberger P, Krause F, von der Twer T (1984) A new type of kinetic critical phenomenon. *J Phys A Math Gen* 17:L105–L109.
37. Mun Si, Konishi Kj, Yoshikawa K (2003) Optimal cordon pricing. *J Urban Econ* 54:21–38.
38. Kraus M (1989) The welfare gains from pricing road congestion using automatic vehicle identification and on vehicle meters. *J Urban Econ* 25:261–281.
39. Solé-Ribalta A, Gómez S, Arenas A (2017) Decongestion of urban areas with hotspot pricing. *Netw Spat Econ* 18:33–50.
40. Zhang J, Pourazarm S, Cassandras CG, Paschalidis IC (2018) The price of anarchy in transportation networks: Data-driven evaluation and reduction strategies. *Proc IEEE* 106:538–553.