

ARTICLE



<https://doi.org/10.1038/s41467-020-15353-2>

OPEN

A simple contagion process describes spreading of traffic jams in urban networks

Meead Saberi¹✉, Homayoun Hamedmoghadam², Mudabber Ashfaq¹, Seyed Amir Hosseini³, Ziyuan Gu¹, Sajjad Shafiei⁴, Divya J. Nair¹, Vinayak Dixit¹, Lauren Gardner⁵, S. Travis Waller¹ & Marta C. González⁶

The spread of traffic jams in urban networks has long been viewed as a complex spatio-temporal phenomenon that often requires computationally intensive microscopic models for analysis purposes. In this study, we present a framework to describe the dynamics of congestion propagation and dissipation of traffic in cities using a simple contagion process, inspired by those used to model infectious disease spread in a population. We introduce two macroscopic characteristics for network traffic dynamics, namely congestion propagation rate β and congestion dissipation rate μ . We describe the dynamics of congestion spread using these new parameters embedded within a system of ordinary differential equations, similar to the well-known susceptible-infected-recovered (SIR) model. The proposed contagion-based dynamics are verified through an empirical multi-city analysis, and can be used to monitor, predict and control the fraction of congested links in the network over time.

¹Research Centre for Integrated Transport Innovation (rCITI), School of Civil and Environmental Engineering, University of New South Wales (UNSW), Sydney, NSW 2032, Australia. ²Department of Civil Engineering, Institute of Transport Studies, Monash University, Melbourne, VIC 3800, Australia. ³School of Electrical and Computer Engineering, K.N. Toosi University of Technology, Tehran, Iran. ⁴Data61, CSIRO, Sydney 2015 NSW, Australia. ⁵Department of Civil Engineering, Johns Hopkins University, Baltimore, MD 21218, USA. ⁶College of Environmental Design, University of California, Berkeley, CA 94720, USA. ✉email: meead.saberi@unsw.edu.au

Traffic congestion in cities, also known as traffic jams, propagate over time and space. Existing approaches to model city traffic often rely on microscopic models with high computational burden as well as excessive parameterization required for calibration^{1–3}. Further, the lack of available transportation network data in many countries, especially those that are less economically developed, poses a challenge for traffic modelers. However, the fast-paced development and deployment of mobile sensors offers the opportunity to generate continuous spatial data, which further enables the estimation of road traffic conditions in real time and at the macroscopic level.

Numerous studies have explored different macroscopic approaches to model the spread of traffic jams in cities^{4–9}, including through the lens of percolation theory^{10,11}, machine-learning methods¹², and queuing theory^{13,14}. However, machine-learning models¹² and models based on queuing theory^{13,14} are unable to capture and quantitatively describe the propagation and dissipation patterns of congestion over time and space. They are instead formulated based on input (arrival) and output (departure) rates to estimate the number of vehicles (i.e., accumulation) in an area, similar to conservation-based models using Macroscopic or Network Fundamental Diagram (MFD or NFD)⁴. In this work, we show that traffic systems exhibit underlying spreading dynamics similar to those observed and applied in other network systems, for example, social networks, population contact networks, or technological networks. Specifically, we propose and empirically demonstrate that traffic congestion in urban networks can be characterized using a simple contagion process, similar to the well-known susceptible-infected-recovered (SIR) model used to describe spread of infectious diseases in a population, wherein traffic spreads and recovers throughout the network over time. A previous study by Wu et al.¹⁵ conjectured that traffic congestion spread can be described with a SIR model, but provided no empirical evidence. To the best of the authors' knowledge, characterizing and modeling congestion propagation and dissipation as a network spreading phenomenon has never been empirically validated.

Urban traffic often exhibits high spatial correlation in which links adjacent to a congested link are more likely to become congested. Additionally, there is a strong temporal correlation in urban congestion, which is known to be driven by the time-dependent profile of travel demand. The spread of congestion at the link level is well theorized and understood with queuing and kinematic wave theories^{16–18}. However, our understanding of congestion propagation dynamics at the network level is still incomplete. Queue spillbacks in networks are shown to be sensitive to link capacities¹⁹, which could remain stable in both under- and over-saturated conditions. Congestion also exhibits fragmentation during recovery²⁰ leading to greater spatial heterogeneity and thus results in a drop of network production^{21–23}. Propagation and dissipation of gridlocks can be characterized by the number of congested links or the length of congestion in the network²³ with propagation occurring often at a much higher rate than dissipation. Also, studies have shown that the ratio of road supply to travel demand could explain the percentage of time lost in congestion as an aggregate measure^{2,22}.

Unlike individual link traffic shockwaves in a two-dimensional time-space diagram, which are categorized as forward or backward moving, network traffic jams evolve in multiple directions over space. Therefore, we propose that a network's propagation and dissipation can be characterized by two average rates, namely, the congestion propagation rate β and the congestion recovery rate μ , which together can predict the number of congested links in the network over time. These two macroscopic characteristics are critical in modeling congestion propagation and dissipation as a simple contagion process²³.

Despite the complex human behavior-driven nature of traffic, we demonstrate that urban network traffic congestion follows a surprisingly similar spreading pattern as in other systems, including the spread of infectious disease in a population or diffusion of ideas in a social network, and can be described using a similar parsimonious theoretical network framework. Specifically, we model the spread of congestion in urban networks by adapting a classical epidemic model to include a propagation and dissipation mechanism dependent on time-varying travel demand and consistent with fundamentals of network traffic flow theory. We illustrate the model to be predictive, and validate the framework using empirical and simulation-based numerical experiments. The proposed model can be used for adaptive and predictive control of congestion in an urban network. By monitoring the network in real time and observing the number of congested links, the model can be applied to develop optimal control strategies with different objectives, such as minimizing the total duration of congestion, minimizing the total number of congested links, and minimizing the recovery time. Similar to other macroscopic model-based control applications^{24–26}, we can employ the proposed SIR-based model to improve the traffic network performance by controlling and optimizing the network input through optimal metering of traffic flow or by increasing the network recovery rate through improved signal timing, bottleneck removal, and capacity expansion.

Results

Identifying congested links. We use empirical data from Google that contains estimated time-dependent traffic speeds on every link in the road network across six different cities in the world, namely Chicago, London, Paris, Sydney, Melbourne, and Montreal (see Methods). We also use simulated data from a calibrated and validated mesoscopic dynamic traffic assignment model of Melbourne (see Methods)¹. Using both empirical and simulation data, we demonstrate that the proposed modeling framework can successfully describe the dynamics of congestion propagation and dissipation in urban networks.

In the proposed network-theoretic framework, nodes represent the intersections (controlled and uncontrolled) and links represent the physical roads between any two intersections. For each link i in the network, we have the time-dependent speed $v_i(t)$. We represent the maximum speed on the link as $v_i^{\max}$. To reveal how congestion propagates and dissipates in a network, we define

$$\lambda_i(t) = \frac{v_i(t)}{v_i^{\max}}, \quad (1)$$

where $\lambda_i(t)$ is the ratio of link speed $v_i(t)$ over link maximum speed $v_i^{\max}$. We then classify each link as in either a congested $s_i = 1$ or uncongested $s_i = 0$ state (also known as free flow) using a threshold ρ as below

$$s_i(t) = \begin{cases} 1, & \lambda_i(t) < \rho, \\ 0, & \lambda_i(t) \geq \rho, \end{cases} \quad (2)$$

where ρ is a pre-specified threshold that represents different congestion levels¹¹. Figure 1 illustrates the identified congested network for different values of ρ at a given t using data from the simulation-based dynamic traffic assignment model of Melbourne. The size of the congested network grows as ρ increases. Alternatively, one can also construct the congested network using traffic density measurements or any other classification method.

Modeling contagion dynamics of network traffic jams. Consider a network with N directed links. At time $t = 0$ every link in the network is in the free flow or uncongested regime, $F(0) = N$

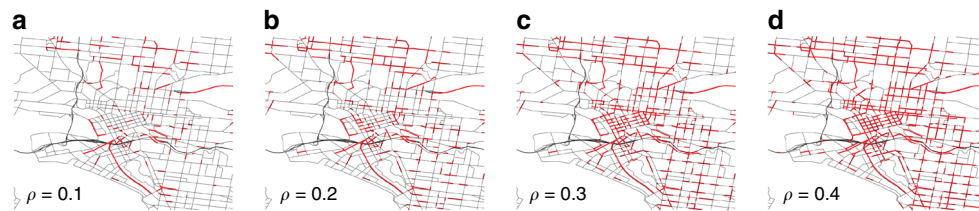


Fig. 1 Spatial distribution of congestion in the Melbourne network. At a given time $t = 8:30$ a.m. maps indicate the congested links (coded red) on the network when **a** $\rho = 0.1$, **b** $\rho = 0.2$, **c** $\rho = 0.3$, and **d** $\rho = 0.4$.

and no link is congested $C(0) = 0$. Let E be the set of links, $j(j \in E)$ of length l_j on which congestion later forms and propagates throughout the network. $F(t)$ represents the number of links that are at free flow regime at time t and $C(t)$ represents the number of congested links at time t based on the specified ρ . For a directed link in the road network with on average k “effective contacts” with other links pointing at its upstream node, β is the rate of congestion propagating to an upstream free flow link in unit time, and is $F(t)/N$ is the probability of a congested link being directly connected downstream of a free flow link. Under the assumption of homogeneous mixing, which refers to an assumption that each link in the network has the same probability of contact with a congested link; a congested link connects to an average of $kF(t)/N$ free flow links in a unit time. Although the assumption of homogeneous mixing is simplistic, it makes the analysis tractable while shown to be predictive at the macroscopic scale. See Supplementary Note 1 for the fundamentals of congestion propagation in simplified traffic systems. See Supplementary Note 2 for the extension of the framework when the homogeneous mixing assumption is relaxed.

Below, we describe the dynamics of congestion propagation with a system of ordinary differential equations (ODEs), which is analogous to the well-known SIR model:

$$\frac{dc(t)}{dt} = -\mu c(t) + \beta kc(t)(1 - r(t) - c(t)), \quad (3)$$

$$\frac{dr(t)}{dt} = \mu c(t), \quad (4)$$

$$\frac{df(t)}{dt} = -\beta kc(t)(1 - r(t) - c(t)), \quad (5)$$

where $c(t)$ represents the fraction of congested links in the network, $f(t)$ is the fraction of free flow links, and $r(t)$ is the fraction of recovered links. Equation 3 describes the rate in which $c(t)$ changes over time given a propagation rate β and recovery rate μ considering that a fraction of congested links will eventually recover as demand for travel diminishes. Equation 4 expresses the rate at which congested links recover given recovery rate μ . Equation 5 represents how the fraction of free flow links $f(t)$ in the network changes over time given $c(t)$ and $r(t)$. Note that $c(t) + r(t) + f(t) = 1$, where $f(t)$ represents links that have remained in a free flow state from $t = 0$. Also, $k\beta/\mu$ represents the average number of newly congested links each already congested link potentially creates, in a fully freely flowing road network. In epidemic modeling, this is often denoted by R_0 and called the “basic reproductive number”. The higher is R_0 , the faster congestion spreads throughout the network. If $R_0 \leq 1$, congestion will not spread in the network and remains a non-persistent local phenomenon. R_0 in an urban network can be estimated at the onset of congestion, before congestion starts propagating in the network. Congestion propagation in an urban network often occurs over a few hours from the onset of congestion and recovers within a few hours after the peak point. If R_0 is known, as soon as congestion forms and starts to

propagate, the proposed SIR model can be used to predict when congestion will peak and how long it takes to recover, which can be used to optimize the implementation of various traffic management and control strategies.

The formulated model simultaneously describes the dynamics of congestion propagation as well as congestion dissipation or recovery in a network, given estimated parameters β and μ , which are dependent on a definite time-dependent travel demand profile as in real-world networks. Analogous representations to the well-known susceptible-infected (SI) and susceptible-infected-susceptible (SIS) model of epidemics can also be formulated to describe the spread of traffic in a network as a two-state model (see Supplementary Note 1). If a SI model is adopted, the estimated $c(t)$ describes propagation of congestion in the network as long as demand for travel continues and congestion eventually propagates to the entire network with no recovery leading to a full gridlock, also known as “complete jam”^{27,28} or “collapse of the network”²⁹. Here, gridlock is defined as a state of the system under which traffic in the entire network or a portion of the network comes to a complete standstill with zero (or minimal) flow²⁰ (see Fig. 2a). If an SIS model is adopted, the congestion propagation dynamics are described such that congestion continues to grow but does not propagate to the entire network and remains invariant after some time leading to a partial network gridlock (see Fig. 2b). While from a theoretical perspective, both SI and SIS models could also be adopted to describe congestion propagation in a network, they fall short of realism. In a partial network gridlock, the number of congested links can grow or shrink depending on whether the gridlock is propagating or resolving itself (dissipating) as what often happens in real-world networks. Therefore, we expect that the SIR model, as a three-state model, provides a more realistic representation as will be empirically demonstrated in the next section (see Fig. 2c).

Empirical evidence. We apply the proposed contagion-based model to empirical data collected from six different large metropolitan cities around the world (see Fig. 3a–d). We explore changes in the fraction of congested links $c(t)$ in the selected networks. The proposed dynamics in Eq. 3–5 are fit to traffic data between the onset and offset of congestion (Fig. 3b) to estimate the propagation rate β and the recovery rate μ using an ordinary least-square (OLS) method with a pattern search algorithm (see Methods) for different values of ρ as in Eq. 2 and assuming an average $k = 3$ for the studied networks based on the distributions of the node degree.

The applied simple contagion process successfully describes the congestion spreading patterns in different cities at a macroscopic level. While the selected cities have significantly different topology and travel demand patterns, their estimated R_0 values are almost the same for $\rho = 0.2$ and only slightly vary for $\rho = 0.3$ (see Fig. 3a). This suggests existence of a universal measure represented by R_0 for the spread of congestion in urban network similar to what is usually observed for infectious diseases. Note

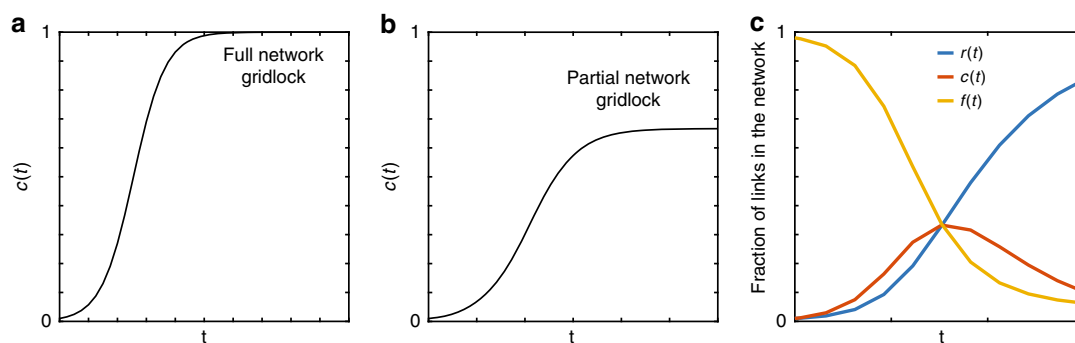


Fig. 2 Describing temporal evolution of congestion. Two-state model of congestion propagation in a network: fraction of congested links $c(t)$ vs. time resulting in **a** full network gridlock described by a variation of the SI model and **b** partial network gridlock described by a variation of the SIS model. **c** Three-state model of congestion propagation and dissipation in a network, analogous to the SIR model subject to a time-varying loading-unloading demand profile.

that $c_{\max}$ and the time where $c_{\max}$ occurs are different between cities. For larger ρ values, the observed difference between the estimated R_0 grows, which is mainly due to the loose definition of congestion when ρ is large. Therefore, we expect to see divergence between cities as ρ increases. The fraction of recovered and free flow links in the studied networks is also illustrated in Fig. 3c, d, and shown to be consistent with the expected outcomes of the theoretical SIR model.

Revealing the underlying dynamics with simulation. We now explore changes in the fraction of congested links $c(t)$ in the network, given the traffic data obtained from the simulation-based dynamic traffic model of Melbourne (Supplementary Note 3) and the identified congested links. See Supplementary Note 4 for a comparison between empirical and simulation-based data from the Melbourne network. Figure 4a shows the evolution of $c(t)$ over time for different values of ρ using simulated data with a calibrated travel demand profile for the morning peak period 6:00–10:00 a.m., followed by a 4-h recovery period with zero demand. Introducing zero demand after a complete network loading is a common approach in network traffic analysis^{2,21}, which allows the network to go under a complete recovery. Many of the interesting properties of network traffic flow can only be observed during a full recovery such as the formation of hysteresis in the network flow-density relationship.

The simulated congestion propagation and dissipation patterns follow the commonly observed spreading patterns in epidemics in which the propagation follows an initial exponential growth regime, followed by a sum of multiple exponential processes during recovery. Smaller values of ρ are used as they better reflect congestion formation compared to larger values of ρ . For example, when $\rho = 0.1$ is being used, the fraction of congested links in the network is almost zero for the first hour of the simulation as congestion has not yet formed in any link across the network. However, when $\rho = 0.9$ is used, nearly 15% of the links in the network are considered congested at the beginning of the simulation. The adopted and formulated SIR model expressed in Eq. 3–5 successfully describes the evolution of $c(t)$ over time as shown in Fig. 4b. The model is applied to traffic data between the congestion onset and offset times at which the traffic jam starts propagating and almost finishes dissipating, respectively.

To further reveal the congestion spreading patterns, we also conduct a simulation with 1 h of peak demand loading followed by several hours of recovery with zero demand, consistent with an analysis previously reported by Olmos et al.². Here we focus on a target group of simulated vehicles that enter the network within the peak hour (8:00–9:00 a.m.), specified as the hour immediately

before the peak point on the demand profile. See Fig. 5a in which congestion propagation in the network can be approximated by an initial growth followed by a decay. Figure 5b, c illustrates the changes in the estimated parameters β , μ , and R_0 for a range of ρ values. It is counter-intuitive that when ρ increases, the rates of congestion propagation and dissipation both decrease exponentially. However, R_0 increases when ρ increases as expected. In fact, propagation and dissipation rates here must be interpreted relatively. Therefore, their individual values may not provide an absolute indication of the extent that congestion is spreading. Instead, it is the ratio of β over μ , or more accurately R_0 , that has a physical meaning. This is analogous to queuing theory in which the size of a queue depends on the difference between the arrival and departure curves rather than the individual arrival and departure rates. Here, R_0 can also be seen as a representation of the rate in which the network shockwave evolves. Further, R_0 follows an interesting linear relationship with ρ , as shown in Fig. 5c.

Local congestion propagation dynamics. In order to further investigate the spreading properties of congestion, we demonstrate the propagation of congestion from congested links to their upstream links in the network using simulated traffic data. Specifically, we reveal the formation of congested clusters at the upstream of congested links. For any directed link i connecting its source node to its target node, let us define the corresponding congested upstream cluster as the subset of nodes on the network that can reach link i 's source node via at least one directed path entirely consisted of congested links; the cluster includes the source node. For a fixed ρ , we individually consider each congested link, and calculate the size of its congested upstream cluster at different points in time. For each time step t , we generate a null model by randomly drawing $\lambda_i(t)$ of each link i from the same distribution of the link relative speeds as in the simulated data at the same time step. To measure the significance of the impact of congested links on their neighborhood, we compare the size of the congested upstream cluster associated with each link on the simulated network with that of the generated null model, over time (see Fig. 6). Results indicate the spatial distribution of link λ 's over the simulated network is significantly different from the null model. In particular, there are significantly more congested upstream clusters of larger size in the simulated traffic network compared to the null model. In other words, in the traffic network, links with lower $\lambda_i(t)$ tend to emerge inside small- to medium-sized clusters, while this phenomenon is not observed in the null model. Furthermore, the size of the congested upstream clusters in the simulated network changes

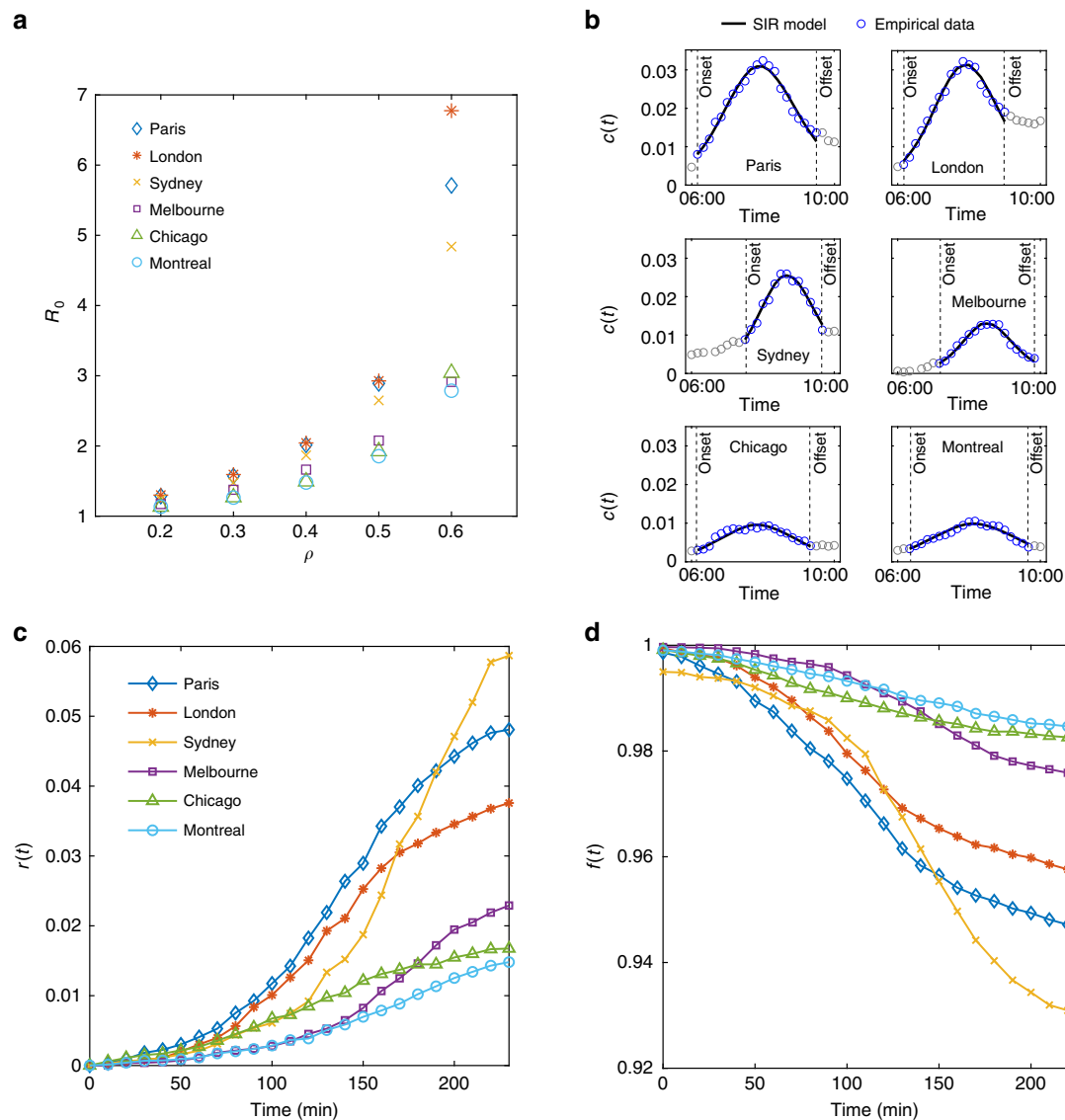


Fig. 3 Empirical evidence on the congestion propagation and dissipation dynamics from six different cities. **a** Relationship between R_0 and ρ . For smaller values of ρ , there is a consistency in the observed dynamics across different cities despite their differences in network structure, demand, and traffic patterns. R_0 remains roughly invariant for the same value of ρ for smaller ρ values, suggesting the existence of a universal measure for congestion propagation; **b** proportion of congested links in the network over time $c(t)$ for each city from 6:00 to 10:00 a.m. when $\rho = 0.2$. Congestion onset and offset is marked with dashed lines. Note that the y-axis has a fixed range $[0, 0.035]$ and a subset of data between the congestion onset and offset times are utilized to fit the SIR model; **c** evolution of fraction of recovered links $r(t)$ in the network over time; **d** evolution of the fraction of free flow links $f(t)$ in the network over time.

over time. It starts from low values around $t = 0$, and gradually increases until approximately $t = 200$ min when it peaks for most of the links, and then decreases toward zero until around $t = 400$ min, when the network is empty. However, for the null model the size of the clusters keeps fluctuating over time with large variations, and the size of the congested upstream clusters only reaches half of what is observed in the simulated Melbourne traffic network. Consistent with the physics of traffic flow and kinematic wave theory, this verifies the hypothesis that congestion at the link level spreads over the network via upstream links. We illustrate the difference in the distribution of the congested upstream cluster size in the simulation network to that of the null model at time $t = 180$ min as an example (see Fig. 6c for $\rho = 0.5$ and Fig. 6f for $\rho = 0.7$). The observed difference between the distributions of the cluster size in the simulated data versus the null case confirms that congestion follows a non-random spatial spreading pattern.

Connection with travel demand. Here, we examine the impact of changing demand on the propagation and dissipation dynamics of congestion. For the same network with 1 h of loading followed by 4-h recovery, we have conducted multiple simulation runs with a scaling factor of $\zeta = 0.75, 1.0, 1.25, 1.5, 1.75$, and 2.0 to decrease or increase demand while keeping the origin–destination demand patterns the same. We then solve the proposed ODE model and estimate the model parameters β and μ as described in Supplementary Note 5 and Methods section. When demand increases, the fraction of congested links also increases for the same network, and as a result, complete recovery of congestion takes longer while the start of the recovery phase remains the same. This is reflected in the reduction of μ in response to the increase in demand for any given ρ (see Fig. 7a for the evolution of $c(t)$ over time for various demand levels). Counter-intuitively, increasing demand also results in reduction of β for a given ρ (see Fig. 7b, c). This should not be interpreted as a slower propagation

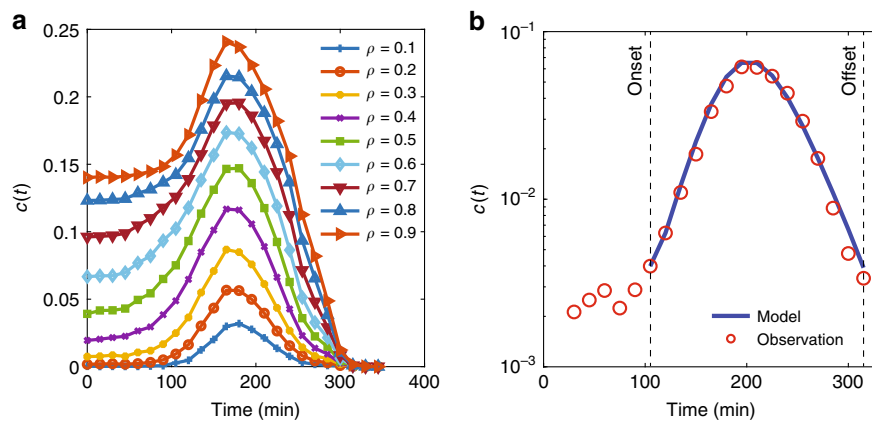


Fig. 4 Propagation and dissipation of congestion in Melbourne network. **a** Evolution of fraction of congested links in the network over time for various ρ subject to realistic and calibrated 4-h demand followed by a 4-h recovery period with zero demand starting at $t = 240$ min. **b** Fitted SIR model to dynamics of the system with estimated parameters $\beta = 0.0577$ and $\mu = 0.0812$ when $\rho = 0.2$ and $k = 2.12$. Data are obtained from the simulation-based DTA model. The vertical axis associated with $c(t)$ is scaled logarithmically.

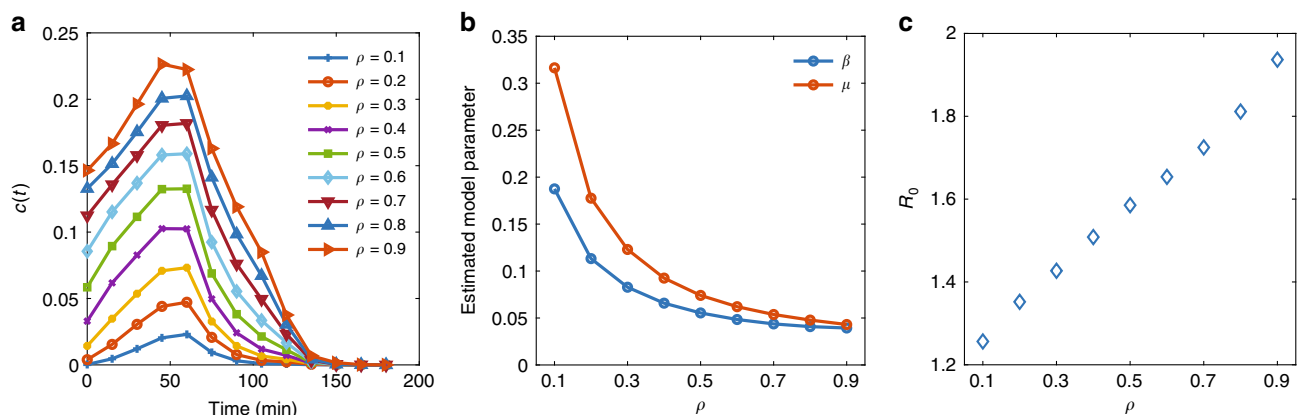


Fig. 5 Spread of traffic congestion in Melbourne network. Propagation and dissipation of congestion in the Melbourne network subject to 1-h peak demand, followed by a recovery period with zero demand starting at $t = 60$ min: **a** evolution of fraction of congested links in the network over time for different values of ρ , **b** estimated model parameters β and μ for various ρ , **c** sensitivity of R_0 to ρ . Here, $R_0 > 1$ suggests that congestion is spreading for every value of $\rho > 0$.

of congestion. In fact, β is not independent of μ . The two macroscopic characteristics vary interdependently, so $c(t)$ reaches its peak value at $t = 75$ min according to the demand profile. However, R_0 increases when demand increases for any given ρ with a clear indication that the size of the network shockwave will grow larger and it takes longer to recover. R_0 has a linear relationship with ρ in which the slope also follows an almost linear relationship with demand (see Fig. 7d, e). Results suggest that there is also a three-dimensional relationship between R_0 , ρ , and demand as illustrated in Fig. 7f, in which for smaller values of ρ , the relationship between R_0 and demand is almost linear. The relationship with demand is critical for applying the proposed model to travel demand management and traffic control in real world. The proposed model can be used for perimeter control of a sub-network within a larger network, in which the relationship with demand guides which R_0 should be used given a fixed ρ value, and vice versa. When R_0 is known, the proposed model can then predict when the congestion will peak and when it fully recovers. Identifying the time of the congestion peak and recovery duration can be used to identify the time-optimal traffic control strategies (see Supplementary Note 6). What is missing here is the observed relationship with demand from empirical data. Travel demand is difficult to observe in real world. Therefore, our analysis here remains limited to the simulation environment. Passive

mobile phone use data could potentially be used to obtain travel demand over time and relate with congestion propagation and dissipation in the network, which remains an interesting direction for future research.

Discussion

This study has shown that propagation and dissipation of traffic jams in cities can be described by a simple contagion process, which is formulated as a system of ODEs, analogous to the well-known SIR model, dependent on two macroscopic network traffic flow characteristics, namely congestion propagation and congestion recovery rate. The paper has shown both empirically and simulation based that congestion propagation is indeed a spreading phenomenon. The extent to which congestion builds up in a network and how fast it recovers are shown to be dependent on the ratio of the propagation over recovery rates represented by the basic reproductive number R_0 . Using data from a simulation-based dynamic traffic assignment model of Melbourne, we showed that time-dependent travel demand profile has, as expected, an impact on the dynamics of congestion propagation and dissipation in a network. In fact, R_0 increases when demand increases for any given ρ with a clear indication that the size of the traffic jam will grow larger and it takes longer to recover in response to travel demand. Moreover, empirical data from six different cities are

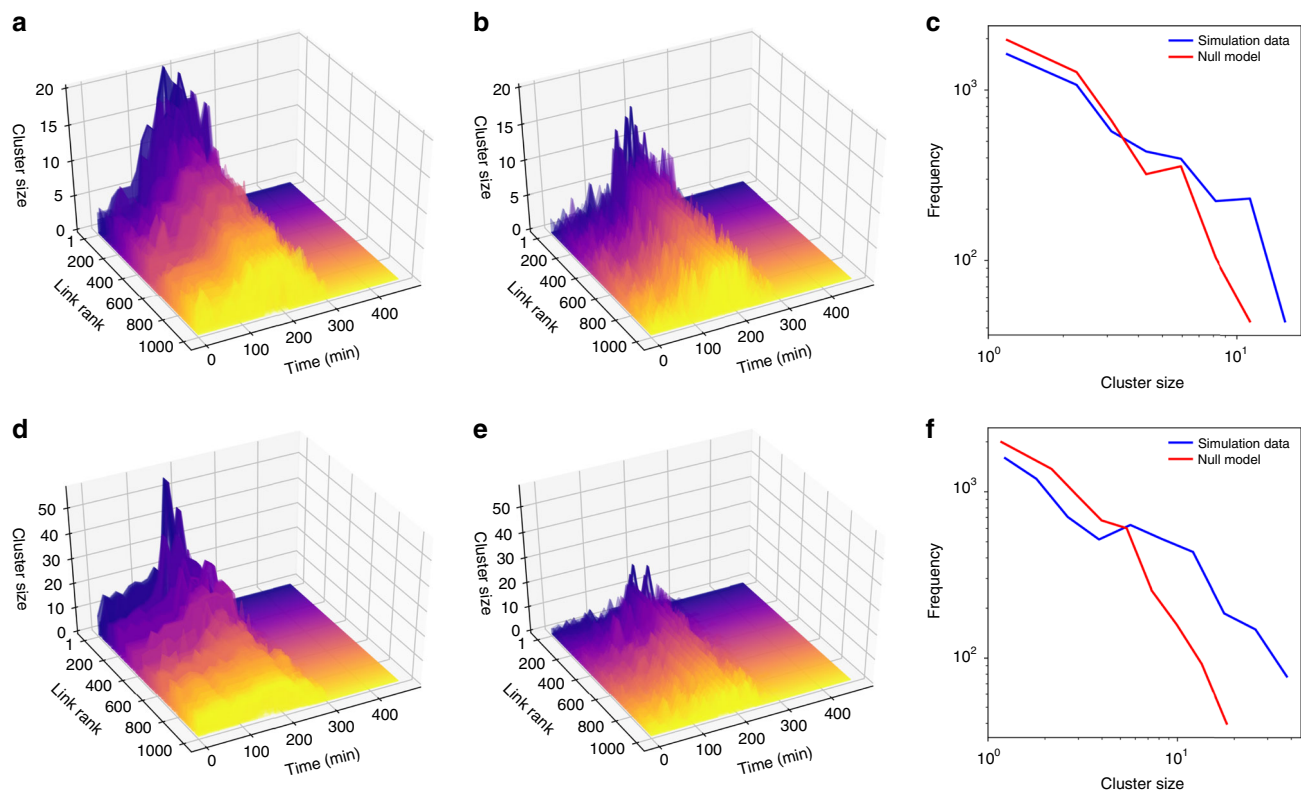


Fig. 6 Evolution of the congested upstream clusters associated with congested links. Here, **a** and **d** show the change over time in the size of the congested clusters associated with links in Melbourne's traffic network, while **b** and **e** depict the same results for the null model generated independently for each time step by drawing random $\lambda_i(t)$'s for each link i from the same distribution as in the simulated data; links are sorted according to the maximum size of their congested upstream cluster and the results only show the top 1000 links. The distribution of the size of the congested upstream clusters for the Melbourne traffic simulation model and its null model counterpart at $t = 180$ is compared for **c** $\rho = 0.5$ and **f** $\rho = 0.7$. Larger congested upstream clusters are observed more frequently in the Melbourne traffic network model compared to the null model generated with the same structure and the same distribution of the link relative velocity.

analyzed to verify the validity of the proposed contagion-based model; see also Supplementary Note 7 on the case of a very small urban road network. We observed that different cities surprisingly exhibit very similar R_0 for small ρ values, indicating that at the macroscopic level, they tend to have consistent congestion propagation dynamics. A limitation of this study is that in the classical SIR model when an infected individual recovers, he/she will not be infected again (or removed). This clearly does not apply directly to traffic networks in which a link may recover and become congested again after a short period. However, the assumption may still hold if the temporal aggregation of analysis is large enough to prevent links to go under multiple cycles of loading and recovery in a short sequence of time steps. Given the unimodality of the travel demand profile during a peak period (morning or afternoon), it may not be unreasonable to assume, at a macroscopic level, that when a link recovers from congestion, it will not become congested again until the next peak period. Nevertheless, relaxation of this assumption and introduction of other traffic states considering multiple cycles of loading and recovery should be an item of research priority in the future.

Methods

Traffic simulation model. Simulation of the Melbourne network is conducted in AIMSUN, a commercially available traffic simulation software. The model is a mesoscopic simulation-based dynamic traffic assignment, which has been calibrated and validated for the morning peak period 6:00–10:00 a.m. A large number of input (demand and supply) parameters need to be calibrated before the simulation outcomes are used. Details of the calibration and validation process can be found in ref. ¹ and Supplementary Note 3.

Google traffic data. Google traffic data on 27 June 2018 are acquired and used, which consist of speeds in the unit of km/h for each link in the network with a unique identifier. The time frame for congestion modeling varies from city to city, which is determined based on the profile of the fraction of congested links, so as to get a fitted model between the onset and offset times of congestion. Nevertheless, the total time window remains the same as in our simulation (i.e., 6:00–10:00 a.m.). Links with missing data for the entire day are discarded in our analysis, but the number of such links is negligible and does not affect the results. A direction for future research is to extend the analysis to multiple days to study the day-to-day spreading dynamics of network traffic.

Google traffic data record the speeds of road sections at continuous time intervals. The data were estimated using floating GPS points of mobile users once they used Google services such as Google Maps. The sampling interval of the data is 10 min, which provides an appropriate temporal resolution for our analysis. Cities that are selected include Melbourne, Sydney, London, Paris, Chicago, and Montreal, which all have a complex urban transport network with a dense population, thereby ensuring the availability of GPS data in large volumes. Note that the reliability of estimations is directly proportional to the amount of data. The calculation of $v_i^{\max}$ takes into account the maximum speed observed on link i in a cycle of one day (i.e., 24 h). Using this measure along with $v_i(t)$, the ratio $\lambda_i(t)$ is calculated for the entire network. To be consistent, only the morning peak of each city is considered in our analysis, which all lies in the time period between 6:00 and 10:00 a.m., although being subject to some differences from city to city.

Parameter estimation with global pattern search. To estimate the parameters of the proposed model, we follow a similar approach presented in ref. ³⁰. The estimation is performed using an OLS method with a pattern search algorithm. The estimation is formulated as a minimization problem in which the pattern search seeks to find the model parameters that minimizes the root mean-squared error (RMSE) calculated as the deviation between the observed and modeled $c(t)$ over the study time period. We have applied the global pattern search algorithm as a derivative-free global optimization method in which the modeled curves for $c(t)$ are fit to the traffic data with the objective function to minimize the RMSE. The pattern search algorithm falls under the general category of global

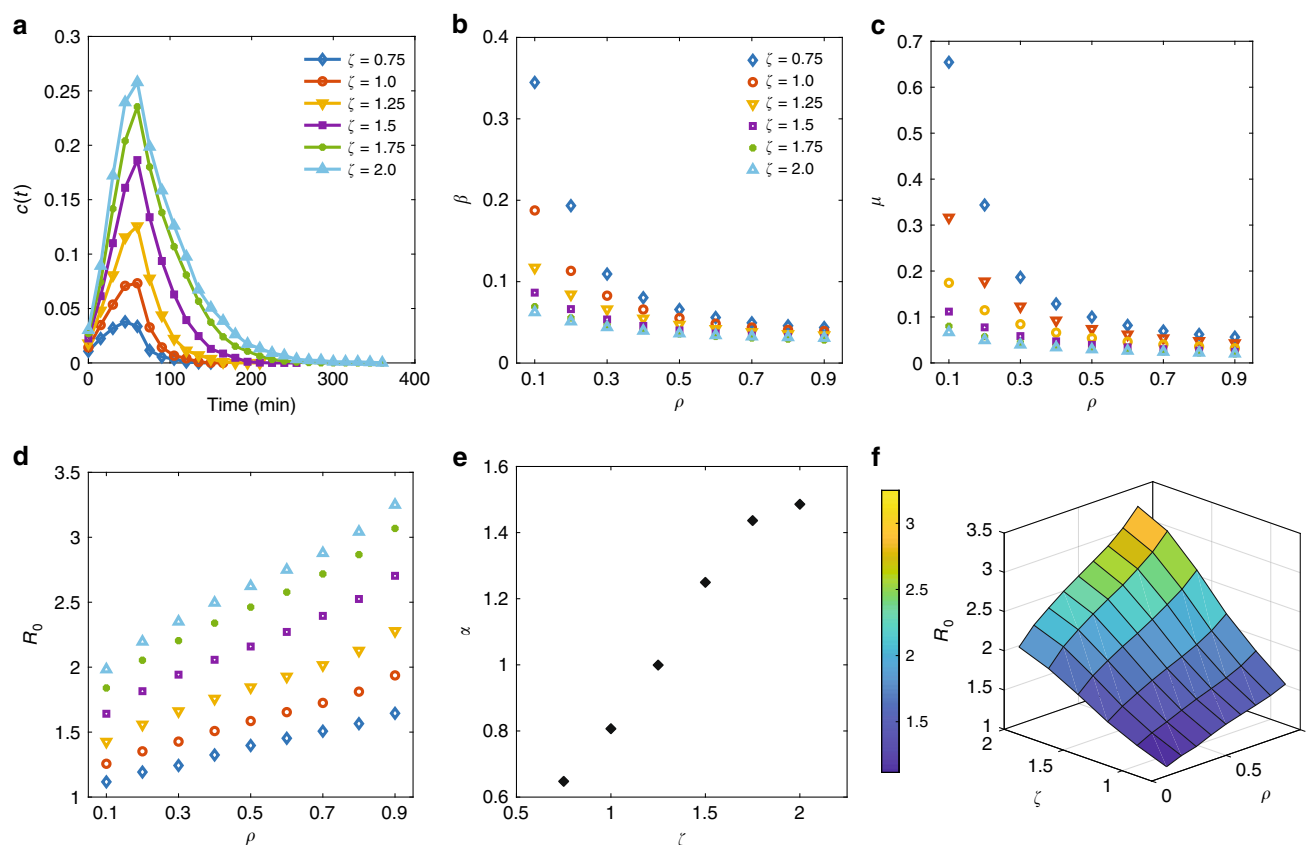


Fig. 7 Effect of travel demand on congestion dynamics. Propagation and dissipation of congestion in the Melbourne network subject to 1-h demand followed by a recovery period: **a** evolution of fraction of congested links in the network over time for various demand levels given $\rho = 0.3$; **b** and **c** estimated model parameters β and μ for various ρ and demand levels; **d** R_0 as a function of ρ for various demand levels; **e** sensitivity of α to ζ . Here, α represents the slope of a linear fit to the data shown in **d**. **f** Three-dimensional illustration of the relationship between R_0 , ρ , and ζ .

optimization methods in which an initial mesh is first specified in the solution space given an initial guessed solution. The algorithm computes the objective function value at each mesh point until it finds a point whose value is smaller than the objective function value at the initial solution point. The algorithm then updates the mesh size and re-computes the objective function value at each mesh point. This will iteratively continue until one of the stopping criteria is met, such as the mesh size getting smaller than a mesh tolerance threshold or if the maximum number of iterations is reached. For details about the pattern search algorithm, see ref. ³¹. Also see Supplementary Table 1 in Supplementary Note 5 for the estimated model parameters and associated RMSE for different values of ρ .

Data availability

For contractual and privacy reasons, we cannot make the empirical data from Google available. However, all data from the simulation-based dynamic traffic assignment model that are needed to replicate the findings reported in the paper and Supplementary Information are available at <https://github.com/meeadsaberi/trafficspreading>.

Code availability

The simulation-based dynamic traffic assignment model of Melbourne is also available as an open access model, which can be downloaded via <https://github.com/meeadsaberi/dynamel>.

Received: 3 June 2019; Accepted: 2 March 2020;
Published online: 07 April 2020

References

- Shafiei, S., Gu, Z. & Saberi, M. Calibration and validation of a simulation-based dynamic traffic assignment model for a large-scale congested network. *Simul. Modell. Pract. Theory* **86**, 169–186 (2018).
- Olmos, L. E., Çolak, S., Shafiei, S., Saberi, M. & Gonzalez, M. C. Macroscopic dynamics and the collapse of urban traffic. *Proc. Natl. Acad. Sci. USA* **115**, 12654–12661 (2018).
- Jiang, S., Yang, Y., Gupta, S., Veneziano, D., Athavale, S. & González, M. C. The TimeGeo modeling framework for urban mobility without travel surveys. *Proc. Natl. Acad. Sci. USA* **113**, E5370–E5378 (2016).
- Geroliminis, N. & Daganzo, C. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. *Transp. Res. B* **42**, 759–770 (2008).
- Gayah, V. & Daganzo, C. Clockwise hysteresis loops in the macroscopic fundamental diagram: an effect on network instability. *Transp. Res. B* **45**, 643–655 (2011).
- Saberi, M. & Mahmassani, H. Exploring properties of network-wide flow-density relations in a freeway network. *Transp. Res. Res.* **2315**, 153–163 (2012).
- Saberi, M. & Mahmassani, H. Hysteresis and capacity drop phenomena in freeway networks: empirical characterization and interpretation. *Transp. Res. Res.* **2391**, 44–55 (2013).
- Ji, Y. & Geroliminis, N. On the spatial partitioning of urban transportation networks. *Transp. Res. Res.* **46**, 1639–1656 (2012).
- Saeedmanesh, M. & Geroliminis, N. Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. *Transp. Res. Procedia* **23**, 962–979 (2017).
- Zeng, G., Li, D., Guo, S., Gao, L., Gao, Z., Stanley, H. E. & Havlin, S. Switch between critical percolation modes in city traffic dynamics. *Proc. Natl. Acad. Sci. USA* **116**, 23–28 (2019).
- Li, D., Fu, B., Wang, Y., Lu, G., Berezin, Y., Stanley, H. E. & Havlin, S. Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proc. Natl. Acad. Sci. USA* **112**, 669–672 (2015).
- Asif, M. T., Dauwels, J., Goh, C. Y., Oran, A., Fathi, E., Xu, M., Dhanya, M. M., Mitrovic, N. & Jaillet, P. Spatiotemporal patterns in large-scale traffic speed prediction. *IEEE Trans. Intell. Transp. Syst.* **15**, 794–804 (2014).
- Vickrey, W. Congestion theory and transport investment. *Am. Econ. Rev.* **59**, 251–260 (1969).
- Arnott, R. A bathtub model of downtown traffic congestion. *J. Urban Econ.* **76**, 110–121 (2013).

15. Wu, J., Gao, Z. & Sun, H. Simulation of traffic congestion with SIR model. *Mod. Phys. Lett. B* **18**, 1537–1542 (2004).
16. Lighthill, M. J. & Whitham, J. B. On kinematic waves. I: Flow movement in long rivers; II: a theory of traffic flow on long crowded roads. *Proc. R. Soc. A* **229**, 281–345 (1955).
17. Richards, P. I. Shockwaves on the highway. *Oper. Res.* **4**, 42–51 (1956).
18. Newell, G. F. A simplified theory of kinematic waves. I: general theory; II: Queuing at freeway bottlenecks; III: Multi-destination flows. *Transp. Res. B* **27**, 281–313 (1993).
19. Daganzo, C. Queue spillovers transportation networks with a route choice. *Transp. Sci.* **32**, 3–11 (1998).
20. Daganzo, C. On the macroscopic stability of freeway traffic. *Transp. Res. B* **45**, 782–788 (2011).
21. Mahmassani, H., Saberi, M. & Zockaie, A. Urban network gridlock: theory, characteristics, and dynamics. *Transp. Res. C* **36**, 480–497 (2013).
22. Çolak, S., Lima, A. & González, M. Understanding congested travel in urban areas. *Nat. Commun.* **7**, 10793 (2016).
23. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
24. Mohajerpour, R., Saberi, M., Vu, H. L., Garoni, T. & Ramezani, M. H_∞ robust perimeter flow control in urban networks with partial information feedback. *Transport. Res. B* <https://doi.org/10.1016/j.trb.2019.03.010> (2020).
25. Han, Y., Ramezani, M., Hegyi, A., Yuan, Y. & Hoogendoorn, S. Hierarchical ramp metering in freeways: an aggregated modeling and control approach. *Transp. Res. C* **110**, 1–19 (2020).
26. Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N. & Papageorgiou, N. Multiple concentric gating traffic control in large-scale urban networks. *IEEE Trans. Intell. Transp. Syst.* **16**, 2141–2154 (2015).
27. Daganzo, C. F. The nature of freeway gridlock and how to prevent it. In *Proc. 13th International Symposium on Transportation and Traffic Theory* 629–646 (Elsevier Science Publishing, 1996).
28. Gayah, V. & Daganzo, C. F. Clockwise hysteresis loops in the macroscopic fundamental diagram: an effect of network instability. *Transp. Res. B* **45**, 643–655 (2011).
29. Daganzo, C. F., Gayah, V. & Gonzales, E. Macroscopic relations of urban traffic variables: bifurcations, multivaluedness and instability. *Transp. Res. B* **41**, 278–288 (2011).
30. Marinov, T. T., Marinova, R. S., Omojola, J. & Jackson, M. Inverse problem for coefficient identification in SIR epidemic models. *Comput. Math. Appl.* **67**, 2218–2227 (2014).
31. Audet, C. & Dennis, J. E. Analysis of generalized pattern searches. *SIAM J. Optim.* **13**, 889–903 (2003).

Acknowledgements

We are thankful for the fruitful discussions with Dr. David Rey at the University of New South Wales and Dr. Mohsen Ramezani from the University of Sydney.

Author contributions

M.S., H.H., L.G., and M.G. designed the study. M.S., H.H., M.A., and S.A.H. implemented the method. M.A., Z.G., and S.S. conducted the simulations. D.J.N. and V.D. supported with empirical data acquisition and analysis. M.S., H.H., M.A., S.A.H., Z.G., L.G., S.T.W., and M.G. analyzed the results and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-15353-2>.

Correspondence and requests for materials should be addressed to M.S.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020