Article

Spatiotemporal dynamics of traffic bottlenecks yields an early signal of heavy congestions

Jinxiao Duan [1,2], Guanwen Zeng[2,3], Nimrod Serok[4], Daqing Li[3], Efrat Blumenfeld Lieberthal [4] ✉, Hai-Jun Huang [1] ✉ & Shlomo Havlin [2] ✉

Heavy traffic jams are difficult to predict due to the complexity of traffic dynamics. Understanding the network dynamics of traffic bottlenecks can help avoid critical large traffic jams and improve overall traffic conditions. Here, we develop a method to forecast heavy congestions based on their early propagation stage. Our framework follows the network propagation and dissipation of the traffic jams originated from a bottleneck emergence, growth, and its recovery and disappearance. Based on large-scale urban traffic-speed data, we find that dissipation duration of jams follows approximately power-law distributions, and typically, traffic jams dissolve nearly twice slower than their growth. Importantly, we find that the growth speed, even at the first 15 minutes of a jam, is highly correlated with the maximal size of the jam. Our methodology can be applied in urban traffic control systems to forecast heavy traffic bottlenecks and prevent them before they propagate to large network congestions.

Urban traffic congestion[1,2] is an everyday troubling phenomenon where many large traffic jams are triggered by uncertain congestion sources, also known as traffic bottlenecks[3]. The challenges of understanding complex congestion propagation have stimulated extensive traffic flow approaches to modeling and understanding urban traffic dynamics. Some of them investigated the spontaneous occurrences of congestion, including the kinematic wave theory[4,5], the cellular automaton models[6,7], and the three-phase traffic theory[8–10]. Attention is also paid to understanding the jam formation for the known causes, including the queue model[11], the lane-changing model[12–14], and the cell transmission model[15,16] (see Supplementary Note 1). In recent years, emerging theories from other natural systems have been borrowed and applied to traffic systems for network's propagation of traffic congestions, such as the cascading failure models[17–20], epidemic models[21–23], and congestion tree method[24,25]. Studies of traffic condition prediction[26–34] and travel demand control[35–43] also considered congestion propagation and suggested that understanding such spatiotemporal dynamics, especially their bottlenecks, could be effective for forecasting them and thus have the potential to prevent congestions from spreading[44].

A recent study of ref. 21 formulated and validated the macroscopic propagation and dissipation of overall traffic congestions by a susceptible-infected-recovered model. Since macroscopic traffic congestions in the overall network are comprised of numerous fragmented congestion components[45] that could be created by specific bottlenecks, overall traffic conditions could be improved by identifying the spatiotemporal dynamics of bottlenecks. Identifying traffic bottlenecks will not only help avoid themselves but also alleviate other associated congested roads due to network propagation. However, existing bottleneck methods[46–50] characterizing the queues and delays associated with a specific traffic bottleneck are mostly based on the single-car particles in a simplified corridor[51]. Such methods may not be suitable for describing network dynamics of urban traffic bottlenecks. This is because urban congestions often exhibit strong spatial dependencies in which multiple-dimensional road segments adjacent to a bottleneck have a greater risk of becoming congested[1,20,21,45,52]. In

[1]School of Economics and Management, Beihang University, Beijing 100191, China. [2]Department of Physics, Bar-Ilan University, Ramat Gan 52900, Israel. [3]School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China. [4]Azrieli School of Architecture, Tel Aviv University, Tel Aviv 6997801, Israel. ✉e-mail: efratbl@tauex.tau.ac.il; haijunhuang@buaa.edu.cn; havlins@gmail.com

the present study, we introduce a simple dynamic framework and demonstrate empirically that it can quantitatively capture the network's propagation and dissipation of the congested components connected to specific traffic bottlenecks.

Many studies have provided important foundations to define or empirically identify urban traffic bottlenecks. The primary methods are based on the congestion level of a road segment itself. For example, a road with an average velocity lower than 20 km/h is regarded as the traffic bottleneck[3]. Network and percolation approaches[24,25,45,52–55] have been recently developed to identify bottlenecks of traffic congestions by considering spatiotemporal traffic dynamics. For instance, a recent method has combined congestion propagation into a "tree" structure[24,25] to identify the origins of congestions and traffic jams that are likely to stem from them in real-time traffic. Such methods referred to the earliest congested street as the "trunk" (or bottleneck), and the neighboring congested streets that developed in the following times represent the "branches" of the trunk. They found that bottlenecks associated with heavy congestion usually do not reappear on different days and hours. Luan et al.[26] used Bayesian methods to infer the congestion propagation from a given congestion bottleneck. They found that a change in the location of congestion source can lead to distinctly different congestions. Thus, a suitable forecasting method is highly necessary to identify bottlenecks that will yield large and costly traffic jams. However, methodologies that address the identification of the most critical bottlenecks in order to stop them from growing to significant congestion components, are still missing.

In this paper, we quantitatively investigate the congestion propagation and dissipation which originate from specific traffic bottlenecks in traffic networks, and provide an early signal for forecasting the oncoming significant congestions. To this end, we follow the entire spatiotemporal process of each congestion component from its emergence as a bottleneck until its disappearance. We distinguish the evolution of traffic bottlenecks for each congestion component between the growth stage and recovery stage. We follow congestion dynamics that emerged from all bottlenecks in 1 month of traffic networks in two large cities in China, and find that the recovery duration of jams associated with each bottleneck follows a power-law distribution with very similar exponents in different days. Interestingly, we observed that the recovery duration of congestion components is typically nearly twice compared to its growth duration. Our study suggests an effective early predictive method of large jams. This method is based on our finding that the sizes of congestion components are highly correlated with their growth speed. In particular, we find that the growth speed of jams in their very early propagation stage, is also highly correlated with the maximal size of the jams, and therefore, this early speed can be used to predict major bottlenecks with high accuracy.

Results
Dynamics of traffic bottlenecks
We analyze real-time traffic datasets that include the time-dependent velocity of each road segment for one month in two large cities in China, Beijing and Shenzhen. The velocity matched to each road is aggregated from numerous trajectory records recorded by GPS devices in floating cars, with a resolution of 1 min[54]. The studied time periods are 00:00–24:00 for 30 days during October 2015 (for more details of the dataset, see Methods, Supplementary Note 2 and Supplementary Figs. 1 and 2). We identify the traffic bottlenecks based on the jam tree structure[24] that assigned the earliest congested road downstream of the congestion component as the traffic bottleneck. The new congestions in neighboring upstream roads are associated with the bottleneck if they become congested no longer than a predefined time θ after the bottleneck or an existing downstream neighboring congestion in the component (see Methods).

To trace the spatiotemporal propagation and dissipation of traffic bottlenecks in empirical traffic data, we follow a bottleneck jam from its emergence to disappearance and calculate the number of existing congested roads $S(t)$ associated with this specific bottleneck over its evolution (see Fig. 1a and Methods). This dynamic network process of congestions from a bottleneck can distinguish the time evolution of size S between two stages: growth and recovery stages. The peak time t_P when size $S(t)$ of congestions connected to a bottleneck reaches maximum S_P is regarded as the end of growth stage and the beginning of the recovery stage, see Fig. 1. We define the entire lifespan T of a
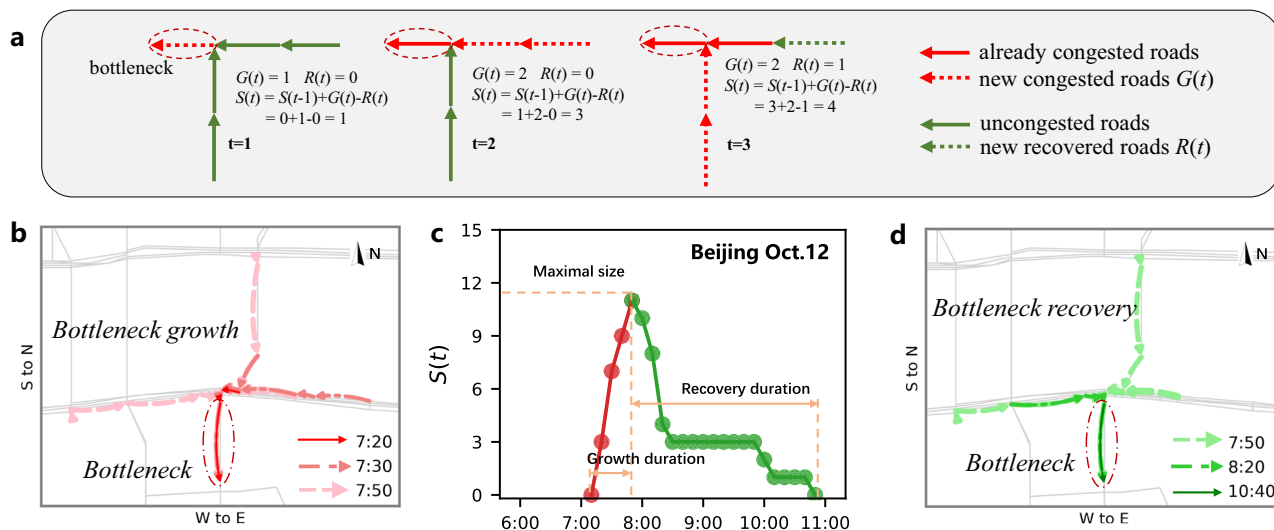

Fig. 1 | Demonstration of jam propagation and dissipation from a typical bottleneck in traffic network. a Demonstration of congestion size S associated with a bottleneck in an illustrated network. **b** Several snapshots of the congested road segments during the growth stage at 7:20, 7:30, and 7:50 for a specific but typical bottleneck on Monday, October 12, 2015, in Beijing. The congestions originated from the bottleneck (circled red solid link), and gradually developed to more congestion (light red dashed links) in the upstream neighboring streets. The arrows of the links are the directions of traffic flow. **c** The size S changes over time. This congestion component lasted for more than 3 h after its bottleneck emerged. Its maximal size S_P reached 11 road segments at 7:50, and then the size S began to decrease until it completely vanished at 10:50. Note that the recovery duration is much longer than the growth, which is a common feature of the bottleneck dynamics (see Fig. 2). **d** Several snapshots of the congested road segments during the recovery stage at 7:50, 8:20, and 10:40. The congestion component gradually dissolves after its maximal size (light green dashed links) as the road segments become no longer congested.
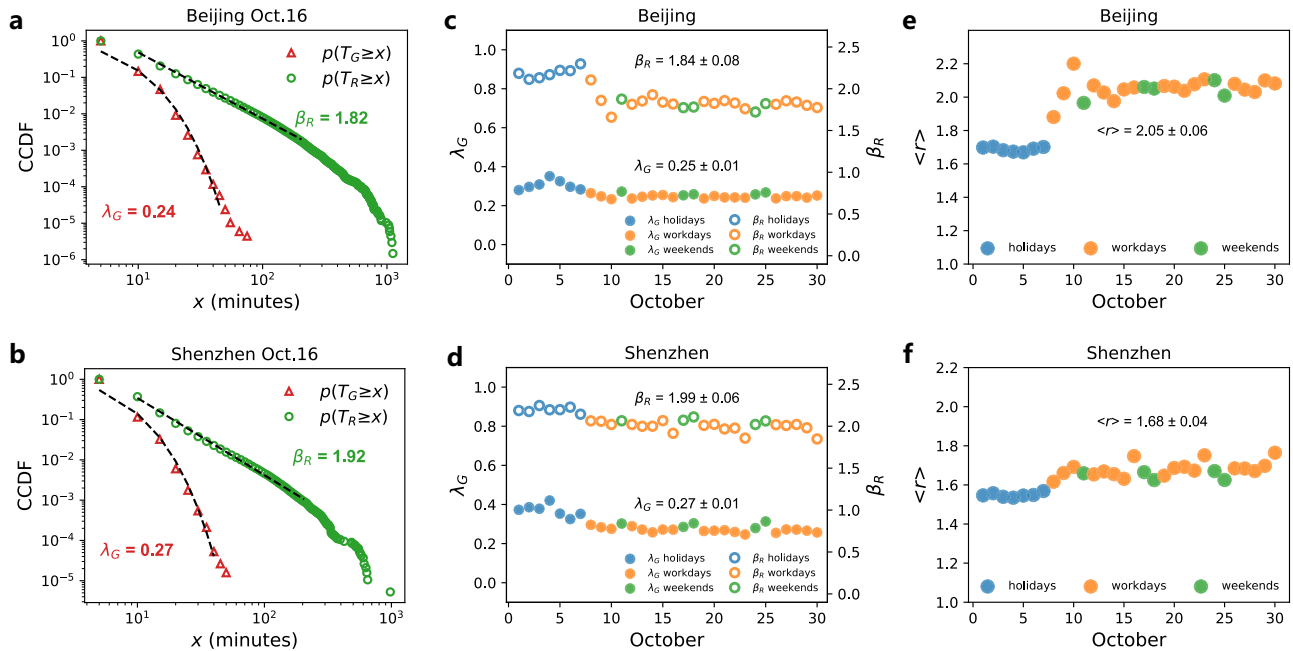
Fig. 2 | Distributions of growth duration and recovery duration. The CCDF of growth duration T_G follows approximately an exponential distribution, and the CCDF of recovery duration T_R follows approximately a power-law distribution. The data are taken for Friday, October 16, 2015, for (**a**) Beijing and (**b**) Shenzhen. We accumulate the data for all days in 1 month together and find CCDF of the growth duration above 40 min could be described by exponential distributions with smaller exponent λ_G (Supplementary Fig. 24). The exponents λ_G and β_R for each day in the entire month for (**c**) Beijing and (**d**) Shenzhen. The values of the mean ratio

$<r>$ for each day in the entire month for (**e**) Beijing and (**f**) Shenzhen. The days between October 1 and October 7 are the seven holidays of China's National Day. October 11, 17, 18, 24, and 25 are the regular weekends, and the other 18 days are workdays. The mean values and standard deviations in subfigures (**c**–**f**) are calculated for 18 workdays. As the model selection analyses (Supplementary Note 8 and Supplementary Table 1) suggest, the p value and the well-fitted curves support that the exponential distribution describes well for the main part of T_G.

bottleneck dynamics as the sum of the duration T_G of the growth stage and the duration T_R of the recovery stage, i.e.,

$$T = T_G + T_R \qquad (1)$$

here, the growth duration T_G is the time interval between the earliest time when a bottleneck emerges and the time it reaches its maximal size, S_P, and the recovery duration T_R is the time interval between the peak time and the time when the bottleneck jam has been completely dissolved and the traffic flows are uncongested again.

Figure 1 demonstrates the real-time spatiotemporal dynamics of a typical bottleneck on Monday, October 12, 2015, in Beijing, where three snapshots of the congested roads connected to the bottleneck during growth stage are displayed in Fig. 1b and three snapshots of the congested roads connected to the bottleneck during recovery stage are displayed in Fig. 1d. During the growth stage, increasing of the newly congested road segments associated with the bottleneck is faster than the recovered road segments, and this jam component propagates to its maximal size $S_P = 11$ at 7:50 (Fig. 1b, c). After the peak time, the congestions began to dissolve and the size of the jam component started to decrease, reaching size $S = 4$ at 8:20, and size $S = 1$ at 10:40. By 10:50, the jam component dissolved completely and resulted in uncongested traffic. Once the traffic congestion created by a specific bottleneck has been dissolved, a new bottleneck may emerge from the same place to initiate another congestion component (more patterns of bottleneck dynamics are given in Supplementary Figs. 3 and 4).

Additionally, we find that the network's propagation and dissipation of traffic congestions originated from a specific bottleneck, can be quantitatively described by a simple dynamical equation

$$\frac{dS(t)}{dt} = G(t) - R(t) \qquad (2)$$

which is an extension of the classical bottleneck model[37,47,56] that characterizes the queues from a bottleneck as the accumulations of single-car particles in a simplified corridor (see Supplementary Note 3 and Supplementary Fig. 5). Here, $G(t)$ and $R(t)$ are respectively the number of newly developed congested roads and the number of newly recovered roads associated with the bottleneck at time window t (see Supplementary Note 4 and Supplementary Figs. 6 and 7). Therefore, the existing size $S(t)$ in the congestion component at time window t is the integration of the difference between $G(t)$ and $R(t)$ from the time t_B when a bottleneck B occurred up to current time window t, given by

$$S(t) = \int_{t_B}^{t} (G(u) - R(u)) du \qquad (3)$$

Distributions of growth and recovery durations

To explore the evolution durations of growth and recovery stages, we follow the dynamics (or lifespan) of each traffic bottleneck in the entire road networks of Beijing and Shenzhen during 1 month. To this end, we compute the vectors of growth duration $T_G = \{T_G^1, \cdots, T_G^N\}$ and recovery duration $T_R = \{T_R^1, \cdots, T_R^N\}$. Here, T_G^i and T_R^i are respectively the growth and recovery durations of bottleneck i, and N is the total number of bottlenecks, which is over 670,000 during a typical workday in Beijing and over 180,000 in Shenzhen. The distributions of durations T_G and T_R on Friday, October 16, 2015, in both cities are presented in Fig. 2a, b. The results show that the distribution of T_R is significantly broader than that of T_G. Bottlenecks spend at most 100 min to reach the maximal size S_P, while they might spend up to 1000 min to recover to an uncongested state. The complementary cumulative distribution function (CCDF) of growth duration T_G is well

approximated by the exponential distribution

$$p(T_G \geq x) \sim e^{-\lambda_G x} \qquad (4)$$

with exponents λ_G of 0.24 for Beijing and 0.27 for Shenzhen on Oct. 16. The CCDF of recovery duration T_R is well approximated by the power law distribution

$$p(T_R \geq x) \sim x^{-\beta_R} \qquad (5)$$

with β_R of 1.82 for Beijing and 1.92 for Shenzhen. We fit both the CCDF (Supplementary Figs. 8–11) and probability distribution (Supplementary Figs. 12 and 13) to describe the main part of T_G and T_R, and find that they have similar distribution patterns in different days. In addition, CCDF of the ratio $r = T_R/T_G$ of each congestion component follows also a power-law distribution $p(r \geq x) \sim x^{-\beta_r}$ with exponents β_r around 2.1 ± 0.09 on workdays in Beijing and 2.16 ± 0.05 on workdays in Shenzhen (see Supplementary Figs. 14 and 15 for CCDF of r, and Supplementary Figs. 16, 17 and 18 for distribution of r). For sensitivity analysis, we absorb very short roads into the junctions or consider congestion length in the definition of congestion size $S(t)$, and find that the CCDFs of the T_G and T_R are robust (Supplementary Note 5 and Supplementary Figs. 19, 20 and 21). We also study the bottleneck dynamics by smoothing the traffic conditions in different time intervals, and find CCDFs of T_G and T_R follow similar distributions (Supplementary Fig. 22).

The average ratio $<r>$ for all the congestion components is evaluated by

$$<r> = \frac{1}{N} \sum_{1 \leq i \leq N} r_i = \frac{1}{N} \sum_{1 \leq i \leq N} \frac{T_R^i}{T_G^i} \qquad (6)$$

where the values of $<r>$ are 2.06 for Beijing and 1.74 for Shenzhen on Oct. 16. This shows that, on average the recovery duration of a congestion component is 2.06 times of its growth duration in Beijing, and 1.74 times in Shenzhen. This demonstrates the strong asymmetry between network's propagation and dissipation of bottlenecks, which could be explained by the theory of phase transition in traffic network flows. The propagation of urban traffic congestion has been regarded as the percolation in the system[54,57], and the traffic systems have exhibited phase transitions as the traffic congestion components are close to the critical point in the self-organized process[7,45]. It has been found that the recovery of the systems against the perturbations can be slowing down, when the system is close to such critical point, also called "critical slowing down"[58–61]. As the results in our manuscript, the power-law distribution of the congestion dissipation indicates that the traffic systems might be near the critical points of the self-organized process. Thus, the slower recovery of the system could be the result of the "critical slowing down" mechanisms[62].

Comparing growth and recovery stages on different days, we find that the exponential exponents λ_G of growth duration and power law exponents β_R of recovery duration are stable for the same type of days (Fig. 2c, d). This may indicate the stable self-adaptability of the traffic system in the dynamics of congestion evolutions[45] (See Supplementary Note 6 for further discussions of the exponents). In Fig. 2e, f, we compare the average ratio $<r>$ between the recovery duration and growth duration on 30 days during October 2015 in both Beijing and Shenzhen. The results show that $<r>$ has a stable value for a specific type of day. In both cities, values of $<r>$ on workdays are larger than those on holidays. This means that on workdays, the recovery stage of the congestion components takes much longer time than their growth stage compared to holidays. The values of $<r>$ could be influenced by the heavy tails of the recovery duration (Supplementary Note 7 and Supplementary Fig. 23). The "critical slowing down" effect[58] becomes more significant for those bottlenecks with large recovery duration.

This could be related to the fact that nonlinear mechanisms exist between the congestion propagation and dissipation processes given the complex urban traffic network topologies and the travelers' real-time self-adaptive route choice behaviors[58,63,64].

Correlation between spatial and temporal dynamics

We further explore the relationship between temporal and spatial dynamics of traffic bottlenecks, by examining the correlation between the maximal size S_P and the growth speed of the congestion components. For that, we define and calculate the average growth speed of congestion component associated with each bottleneck. The average growth speed V_A is defined as the maximal size S_P divided by its growth duration T_G, i.e.,

$$V_A = \frac{S_P}{T_G} \qquad (7)$$

A larger V_A means that a bottleneck induces more associated congestions per unit time. Here, V_A is calculated by the average number of the increased congested road segments in every 5 min. Figure 3a, b shows the relationship between S_P and V_A on October 16, 2015, in Beijing and Shenzhen (for similar results on other days, see Supplementary Figs. 25 and 26). These box plots of S_P are classified into different groups by the values of V_A. It is seen that congestion components having a larger speed V_A are more likely to reach a larger congestion size (In Supplementary Fig. 27, the positive relationship of V_A and S_P is also observed by plotting the density surfaces). The linear correlation between the maximal size $S_P = \{S_P^1, \cdots, S_P^N\}$ and the average growth speed $V_A = \{V_A^1, \ldots, V_A^N\}$ of these congestion components is also characterized by the high values of Pearson correlation coefficients ρ_{S_P, V_A}. When omitting isolated bottlenecks with $S_P = 1$, the value of ρ_{S_P, V_A} on Oct. 16, 2015 is 0.75 for Beijing and 0.79 for Shenzhen, indicating that S_P and V_A have a strong positive linear correlation.

Figure 3c, d shows that in both cities, the values of the correlation ρ_{S_P, V_A} are stable for the same type of days. The correlation coefficients are larger on holidays than on workdays, and also larger in Shenzhen than in Beijing. Note also that in the middle of holidays (days Oct. 3–5) the correlations are higher due to less traffic demand. We also find that the bottlenecks which create large congestions are less probable to sustain for a very long duration (Supplementary Figs. 28 and 29). This could be because drivers are strongly averse to very large congestions in the self-organization process of their game behavior, and would like to make adaptive choices to deviate from the long queues[65,66]. Moreover, well-developed traffic control technologies could have also focused on alleviating these heavy congestions[67]. Even so, as can be seen in the figures (Supplementary Figs. 28 and 29), very large congestions can still be sustained for hours.

An early signal for forecasting heavy congestions

The average growth speed has been found in previous subsection to be correlated with the jam size. However, this correlation can only be determined at the end of the growth stage and thus it is not suitable for predicting heavy bottlenecks before their full growth.

In the present section, we develop an effective early signal for forecasting heavy traffic congestions based on the initial growth stage of jams. For that, we explore if the maximal size S_P is correlated to the earliest attributes of the congestion propagation, i.e., the initial growth speed V_{T_I}, defined as the value of the early size S_{T_I} divided by its initial growth time T_I at an early growth stage. Thus, V_{T_I} is expressed by

$$V_{T_I} = \frac{S_{T_I}}{T_I} \qquad (8)$$

where T_I is the time interval between the emergence of a bottleneck and a selected early time before the congestion component reaches its
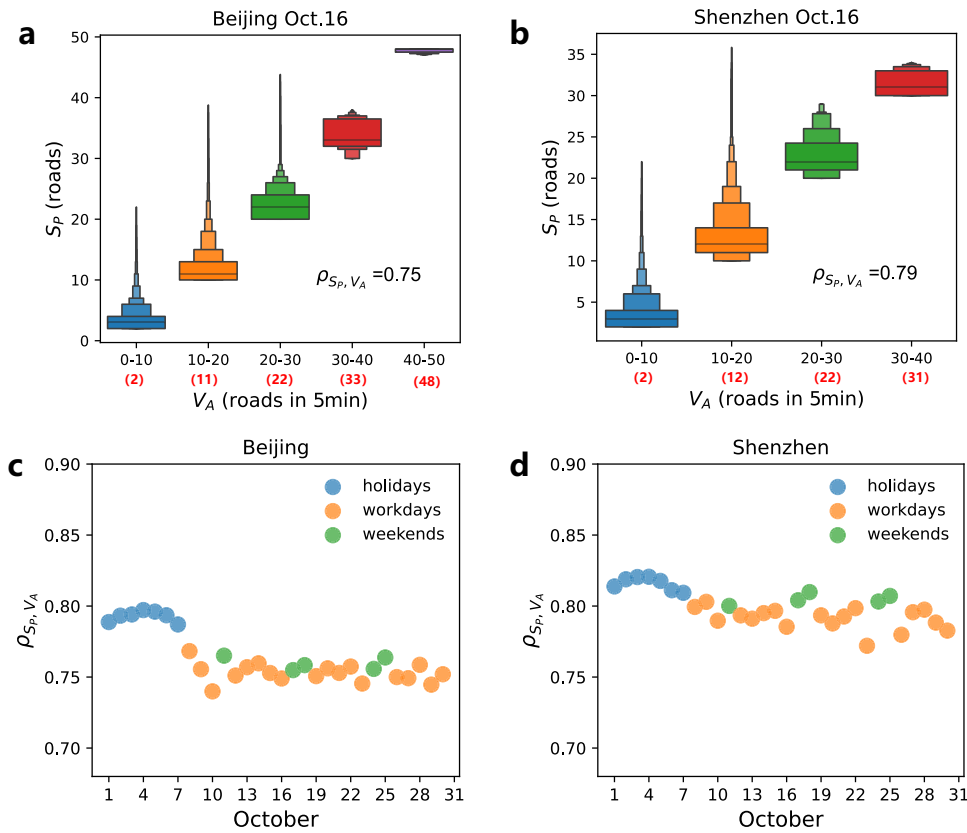
Fig. 3 | Correlation between the spatial maximal size S_P and the temporal growth speed of congestion components. Box plots of size S_P grouped by average growth speed V_A on Friday, October 16, 2015, in (**a**) Beijing and (**b**) Shenzhen. Here, V_A means the average number of increased congested road segments every 5 min during the entire growth stage, the black numbers on the x-axis are the range of average growth speed V_A, and the red numbers in the bracket are the median value of each group. The black lines at the bottom, middle and top of the largest box are respectively 25%, 50%, and 75% percentiles of S_P. The tails of the distributions could be more clearly observed by more quantiles in 0–25% and 75–100% of S_P. The outliers are not displayed in the box plots. The growth duration T_G increases with S_P on average showing very large fluctuations, as shown in Supplementary Fig. 30. Pearson correlation ρ_{S_P,V_A} for the 30 days during October 2015 in (**c**) Beijing and (**d**) Shenzhen. Holidays include the 7 days in China's National Day, and the workdays include 18 days. Bottlenecks with $S_P \geq 2$ are counted here. For sensitivity analysis, we absorb very short roads or consider congestion length in the definition of congestion size $S(t)$, and find that the growth speed of congestion components is still strongly and positively correlated to their maximal congestion size (Supplementary Note 5 and Supplementary Fig. 31).

maximal size, and S_{T_I} is the number of congested roads connected to the bottleneck at the chosen early growth duration T_I.

As seen in Fig. 4a–c, the box plots of S_P grouped by V_{T_I} (T_I ranges from 5 to 15 minutes) show that on October 16, 2015, Beijing, the maximal size S_P increases with V_{T_I} at the first 5, 10, or 15 min of the bottleneck dynamics (box plots for Shenzhen show similar patterns, see Supplementary Fig. 32). This indicates that congestion components with a larger initial growth speed even at the earliest stage are also more likely to reach a larger congestion size. The correlation coefficients $\rho_{S_P,V_{T_I}}$ between the maximal size $S_P = \{S_P^1, \cdots, S_P^N\}$ and the initial growth speed $V_{T_I} = \{V_{T_I}^1, \cdots, V_{T_I}^N\}$ are respectively 0.66, 0.76, and 0.87 for $T_I = 5$, $T_I = 10$ and $T_I = 15$, which also supports our claim that the initial growth speed has a strong positive correlation with the maximal size S_P (Fig. 4a–c). In addition, we calculate and compare the linear correlation $\rho_{S_P,V_{T_I}}$ between the maximal size S_P and the initial growth speed V_{T_I} for different T_I in workdays and holidays. As seen in Fig. 4d, e, the values of $\rho_{S_P,V_{T_I}}$ for the two cities are around 0.65 to 0.75 for $T_I = 5$ min, and increase to a higher plateau, around 0.85 to 0.95, when V_{T_I} is obtained for 15 min and above. Note that correlations $\rho_{S_P,V_{T_I}}$ in holidays are higher due to less traffic demand, which may lead to better prediction results. This suggests that the initial growth speed in the earliest 15 min, could be an effective predictor for the maximal size of congestions originated from a bottleneck, since the change in correlations is very little when T_I is above 15 min.

Forecasting performance for major bottlenecks

Before examining the performance of predicting major bottlenecks based on initial growth speed V_{T_I}, we evaluate the recurrence of the bottlenecks on two different days. This is done based on the Jaccard index J that assesses the overlap (see Supplementary Fig. 33) of bottlenecks or their associated congestion components between 2 workdays. As shown in Supplementary Figs. 34, 35, and 36, the overlap of associated congestions is higher than the overlap of bottlenecks. This could be because the congestion propagation mechanism is fundamental[68]. Thus, the overall congestion regions could include similar propagation pathways due to the congestion propagation mechanism (Supplementary Note 9). Importantly, both Jaccard indexes J_B for bottlenecks and J_C for their associated congestions decrease rapidly with increasing jam size. J_B decreases to around 0.05 and J_C decreases to around 0.2 when S_P is greater than or equal to 20, meaning that only around 5% of all the heavy bottlenecks and around 20% of their associated congestions (i.e., $S_P \geq 20$) that appear in both days are the same. Note that, similarities of large congestions are still very low even when we relax the overlap definition (Supplementary Note 10 and Supplementary Figs. 33 and 37). This could be because the traffic flows are the result of the dynamic self-organization of the game behaviors of numerous travelers[69–71]. In the day-by-day game process, due to risk aversion, travelers have been found to make adaptive choices and deviate from the congestions in previous days[72–75]. Thus,
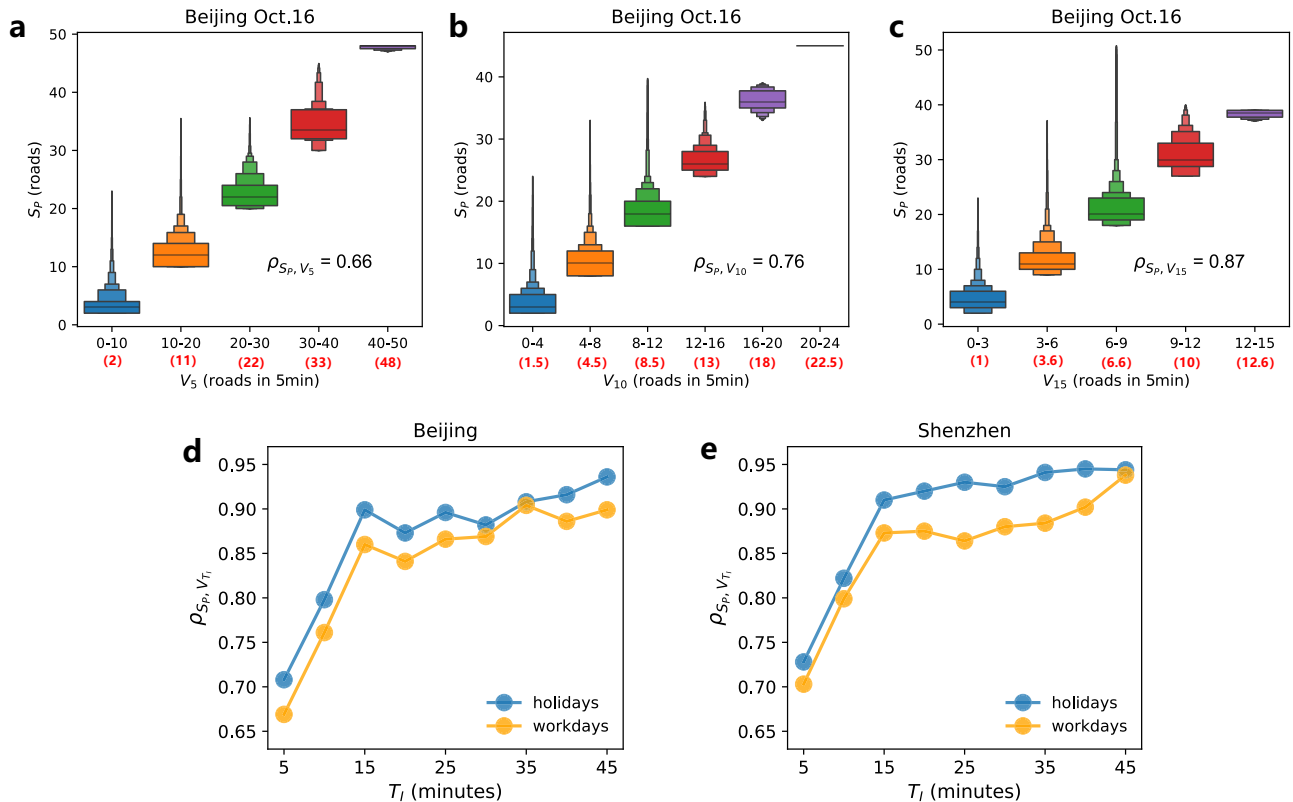
Fig. 4 | Correlation between the maximal size S_P and the initial growth speed V_{T_I} at the earliest stage. Box plots of size S_P grouped by initial growth speed V_{T_I} at the earliest (**a**) 5, (**b**) 10 and (**c**) 15 min, on October 16, 2015, in Beijing. Here, V_{T_I} means the number of increased congested road segments every 5 minutes during the early growth stage, the black numbers on the x-axis are the range of initial growth speed V_{T_I}, and the red numbers in the bracket are the median value of each group. The outliers are not displayed in the box plots. Pearson correlation $\rho_{S_P, V_{T_I}}$ between maximal size S_P and the initial growth speed V_{T_I} increases with increasing the initial growth duration T_I, for (**d**) Beijing and (**e**) Shenzhen. The correlations between S_P and V_{T_I} are calculated for the congestion components with growth duration longer than or equal to T_I (see Methods). Bottlenecks with $S_P \geq 2$ are counted here. Note that above 15 min the correlations almost do not increase. Holidays include the 7 days in China's National Day, and the workdays include 18 days.

large congestion events could be less recurrent than expected, and heavy bottlenecks or their associated congestions in one day cannot represent much of those in the other days.

We define a major bottleneck (i.e., $y = 1$) as the bottleneck that developed into a congestion component with size S_P greater than or equal to a pre-defined threshold S_L, and regard it as a minor bottleneck (i.e., $y = 0$) otherwise. To test the power of predicting major bottlenecks by V_{15}, we trained the binary Probit model based on dynamics of all the bottlenecks on Monday, October 12, 2015, in Beijing (see Methods for details of the prediction method). We applied the trained model to predict whether bottlenecks on another workday, i.e., October 16, 2015, could yield congestion components greater than or equal to $S_L = 20$. We find that when the false positive rate (FPR) is set to 5%, the accuracy of detecting major bottlenecks is around 88%. The detected and the undetected major bottlenecks displayed in Fig. 5 indicate that most of the major bottlenecks in rush and non-rush hours are different, and an appropriate prediction framework at the early propagation stage is thus highly needed. Additionally, examples of false positive bottlenecks are also located in significantly different road segments in rush and non-rush hours (see Supplementary Fig. 38).

Performances of predicting the major bottlenecks ($S_L = 20$) are assessed next for different values of FPR, for V_5, V_{10}, V_{15}, and V_{20}. In Fig. 6a, b, the receiver operating characteristic (ROC) curves[76] are obtained and plotted based on the prediction results for Friday, October 16, 2015, in both Beijing and Shenzhen. The ROC curves indicate that the predictions based on initial growth speed within 15 minutes of the congestion components are close to the classifier (0,1) for both cities (ROC curves of other days are similar and presented in

Supplementary Figs. 41 and 42). We also computed the area under the curve (AUC) for 17 workdays in both cities, as presented in Fig. 6c. The values of AUC are stable for different days and increase when a longer duration T_I is used to calculate the predictor, i.e., the initial growth speed. These results suggest that initial growth speed of the congestion components within its first 15 minutes is a very reliable predictor for forecasting major congestions. Beyond 15 min, the performance of the predictions does not improve significantly. Notably, as seen in Supplementary Note 12 and Supplementary Fig. 43, the prediction time has a trade-off with prediction accuracy. In practice, the predictors V_5, V_{10}, and V_{15} could be comprehensively applied to early identify the heavy bottlenecks before maximal size or full jam duration.

To further assess the prediction performance of the growth speed in the earliest 15 min, we forecast whether the bottlenecks will develop to be greater than or equal to different threshold size S_L. As seen in Fig. 6d and e, the ROC curves of predicting major bottlenecks for different S_L are similar and very close to the classifier (0,1) on October 16, 2015 in both cities. We also computed in Fig. 6f, the AUC values versus the predefined threshold size S_L for 17 workdays in both cities. It is seen that the values of the AUC are stable and very high (around 0.95) for all values of S_L. These results strongly support that the initial growth speed is a reliable and excellent measure for detecting major bottlenecks at their early stage of propagation. Thus, our prediction framework helps to understand, forecast, and potentially prevent major bottlenecks by identifying them at their early propagation stages. Integrating this framework into existing traffic control systems, may improve traffic conditions and prevent bottlenecks from reaching their full potential.
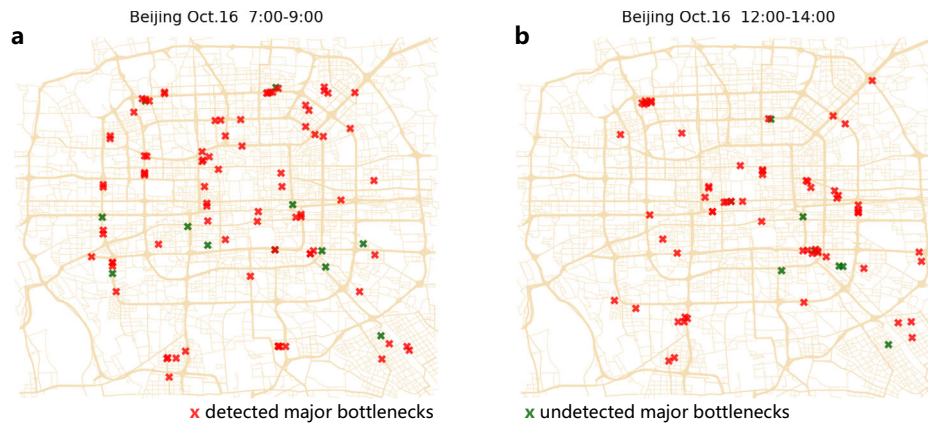
Fig. 5 | The detected major bottlenecks (true positive) and the undetected major bottlenecks (false negative) when the FPR is set as 5%. Major bottlenecks ($S_L = 20$) during (a) rush hours (7:00 a.m.–9:00 a.m.) and (b) non-rush hours (12:00 p.m.–14:00 p.m.) on October 16, 2015, in Beijing. The red stars are the true positive bottlenecks that are predicted as major bottlenecks and actually grew to be

a major size. The green stars are the false negative bottlenecks that are predicted to be minor bottlenecks but actually developed to be major. The CCDF of maximal jam size S_P is seen in Supplementary Note 11 and Supplementary Fig. 39. The CCDF of jam duration T for the large congestion components is seen in Supplementary Note 11 and Supplementary Fig. 40.

Discussion

In summary, we analyze the spatiotemporal dynamics of traffic bottlenecks during 1 month in two Chinese metropolises. The empirical results demonstrate that our framework of dynamics of traffic bottlenecks can successfully characterize both the network's propagation and dissipation of urban traffic bottlenecks. We find that the propagation and dissipation processes of congestion components develop asymmetrically and the duration of recovery stage approximately follows a power-law distribution. In addition, the recovery duration of the bottlenecks is found to be typically nearly twice the time it takes to fully grow. Our findings can provide data explanations for the well-established models and help extend mesoscopic traffic models in urban networks. For example, the scaling patterns of the jam durations in real urban networks can well explain and expand the 1-dimensional scaling results modeled in existing studies[77,78].

Moreover, the empirical results suggest that our method is efficient in forecasting heavy traffic bottlenecks, based on the finding that the spatial and temporal dynamics of traffic congestions are strongly correlated. Our empirical analysis shows that the growth speed of congestion components is highly correlated with the maximal size of their associated congestions. Most importantly, we find that the initial growth speed, even at the earliest 15 min of the jam's growth, is useful in predicting whether a traffic bottleneck will develop into a major congestion component. Our predictive framework can provide an alerting indicator that could help urban planners improve real-time traffic control by addressing the sources of destructive traffic jams before they develop into their maximal size. The early real-time traffic control began with the implementation of the traffic signal control system SCOOT to calculate the size of the queue and time of clearance given the cyclic flow profiles[79]. With artificial intelligence technologies, and connected-automated vehicles techniques[80–82], the model predictive control framework has been proposed to sense and control traffic conditions while agents interact with ones in the vicinity[67,83–86]. Meantime, the spatial congestion propagation could significantly influence the coordinated control performances among different signalized intersections[87–89] (Supplementary Note 13). Based on our identified spatiotemporal propagation mechanisms, the real-time control systems could consider interrelations among different traffic regions and perform better coordinated control for the most congested areas before they propagate to a global gridlock jam.

To better alleviate traffic congestions, our framework and findings can also be applied in the well-developed real-time navigation

systems[90,91] or demand management strategies[63,64,69,71,73,75,92]. Urban traffic control systems should be connected to navigation apps to warn travelers, in near real-time, to avoid the roads that may lead to a heavy traffic bottleneck[63,64,69,93]. On one hand, warning travelers with alternative routes to be far away from the predicted heavy congestion regions[73,75] could alleviate the traffic flows in predicted areas. On the other hand, rescheduling departure time of the potential travelers with no alternative routes by road pricing strategies[71] could reduce traffic demand in the predicted area. These methods of demand management have also been found effective in alleviating congestion in earlier studies[63,64,93]. Moreover, coordinated management from the supply side could also help to alleviate the traffic on the predicted heavy congestions[69]. While such interventions may cause congestions in other locations, our real-time updated framework can alert the control system about new heavy bottlenecks before they emerge. Considering our propagation mechanism in the above well-established technologies, the real-time traffic management and organization would be better developed in the near future. This will provide not only accurate predictions but also effective mitigation of heavy congestions.

Methods
Dataset
The Beijing road network has over 52,000 road segments (links) and 27,000 intersections (nodes), and the Shenzhen road network has over 22,000 road segments (links) and 12,000 intersections (nodes). The empirical datasets contain time-dependent velocity records of both cities for 30 days during October 2015 with a resolution of 1 minute. For a road with an outlier of velocity record, we handle it as congested state if it connects its upstream and downstream roads that are congested in the predefined chronological order (Supplementary Note 2 and Supplementary Fig. 1). The distribution of road velocity suggests that the road velocity records are generally stable over 5-minute interval (Supplementary Note 2 and Supplementary Fig. 2). Large-scale network topology, huge traffic flow, serious traffic jams, diverse traffic bottlenecks, as well as the high-quality datasets make these two cities ideal for the analysis of bottleneck dynamics.

Bottlenecks and their associated congestions
Since traffic congestion state at time t can be described by the average relative velocity during a short time interval[52], we use the average relative velocity calculated in every 5 minutes to represent the
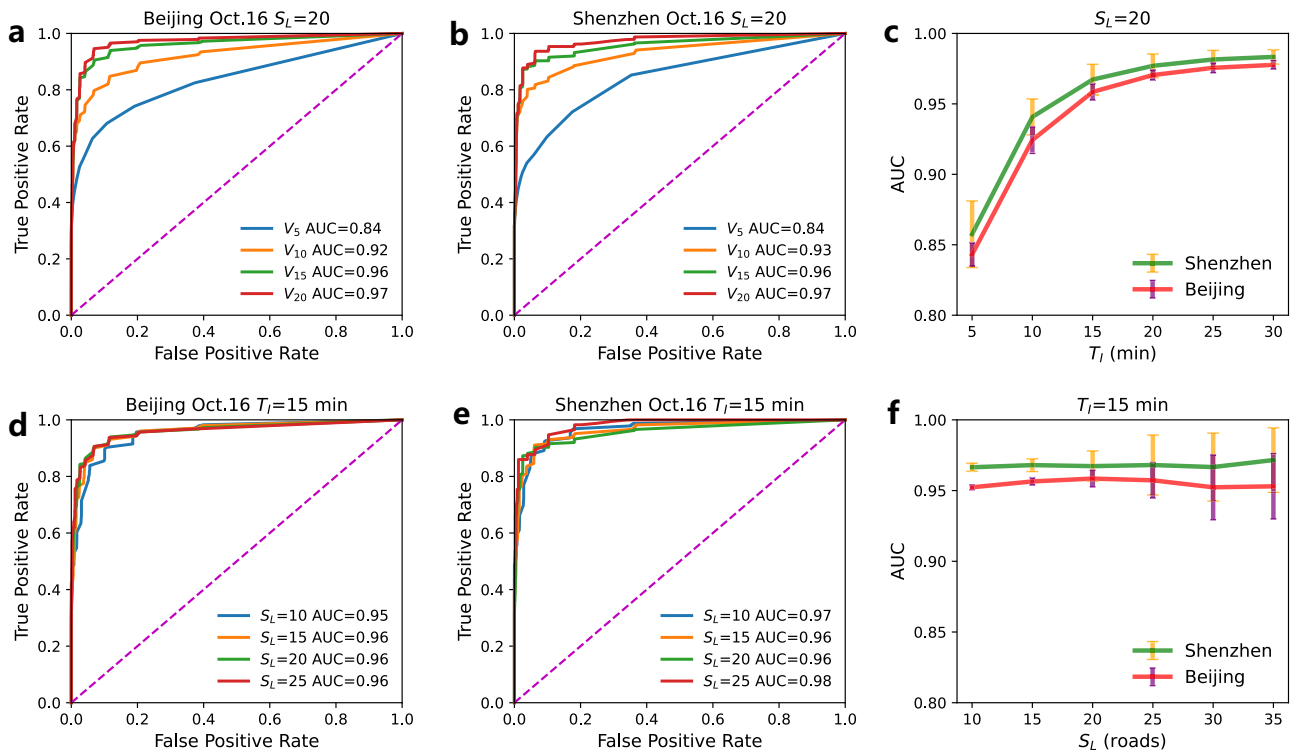
Fig. 6 | Performance of predicting major bottlenecks based on initial growth speed. ROC curves of predicting whether a bottleneck would grow up into a major congestion size of 20 road segments or above, based on V_{T_I} within the earliest 5, 10, 15, and 20 min of their propagation. The prediction results shown are for October 16, 2015, in (**a**) Beijing and (**b**) Shenzhen. **c** The AUC value increases when the predictor V_{T_I} is calculated based on a longer initial duration T_I. The error bars are calculated based on the prediction results of 17 workdays. The small standard deviations show that the values are stable. Note the almost plateau seen for $T_I \geq 15$ minutes. The ROC curves of using V_{T_I} within 15 min to predict whether a bottleneck develops into at least 10, 15, 20, and 25 congested roads. The curves are based on the prediction results for October 16, 2015, in (**d**) Beijing and (**e**) Shenzhen. **f** The values of AUC for predicting major bottlenecks by V_{T_I} within 15 min with different size thresholds S_L. The error bars are calculated based on the prediction results in 17 workdays. The small standard deviations show that the prediction performances are stable. The prediction results are also found stable when training the model on different days. For sensitivity analysis, we absorb very short roads or consider congestion length in the definition of congestion size $S(t)$, and find robust prediction performance (Supplementary Note 5 and Supplementary Fig. 44).

congestion weight $W_e(t)$ of road e at time window t, i.e.,

$$W_e(t) = \frac{U_e(t)}{U_e^{95}} \tag{9}$$

Here, $U_e(t)$ is the velocity of road e at time window t, and U_e^{95} is 95% percentile of velocity records of road e, which approximates the standard maximal velocity on this road. In this paper, road e is regarded as congested if $W_e(t) < 0.5$, and non-congested otherwise[26].

After having the congestion weight $W_e(t)$, we identify bottlenecks using the jam tree structure defined by ref. [24]. We identified the bottlenecks, and the newly congested or the recovered roads from them as the following steps: *i*) Obtaining the time-interval matrix of the continuous congestion states. $c_e(t)$ is the time interval during which road e has been continuously congested up to the time window t. *ii*) Identification of bottleneck. The bottleneck is the road segment that has been continuously congested for the longest interval $c_e(t)$ in downstream of a congestion component. *iii*) Identification of newly congested road segments. A neighboring upstream congested road is regarded as a new congestion associated with a bottleneck at time window t if it became congested no later than a predefined time interval θ (set as 10 min generally) after the bottleneck or another congested road connected to the bottleneck. *iv*) Identification of recovered road segments. A road is regarded as recovered from the bottleneck at time window t, if it (or another congested road that connected it to the bottleneck) returned to an uncongested state (see Supplementary Note 4 and Supplementary Fig. 6). The existing number

of congested roads, i.e., size $S(t)$, includes newly congested links and removes those recovered. In the recovery process, the existing size S is decreasing over time since the newly recovered congested roads occur more quickly than newly developed congestions (Supplementary Note 4 and Supplementary Fig. 7). In our algorithm, if the recovery of the bottleneck is earlier than other neighboring branches, the branches that belonged to the recovered bottleneck are not considered anymore to be associated with the original (recovered) bottleneck.

Pearson correlation
The Pearson coefficient represents the degree of linear correlation between two variables. To observe the linear correlation between the maximal size S_P and the average growth speed V_A or the initial growth speed V_{T_I}, we calculated Pearson correlation between them. For two variables $X = \{x_1, x_2, \cdots, x_n\}$ and $Y = \{y_1, y_2, \cdots, y_n\}$, the Pearson correlation $\rho_{X,Y}$ is

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{10}$$

where n is the number of samples, μ_X and μ_Y are the mean values of X and Y, and σ_X and σ_Y are the standard deviations of X and Y. $\rho_{X,Y}$ is in the range of -1 to 1. $\rho_{X,Y} = \pm 1$ means variables X and Y have the highest linear correlation. When calculating $\rho_{S_P, V_{T_I}}$ for congestion components with growth duration longer than or equal to T_I, the samples are downsized in a short T_I and upsized in a long T_I by observations obtained from different resolutions.

Jaccard index

The Jaccard index is used for measuring repetition between the bottlenecks or their associated congestions on different days. For two sample sets X and Y, the Jaccard index is defined as the ratio between the size of the overlap and the size of the union of the observed sample sets, i.e.,

$$J = \frac{|X \cap Y|}{|X \cup Y|} = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \tag{11}$$

where J is in the range of 0 to 1. $J = 1$ means sets X and Y are wholly overlapped and $J = 0$ means sets X and Y are wholly different.

Prediction method

To conduct the prediction, we train the binary Probit model (a basic binary classifier) based on the evolutions of traffic bottlenecks on Monday, October 12, 2015, and use the initial growth speed V_{T_I} till an early growth time T_I to predict the probability P_i that a bottleneck on another workday may develop into a major congestion component. The utility function (y_i^*) of bottleneck i with the explanatory variable V_{T_I} is expressed as

$$y_i^* = a_1 + a_2 V_{T_I}^i + \varepsilon_i, y_i = 1 \; if \; y_i^* > 0 \tag{12}$$

where a_1 is a constant, a_2 is a coefficient, and ε_i is the error term assumed to follow a standard normal distribution. Based on the estimated $\hat{a}_1$ and $\hat{a}_2$, the probability for $y_i = 1$ is expressed as

$$p_i(y_i = 1 | V_{T_I}) = p_i(y_i^* > 0) = p_i(\varepsilon_i > -(\hat{a}_1 + \hat{a}_2 V_{T_I}^i)) = \Phi(\hat{a}_1 + \hat{a}_2 V_{T_I}^i) \tag{13}$$

where Φ is the cumulative density function (*cdf*) of the standard normal distribution. Prediction based on V_{T_I} is conducted in the following way: the probability P_i of whether a bottleneck would become major till the time T_I, is calculated as the highest possibility that a bottleneck may become major based on V_{T_I} till time T_I, i.e., $P_i = \max\{p_i(y_i = 1 | V_t)\}, t \le T_I$. If the probability P_i reaches a threshold P^{thre}, a bottleneck is classified as major. The prediction result γ_i of a bottleneck could be true positive (TP), false positive (FP), true negative (TN), or false negative (FN), i.e.,

$$\gamma_i = \begin{cases} TP, P_i \ge P^{thre} \text{ and } y_i = 1 \\ FN, P_i < P^{thre} \text{ and } y_i = 1 \\ TN, P_i < P^{thre} \text{ and } y_i = 0 \\ FP, P_i \ge P^{thre} \text{ and } y_i = 0 \end{cases} \tag{14}$$

When obtaining the prediction results γ for different threshold P^{thre}, the ROC curve can be plotted by the results of the true positive rate (TPR) versus the false positive rate (FPR). The x-axis is the FPR, representing the ratio between the number of minor bottlenecks wrongly categorized as positive and the total number of actual minor bottlenecks. The y-axis is the TPR, representing the ratio between the number of major bottlenecks correctly categorized as positive and the total number of actual major bottlenecks. i.e.,

$$FPR = \frac{N_{FP}}{N_{FP} + N_{TN}} \tag{15}$$

$$TPR = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{16}$$

where N_{FP} is the number of FP observations, N_{TN} is the number of TN observations, N_{TP} is the number of TP observations, and N_{FN} is the number of FN observations. The AUC represents the performance of separability that the classifier can distinguish between the major and minor bottlenecks. The curve through (0,1) and AUC = 1 represent that the classifier has the best performance to separate each bottleneck correctly.

Data availability

The necessary data generated in this study have been deposited in the Github database https://github.com/JinxiaoDuan/DynamicsOfBottlenecks to enable the reproducibility of the results in the paper. The raw velocity data are protected and unavailable due to data privacy laws.

Code availability

We provide the source codes of our algorithms in GitHub https://github.com/JinxiaoDuan/DynamicsOfBottlenecks to enable the reproducibility of our findings.

References

1. Mahmassani, H. S., Saberi, M. & Zockaie, A. Urban network gridlock: theory, characteristics, and dynamics. *Procedia-Soc. Behav. Sci.* **80**, 79–98 (2013).
2. Saeedmanesh, M. & Geroliminis, N. Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. *Transp. Res. Procedia* **23**, 962–979 (2017).
3. Long, J., Gao, Z., Ren, H. & Lian, A. Urban traffic congestion propagation and bottleneck identification. *Sci. China Ser. F Inf. Sci.* **51**, 948–964 (2008).
4. Lighthill, M. J. & Whitham, G. B. On kinematic waves II. A theory of traffic flow on long crowded roads. *Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci.* **229**, 317–345 (1955).
5. Richards, P. I. Shock waves on the highway. *Oper. Res.* **4**, 42–51 (1956).
6. Biham, O., Middleton, A. A. & Levine, D. Self-organization and a dynamical transition in traffic-flow models. *Phys. Rev. A* **46**, R6124 (1992).
7. Nagel, K. & Schreckenberg, M. A cellular automaton model for freeway traffic. *J. Phys. I* **2**, 2221–2229 (1992).
8. Kerner, B. S. Experimental features of self-organization in traffic flow. *Phys. Rev. Lett.* **81**, 3797 (1998).
9. Kerner, B. S. & Rehborn, H. Experimental properties of phase transitions in traffic flow. *Phys. Rev. Lett.* **79**, 4030 (1997).
10. Kerner, B. S., Klenov, S. L. & Wolf, D. E. Cellular automata approach to three-phase traffic theory. *J. Phys. A Math. Gen.* **35**, 9971 (2002).
11. Arnott, R. & Small, K. The economics of traffic congestion. *Am. Sci.* **82**, 446–455 (1994).
12. Nagel, K., Wolf, D. E., Wagner, P. & Simon, P. Two-lane traffic rules for cellular automata: a systematic approach. *Phys. Rev. E* **58**, 1425 (1998).
13. Hidas, P. Modelling lane changing and merging in microscopic traffic simulation. *Transp. Res. Part C: Emerg. Technol.* **10**, 351–371 (2002).
14. Daganzo, C. F. A behavioral theory of multi-lane traffic flow. Part II: merges and the onset of congestion. *Transp. Res. Part B: Methodol.* **36**, 159–169 (2002).
15. Daganzo, C. F. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transp. Res. Part B: Methodol.* **28**, 269–287 (1994).
16. Daganzo, C. F. The cell transmission model, part II: network traffic. *Transp. Res. Part B: Methodol.* **29**, 79–93 (1995).
17. Li, D., Jiang, Y., Kang, R. & Shlomo, H. Spatial correlation analysis of cascading failures: congestions and blackouts. *Sci. Rep.* **4**, 5381 (2014).
18. Wu, J., Sun, H. & Gao, Z. Cascading failures on weighted urban traffic equilibrium networks. *Phys. A: Stat. Mech. Appl.* **386**, 407–413 (2007).
19. Cwilich, G. & Buldyrev, S. V. Cascading traffic jamming in a two-dimensional Motter and Lai model. *Phys. Rev. E* **106**, 024303 (2022).

20. Duan, J., Li, D. & Huang, H.-J. Reliability of the traffic network against cascading failures with individuals acting independently or collectively. *Transp. Res. Part C: Emerg. Technol.* **147**, 104017 (2023).

21. Saberi, M. et al. A simple contagion process describes spreading of traffic jams in urban networks. *Nat. Commun.* **11**, 1616 (2020).

22. Wu, J., Gao, Z. & Sun, H. Simulation of traffic congestion with SIR model. *Mod. Phys. Lett. B* **18**, 1537–1542 (2004).

23. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).

24. Serok, N., Havlin, S. & Lieberthal, E. B. Identification, cost evaluation, and prioritization of urban traffic congestions and their origin. *Sci. Rep.* **12**, 13026 (2022).

25. Nguyen, H., Liu, W. & Chen, F. Discovering congestion propagation patterns in spatio-temporal traffic data. *IEEE Transactions on. Big Data* **3**, 169–180 (2016).

26. Luan, S., Ke, R., Huang, Z. & Ma, X. Traffic congestion propagation inference using dynamic Bayesian graph convolution network. *Transp. Res. Part C Emerg. Technol.* **135**, 103526 (2022).

27. Asif, M. T. et al. Spatiotemporal patterns in large-scale traffic speed prediction. *IEEE Trans. Intell. Transp. Syst.* **15**, 794–804 (2013).

28. Wu, Y., Tan, H., Qin, L., Ran, B. & Jiang, Z. A hybrid deep learning based traffic flow prediction method and its understanding. *Transp. Res. Part C Emerg. Technol.* **90**, 166–180 (2018).

29. Lei, W., Alves, L. G. & Amaral, L. A. N. Forecasting the evolution of fast-changing transportation networks using machine learning. *Nat. Commun.* **13**, 1–12 (2022).

30. Ren, Y., Ercsey-Ravasz, M., Wang, P., González, M. C. & Toroczkai, Z. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. *Nat. Commun.* **5**, 5347 (2014).

31. Anbaroglu, B., Heydecker, B. & Cheng, T. Spatio-temporal clustering for non-recurrent traffic congestion detection on urban road networks. *Transp. Res. Part C: Emerg. Technol.* **48**, 47–65 (2014).

32. Avila, A. M. & Mezić, I. Data-driven analysis and forecasting of highway traffic dynamics. *Nat. Commun.* **11**, 1–16 (2020).

33. Hanna, R., Kreindler, G. & Olken, B. A. Citywide effects of high-occupancy vehicle restrictions: evidence from "three-in-one" in Jakarta. *Science* **357**, 89–93 (2017).

34. Bellocchi, L. & Geroliminis, N. Unraveling reaction-diffusion-like dynamics in urban congestion propagation: Insights from a large-scale road network. *Sci. Rep.* **10**, 1–11 (2020).

35. Yildirimoglu, M., Ramezani, M. & Geroliminis, N. Equilibrium analysis and route guidance in large-scale networks with MFD dynamics. *Transp. Res. Procedia* **9**, 185–204 (2015).

36. Olmos, L. E., Çolak, S., Shafiei, S., Saberi, M. & González, M. C. Macroscopic dynamics and the collapse of urban traffic. *Proc. Natl Acad. Sci.* **115**, 12654–12661 (2018).

37. Ben-Akiva, M., De Palma, A. & Kanaroglou, P. Dynamic model of peak period traffic congestion with elastic arrival rates. *Transp. Sci.* **20**, 164–181 (1986).

38. Huang, H.-J. & Lam, W. H. Modeling and solving the dynamic user equilibrium route and departure time choice problem in network with queues. *Transp. Res. Part B: Methodol.* **36**, 253–273 (2002).

39. Ran, B. & Boyce, D. *Modeling Dynamic Transportation Networks: an Intelligent Transportation System Oriented Approach*. (Springer Science & Business Media, 2012).

40. Kasmalkar, I. G. et al. When floods hit the road: resilience to flood-related traffic disruption in the San Francisco Bay Area and beyond. *Sci. Adv.* **6**, eaba2423 (2020).

41. Tian, L.-J., Yang, H. & Huang, H.-J. Tradable credit schemes for managing bottleneck congestion and modal split with heterogeneous users. *Transp. Res. Part E: Logist. Transp. Rev.* **54**, 1–13 (2013).

42. Daganzo, C. F. Queue spillovers in transportation networks with a route choice. *Transp. Sci.* **32**, 3–11 (1998).

43. Arnott, R., De Palma, A. & Lindsey, R. A structural model of peak-period congestion: a traffic bottleneck with elastic demand. *Am. Econ. Rev.* **83**, 161–179 (1993).

44. Kerner, B. S. & Klenov, S. L. Microscopic theory of spatial-temporal congested traffic patterns at highway bottlenecks. *Phys. Rev. E* **68**, 036130 (2003).

45. Zhang, L. et al. Scale-free resilience of real traffic jams. *Proc. Natl Acad. Sci. USA* **116**, 8673–8678 (2019).

46. Kearney, M. J. Exactly solvable cellular automaton traffic jam model. *Phys. Rev. E* **74**, 061115 (2006).

47. Arnott, R., De Palma, A. & Lindsey, R. Economics of a bottleneck. *J. Urban Econ.* **27**, 111–130 (1990).

48. Daganzo, C. F. & Laval, J. A. Moving bottlenecks: a numerical method that converges in flows. *Transp. Res. Part B: Methodol.* **39**, 855–863 (2005).

49. Newell, G. F. A simplified theory of kinematic waves in highway traffic, part I: general theory. *Transp. Res. Part B: Methodol.* **27**, 281–287 (1993).

50. Arnott, R., De Palma, A. & Lindsey, R. Properties of dynamic traffic equilibrium involving bottlenecks, including a paradox and metering. *Transp. Sci.* **27**, 148–160 (1993).

51. Li, Z.-C., Huang, H.-J. & Yang, H. Fifty years of the bottleneck model: a bibliometric review and future research directions. *Transp. Res. Part B: Methodol.* **139**, 311–342 (2020).

52. Li, D. et al. Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proc. Natl Acad. Sci.* **112**, 669–672 (2015).

53. Hamedmoghadam, H., Jalili, M., Vu, H. L. & Stone, L. Percolation of heterogeneous flows uncovers the bottlenecks of infrastructure networks. *Nat. Commun.* **12**, 1254 (2021).

54. Zeng, G. et al. Switch between critical percolation modes in city traffic dynamics. *Proc. Natl Acad. Sci. USA* **116**, 23–28 (2019).

55. Hamedmoghadam, H., Zheng, N., Li, D. & Vu, H. L. Percolation-based dynamic perimeter control for mitigating congestion propagation in urban road networks. *Transp. Res. Part C Emerg. Technol.* **145**, 103922 (2022).

56. Vickrey, W. S. Congestion theory and transport investment. *Am. Econ. Rev.* **59**, 251–260 (1969).

57. Zeng, G. et al. Multiple metastable network states in urban traffic. *Proc. Natl Acad. Sci. USA* **117**, 17528–17534 (2020).

58. Dakos, V. et al. Slowing down as an early warning signal for abrupt climate change. *Proc. Natl Acad. Sci.* **105**, 14308–14312 (2008).

59. Scheffer, M. Foreseeing tipping points. *Nature* **467**, 411–412 (2010).

60. Scheffer, M. et al. Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).

61. Veraart, A. J. et al. Recovery rates reflect distance to a tipping point in a living system. *Nature* **481**, 357–359 (2012).

62. Al-Madani, H. M. Dynamic vehicular delay comparison between a police-controlled roundabout and a traffic signal. *Transp. Res. Part A: Policy Pract.* **37**, 681–688 (2003).

63. Akbarzadeh, M. & Estrada, E. Communicability geometry captures traffic flows in cities. *Nat. Hum. Behav.* **2**, 645–652 (2018).

64. Lee, M., Barbosa, H., Youn, H., Holme, P. & Ghoshal, G. Morphology of travel routes and the organization of cities. *Nat. Commun.* **8**, 2229 (2017).

65. De Palma, A., Lindsey, R. & Picard, N. Risk aversion, the value of information, and traffic equilibrium. *Transp. Sci.* **46**, 1–26 (2012).

66. Liu, P. & Liu, Y. Optimal information provision at bottleneck equilibrium with risk-averse travelers. *Transp. Res. Rec.* **2672**, 69–78 (2018).

67. Thunig, T., Kühnel, N. & Nagel, K. Adaptive traffic signal control for real-world scenarios in agent-based transport simulations. *Transp. Res. Procedia* **37**, 481–488 (2019).

68. Zhao, J., Li, D., Sanhedrai, H., Cohen, R. & Havlin, S. Spatio-temporal propagation of cascading overload failures in spatially embedded networks. *Nat. Commun.* **7**, 10094 (2016).

69. Çolak, S., Lima, A. & González, M. C. Understanding congested travel in urban areas. *Nat. Commun.* **7**, 10793 (2016).

70. Roughgarden, T. *Selfish Routing and the Price of Anarchy*. (MIT Press, London, 2005).

71. Schröder, M., Storch, D.-M., Marszal, P. & Timme, M. Anomalous supply shortages from dynamic pricing in on-demand mobility. *Nat. Commun.* **11**, 1–8 (2020).

72. Dussutour, A., Fourcassié, V., Helbing, D. & Deneubourg, J.-L. Optimal traffic organization in ants under crowded conditions. *Nature* **428**, 70–73 (2004).

73. Yeung, C. H., Saad, D. & Wong, K. M. From the physics of interacting polymers to optimizing routes on the London Underground. *Proc. Natl Acad. Sci.* **110**, 13717–13722 (2013).

74. Cantarella, G. E. & Cascetta, E. Dynamic processes and equilibrium in transportation networks: towards a unifying theory. *Transp. Sci.* **29**, 305–329 (1995).

75. Vazifeh, M. M., Santi, P., Resta, G., Strogatz, S. H. & Ratti, C. Addressing the minimum fleet problem in on-demand urban mobility. *Nature* **557**, 534–538 (2018).

76. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).

77. Nagatani, T. Self-organized criticality and scaling in lifetime of traffic jams. *J. Phys. Soc. Jpn.* **64**, 31–34 (1995).

78. Petri, G., Expert, P., Jensen, H. J. & Polak, J. W. Entangled communities and spatial synchronization lead to criticality in urban traffic. *Sci. Rep.* **3**, 1–8 (2013).

79. Robertson, D. I. & Bretherton, R. D. Optimizing networks of traffic signals in real time-the SCOOT method. *IEEE Trans. Veh. Technol.* **40**, 11–15 (1991).

80. Martin, R. AV futures or futures with AVs? Bridging sociotechnical imaginaries and a multi-level perspective of autonomous vehicle visualisations in praxis. *Hum. Soc. Sci. Commun.* **8**, 68 (2021).

81. Gazis, D. C. Mathematical theory of automobile traffic. *Science* **157**, 273–281 (1967).

82. Mansourianfar, M. H., Gu, Z. & Saberi, M. Distance-based time-dependent optimal ratio control scheme (TORCS) in congested mixed autonomy networks. *Transp. Res. Part C Emerg. Technol.* **141**, 103760 (2022).

83. De Oliveira, L. B. & Camponogara, E. Multi-agent model predictive control of signaling split in urban traffic networks. *Transp. Res. Part C Emerg. Technol.* **18**, 120–139 (2010).

84. Srinivasan, D., Choy, M. C. & Cheu, R. L. Neural networks for real-time traffic signal control. *IEEE Trans. Intell. Transp. Syst.* **7**, 261–272 (2006).

85. Feng, S. et al. Dense reinforcement learning for safety validation of autonomous vehicles. *Nature* **615**, 620–627 (2023).

86. Guo, Q., Li, L. & Ban, X. J. Urban traffic signal control with connected and automated vehicles: a survey. *Transp. Res. Part C Emerg. Technol.* **101**, 313–334 (2019).

87. Hegyi, A. Model Predictive Control for Integrating Traffic Control Measures. (Netherlands TRAIL Research School, Netherlands, 2004).

88. Ramezani, M., de Lamberterie, N., Skabardonis, A. & Geroliminis, N. A link partitioning approach for real-time control of queue spillbacks on congested arterials. *Transp. B: Transp. Dyn.* **5**, 177–190 (2017).

89. Barmpounakis, E. & Geroliminis, N. On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment. *Transp. Res. Part C: Emerg. Technol.* **111**, 50–71 (2020).

90. Liang, X., Du, X., Wang, G. & Han, Z. A deep reinforcement learning network for traffic light cycle control. *IEEE Trans. Veh. Technol.* **68**, 1243–1253 (2019).

91. Cao, Z., Jiang, S., Zhang, J. & Guo, H. A unified framework for vehicle rerouting and traffic light control to reduce traffic congestion. *IEEE Trans. Intell. Transp. Syst.* **18**, 1958–1973 (2016).

92. Schläpfer, M. et al. The universal visitation law of human mobility. *Nature* **593**, 522–527 (2021).

93. Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E. & Rus, D. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proc. Natl Acad. Sci.* **114**, 462–467 (2017).

Acknowledgements

Author contributions

J.D. and S.H. designed the research. J.D., S.H., H.-J.H., N.S., E.B.L., and D.L. developed the method. J.D. and G.Z. performed the computation. All the authors analyzed the results. J.D., E.B.L., N.S, S.H., H.-J.H., and D.L. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

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