Scaling law of real traffic jams under varying travel demand

Rui Chen, Yuming Lin*, Huan Yan, Jiazhen Liu, Yu Liu and Yong Li

*Correspondence: linyuming9@mail.tsinghua.edu.cn
1Department of Electronic Engineering, Tsinghua University, Haidian District, Beijing, China

Abstract
The escalation of urban traffic congestion has reached a critical extent due to rapid urbanization, capturing considerable attention within urban science and transportation research. Although preceding studies have validated the scale-free distributions in spatio-temporal congestion clusters across cities, the influence of travel demand on that distribution has yet to be explored. Using a unique traffic dataset during the COVID-19 pandemic in Shanghai 2022, we present empirical evidence that travel demand plays a pivotal role in shaping the scaling laws of traffic congestion. We uncover a noteworthy negative linear correlation between the travel demand and the traffic resilience represented by scaling exponents of congestion cluster size and recovery duration. Additionally, we reveal that travel demand broadly dominates the scale of congestion in the form of scaling laws, including the aggregated volume of congestion clusters, the number of congestion clusters, and the number of congested roads. Subsequent micro-level analysis of congestion propagation also unveils that cascade diffusion determines the demand sensitivity of congestion, while other intrinsic components, namely spontaneous generation and dissipation, are rather stable. Our findings of traffic congestion under diverse travel demand can profoundly enrich our understanding of the scale-free nature of traffic congestion and provide insights into internal mechanisms of congestion propagation.

Keywords: Traffic congestion; Travel demand; Scaling law; Complex systems

1 Introduction
Traffic congestion is characterized by slower vehicle speeds, longer trip times and increased queuing of vehicles, often occurring when travel demand exceeds the road capacity [1]. In 2022, the total cost of traffic congestion was over $81 billion in the US and £9.5 billion in the UK [2], making it a severe problem hindering urban development. Studies have shown that under various internal or external disturbances, minor congestion can potentially evolve into large-scale cascading reactions similar to a domino effect [3–5], with unpredictable and severe consequences. In light of this, it is highly crucial to explore factors that affect traffic congestion and understand its intrinsic characteristics, which can contribute significantly to enhancing traffic planning [6, 7] and traffic management [8, 9].
There have been adequate studies that approached the issue of congestion from the perspectives of demand and supply, focusing on factors such as travel demand [10–15], road network supply [16–19], and network topology [20–23]. Other studies have examined how resident travel patterns and transportation mode choices may influence the development of congestion [24–26], or how the socio-economic attributes of cities, like population density [27–29] and land area [30–34], can shape the congestion. Besides, external factors such as climate conditions also play a role in traffic congestion [35–37]. However, these studies primarily employ simulation or machine learning methods, which exhibit limitations in gaining insight into the fundamental principles governing urban traffic congestion.

To investigate further mechanisms governing urban traffic congestion in real-life scenes, more researchers have recently turned to physical models from empirical data. The percolation theory is widely used to model traffic congestion, providing valuable insight through the analysis of critical percolation threshold [1, 38, 39]. The framework of urban scaling law [40] is also promising since the scale-free distribution has been observed in many urban subsystems [41–43], as well as in the congestion percolation transition [44, 45]. The concept reveals the self-similarity of a system, wherein the system exhibits similar behavioral patterns at different spatial or temporal scales [46, 47]. This self-similarity means rules on different spatial regions and time scales follow the same statistic, which has wide-ranging applications in exploring nonlinear behaviors and self-organizing phenomena in complex systems [41, 48–50]. In traffic congestion propagation within road networks, jammed clusters can range from an entire area to a small section of road, lasting from minutes to hours. And the urban scaling law can unveil their inherent nature by studying their basic variables: cluster size for their spatial size and duration for their temporal size [51–54]. One of the critical research by Zhang et al. verified that the distribution of congestion cluster size and duration shows a scale-free property, independent of microscopic details [53]. When the distribution of these metrics follows a power-law pattern, it indicates the presence of similar congestion phenomena at various scales, rather than being isolated incidents [52, 55]. Furthermore, this pattern often reflects the self-organizing nature and nonlinear behavior within complex systems [56, 57]. Therefore, by identifying this power-law distribution, we can gain a better understanding of the dynamic characteristics of traffic networks and how congestion emerges and spreads within the road networks. This, in turn, facilitates more accurate congestion prediction and management, enhances the efficiency of road networks, and supports the development of more effective traffic policies.

However, despite Zhang et al.’s observations [53] claim that the power-law exponents of size and duration distributions are stable on different workdays in different cities, there are still some critical issues that need to be explored. Firstly, the study mentioned a higher exponent during the significantly decreased travel demand but did not further analyze possible influence from demand. Furthermore, the study primarily focused on the overall distribution of congestion and did not delve into the propagation of congestion, which is critical for understanding possible cascading diffusion under strong traffic demand [58]. Based on these considerations, our study aims to fill this gap by focusing primarily on the scale-free property of traffic congestion and its propagation characteristics in different travel demand levels.

Studying the impact of traffic demand on congestion may face challenges from the relative stability of travel demand in the city. Albeit there are some fluctuations in traffic
volumes between weekends, holidays, and weekdays [53], the intensity of these variations is often insignificant. While the long-term variations are usually accompanied by changes in land use and urban morphology [59], making direct comparisons imprudent. The outbreak of COVID-19 and the corresponding lockdown, as a black swan event [60, 61], have provided us with excellent empirical data. During the pandemic of COVID-19, travel demand has undergone tremendous changes in the same urban setting, making it a natural experiment to study the patterns and variations of traffic congestion in different travel demands.

We conducted the research using real traffic data of Shanghai City from March 1st, 2022, to July 1st, 2022, covering an entire period of the pandemic cycle with dramatically decreased and slowly recovered travel demand. Our result shows that the traffic resilience, described by the distribution exponents of congestion cluster size and recovery duration [53], linearly decreases as travel demand increases. Besides the exponents of distributions, travel demand also brings about a significant impact on the scale of congestion, with a sub-linear scaling relation to the aggregate volume and a modified scaling-like relation to the number of congestion clusters. In microscopic consecutive periods, our results suggest that rising demand will lead to a sharp worsening of congestion, demonstrated by the super-linear scaling relation between the number of congested roads and vehicles on the roads. Subsequent analysis demonstrates that cascade diffusion dominates such sharp worsening, while other propagation components like spontaneous generation and dissipation are rather stable. Our results show the critical role of travel demand, enable better prediction and management of urban traffic and subsequently improve traffic efficiency and sustainability.

2 Results
2.1 Scale-free distributions of traffic congestion cluster
We conducted our research using real traffic data from Shanghai covering the pandemic period from March 1st, 2022 to July 1st, 2022. Shanghai is one of the largest cities in China, characterized by an intricate road network and substantial traffic flow. From March 28th to June 1st, 2022, Shanghai implemented a lockdown policy to stop the pandemic [62], including suspension of the public transportation system and strict traffic permit control for private vehicles. The fluctuations in travel demand resulting from the pandemic and policies facilitated the availability of non-routine large-scale traffic data, providing us with real traffic data under varying travel demands.

Considering congestion in urban areas often forms clusters through propagation, we first construct spatiotemporal congestion clusters [53] for subsequent analysis. Specifically, the road network of Shanghai is abstracted into a directed graph where each node represents a road segment, edges represent the adjacency relationships between road segments, and the direction of edges aligns with the direction of traffic flow. Therefore, congestion can be viewed as a spatiotemporal cluster on the graph, where all spatial or temporal connected congested road segments are assigned to the same jammed cluster, constructing multiple spatiotemporal jammed clusters within the road network of Shanghai cross time (see Additional file 1, Sect. 2 for the details). All the red links within the shaded area of Fig. 1A belong to the same cluster, an example of a spatiotemporal jammed cluster. Note that, when a jammed cluster splits into two or more subgraphs at a certain time, all...
nodes within all subgraphs still belong to the same cluster since they are temporally connected. This construction of jammed clusters intuitively reflects the propagation of traffic congestion in both time and space, laying the foundation for subsequent analysis.

The schematic representation of the second-largest jammed cluster in Shanghai on March 1st is depicted in Fig. 1B. The horizontal axis represents time, and the vertical axis depicts the number of congested road segments within this cluster at each time $t$. To illustrate that the congestion is a deduction of traffic capacity and to align with the resilience triangular, the $y$-axis is reversed to put the larger number at a lower location. The recovery duration $T$ of the jammed cluster is defined as the duration between the first congested road segment of this cluster occurring $t_0$ and the last congested segment dissipating $t_1$. Furthermore, the shaded region in Fig. 1B represents the temporal accumulation of congested road segments, defined as the cluster size $S$. The specific jammed cluster gradually increases to approximately 60 and then recovers to zero, leading to an accumulative $S$ that speeds up and then slows down. The temporal evolution of the size $S$ of the three largest jammed clusters on March 1st is shown in Fig. 1C, showing distinct characteristics. The largest jammed cluster in pink persists throughout the day and dissipates completely after 18:00. While the jammed cluster in blue emerges during the morning rush hour, and the jammed cluster in black develops during the evening rush hour.
To further explore the impact of travel demand, we utilize the number of origin-destination (OD) pairs, denoted as $M$, as a representation of daily travel demand (Fig. 2A). It can be observed that following the outbreak of the pandemic, the travel demand in Shanghai experienced a significant decrease of 92%, remained at a lower level during the lockdown, and gradually recovered to a near pre-pandemic level after the policy was lifted. Under such dramatically varying travel demand in a relatively short period, our dataset can reflect the impact of travel demand on the traffic road system.

To further understand the spatial and temporal characteristics of a single jammed cluster, we look into its size $S$ and duration $T$. The $S$ captures how spatially the congestion ranges and the $T$ captures how temporally it lasts. Considering that a jammed cluster contains many stochastic factors, it is more important to examine their distribution on the variation of traffic demand $[44, 45, 53]$. As shown in Fig. 2B,C, we observe significant variations in both size and recovery duration of jammed clusters, which aligns with previous research. We confirm that both cluster size $S$ and recovery duration $T$ follow a power-law distribution, specifically

$$p(S) \sim S^{-\alpha_S} \quad \text{and} \quad p(T) \sim T^{-\alpha_T} (\alpha_S, \alpha_T > 0), \quad (1)$$
with parameters $\alpha_S \in [2.40, 2.71]$ and $\alpha_T \in [3.04, 3.48]$ respectively. The results are very close to the 2.3 and 3.1 in Beijing and Shenzhen reported by Zhang et al. [53].

The ubiquity of scale-free distributions across different days suggests that despite varying demand, traffic jams exhibit the same self-organizing mechanisms. However, it is equally noteworthy that scaling exponents $\alpha_S$ and $\alpha_T$, representing traffic resilience, exhibit significant differences across travel demand. During periods of high travel demand, the scaling exponents are smaller. Whereas, during periods of low travel demand, the scaling exponent of the distribution is larger (Fig. 2B,C). Note that smaller $\alpha_S$ and $\alpha_T$ mean more prevalent larger jammed clusters and longer recovery duration, and the congestion clusters are less vulnerable and harder to deal with. Therefore, it is reasonable to conclude that higher travel demand is related to worse traffic congestion.

### 2.2 Travel demand and scaling exponents of congestion clusters

Since the scaling exponents capture the full picture of congestion, we further investigate the correlation between the scaling exponents and the daily travel demand. As depicted in Fig. 3A,B, a strong negative linear correlation exists between the scaling exponents and the travel demand $M$, which holds true for both cluster size $\alpha_S$ (Fig. 3A) and duration $\alpha_T$ (Fig. 3B):

\[
\alpha_S = 2.7134 - 2.485 \times 10^{-7} M \quad (R^2 = 0.91),
\]

\[
\alpha_T = 3.5078 - 3.844 \times 10^{-7} M \quad (R^2 = 0.94).
\]

These findings indicate that daily travel demand directly influences the magnitude of congestion distributions. Specifically, an increase of one unit in travel demand results in a corresponding decrease of $2.485 \times 10^{-7}$ in the scaling exponent of jammed cluster size. Likewise, with every unit increase in travel demand, we observe a decline by $3.844 \times 10^{-7}$ in the scaling exponent for recovery duration. This means that the power-law exponent is not stable as reported before [53], and the variation can be observed in the case of large changes in travel demand.

Since the travel demand $M$ represents a daily variation, we further investigate its relationship with the total congestion size $S_{\text{total}}$, which is defined as the sum of all congestion cluster size $S$ and represents a cumulative measure of daily congestion. As demonstrated in Fig. 3C, the $S_{\text{total}}$ exhibits a power-law growth with respect to $M$, following the relationship:

\[
S_{\text{total}} = S_0 \cdot M^\lambda,
\]

where $S_0 = 1.00$, $\lambda = 0.78$ ($R^2 = 0.94$, 95% CI = [0.75, 0.81]). The observed sub-linear scaling exponent $\lambda < 1$ suggests the growth rate of congestion is relatively slower than the corresponding increase of demand, implying the possible existence of self-organizing properties within the traffic network system. When travel demand grows, for instance, traffic flow can be distributed throughout various road segments and time periods by changes to routes and timetables, resulting in a comparatively slower growth in congestion.

Furthermore, by leveraging the quantitative relationship between $M$, $\alpha_S$ and $S_{\text{total}}$, we can mathematically derive the relationship between the number of jammed clusters $N$.
and travel demand $M$. Assuming the size of a single cluster is denoted by $S$, the value of summing up all daily values of $S_{\text{total}}$ can be represented as

$$S_{\text{total}} = N \cdot E[S], \quad (4)$$

where $E[S]$ represents the expected value of a single cluster size. Given the power-law probability distribution of $p(S) \sim S^{-\alpha_S}$, we have

$$E[S] = \int_1^\infty S \cdot p(S) \cdot dS = \frac{1}{\alpha_S - 2}, \quad \alpha_S > 2. \quad (5)$$

Given Eq. (2)(3)(4)(5), the relationship between $N$ and $M$ is obtained as

$$N = (\alpha_S - 2) \cdot S_0 \cdot M^\lambda. \quad (6)$$

It combines a power-law component $M^\lambda$ and a linear component $\alpha_S - 2$. From Fig. 3D, it is evident that the observed data aligns closely with the theoretical values ($R^2 = 0.96$).

It can be observed in Fig. 3D that as the travel demand increases, the number of jammed clusters also increases but at a steadily diminishing rate, similar to the sub-linear growth of $S_{\text{total}}$. This result indicates that the increase in travel demand initially leads to rapidly
growing congestion when the road segments reach their capacity limit. However, as travel demand continuously increases, the growth rate of $N$ slows down. One possible explanation is that as more congestion clusters occur, small clusters are more likely to merge and limit the total number of clusters. Moreover, noting that our analysis represents the congestion over a day, the temporal variations of travel demand may also contribute to the result. The growth in travel demand during peak periods may be more pronounced, leading to a sharp increase in $N$. However, residents may correspondingly adjust their travel periods to avoid peak, which results in a slowdown of congestion growth.

2.3 Scaling laws for the propagation of congestion

The linear relationship between demand and scaling exponents implies that travel demand directly influences the micro-mechanism of congestion cluster formation. Therefore, we examined the microscopic level of congestion propagation, aiming to enhance the understanding of this dynamic process.

To examine congestion in successive periods, we track the total number of congested roads, denoted as $C$, at five-minute intervals between 6:00 a.m. and 11:00 p.m. For fine-grained traffic demand, we calculate the total number of vehicles on the road network $V$ corresponding to each time interval, using a moving average of three intervals for noise reduction. Unlike $S_{total}$ and $M$ for daily aggregation, $C$ and $V$ represent the congestion and demand at any given time slot respectively, forming the microscopic perspective of congestion. This facilitates a more detailed observation of the travel demand and congestion over consecutive periods.

Figure 4B illustrates the scatter plot of $C$ versus $V$ on double logarithmic axes, revealing an obvious power-law relation with the exponent $\gamma$ of approximately 2.10 ($R^2 = 0.89, 95\% CI = [2.06, 2.12]$):

$$C = k \cdot V^{\gamma}.$$  \hspace{1cm} (7)

Note that in contrast to the sub-linear scaling of $S_{total}$ and $M$ during the whole day, congested roads $C$ increase with vehicles on road $V$ with a super-linear scaling pattern, which means slight increases in travel demand can rapidly elevate the number of congested roads and exacerbate traffic congestion. It is also interesting that the scaling exponent $\gamma$ falls between 2 and 3 for congested road networks, similar to other complex networks such as social networks [63–65], urban road networks [66, 67], and so on.

To further examine the cascading propagation of congestion, we differentiate $C$ into three distinct components based on the real scenarios of traffic congestion and relevant literature: 1) congestion generated through cascade diffusion, denoted by $F$, which is the accumulation number of congested roads that are connected to pre-existing congestion; 2) congestion generated spontaneously, denoted by $I$, which is the accumulation number of congested roads that arise spontaneously; 3) dissipation, denoted by $R$, which is the accumulation number of roads that were congested but subsequently recovered. The distinction of $C$ into diffusion ($F$), spontaneous congestion ($I$), and recovery ($R$) is based on real scenarios. Typical congestion would start from a spontaneously congested road ($I$) associated with sporadic factors such as traffic accidents and construction projects [1, 68, 69]. Then, the congestion would diffuse ($F$) with traffic flow and road network topology, which may persist for extended periods during peak hours [23, 70]. Finally, the congested roads
Figure 4. Power law relationship of road congestion on March 1st in Shanghai. (A) A demonstration of cascading progress in the traffic network. Color represents the state of road segments, including spontaneous congestion (pink), cascading diffusion (red), existing congestion (orange), and dissipation (blue). (B) Power-law relationship between the total number of congested roads $C$ and the travel demand $V$ (represented by the number of vehicles recorded) in consecutive periods. (C) Illustration of the scaling relationship between the number of congested roads generated by diffusion, denoted as $f$ and the total number of congested roads $C$. (D) Illustration of the scaling relationship between the number of spontaneously generated congested roads $I$ and the total number of congested roads $C$. (E) Illustration of the scaling relationship between the number of dissipating roads $R$ and the total number of congested roads $C$.

Traffic congestion exhibits such dynamic spatiotemporal patterns, and analyzing the temporal and spatial changes of different types of congestion helps provide a more comprehensive understanding of congestion patterns. This, in turn, allows for a more precise determination of measures needed to alleviate congestion. Figure 4A illustrates a sample of cascading progress and the corresponding components. Naturally, the sum of these three parts equals $C$:

$$C = F + I - R. \quad (8)$$

Since the diffusion component $F$ is an extension from pre-existing congestion and inherently has an incremental property, we reformulate it using the newly added diffusive congested roads, denoted by $f$. The $f$ directly captures the propagation rate of congestion at a given time, offering more information about congestion [4, 71–73].
be reformulated as:

\[ C = \sum f + I - R, \] (9)

We then examined the relationship between \( C \) and its three components \( f, I, R \) from our empirical data (Fig. 4C,D,E). The results show that all three components show a power law to the total number of congested roads on any given day.

\[ f = k_I C^{\beta_f}, \quad I = k_I C^{\beta_I}, \quad R = k_R C^{\beta_R}. \] (10)

Taking data on March 1st as an example, results shows that \( f \) obeys sub-linear growth to \( C \) with \( \beta_f = 0.514 \pm 0.011, (R^2 = 0.92) \). While \( I \) and \( R \) is approximately linear to \( C \) with \( \beta_I = 0.998 \pm 0.003, (R^2 = 0.88) \) and \( \beta_R = 0.997 \pm 0.007, (R^2 = 0.89) \). The result is stable over different dates and travel demand (refer to Additional File 1, Sect. 5 for more information).

The results show that both \( I \) and \( R \) demonstrate almost linear relationships with \( C \), meaning they keep proportional to \( C \) despite dates and demand. For spontaneously generated component \( I \) that serves as seeds of congestion clusters, this proportion can be related to relatively stable road capacity and recurring congestion [74]. Regarding the dissipation component \( R \), the constant proportion suggests the dissipation is independent of present congestion and can be simply described by a dissipation rate [3]. As for the sub-linear scaling of diffusion \( f \), one plausible conjecture is that the network topology may impose constraints besides the intrinsic diffusion dynamics, similar to what has been observed in urban amenities [42, 67]. Further investigation into the topology structure may provide more insight into it.

Based on the existing results, we examine how the three components determine the power-law relationship of \( C \) and \( V \) by determining key coefficients \( k \) and \( \gamma \). Starting with the difference in congested roads in different demand scenarios, we have a differential form of the congestion propagation:

\[ \frac{dC}{dV} = f + \frac{dI}{dV} - \frac{dR}{dV}. \] (11)

While Eq. (11) cannot be solved for an analytic expression, an approximation can still be derived under the empirical data showing \( \beta_f \approx 1 \) and \( \beta_R \approx 1 \). Based on Eq. (10), (11), we can obtain

\[ k \approx \frac{k_f (1 - \beta_f)}{1 - k_I + k_R}, \quad \gamma \approx \frac{1}{1 - \beta_f}, \] (12)

with both \( k \) and \( \gamma \) are determined by the coefficients and exponents of components (See Additional File 1 Sect. 4 for more information).

Figure 5A,B,C show that the values of \( \beta_f, \beta_I, \) and \( \beta_R \) exhibit relatively stable distributions in all observed days \( (\beta_f = 0.515 \pm 0.002, \beta_I = 0.996 \pm 0.002, \beta_R = 0.996 \pm 0.002). \)

Considering the drastic changes in daily demand \( M \), the observation suggests a universal connection in congestion components over consecutive time. The observed empirical data of \( k \) and \( \gamma \) is illustrated in Fig. 5D,E, alongside their corresponding theoretical values obtained through Eq. (12). Notably, Eq. (12) indicates that both \( k \) and \( \gamma \) are solely correlated with \( \beta_f \), demonstrating the dominant role of cascade diffusion in the growth of
congestion. The reciprocal-like relationship between $\gamma$ and $\beta_f$ demonstrates how a sub-linear growth of cascading diffusion leads to super-linear growth of total congested roads. These scaling laws enable us to comprehend the micro-scale mechanism underlying congestion across different travel demand, making it conceivable that universally applicable strategies for congestion management may exist.

3 Discussion
Traffic demand has always been one of the key factors when examining traffic congestion [10–15]. Our primary contribution lies in discovering a universal scaling law that constrains the spatiotemporal scale and propagation patterns of congestion in cities under varying travel demand. In this research, we identify the crucial role of travel demand in the scaling laws of urban traffic congestion. Based on the real traffic data during the COVID-19 pandemic in Shanghai, we extend our research on the power-law scaling of urban traffic congestion and reveal the relationship between traffic congestion and travel demand at the macroscopic congestion cluster and the microscopic congestion propagation.

Compared to the study by Zhang et al. [53], we have utilized more recent traffic data to verify the power-law distribution in scenarios with varied traffic demand. Our results illustrate that changes in traffic demand directly affect traffic resilience, with a negative linear relationship that has not been reported before. Such results suggest that congestion can be predicted and intervened by managing travel demand, and proper mitigation resources and policies should be designed for different demand levels. Furthermore, the
sub-linear power-law relationship between traffic demand and the total congestion size demonstrates resilience within the transportation system [75]. More efficient and frugal methods and policies could be developed under the guidance of such resilience. It is also interesting that the scaling exponent $\gamma$, between the number of congested roads and the number of vehicles, falls between 2 and 3. One of the possible explanations is the preferential attachment mechanism commonly used to model a dynamic network. A small fraction of congestion bottlenecks within the traffic network attracted newly congested segments driven by demand, establishing the super-linear growth pattern. The initial bottlenecks and subsequent attachments form the cascading propagation. Although we have not conducted any direct analysis to support such a hypothesis, we believe that future modeling based on it may provide a detailed analysis of the internal mechanisms.

At the micro level, the super-linear growth pattern of congested roads highlights the potential of significant congestion arising from minor increases in travel demand, in line with the cascading dynamics of congestion [58]. Using the cascading failure framework [4], we discover a stable proportion of spontaneous generation and dissipation across dates and demand levels, implying it can be hard to ease congestion by speeding dissipation or reducing spontaneous generation. However, the sub-linear scaling relation between diffusion and congested roads indicates that diffusion increases relatively slower, and blocking cascading diffusion through appropriate early intervention could be an essential path to avoid cascading diffusion and change congestion distribution. Our result is of great theoretical interest since it can link the cascade diffusion to the scaling law framework and illustrate that the cascade diffusion rate is decisive to the power-law of congestion.

The two distinct power-law growth exponents we have observed at the cluster level and road level hold significant implications. Although total congestion cluster size is sub-linearly scaling with daily travel demand, the congested road number is a super-linear scaling to short-term travel demand. Similar paradoxical scaling laws have been found in other areas of urban research [76]. For instance, there is a sub-linear or linear relationship between building area and population in various administrative districts of Shanghai. However, a super-linear scaling law emerges between building area and population among more administrative districts over time [77]. Studies have also revealed that different scaling relationships can arise due to variations in the definition of cities [78–80] or differences among groups [81]. This difference means we should consider targeted measurements and management on different space or time scales.

Specifically, the sub-linear growth at the daily clusters suggests that people have more flexibility and adaptability to avoid peak hours by choosing alternative transportation modes or changing their schedules, resulting in a lower growth rate than the travel demand. Therefore, it is essential to consider people's flexibility and adaptability for long-term planning to alleviate traffic pressure. Persistent measures like encouraging alternative transportation choices, optimizing public transportation systems, and implementing differentiated road tolls help alleviate overall congestion pressure and ensure efficient, safe, and sustainable urban transportation development. However, the super-linear growth within consecutive time intervals is related to the cascade diffusion, leading to the rapid emergence of road bottlenecks and congestion spread. Hence, short-term response and quick interventions are necessary, including monitoring fluctuations in traffic flow on highly connected roads and promptly identifying potential bottlenecks. Such dis-
Distinct managing strategies are based on the different characteristics in different time scales, leading to distinct traffic optimization directions.

Regarding the observed power-law growth pattern of congestion with traffic demand, our data only provide observations from one city focusing on the demand changes. Due to constraints related to data availability and research resources, our research exclusively utilized traffic data from Shanghai during the COVID-19 pandemic. Unlike minor fluctuations in traffic demand under normal conditions, this unique dataset provided us with an opportunity to investigate congestion patterns throughout the entire process of a significant decrease in traffic demand to recovery, thereby establishing a link between traffic demand and congestion. With more urban data, we can verify if similar congestion patterns exist in other cities in future work. Besides, although we utilized data including the period before the lockdown (March 1st to March 28th) and after the lockdown (June 1st to July 1st), the COVID-19 pandemic may have drastically altered our cities and making the comparison using multi-year historical data an important aspect. Furthermore, the authorities implemented traffic control measures like checkpoints during the lockdown, which could potentially impact the occurrence and propagation of traffic congestion. Although most lockdown policies are mainly achieved by reducing travel demand, which is covered in our study, the subsequent effect on other aspects, like travel mode preference [82], should be carefully examined in the future. Other factors, such as road network topology and traffic flow heterogeneity, can also be included to describe the congestion dynamics. Unexpected weather conditions, road incidents, ongoing construction, and road capacity can all affect the relationship between travel demand and traffic congestion. In future research, our analytical framework can be extended to encompass a broader range of urban contexts, enabling its application to multiple cities. It is promising that big traffic data will continue to enhance our understanding of traffic congestion and contribute to effective strategies for managing transportation systems.

4 Materials and methods
4.1 Traffic dataset
Our research is primarily based on travel Origin-Destination (OD) data and road congestion index data (0 for congested, and 1 for normal traffic flow) with a resolution of 5 min in Shanghai. The dataset covers a time period of 123 days from March 1st to June 1st, 2022, collected through floating car records in Shanghai. The data are sourced from a large map service provider. In addition, we have collected topological data on the road network of Shanghai from the OpenStreetMap [83], which includes the adjacency relations among major roads in Shanghai. The dataset contains over 32,000 road segments, along with the lists of outbound and inbound roads for each segment.

4.2 Definition of spatiotemporal jammed clusters
Based on the topological information of the road network in Shanghai, we define a spatiotemporal network $G_T$. $G$ is a directed connected graph, representing a snapshot of the graph at time $T$. And $T$ is a period comprising a set of consecutive time slots. The edges in $E$ represent road segments and the nodes in $V$ represent the adjacencies of the road segments. The direction of edges is the same as the direction of congestion propagation, and the opposite direction of traffic flow. In the first snapshot, denoted as $t_0$, all connected road segments form the initial jammed cluster. For each subsequent snapshot $t$, the newly
Table 1 List of Symbols and Description

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>Cluster size</td>
</tr>
<tr>
<td>$T$</td>
<td>Recovery Duration</td>
</tr>
<tr>
<td>$M$</td>
<td>Total daily travel demand</td>
</tr>
<tr>
<td>$N$</td>
<td>Daily number of jammed clusters</td>
</tr>
<tr>
<td>$C$</td>
<td>The total number of congested roads</td>
</tr>
<tr>
<td>$k$</td>
<td>Normalization constant of $C$</td>
</tr>
<tr>
<td>$f$</td>
<td>Number of newly-added diffusive congested roads</td>
</tr>
<tr>
<td>$I$</td>
<td>Number of spontaneously generated congested roads</td>
</tr>
<tr>
<td>$R$</td>
<td>Number of dissipating roads</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_S$</td>
<td>Scaling exponent of $S$</td>
</tr>
<tr>
<td>$\alpha_T$</td>
<td>Scaling exponent of $T$</td>
</tr>
<tr>
<td>$S_{total}$</td>
<td>Total number of cluster size</td>
</tr>
<tr>
<td>$V$</td>
<td>Travel demand represented by vehicles on the road</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Scaling exponent of $C$</td>
</tr>
<tr>
<td>$\beta_f$</td>
<td>Scaling exponent of $f$</td>
</tr>
<tr>
<td>$\beta_I$</td>
<td>Scaling exponent of $I$</td>
</tr>
<tr>
<td>$\beta_R$</td>
<td>Scaling exponent of $R$</td>
</tr>
</tbody>
</table>

Congested road segments connected to the previous snapshot are grouped into the same jammed cluster. Within the time period from $t_0$ to $t_n$, congested and connected road segments are included in the same jammed cluster. Therefore, the jammed cluster at snapshot $t$ represents the spatio-temporal distribution of road congestion at time $t$. It should be noted that if two jammed clusters are connected during the diffusion process, they are considered as the same jammed cluster. Ultimately, multiple jammed clusters are obtained within the time period from $t_0$ to $t_n$, representing the distribution of congestion throughout the day (refer to Additional File 1, Sect. 2 for detailed definition). The full list of symbols is shown in Table 1.

4.3 Parameter fitting methods

The fitting of the power-law distribution is accomplished through the powerlaw package [84] in Python. We used the ordinary least square (OLS) method to fit the data and estimated parameters of the power function as previous literature [53, 85]. The power function is defined as $y = ax^b$, where $a$ and $b$ are parameters, $x$ is the independent variable, and $y$ is the dependent variable. The principle of fitting involves minimizing the error between the power function and real data utilizing ordinary least squares estimation.

$$\min_{a,b} \sum_i (\log(y_i) - b \log(x_i) - \log(a))^2. \quad (13)$$

The selection of OLS method is grounded in the minimal variance and absence of heterogeneous fluctuations in our dataset, where OLS reliably delivers reasonable and easily interpretable outcomes. However, in scenarios characterized by substantial variance and heteroscedasticity, alternative methods such as probabilistic models [86] ought to be explored. Please refer to Additional File 1, Sect. 3 for more details.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1140/epjds/s13688-024-00471-4.

Additional file 1. We report additional details on methods and results. (PDF 204 kB)

Acknowledgements

Not applicable.

Author contributions

R. Chen, Y. Lin, H. Yan and Y. Li conceived the idea and designed the experiments; R. Chen and Y. Lin analyzed the data and performed the experiments; R. Chen, Y. Lin and Y. Li wrote the paper. All authors reviewed and edited the paper.
Funding
Supported by the National Natural Science Foundation of China (Grant No. 62302261).

Data availability
The data that support the findings of this study are available from Amap-Alibaba Group but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Amap-Alibaba Group.

Declarations
Competing interests
The authors declare that they have no competing interests.

Received: 12 September 2023   Accepted: 3 April 2024   Published online: 11 April 2024

References

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