Downloaded from http://rsos.royalsocietypublishing.org/ on October 24, 2016

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Solé-Ribalta A, Gómez S, Arenas A. 2016 A model to identify urban traffic congestion hotspots in complex networks. *R. Soc. open sci.* **3**: 160098. http://dx.doi.org/10.1098/rsos.160098

Received: 8 February 2016
Accepted: 8 September 2016

Subject Category:
Physics

Subject Areas:
Complexity

Keywords:
complex networks, congestion model, phase transitions physics

Author for correspondence:
Albert Solé-Ribalta
e-mail: albert.sole@urv.cat

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3491559.

THE ROYAL SOCIETY PUBLISHING

A model to identify urban traffic congestion hotspots in complex networks

Albert Solé-Ribalta[1,2], Sergio Gómez[1] and Alex Arenas[1,3]

[1]Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain
[2]Internet Interdisciplinary Institute, Universitat Oberta de Catalunya, 08018 Barcelona, Catalonia, Spain
[3]IPHES, Institut Catala de Paleoecologia Humana i Evolucio Social, 43007 Tarragona, Spain

AA, 0000-0003-0937-0334

The rapid growth of population in urban areas is jeopardizing the mobility and air quality worldwide. One of the most notable problems arising is that of traffic congestion. With the advent of technologies able to sense real-time data about cities, and its public distribution for analysis, we are in place to forecast scenarios valuable for improvement and control. Here, we propose an idealized model, based on the critical phenomena arising in complex networks, that allows to analytically predict congestion hotspots in urban environments. Results on real cities' road networks, considering, in some experiments, real traffic data, show that the proposed model is capable of identifying susceptible junctions that might become hotspots if mobility demand increases.

1. Introduction

Urban life is characterized by a huge mobility, mainly motorized. Amidst the complex urban management problems, there is a prevalent one: traffic congestion. Several approaches exist to efficiently design road networks [1] and routing strategies [2]; however, the establishment of collective actions, given the complex behaviour of drivers, to prevent or ameliorate urban traffic congestion is still at its dawn. Usually, congestion is not homogeneously distributed around all city areas but there are salient locations where congestion is settled. We call this locations congestion hotspots. These hotspots usually correspond to junctions and are problematic for the efficiency of the network as well as for the health of pedestrians and drivers. It has been shown [3] that drivers queuing in a traffic jam are the most affected individuals to car

© 2016 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

Downloaded from http://rsos.royalsocietypublishing.org/ on October 24, 2016

exhaust pollution inhalation. In addition, these hotspots are usually located in the city centre, magnifying the problem [4]. Assuming that congestion is an inevitable consequence of urban motorized areas, the challenge is to develop strategies towards a sustainable congestion regime at which delays and pollution are under control. The first step to confront congestion is the modelling and understanding of the congestion phenomena.

The modelling of traffic flows has been a prevalent hot topic since the late 1970s when Gipps' model appeared [5]. Gipps' model and other car-following models [6,7] have evidenced the necessity of modelling traffic flows to improve road network efficiency and also have shown how congestion severely affects the traffic flows. Since 10 years ago the complex networks community has also proposed stylized models to analyse the problem of traffic congestion in networks and design optimal topologies to avoid it [8–21]. The focus of attention of the previous works was the onset of congestion, which corresponds to a critical point in a phase transition, and how it depends on the topology of the network and the routing strategies used. However, the proper analysis of the system after the onset of congestion has remained analytically slippery. It is known that when a transportation network reaches congestion, the system becomes highly nonlinear, large fluctuation exists and the travel time and the number of vehicles queued at a junction diverge [16]. This phenomenon is equivalent to a phase transition in physics, and its modelling is challenging [22–24]. Here, we propose an idealized model to predict the behaviour of transportation networks after the onset of congestion. The presented model is analytically tractable and can be iteratively solved up to convergence. To the best of our knowledge, this is the first analytical model that is able to give predictions beyond the onset of congestion. We present the model in terms of road transportation networks but it could also be applied to analyse other types of transportation networks, such as computer networks, business organizations or social networks.

2. Transportation balance equations

To identify congestion hotspots in urban environments, we propose a model based on the theory of critical (congestion) phenomena on complex networks. The model, that we call the *microscopic congestion model* (MCM), is a mechanistic model (yet simple) and analytically tractable. It is based on assuming that the growth of vehicles observed at each congested node of the networks is constant. This usually happens in real transportation networks at the stationary state. The assumption allows us to describe, with a set of balance equations (one for each node), the increment of vehicles in the junction queues and the number of vehicles arriving or traversing each junction from neighbouring junctions. Mathematically, the increment of the vehicles per unit time at every junction i of the city, Δq_i, satisfies the following balance equation

$$\Delta q_i(t) = g_i(t) + \sigma_i(t) - d_i(t), \tag{2.1}$$

where $g_i(t)$ is the average number of vehicles entering junction i from the area surrounding i at time t, $\sigma_i(t)$ is the average number of vehicles that arrive to junction i from the adjacent links of that junction and $d_i(t) \in [0, \tau_i]$ corresponds to the average number of vehicles that actually finish in junction i or traverse towards other junctions. Note that the value of d_i is upper-bounded by the maximum number of vehicles τ_i that can traverse junction i in a time step. This simulates the physical constraints of the road network. A graphical explanation of the variables of the model is shown in figure 1.

The system of equations (2.1) defined for every node i, is coupled through the incoming flux variables $\sigma_i(t)$, that can be expressed as

$$\sigma_i(t) = \sum_{j=1}^{S} P_{ji}(t) p_j(t) d_j(t), \tag{2.2}$$

where $P_{ji}(t)$ accounts for the routing strategy of the vehicles (probability of going from j to i), $p_j(t)$ stands for the probability of traversing junction j but not finishing at j and S is the number of nodes in the network.

For each junction i, the onset of congestion is determined by $d_i = \tau_i$, meaning that the junction is behaving at its maximum capability of processing vehicles. Thus, for any flux generation (g_i), routing strategy (P_{ij}) and origin–destination probability distribution, equations (2.1) can be solved using an iterative approach to predict the increase of vehicles per unit time at each junction of the network ($\Delta q_i(t)$) (see §3). The only hypothesis we use is that the system dynamics has reached a stationary state in which the growth of the queues is constant. It is worth commenting here that the MCM model considers a fixed average of new vehicles entering the system g_i. However, g_i certainly changes during daytime, with increasing values in rush hours and lower values during off-peak periods. MCM can easily consider evolving values of g_i provided the time scale to reach the stationary state in the MCM (which is usually

Downloaded from http://rsos.royalsocietypublishing.org/ on October 24, 2016

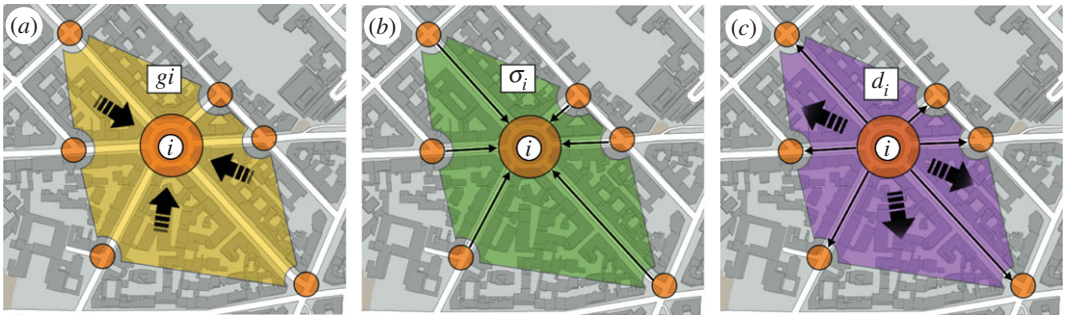

Figure 1. Illustration of the variables of the MCM model. (*a*) Vehicles entering junction *i* from the area surrounding *i*. (*b*) Vehicles entering junction *i* from its neighbouring junctions. (*c*) Vehicles leaving junction *i*, either to go to other neighbouring junctions or to finishing the trip in its surrounding area.

of the order of minutes in real traffic systems) is shorter than the rate of change in the evolution of g_i (which is usually of the order of hours for the daily peaks).

3. Microscopic congestion model

Let node i denote a road junction, edge a_{ij} the road segment between junctions i and j, N_i^{in} and N_i^{out} the sets of ingoing and outgoing neighbouring junctions of junction i, respectively, and S the number of junctions in the road network of the city. Incoming vehicles to node i at each time step can be of two types: those coming from other junctions N_i^{in} and those that start their trip with node i as their first crossed junction. We consider this second type of incoming vehicles as generated in node i. Our MCM describes at the stationary state the increment of the vehicles per unit time at every junction i of the city, Δq_i, as

$$\Delta q_i = g_i + \sigma_i - d_i. \tag{3.1}$$

As described above, we decompose the incoming flux of vehicles σ_i to node i in the stationary state as

$$\sigma_i = \sum_{j \in N_i^{\text{in}}} P_{ji} p_j d_j. \tag{3.2}$$

As vehicles just generated in a certain node are not affected by the congestion in the rest of the network, we separate their contributions in the computation of probabilities p and P. Thus, we decompose p_i as

$$p_i = p_i^{\text{gen}} p_i^{\text{loc}} + (1 - p_i^{\text{gen}}) p_i^{\text{ext}}, \tag{3.3}$$

where the first term accounts for vehicles generated in node i (p_i^{gen}) whose destination is not i (p_i^{loc}) and the second term accounts for vehicles not generated in i whose destination is not i (p_i^{ext}). Supposing trips consist in travelling through two or more junctions, we have that $p_i^{\text{loc}} = 1$. Probability p_i^{gen} is equal to the fraction of vehicles generated in i with respect to the total number of incoming vehicles

$$p_i^{\text{gen}} = \frac{g_i}{g_i + \sigma_i}. \tag{3.4}$$

Considering the distribution of origins, destinations, the routing strategy and the congestion in the network, probability p_i^{ext} can be expressed in terms of the effective node betweenness $\tilde{B}_i$ and the effective vehicle arrivals $\tilde{e}_i$ (the number of vehicles with destination node i that arrive to node i at each time step)

$$p_i^{\text{ext}} = \frac{\tilde{B}_i}{\tilde{B}_i + \tilde{e}_i}. \tag{3.5}$$

The effective betweenness $\tilde{B}_i$ of a node i accounts for the expected number of vehicles each node i receives per unit time considering the routing algorithm and the overall congestion of the network. See §3.2 for an extended description and computation of the effective node betweenness $\tilde{B}_i$ and the effective vehicle arrivals $\tilde{e}_i$.

Downloaded from http://rsos.royalsocietypublishing.org/ on October 24, 2016

In the same spirit, we decompose the probability P_{ji} that a vehicle waiting in node j goes to node i as

$$P_{ji} = p_j^{\mathrm{rgen}} P_{ji}^{\mathrm{loc}} + (1 - p_j^{\mathrm{rgen}}) P_{ji}^{\mathrm{ext}}. \tag{3.6}$$

The first term corresponds to the routed vehicles generated in node j (p_j^{rgen}) that go to node i (P_{ji}^{loc}) and the second term to the routed vehicles not generated in j that go to node i (P_{ji}^{ext}). Similarly as before, p_j^{rgen} can be expressed as the rate between the vehicles generated in j and the total number of routed vehicles

$$p_j^{\mathrm{rgen}} = \frac{g_j}{g_j + \sigma_j p_j^{\mathrm{ext}}}, \tag{3.7}$$

and P_{ji}^{loc} and P_{ji}^{ext} can be computed in terms of the normalized effective edge betweenness of the network

$$P_{ji}^{\mathrm{loc}} = \frac{\tilde{E}_{ji}^{\mathrm{loc}}}{\sum_{k=1}^{S} \tilde{E}_{jk}^{\mathrm{loc}}}, \tag{3.8}$$

and

$$P_{ji}^{\mathrm{ext}} = \frac{\tilde{E}_{ji}^{\mathrm{ext}}}{\sum_{k=1}^{S} \tilde{E}_{jk}^{\mathrm{ext}}}, \tag{3.9}$$

where the computation of $\tilde{E}_{ji}^{\mathrm{loc}}$ only considers paths that start on node j and $\tilde{E}_{ji}^{\mathrm{ext}}$ only considers paths that do not start on node j. Equivalently to the effective node betweenness $\tilde{B}_i$, computation of $\tilde{E}_{ji}^{\mathrm{loc}}$ and $\tilde{E}_{ji}^{\mathrm{ext}}$ consider, if required, all congested junctions in the network, as described in a later section, as well as the distribution of the vehicle sources and destinations. Note that the sum of E_{ji}^{loc} and E_{ji}^{ext} corresponds to the classical edge betweenness. Moreover, P_{ji} is an exact expression before and after the onset of congestion.

Eventually, the MCM is composed of a set of S equations ($\Delta q_i = g_i + \sigma_i - d_i$), one for each junction, and, in principle, a set of $2S$ unknowns, Δq_i and d_i for each junction. To see that the system is indeed determined, we need to note that for congested junctions $\Delta q_i > 0$ and, thus, after the transient state, $d_i = \tau_i$. For the non-congested junctions, we have that $\Delta q_i = 0$ and consequently $d_i = g_i + \sigma_i$. In conclusion, for any node i, either $d_i = \tau_i$ or $d_i = g_i + \sigma_i$ which reduces the number of unknowns to S.

To solve the model given a fixed generation rate g_i, we start by considering that no junction is congested and we solve the set of equations equations (3.1)–(3.9) by iteration. It is possible that some nodes exceed their maximum routing rate. If this is the case, we set the node with maximum d_i as congested and we solve the system again. This process is repeated until no new junction exceeds its maximum routing rate.

3.1. Onset of congestion using the microscopic congestion model

Most of the works that consider static routing strategies assume that the generation rate of vehicles is the same for all nodes, $g_i = \rho$. In that case, it is possible to compute the critical generation rate ρ_c such that for any generation rate $\rho > \rho_c$ the network will not be able to route or absorb all the traffic [25–30]. After this point is reached, the number of vehicles $Q(t)$ in the network will grow proportionally with time, $Q(t) \propto t$, as some of the vehicles get stacked in the queues of the nodes. This transition to the congested state is characterized using the following order parameter:

$$\eta(\rho) = \lim_{t \to \infty} \frac{\langle \Delta Q \rangle}{\rho S}, \tag{3.10}$$

where $\langle \Delta Q \rangle$ represents the average increment of vehicles per unit of time in the stationary state. Basically, the order parameter measures the ratio between in-transit and generated vehicles.

In the non-congested phase, the number of incoming and outgoing vehicles for each node can be computed in terms of the node's algorithmic betweenness B_i [25]. In particular,

$$\sigma_i = \rho \left(\frac{B_i}{S-1} + 1 \right), \tag{3.11}$$

where the second term inside the parentheses accounts for the fact that, in our model, vehicles are also queued at the destination node, unlike in [25]. When no junction is congested, we have that $\Delta q_i = 0$ for

Downloaded from http://rsos.royalsocietypublishing.org/ on October 24, 2016

all nodes and consequently

$$d_i = \rho + \sigma_i = \rho\left(\frac{B_i}{S-1} + 2\right). \tag{3.12}$$

A node i becomes congested when it is required to process more vehicles than its maximum processing rate, $d_i > \tau$. Thus, the critical generation rate at which the first node, and so the system, reaches congestion is

$$\rho_c = \min_i \frac{\tau(S-1)}{B_i + 2(S-1)}. \tag{3.13}$$

This is one of the most important analytical results on transportation networks with static routing strategies. In the following, we show that we can recover equation (3.11) before the onset of congestion using our MCM approach. After substitution of the expression of the probabilities in equation (3.2)

$$\sigma_i = \sum_j \frac{\rho(B_j + S - 1)P_{ji}^{\mathrm{loc}} + \sigma_j B_j P_{ji}^{\mathrm{ext}}}{(\rho + \sigma_j)(B_j + S - 1)} d_j, \tag{3.14}$$

and, given we do not have congestion (i.e. $d_j = \rho + \sigma_j$), it simplifies to

$$\sigma_i = \sum_j \frac{\rho(B_j + S - 1)P_{ji}^{\mathrm{loc}} + \sigma_j B_j P_{ji}^{\mathrm{ext}}}{B_j + S - 1}. \tag{3.15}$$

Equation (3.15) in matrix form becomes

$$(I - M)\boldsymbol{\sigma} = \rho\boldsymbol{\pi}, \tag{3.16}$$

where

$$M_{ij} = \frac{B_j P_{ji}^{\mathrm{ext}}}{B_j + S - 1} \tag{3.17}$$

and

$$\pi_i = \sum_j P_{ji}^{\mathrm{loc}}, \tag{3.18}$$

and then

$$\boldsymbol{\sigma} = \rho(I - M)^{-1}\boldsymbol{\pi}. \tag{3.19}$$

This expression can be shown to be equivalent to equation (3.11) by using the following relationship between node and edge betweenness

$$B_i + (S-1) = \sum_j (B_j P_{ji}^{\mathrm{ext}} + (S-1)P_{ji}^{\mathrm{loc}}). \tag{3.20}$$

The right-hand side corresponds to the accumulated fractions of paths that pass through the neighbours of node i and then go to i. Each neighbour contributes with two terms, the paths that go through j coming from other nodes, and the paths that start in j.

3.2. Effective betweenness in congested transportation networks

The effective betweenness $\tilde{B}_i$ of a node i, as defined in [25], accounts for the expected number of vehicles each node i receives per unit time. When the network is not congested and the vehicle generation rate g_i is equal for all nodes, $g_i = \rho$, the number of vehicles each node receives can be obtained using equation (3.11). However, if the network is congested, the traffic dynamics becomes highly nonlinear and the value of σ_i computed in equation (3.11) becomes a poor approximation.

Suppose we focus on a particular congested node j^* of the network. For j^*, being congested means that it is receiving more vehicles that the ones it can process and route. In particular, from the $\sigma_{j^*} + g_{j^*}$ vehicles that arrive to the node, only τ_{j^*} can be processed at each time step.

Therefore, the contribution to the effective betweenness $\tilde{B}_i$ of the paths from a source/destination pair, (s, t), that traverse the congested node j^* before reaching i, must be rescaled by the fraction of processable vehicles

$$s_{j^*} = \frac{\tau_{j^*}}{\sigma_{j^*} + g_{j^*}}. \tag{3.21}$$

When a path traverses multiple congested nodes $j^*, k^*, \ldots$, the remaining fraction of paths that will reach the target node will be the result of the application of the multiple rescalings s_{x^*}.

Downloaded from http://rsos.royalsocietypublishing.org/ on October 24, 2016

The computation of s_{j^*} is not straightforward. In general, σ_i is not known after the onset of congestion and depends on the effective betweenness that requires, at the same time, to know the s_{j^*} fraction for all congested nodes. Thus, an iterative calculation is needed to fit all the parameters at the same time as we do in our MCM.

The effective arrivals $\tilde{e}_i$ account for the number of vehicles with destination node i that arrive at node i at each time step. This value in the non-congested phase can be obtained, considering homogeneous source and destination nodes, as

$$e_i = \rho(S - 1). \tag{3.22}$$

However, congestion also affects the variable e_i and needs to be corrected accordingly using the same procedure presented above.

4. Results

To simulate the traffic dynamics of the road network, we assign a first-in-first-out queue to each junction that simulates the blocking time of vehicles before they are allowed to cross it and continue their trip. We suppose these queues have infinite capacity and a maximum processing rate that simulates the physical constraints of the junction. Vehicle origins and destinations may follow any desired distribution. In this work, we have considered two distributions: a random uniform distribution for the synthetic experiments, and one obtained considering the ingoing and outgoing flux of vehicles of the city of Milan. At each time step (of 1 min duration) vehicles are generated and arrive to their first junction. During the following time steps, vehicles navigate towards their destination following any routing strategy. Here, we have used two different routing strategies: shortest-path and random local search.

4.1. Congestion on synthetic networks

To evaluate the MCM, we have conducted experiments on several synthetic networks and with two different routing strategies: local search strategy and shortest path strategy. In both routing strategies we assume, for simplicity, that all vehicles randomly choose the starting and ending junctions of their journey uniformly within all junctions of the network. Thus, each junction generates new vehicles with the same rate $g_i = \rho$. For shortest path strategy, vehicles follow a randomly selected shortest path towards the destination. Without loss of generality, we fix $\tau = 1$ and analyse the performance of MCM for different values of ρ.

Figure 2 shows the accuracy on predicting the values of the order parameter $\eta = \sum \Delta q_i / \rho S$ and d_i for shortest paths routing strategy. As in [25,30], this order parameter η corresponds to the ratio between in-transit and generated vehicles. In the electronic supplementary material, we extend the evaluation considering local search routing strategies and also evaluate the accuracy on predicting other variables of the model. All experiments show that the MCM achieves high accuracy in predicting the macroscopic and microscopic variables of the stylized transportation dynamics.

4.2. Application to real scenarios

INRIX Traffic Scorecard (http://www.inrix.com/) reports the rankings of the most congested countries worldwide in 2014. USA, Canada and most of the European countries are in the top 15, with averages that range from 14 to 50 h per year wasted in congestion, with their corresponding economic and environmental negative consequences. To demonstrate that the MCM model can be applied to real scenarios to obtain real predictions, in the following we apply the MCM model to the nine most congested cities according to the INRIX Traffic Scorecard (table 1).

We first focus on the city of Milan, the city with largest INRIX value. To evaluate the outcome of the MCM model, we first gather data about the road network topology using Open Street Map (OSM). OSM data represent each road (or way) with an ordered list of nodes which can either be road junctions or simply changes of the direction of the road. We have obtained the required abstraction of the road network building a simplified version of the OSM data which only account for road junctions (nodes). Then, for each pair of adjacent junctions we have queried the real travel distance (i.e. following the road path) using the API provided by Google Maps. The resulting network corresponds to a spatial weighed directed network [31] where the driving directions are represented and the weight of each link indicates the expected travelling time between two adjacent junctions (see electronic supplementary material, figure S21).

Downloaded from http://rsos.royalsocietypublishing.org/ on October 24, 2016

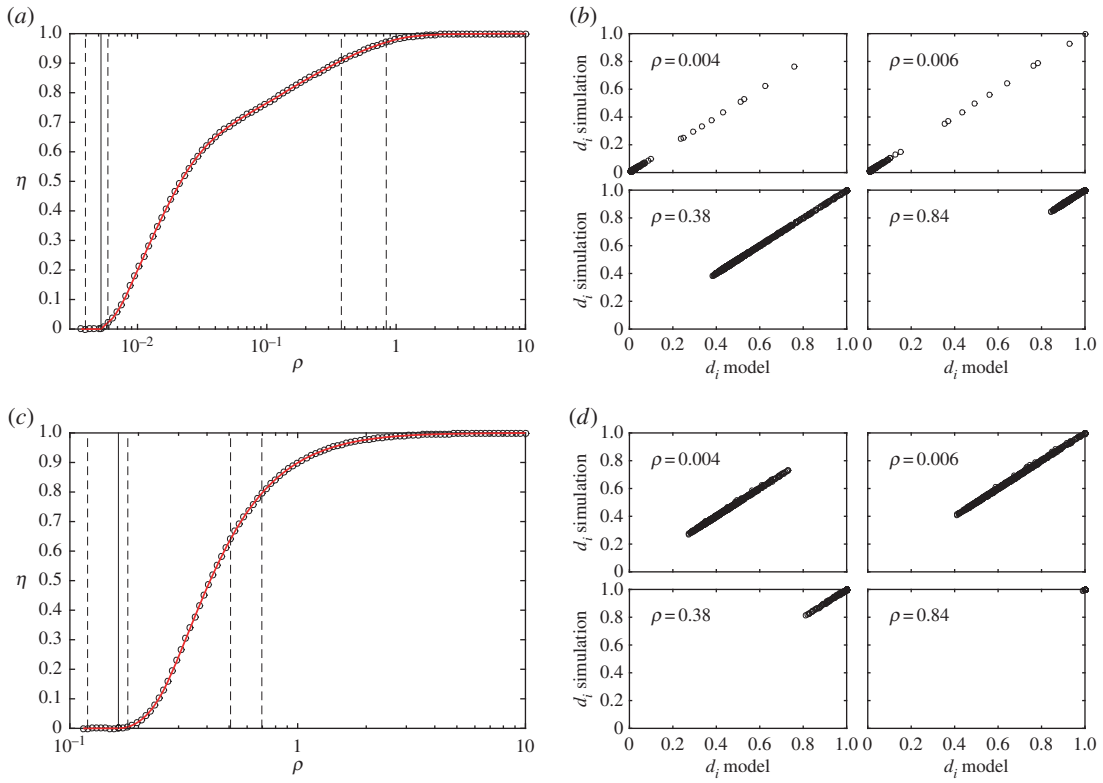

Figure 2. Validation of the microscopic congestion model with Barabási–Albert (*a*,*b*) and Erdős–Rényi (*c*,*d*) networks of 1000 nodes and shortest path routing strategy. In the construction procedure of the Barabási–Albert networks each new node is connected to one existing node in the network. The Erdős–Rényi networks have an average degree of 50. In (*a*,*c*), accuracy in predicting the order parameter η. In (*b*,*d*), correlation between predicted and simulated values of d_i. Vertical solid lines on panels (*a*,*c*) show the predicted critical generation rate ρ_c. Vertical dashed lines show the ρ values where *d* is evaluated on the panels (*b*,*d*).

Table 1. Comparison between the INRIX (12 months) traffic index and the number of hotspots estimated by the proposed model for the most congested cities of the world.

city	INRIX[a]	hotspots	nodes	links
Milano	36.2	108	6924	14 315
London	32.4	93	6378	14 662
Los Angeles	32.2	57	6799	19 368
Brussels	30.5	50	6645	15 624
Antwerpen	28.6	44	6530	15 252
San Francisco	27.9	45	8854	25 530
Stuttgart	21.9	34	8330	19 946
Nottingham	21.6	28	7337	16 723
Karlsruhe	21.3	19	4257	10 379

[a]The INRIX index is the percentage increase in the average travel time of a commute above free-flow conditions during peak hours, e.g. an INRIX index of 30 indicates a 40-min free-flow trip will take 52 min. Each city has been mapped to a graph with the indicated numbers of nodes and links. See text for details and electronic supplementary material, figures S1–S18 for the graph representation of the cities and the geographical representation of the congestion hotspots.

Second, we build up the dynamics of the model analysing real traffic data provided by Telecom Italia for their Big Data Challenge. The data provide, for every car entering the cordon pricing zone in Milan during November and December 2013, an encoding of the car's plate number, time and gate of entrance (a total of 9 183 475 records). This allows us to obtain the (hourly) average incoming and outgoing traffic flow, for each gate of the cordon taxed area.

Downloaded from http://rsos.royalsocietypublishing.org/ on October 24, 2016

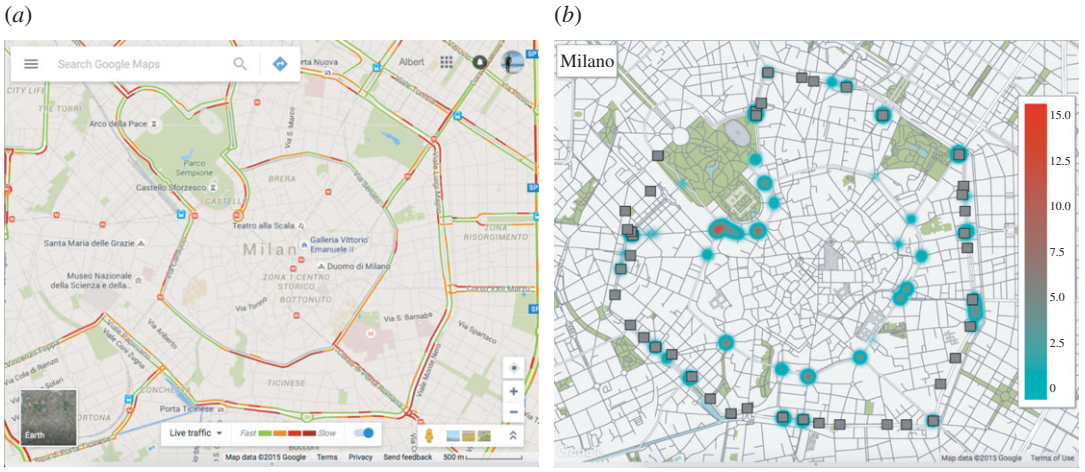

Figure 3. Congestion hotspot analysis of the city of Milan. Panel (*a*) shows the typical situation around 9.00 for a weekday. The image and the data have been obtained with Google Maps. Google Maps displays traffic information considering historical data and real-time car velocity reported by smartphones [32]. Panel (*b*) shows the prediction of the MCM model considering the real road topology obtained using Open Street Map and real traffic data provided by Telecom Italia for their Big Data Challenge. The extracted road network topology is provided in electronic supplementary material, figure S19. For all congestion hotspots the model has predicted, we show its mean increment of the queue size, $\langle \Delta q_i \rangle$.

Given the previous topology and traffic information, we generated traffic compatible with the observations, and evaluated the outcome of the MCM model. Specifically, the simulated dynamics is as follows: for each vehicle entering the Area-C we fix a randomly selected location as destination and use the shortest path route towards it. After the vehicle has arrived to its destination, it randomly chooses an exit door and travels to it also using the shortest path route. This is similar to the well-known *Home-to-Work* travel pattern where vehicles arrive from the outskirts of the city, go to the city centre and then return to the outskirts. Specifically, in our simulation, traffic is generated in the peripheral junctions of the network, goes to a randomly selected junction within the city and then returns back to a randomly selected peripheral junction. We do not consider trips with origin and destination inside the city centre because public transportation systems (e.g. train or subway) usually constitute a better alternative than private vehicles for those trips. The maximum crossing rate of each junction τ_i accounts, among others, for the existence of traffic lights governing the junction, the width of the street as well as its traffic. We have not been able to get this information for the studied cities, and consequently we cannot set to each junction its precise value. Instead, without loss of generality and for the sake of simplicity, we set to all junctions the same maximum crossing rate, $\tau_i = 15$ (an estimation of the average of their real values).

Figures 3 and 4 show the obtained results. Figure 3*b* displays the predicted congestion hotspots on a map of Milan, panel (*a*) of the same figure shows a real traffic situation obtained with Google Maps. We see that the predicted congestion hotspots are located in the circular roads of Milan as well as on the arterial roads of the city; this agrees with the real traffic situation shown in panel (*a*). Figure 4*a* shows the distribution of the mean increments each junction has to deal with. This might be a good indicator to decide about future planning actuations to improve city mobility. However, differently from what is described in [11], the improvement of the throughput of a single junction might not be enough to improve city mobility because this might end up with the collapse of neighbouring junctions (their incoming rate σ_i will increase). This situation is similar to Braess' paradox [33]. Figure 4*b* shows the mean increment of vehicles (in vehicles per minute) for each hour of the weekday. The figure clearly shows the morning and evening rush hours as well as the lunch time.

For the other top nine congested cities, we do not have previous traffic information, neither about the real flux of vehicles nor about the vehicle source and destination distributions (to obtain a fair comparison between all the analysed cities, we have not considered the Telecom traffic data for Milan here). Thus, for each city, we consider homogeneously distributed source and destination locations and the required road traffic to obtain an order parameter η compatible with the congestion effects recorded by INRIX sensing of real traffic. By relating the INRIX value and η, we are assuming that there exists a relation between the fraction of global congestion and the fraction of extra time wasted reported by INRIX. The obtained results are summarized in table 1, which shows that the number of hotspots is

Downloaded from http://rsos.royalsocietypublishing.org/ on October 24, 2016

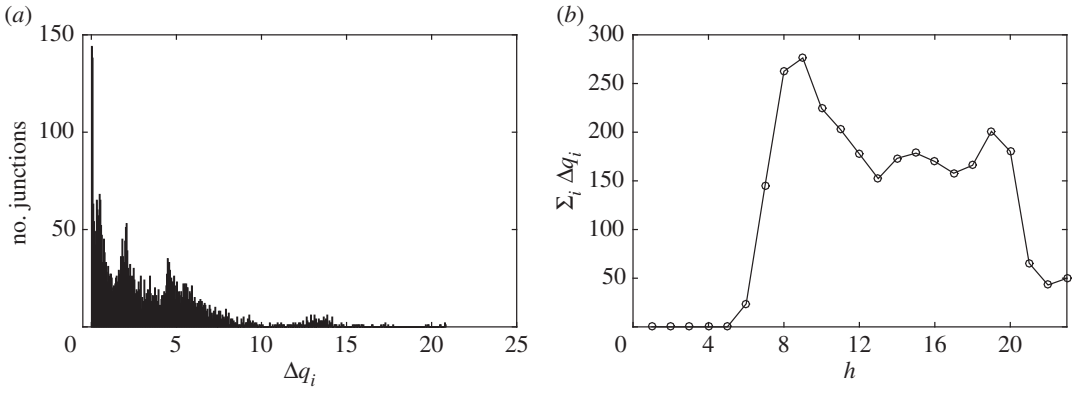

Figure 4. Statistics of Milan congestion hotspots. Panel (*a*) shows the distribution of the vehicle increments (Δq_i) of each congestion hotspot predicted by the MCM. The plot aggregates the MCM predicted congestion considering the average traffic for every weekday and every hour of the day. Panel (*b*) shows the average city congestion in terms of mean increment of vehicle in transit for every hour of the day during a weekday.

correlated with the INRIX value. This shows evidence that the percentage increase in the average travel time to commute between city locations is related to the number of congestion hotspots and the excess of vehicles within the city. The extracted road network topology for all the analysed cities as well as the predicted congestion hotspots is provided in electronic supplementary material, figures S1–S18.

5. Discussion

The previous results show that the MCM can be used to predict the local congestion before and beyond the onset of congestion of a transportation network. To the knowledge of the authors, this is the first analytical model that is able to give predictions beyond the onset of congestion where the system is highly nonlinear, large fluctuation exists and the number of vehicles on transit diverge with respect to time. Our model is based on assuming that the growth of vehicles observed in each congested node of the networks is constant, which allowed us to derive a set of balance equations that can accurately predict macroscopic, mesoscopic and microscopic variables of the transportation network.

Traffic congestion is a common and open problem whose negative impacts range from wasted time and unpredictable travel delays to a waste of energy and an uncontrolled increase of air pollution. A first step towards the understanding and fight of congestion and its related consequences is the analytical modelling of the congestion phenomena. Here, we have shown that the MCM model is detailed enough to give real predictions considering real traffic data and topology. These results pave the way to a new generation of stylized physical models of traffic on networks in the congestion regime that could be very valuable to assess and test new traffic policies on urban areas in a computer-simulated scenario.

Data accessibility. Data available at Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.32mq0 [34].

Authors' contributions. A.S.-R., S.G and A.A. contributed equally to the research and writing of the manuscript; A.S.-R. performed the experiments.

Competing interests. We have no competing interests.

Funding. This work has been supported by Ministerio de Economía y Competitividad (Grant FIS2015-71582-C2-1) and European Comission FET-Proactive Projects MULTIPLEX (grant no. 317532). A.A. also acknowledges partial financial support from the ICREA Academia and the James S. McDonnell Foundation.

References

1. Yang H, Bell MGH. 1998 Models and algorithms for road network design: a review and some new developments. *Transport Rev.* **18**, 257–278. (doi:10.1080/01441649808717016)
2. Bast H, Funke S, Sanders P, Schultes D. 2007 Fast routing in road networks with transit nodes. *Science* **316**, 566–566. (doi:10.1126/science.1137521)
3. Petersson G. 1987 Exposure to traffic exhaust. In *Traffic and children's health*, NHV-Report 1987:2 (eds L Kohler, H Jackson), pp. 117–126. Nordic School of Public Health, European Society for Social Pediatrics. (http://publications.lib.chalmers.se/records/fulltext/local_106658.pdf)
4. Raducan G, Stefan S. 2009 Characterization of traffic-generated pollutants in Bucharest. *Atmósfera* **22**, 99–110.
5. Wilson RE. 2001 An analysis of Gipps's car-following model of highway traffic. *IMA J. Appl. Math.* **66**, 509–537. (doi:10.1093/imamat/66.5.509)

Downloaded from http://rsos.royalsocietypublishing.org/ on October 24, 2016

6. Treiber M, Hennecke A, Helbing D. 2000 Congested traffic states in empirical observations and microscopic simulations. *Phys. Rev. E* **62**, 1805. (doi:10.1103/PhysRevE.62.1805)
7. Newell GF. 2002 A simplified car-following theory: a lower order model. *Transportation Res. Part B: Methodol.* **36**, 195–205. (doi:10.1016/S0191-2615(00)00044-8)
8. Guimerà R, Díaz-Guilera A, Vega-Redondo F, Cabrales A, Arenas A. 2002 Optimal network topologies for local search with congestion. *Phys. Rev. Lett.* **89**, 248701. (doi:10.1103/PhysRevLett.89.248701)
9. Tadić B, Thurner S, Rodgers G. 2004 Traffic on complex networks: towards understanding global statistical properties from microscopic density fluctuations. *Phys. Rev. E* **69**, 036102. (doi:10.1103/PhysRevE.69.036102)
10. Donetti L, Hurtado PI, Muñoz MA. 2005 Entangled networks, synchronization, and optimal network topology. *Phys. Rev. Lett.* **95**, 188701. (doi:10.1103/PhysRevLett.95.188701)
11. Zhao L, Lai Y-C, Park K, Ye N. 2005 Onset of traffic congestion in complex networks. *Phys. Rev. E* **71**, 026125. (doi:10.1103/PhysRevE.71.026125)
12. Ashton DJ, Jarrett TC, Johnson NF. 2005 Effect of congestion costs on shortest paths through complex networks. *Phys. Rev. Lett.* **94**, 058701. (doi:10.1103/PhysRevLett.94.058701)
13. Singh BK, Gupte N. 2005 Congestion and decongestion in a communication network. *Phys. Rev. E* **71**, 055103. (doi:10.1103/PhysRevE.71.055103)
14. Danila B, Yu Y, Marsh JA, Bassler KE. 2006 Optimal transport on complex networks. *Phys. Rev. E* **74**, 046106. (doi:10.1103/PhysRevE.74.046106)
15. Barthélemy M, Flammini A. 2006 Optimal traffic networks. *J. Stat. Mech. Theory Exp.* **2006**, L07002. (doi:10.1088/1742-5468/2006/07/L07002)
16. Dorogovtsev SN, Goltsev AV, Mendes JF. 2008 Critical phenomena in complex networks. *Rev. Modern Phys.* **80**, 1275. (doi:10.1103/RevModPhys.80.1275)
17. Kim K, Kahng B, Kim D. 2009 Jamming transition in traffic flow under the priority queuing protocol. *Europhys. Lett.* **86**, 58002. (doi:10.1209/0295-5075/86/58002)
18. Li G, Reis SD, Moreira AA, Havlin S, Stanley HE, Andrade Jr JS. 2010 Towards design principles for optimal transport networks. *Phys. Rev. Lett.* **104**, 018701. (doi:10.1103/PhysRevLett.104.018701)
19. Ramasco JJ, Marta S, López E, Boettcher S. 2010 Optimization of transport protocols with path-length constraints in complex networks. *Phys. Rev. E* **82**, 036119. (doi:10.1103/PhysRevE.82.036119)
20. Scellato S, Fortuna L, Frasca M, Gómez-Gardeñes J, Latora V. 2010 Traffic optimization in transport networks based on local routing. *Eur. Phys. J. B* **73**, 303–308. (doi:10.1140/epjb/e2009-00438-2)
21. Gleeson J, Melnik S, Ward JA, Porter M, Mucha P. 2012 Accuracy of mean-field theory for dynamics on real-world networks. *Phys. Rev. E* **85**, 026106. (doi:10.1103/PhysRevE.85.026106)
22. De Martino D, Dall'Asta L, Bianconi G, Marsili M. 2009 A minimal model for congestion phenomena on complex networks. *J. Stat. Mech. Theory Exp.* **2009**, P08023. (doi:10.1088/1742-5468/2009/08/P08023)
23. Echenique P, Gómez-Gardeñes J, Moreno Y. 2005 Dynamics of jamming transitions in complex networks. *Europhys. Lett.* **71**, 325. (doi:10.1209/epl/i2005-10080-8)
24. Lü L, Pan L, Zhou T, Zhang Y-C, Stanley E. 2015 Toward link predictability of complex networks. *Proc. Natl Acad. Sci. USA* **112**, 2325–2330. (doi:10.1073/pnas.1424644112)
25. Guimerà R, Díaz-Guilera A, Vega-Redondo F, Cabrales A, Arenas A. 2002 Optimal network topologies for local search with congestion. *Phys. Rev. Lett.* **89**, 248701. (doi:10.1103/PhysRevLett.89.248701)
26. Arenas A, Díaz-Guilera A, Guimera R. 2001 Communication in networks with hierarchical branching. *Phys. Rev. Lett.* **86**, 3196. (doi:10.1103/PhysRevLett.86.3196)
27. De Martino D, Dall'Asta L, Bianconi G, Marsili M. 2009 Ccongestion phenomena on complex networks. *Phys. Rev. E* **79**, 015101. (doi:10.1103/PhysRevE.79.015101)
28. Chen S, Huang W, Cattani C, Altieri G. 2011 Traffic dynamics on complex networks: a survey. *Math. Problems Eng.* **2012**, 732698. (doi:10.1155/2012/732698)
29. Arenas A, Cabrales A, Díaz-Guilera A, Guimerà R, Vega-Redondo F. 2003 *Statistical mechanics of complex networks*, pp. 175–194. Berlin, Germany: Springer.
30. Cabrales A, Danon L, Díaz-Guilera A, Guimerà R, Vega-Redondo F. 2010 Optimal information transmission in organizations: search and congestion. *Rev. Econ. Des.* **14**, 75–93. (doi:10.1007/s10058-008-0067-2)
31. Barthélemy M. 2011 Spatial networks. *Phys. Rep.* **499**, 1–101. (doi:10.1016/j.physrep.2010.11.002)
32. Barth D. 2009 Product manager for Google Maps. See https://googleblog.blogspot.in/2009/08/bright-side-of-sitting-in-traffic.html (accessed 17 December 2015).
33. Youn H, Gastner MT, Jeong H. 2008 Price of anarchy in transportation networks: efficiency and optimality control. *Phys. Rev. Lett.* **101**, 128701. (doi:10.1103/PhysRevLett.101.128701)
34. Sole-Ribalta A, Gomez S, Arenas A. 2016 Data from: A model to identify urban traffic congestion hotspots in complex networks. Dryad Digital Repository. (doi:10.5061/dryad.32mq0).