

Bhardwaj, Ankit; Iyer, Shiva R.; Ramesh, Sriram; White, Jerome; Subramanian, Lakshminarayanan

Article

Understanding sudden traffic jams: From emergence to impact

Development Engineering

Provided in Cooperation with:

Elsevier

Suggested Citation: Bhardwaj, Ankit; Iyer, Shiva R.; Ramesh, Sriram; White, Jerome; Subramanian, Lakshminarayanan (2023) : Understanding sudden traffic jams: From emergence to impact, Development Engineering, ISSN 2352-7285, Elsevier, Amsterdam, Vol. 8, pp. 1-15, <https://doi.org/10.1016/j.deveng.2022.100105>

This Version is available at:

<https://hdl.handle.net/10419/299119>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>



Understanding sudden traffic jams: From emergence to impact

Ankit Bhardwaj^{a,*}, Shiva R. Iyer^{a,1}, Sriram Ramesh^a, Jerome White^{b,2},
Lakshminarayanan Subramanian^a

^a Department of Computer Science, New York University, New York, NY, USA

^b Wadhvani AI, Mumbai, Maharashtra, India

ARTICLE INFO

Keywords:

Road traffic
Traffic jams
Traffic curve

ABSTRACT

Road traffic jams are a major problem in most cities of the world, resulting in massive delays, increased fuel wastage, and monetary and productivity losses. Unlike conventional computer networks, which experience congestion due to excessive traffic, road transportation networks can experience traffic jams over prolonged periods due to traffic bursts over short time scales that push the traffic density beyond a threshold *jam density*. We observe that the emergence of such jams can happen over a very short duration, hence we term them as *sudden traffic jams*. We provide a formalism for understanding the phenomena of sudden traffic jams and show evidence of its existence using loop detector data from New York City. Further, we show the signature of sudden jams when observed at hourly resolution. We also provide a method to compute the traffic curve in a situation where we do not have access to fine-grained flow and density information. With this method, using only hourly speed data from Uber, we compute traffic curves for the road segments in Nairobi, São Paulo, and New York City, which is, by our knowledge, the first attempt to do so for signalized road networks. Running our analysis on the Uber movement speed data for the three cities, we show numerous instances of jams that last several hours, and sometimes as long as 2–3 days. Empirically, we find that Nairobi experiences 3x the mean jam time per road segment as compared to São Paulo and New York City. Based on key development metrics, we find that the ratio of traffic load per road segment for São Paulo, New York City, and Nairobi is approximately 1:2:3. We propose that chaotic driving patterns and traffic mismanagement in the developing world cities lead to tighter traffic curves, more intense jams and overall lower road capacity utilization, which explains the observed data. We posit that the problem of traffic congestion in developing countries cannot be solved entirely by building new infrastructure, but also requires smart management of existing road infrastructure.

1. Introduction

Poor road traffic management can trigger extended periods of traffic congestion, as is witnessed in most parts of the world. As per Texas Transportation Institute's 2009 Mobility report ([Texas Transportation Institute](#)), congestion in the US has increased substantially over the last 25 years with massive amounts of losses pertaining to time, fuel, and money. In the top 10 cities with the worst levels of congestion in the world, the average number of hours wasted per commuter per year is over 150 h ([Friedman, 2020](#)). When the number of hours wasted exceeds about 35 h per year, it is observed to affect the economy negatively ([Badger, 2013](#)). This kind of prolonged traffic congestion persists in many large urban cities ([TomTom International B.V.](#)). Especially in developing regions, with poorly managed road networks and freeways,

this remains an important barrier to economic development in these regions. Reducing traffic jams improves the quality of life also in the form of improved air quality. A report on the air quality study in Delhi attributes nearly 20% of PM concentration in the air to traffic ([Sharma and Dikshit, 2016](#)).

Traffic congestion can be both good and bad ([Badger, 2013](#)). A certain level of congestion, especially in dense urban regions, may indicate economic prosperity and thriving economic development, as in many major cities. However, commuters wasting away their time in jams that last for hours, such as on freeways, is an example of bad congestion. We explore the issue of such bad congestion occurrences that happen mostly due to a sudden burst in traffic density. Contrary to the conventional belief that traffic congestion is triggered

* Corresponding author.

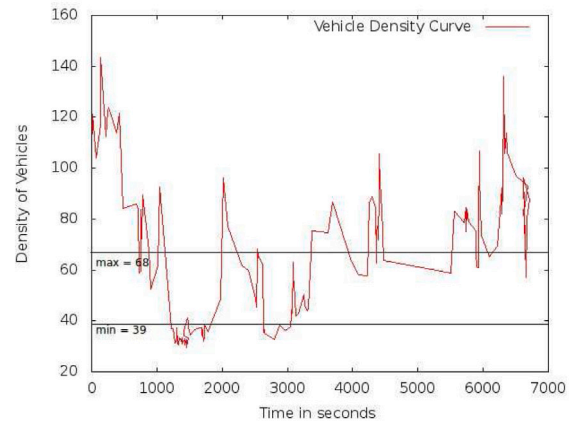
E-mail addresses: bhardwaj.ankit@nyu.edu (A. Bhardwaj), shiva.iyer@nyu.edu (S.R. Iyer), sriram.ramesh@nyu.edu (S. Ramesh), lakshmi@cs.nyu.edu (L. Subramanian).

¹ Equal Contribution.

² Part of this work was done when Dr. Jerome White was affiliated with New York University.



(a) Still from a CCTV camera from the NYC Department of Transportation showing congestion at the end of the Williamsburg bridge on the Manhattan side



(b) Traffic density on the link shown in the image as a function of time

Fig. 1. Traffic on the Williamsburg bridge in New York City.

due to excessive traffic, traffic jams for elongated time periods, such as several hours, can actually be triggered by small traffic bursts over small timescales (Jain et al., 2012; Kerner and Rehborn, 1996). The underlying cause of the traffic jam is not due to the lack of road capacity, but due to a “spiraling effect” triggered by a small burst that pushes the road traffic network to a low-capacity equilibrium point. This equilibrium point is highly stable and the only way to recover from this is to dramatically reduce the input flow into the traffic network and drain the congested network. This phenomenon occurs because traffic links exhibit a traffic curve behavior where the capacity of a link is variable dependent on the traffic density on any link; any input flow beyond the optimal operational rate over a short time that triggers the density beyond a critical threshold automatically triggers a spiraling effect resulting in a traffic jam. We refer to any jam caused this way as a *sudden traffic jam*. It should be noted that this spiraling phenomenon is different from the disturbance propagation phenomenon that results in phantom jams (Treiterer and Taylor, 1966; Treiterer and Myers, 1974) as this is a feature of the traffic curve rather than microscopic perturbations emerging spontaneously.

Traffic collapse results when the traffic density on a link exceeds a certain threshold. The operational free-flow exit rate of the link, which determines how quickly the link is drained, varies with the traffic density (May, 1990). Each traffic link reaches an optimal capacity at a corresponding optimal operating density, beyond which the exit rate rapidly drops. In New York City, consider the exit point of the Williamsburg bridge on the Manhattan side (Fig. 1). A traffic light immediately follows the bridge, and numerous vehicles are regularly stuck there for several minutes, especially during the morning and evening rush hours. The graph shown in Fig. 1(b) is a plot of the traffic congestion at the exit point of the Williamsburg bridge. The y -axis is the measure of the vehicle density, or the traffic density, on the road. The x -axis represents the time in seconds after 14:50:40 when the traffic was observed. By manual inspection of the camera feed, we are able to determine the *min* and *max* densities as well. The *max* density is a density value above which there is complete congestion, whereas the *min* is a value below which there is no congestion at all. In this particular example, the *min* and *max* values were determined to be 68 and 39. The congestion can be mitigated if the input rate of the vehicles into the bridge is controlled before it reaches the tipping point for congestion collapse.

Our contributions in this paper are the following:

- We introduce the concept of sudden traffic jams, provide a formal definition, and illustrate the signature of such jams on speed data averaged at hourly resolutions using loop detector data from New

York City (New York City Department of Transport). Based on this hourly signature, we propose a heuristic definition of jams (called slowdown jams) for hourly resolution speed data. Using hourly resolution Uber Movements speed data (Uber Technologies), we compute the occurrences and statistics of slowdown jams in three cities — Nairobi, São Paulo, and New York City.

- We also provide a method to compute the traffic curve in a situation where we do not have access to fine-grained flow and density information. Using this method, we compute traffic curves for a large number of segments in each of the three cities. To our knowledge, the concept of traffic curves for signalized network conditions was a hypothetical assumption that has not been validated and this is the first attempt to do so.
- Using the derived traffic curves and a simple definition of a traffic jam, we have been able to show how a traffic jam can quickly emerge within a short time period and we also show that once a jam is hit, it often takes a long time to recover from a jam. We have validated these points in our analyses using multiple datasets.

The larger implication of our findings is very significant for development engineering. Developing countries invest large amounts to upgrade road networks in urban cities to address traffic congestion. We picked two cities in developing countries (Kenya and Brazil) and one developed and highly industrialized city (New York City). We show with the findings of our analysis that the ratio of mean time spent in jam per road segment for São Paulo, New York, and Nairobi is approximately 1:1:3. Using key development metrics, we also show that the corresponding ratio of traffic load per road segment is approximately 1:2:3. Thus, we infer that developing world cities like Nairobi and São Paulo have lower road capacity utilization (almost 0.5x) compared to a developed world city like New York City. This paper shows that addressing the traffic congestion problem for urban centers in developing countries is not just about building more roads but actually more about careful traffic management.

We build specifically on one piece of prior work (Jain et al., 2012) which provided a hypothesis for what triggers traffic congestion in developing countries using the concept of traffic curves. The concept of traffic curves, which are very popular in highway traffic networks, had never been used in the context of signalized networks to understand traffic flow. In 2012, there was very limited traffic flow data, especially in the context of developing regions that could really help with the sudden traffic congestion phenomenon. It was hypothesized in the paper by Jain et al. that the sudden traffic congestion phenomenon was a particular problem at several urban choke points in developing

countries, especially with the lack of good traffic management and chaotic driving with high densities of car packing. Once we hit a traffic jam triggered by a sudden burst, it is very hard to recover from it for several hours unless inbound traffic is carefully managed.

The rest of the paper is structured as follows. First, we present relevant literature on traffic jams. Then we provide definitions and formalism of traffic curve and sudden traffic jams, followed by illustrations of sudden jams in the real world. Then we provide the characterization of observed sudden jams for three different cities — New York City, Nairobi and São Paulo. Finally, we conclude the paper with our takeaways.

1.1. Prior art

Modeling of traffic flow has been a subject of study for several decades, since the 1950s (Orosz et al., 2010). There are two primary ways to model traffic flow — macroscopic (continuum) models and microscopic (car-following) models. In the former, a relationship between traffic density and speed is established in the form of a partial differential equation. This allows us to define jams precisely as “stop-and-go” events. However, this requires high-quality data on the traffic density, whereas in practice, most available data from existing instrumentation in cities or probe vehicles or other mobile sources are either travel time data or travel speed data. The latter is even more difficult in practice owing to the large number of parameters required for the model. Zhao et al. (2016) define the concept of resilience in traffic flows to understand how quickly we can recover from jams in the real world. Stathopoulos and Karlaftis (2002) model the amount of time taken for jams to drain as a function of the jam duration. They show that the log-logistic functional form, similar to the form of the traffic curve discussed in Section 3.1, is the best approximation. They also find that if a jam lasts beyond a threshold value of around 20 min, it is most likely caused by external factors and will last a long time. Knorr et al. (2012) propose a strategy to prevent the occurrence of traffic jams by enabling driver-to-driver communications, thus informing drivers of impending congestion ahead. The drivers then slow down and maintain a greater inter-car distance. Through simulations, they report that penetration rates of 10% or less in a city can have a significant influence on traffic flow.

Short-term traffic prediction and forecasting is a related topic that has a long history in the academic community. The challenge is to identify traffic flow patterns at some point in the future; commonly between one and five minutes, but generally between one minute and an hour. Much of the foundational work in short-term traffic comes from models based on time series analysis; Kalman filtering and the ARIMA-family of models, for example. Recent work has taken a machine learning approach, modeling the regression using support vectors, neural networks (Iyer et al., 2020), graphical models (Hu et al., 2016), and pattern matching. For summaries, see Vlahogianni et al. (2014) and Oh et al. (2015).

When classifying traffic state, researchers are essentially trying to bin a snapshot of road condition into a particular state, sometimes known as a *level of service*. Levels of service are sometimes established using well-known standards, such as the Transportation Research Board’s Highway Capacity Manual, but are sometimes more subjective, using an array of human judgment. Porikli and Li (2004) and Lozano et al. (2009), for example, analyze images from traffic cameras to determine roadway conditions; Sen et al. (2013) use video feeds to accurately estimate traffic speed and density; Roy et al. (2011) use strategically placed Wi-Fi transmitters to monitor traffic state. Rather than determine traffic state specifically, Yang et al. (2014) use loop detector data to detect changes in flow. What is common across such work, however, is the need to obtain volume data from the underlying data set. That is, an estimate of the density of traffic at a given time. In this work, such information is not provided, nor can it be inferred;

thus, existing techniques from the “density” community are not directly applicable.

One alternative is to use speed information to determine the traffic state. Doing so requires a different process for establishing levels of service, as what speed means to traffic jams is relative to a notion of free flow at a given point. In the absence of this information, researchers either have to establish a meaning of speed-versus-jam using human judgment (Pattara-Atikom et al., 2006), or use novel techniques to create their own (Yoon et al., 2007). Irrespective of the nature of the data — whether it contains speed, volume, or a combination of both — existing work has largely focused on highway data. A notable exception being Yoon et al. (2007), who consider a suburban stretch containing traffic signals.

2. Materials

2.1. New York City department of transportation data

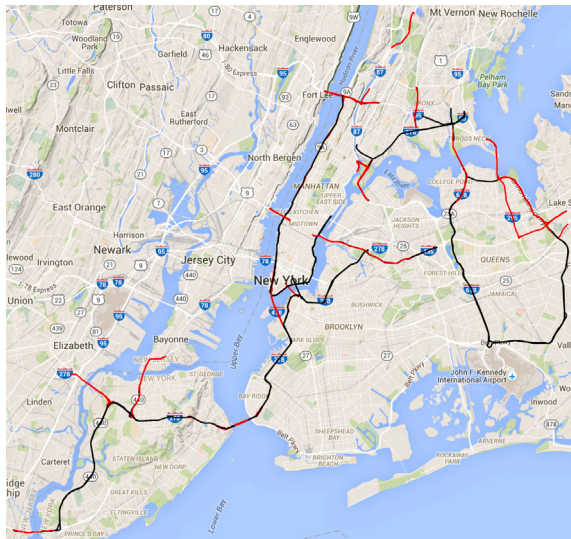
This data comes from the City of New York Department of Transportation (NYCDOT). The department provides a website (New York City Department of Transport) that publishes traffic information for various traffic segments throughout the five boroughs of New York City; there are currently 153 such segments. Each segment has a name, a unique identifier, and location data. The name is a string describing the segment as most resident travelers would, “FDR, north 25th at 63rd street,” for example. The location data consists of a polyline, suitable for understanding the geography of the segment and placing it on a map.

Along with basic information, measurements come with a timestamp and two types of speed data: the average vehicular speed over the segment, and the average time required to traverse the segment. In theory, each signal is continuous—its timestamp representing a snapshot. In practice, and due to limitations in polling, the information is updated each minute. Thus, although the timestamp is at the granularity of seconds, it can be resampled to the granularity of minutes without loss of generality.

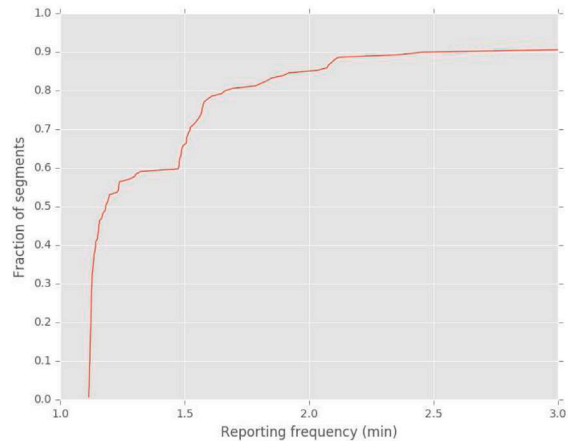
However, the reporting reliability at each segment varies, so not all segments have reports for each minute. Further, periodic system downtime, on both the reporting and collection end, prevented complete continuous polling. For a section of the data collected between November 2014 through April 2016, on average, a segment consists of approximately 513,000 measurements instead of the expected 740,000 measurements. Inter-polling statistics — the average amount of time between traffic measurements — is one way of describing a segments’ reliability. For the above-mentioned section of the data, the average inter-polling time across all segments was approximately 7.85 minutes, however that number is dominated by a few very unreliable segments as the standard deviation is almost 63 minutes. In the best case, there is a measurement every 66.88 seconds; in the worst, every 12.72 h. Fig. 2(b) presents a distribution of average reporting times across the signals.

2.2. Uber movements speed data for Nairobi, New York City and São Paulo

We analyzed the Uber Movements data for Nairobi city in Kenya, New York city in the United States of America, and São Paulo in Brazil to study sudden traffic jams (Uber Technologies). The data has hourly average speeds in all road segments in the city. Every segment or way in the dataset is defined as the stretch of road between two junctions. Each road segment has multiple nodes which are used to characterize the stops along the way, and the dataset consists of average speed values every hour for every pair of source and destination nodes. To calculate the average speed in a segment, all the individual speeds across all consecutive pairs of source–destination nodes in the segment were averaged. Each segment corresponds to a *way* in OpenStreetMap (OpenStreetMap contributors, 2017), denoted by a *way_id*, and therefore we are able to look up these segments on a map.



(a) Measured road segments in across the New York City area. Segments in black are those with reporting times of 75 seconds or less. (use color in print)



(b) Distribution of reporting times.

Fig. 2. Available segments in the loop detector data feed from NYC DoT from November 2014 to April 2016. The left plot shows the segments on the map, color-coded by the reporting frequency, and the right side shows the distribution of reporting frequencies.

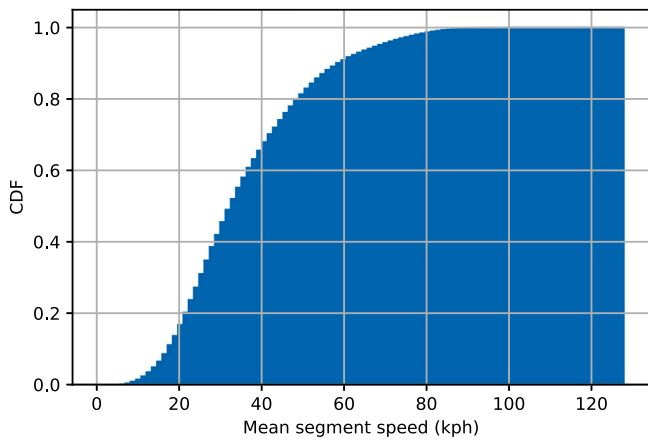


Fig. 3. Cumulative distribution of hourly traffic speeds across all the segments in the city of Nairobi.

We analyzed the dataset from January 2018 to March 2020. The cities of Nairobi, New York, and São Paulo have respectively 4949; 35,602, and 89,121 segments respectively. Fig. 3 shows the cumulative distribution of mean segment speeds across all 4949 segments for the city of Nairobi. From the data of Nairobi city, we notice that only 12% of the average speed exceeds the speed limit of 50 km/h (Hodge) and the average length of such a road segment is about 80 m long (Nesbitt and Dara-Abrams, 2017). We see similar behavior of speed distributions for NYC and São Paulo. After sampling the mean speed time series of each segment using a Poisson distribution with $\lambda = 8$ (for removing temporal correlations), we put an assertion on the segments to have at least 20 sampled points for generating reliable speed distribution curves. After applying the above condition, we are left with 3063; 25,371, and 53,658 segments for Nairobi, New York City, and São Paulo respectively. In our downstream analysis, we take these segments and compute traffic jams based on an hourly-resolution definition.

While one would like to validate the Uber Movements data using other well-known sources of traffic data like Google Maps, etc, this

exercise presents several hurdles in practice. We investigated the possibility of “ground truthing” the Uber movements dataset using google maps. The most notable obstacle in this attempt is that Google maps don’t provide access to historical data. The Uber movement data for all cities ranges from 2018–2020, and comparing it with the google maps data at the time of writing this article requires assuming that the traffic patterns have remained constant over the last 2–3 years. We consider this assumption questionable since 2–3 years is sufficient time for the road network and corresponding traffic conditions to change, especially in the context of developing countries. Another challenge is that Google distance matrix API (google maps API) only gives travel-time estimates between origin and destination points and does not give average speed per road segment. The Uber movement data also provides aggregated and bracketed travel time estimates, but the travel time data has not been used much in our analysis. While one could validate the Uber travel time data using google maps after taking the earlier assumption, one would have to further assume that the correctness of Uber travel time data implies the correctness of Uber speed data. Similar challenges emerge when other sources of data are considered for validation.

In fact, there has been a prior study validating Uber travel time data against Google maps travel time predictions (Wu, 2018), that makes the constant traffic pattern assumption, albeit over a smaller time duration of 6 months for the city of Sydney, Australia. Wu (2018) notes that travel time from Google and from Uber is generally similar but the observations from the Uber data are systematically lower than Google’s predictions. Differences in the data collection method (actual trip time for Uber vs prediction for Google), as well as the corresponding subtle intents and objectives (like faster trips for Uber vs reliable estimation for Google), have been attributed as possible causes of this behavior. Wu (2018) indicates that the Uber travel time data is likely to be correct and potentially valid. Thus, by extension, we assume the correctness of the Uber speed data in all our downstream analyses.

3. Theory and empirical approximations

3.1. Traffic curve and traffic collapse

A transportation network is a collection of segments or links, where a segment/link consists of a set of geographic coordinates representing

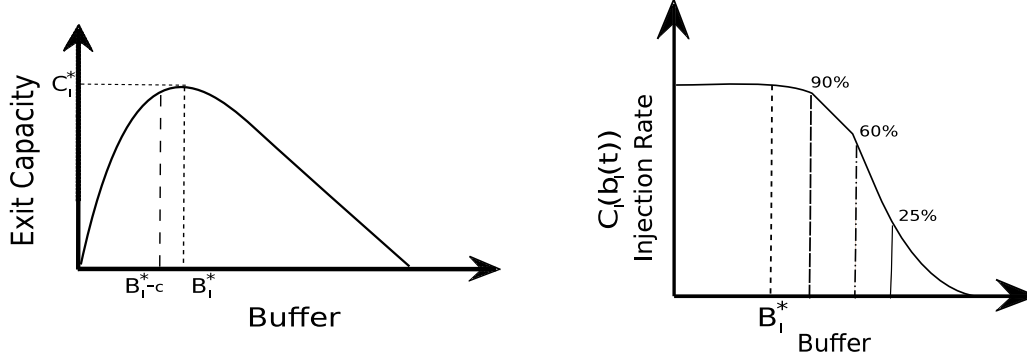


Fig. 4. Traffic curves showing the instantaneous exit rate (left) and the maximum exit rate (right) as functions of the buffer size.

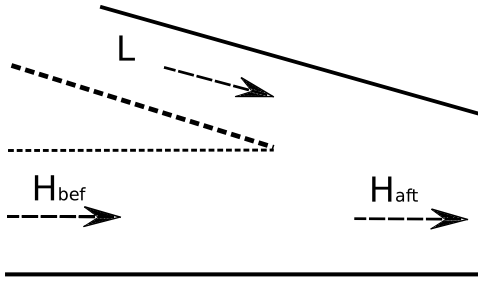


Fig. 5. Merging of two freeways.

a polyline and a collection of *observations*. Observations are a family of average speeds along and/or travel time across the polyline at an instance in time: $\{(x, y)_t\}_{t \in \mathbb{N}}$, where x and y are speed and travel time, respectively. For convenience and consistency with data, travel time is dropped from the formalization, allowing the speed at time t for a given segment x to be denoted as simply x_t . Every finite stretch of road, a *link*, can be associated with a *traffic density*, or the fraction of the link capacity that is occupied by vehicles, at a given time. This may equivalently be expressed by the *buffer size or buffer capacity* (B_ℓ), which is the number of vehicles in the link. The *exit rate or exit capacity* (C_ℓ) of a link is defined as the number of vehicles exiting the link per unit time. The traffic curve captures the variation between these two parameters (Jain et al., 2012). At high traffic densities (indicating traffic jams), links have very low operational exit capacities and at low densities, the exit rate varies linearly with the density (Fig. 4). We define the optimal operating points of a traffic curve based on *optimal exit rate* C_ℓ^* where the exit rate is the highest and the corresponding B_ℓ^* , the *optimal buffer size at C_ℓ^** . Based on the traffic curve, one can define the maximum exit rate of a link as a function of the current buffer capacity as shown in Fig. 4. The maximum exit rate is the maximum number of vehicles that can exit the buffer per-unit time, $C_\ell(B_\ell(t))$, which is also equal to the maximum sustainable flow rate at equilibrium for a given buffer capacity.

Now, consider the case where the input rate is larger than the optimal exit rate for a short time period, causing the link buffer to grow. Once the buffer size increases beyond the optimal value B_ℓ^* , the exit rate begins to decrease, leading to a more rapid increase of the buffer size, which further perpetuates the cycle, until a point is reached when the buffer is full and the exit rate is at its lowest possible value. This is called a *traffic collapse*. A very common real-world example is two freeways merging into a single freeway. A simple example is illustrated in Fig. 5, where vehicles in L are merging with the stream of vehicles on H . This simple example can be viewed in two ways: two lanes in the same freeway merging into a single lane or two separate freeways merging or a single lane merging into a freeway. To visualize this problem from the perspective of traffic curves, consider three links

in the setup: (a) H_{bef} representing, a small segment of H (covering a short distance of up to 0.5 miles) before the merge point; (b) a small segment L before the merge point; (c) H_{aft} , representing a small segment of H after the merge point. Each of the links can be associated with their corresponding traffic curves. Since we are dealing with a discrete version approximation using traffic curves, we should choose reasonable lengths to have meaningful buffer values for the links. The above traffic merging can be viewed using a simple 3-link topology where H_{bef} and L merge into H_{aft} , and each segment has an associated traffic curve. We primarily concentrate on two specific parameters of H_{aft} : $C_l^*(H_{aft})$ and $B_l^*(H_{aft})$. If the sum total of the exit rates of H_{bef} and L is always less than the optimal exit rate $C_l^*(H_{aft})$, then the merging never faces a congestion problem. If, however, the sum of the input rates of L and H_{bef} is larger than $C_l^*(H_{aft})$, then the buffer size of H_{aft} grows. If the buffer of H_{aft} grows beyond $B_l^*(H_{aft})$, then the exit rate of H_{aft} begins to drop thereby, triggering the spiraling effect.

3.2. Identifying sudden jams

The phenomenon of traffic collapse as defined in the previous section leads to a *traffic jam*, which is a prolonged state of very slow movement of vehicles on the road. A road segment is said to be in a state of *complete jam* when the density is maximum and the average speed of vehicles in the segment is 0 i.e. the vehicles have come to a standstill. Above a certain threshold of density, or equivalently, below a certain threshold of the average speed of the flow of vehicles in the segment, the segment can be said to be *approaching* a jam. The choice of either threshold is based on the range of possible speeds in the segment. We define *sudden jam* qualitatively as the state when we approach a jam quickly. That is, if the traffic collapse happens over a very short time period (typically within a matter of minutes), then we call it a sudden jam. This can happen due to a rapid build-up in the buffer size, in turn resulting in a rapid drop in vehicle speeds in the segment. Sudden jams are very common in congested cities all over the world (Jain et al., 2012).

Formally, a sudden jam may be defined as follows. Consider a point in time t in segment i and T_i be the corresponding time-series array of observation times. Let id_t be the index of time t in T_i . Consider two positive integers, m and n where $m < n$. The *prediction window* is the interval $(t, T_i(id_t + m))$, and the *target window* is $[T_i(id_t + m), T_i(id_t + n)]$. Further, let the *observation window* be an interval prior to t that is the same size as the target window based on the number of observations: $[T_i(id_t - (n - m)), t]$. A *sudden jam* at time t is a condition in which acceleration between the target and observation windows is less-than, or equal to, some threshold α . Eq. (1) shows the mathematical definition of the sudden jam function.

$$s_t(x, m, n) = \begin{cases} 1 & \text{if } \frac{1}{(T_i(m-t) \times (n-m))} \left(\sum_{i=t+m}^{t+n} x_i - \sum_{i=t-(n-m)}^t x_i \right) \leq \alpha \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

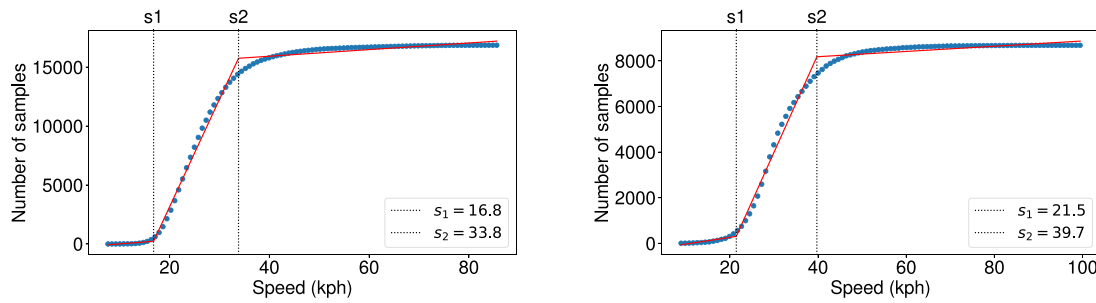


Fig. 6. Examples of most segments (90%) where there were two clear break points in the CDF of observed speeds. We named these breakpoints s_1 and s_2 . The s_1 speed would be approximately the point at which the traffic crosses the “threshold” density for a jam. These breakpoints were obtained by fitting a piecewise linear model to the CDF function.

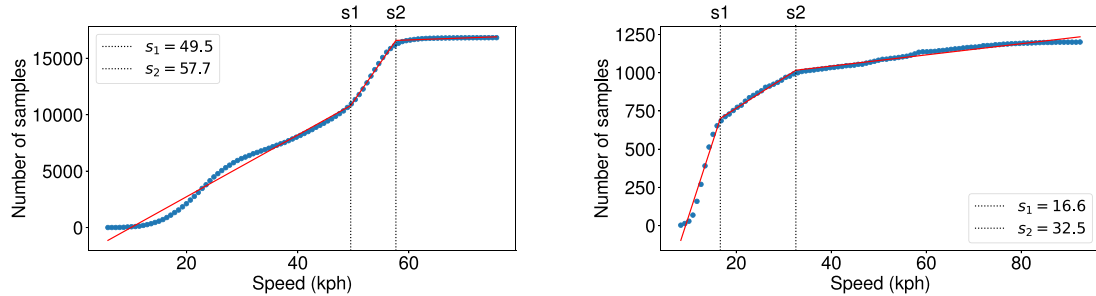


Fig. 7. Examples of some segments (10%) where the speed CDFs were different. There were no two clear breakpoints, hence not a well-defined threshold density to define a jam. Note that this is *not* an artifact of the amount of available data in these segments — some of these segments had even more data available than those in the first set.

Essentially, there is a sudden jam if the average speed of the observation window differs from the average speed of the target window, hence the division by the window size $n - m$. The additional factor in the denominator, $T_i(m) - t$, is the temporal size of the target window, converting the derivation to one of acceleration, and allowing α to be expressed as gravitational units. This not only allows the metric to be consistent with the literature on sudden braking events (Harbluk et al., 2007; Simons-Morton et al., 2009), but removes biases toward segments with faster free-flows. By “normalizing” with respect to deceleration, the impact that the slow down has on the passenger remains relatively constant. We apply this definition in our New York dataset, where we have speed data from very selected road segments, especially freeways, bridges, and tunnels, every minute, collected using loop detector instrumentation, from the local transportation department (NYC DOT) (New York City Department of Transport). The results are shown in the next section Section 4.

3.3. Speed-distribution curves

We explore the distribution of mean segment speeds in the Uber movements dataset by plotting the cumulative distribution of speed time series. Given the hourly speed data in a road segment over a period of time, we first sample the speed time series using a Poisson distribution with $\lambda = 8$ to remove temporal correlation effects in the speed data. The value of λ was chosen based on the length of observed regions of low-speeds (or jams). Then we plot a distribution of all the measured/observed average vehicle speeds as a cumulative density, which shows the distribution of speeds across all the segments observed over the entire time period of two years. For well-behaved segments shown in Fig. 6, which make up almost 90% of all segments, we obtain an S-shaped curve, while for the remaining segments, as shown in Fig. 7, we see irregular shapes.

For the well-behaved segments, it is clear that there are two turning points, which show significant transitions in the traffic behavior of the segment. To obtain the corresponding values of speeds for these points, we fit a 3-way piecewise linear function to this CDF using the *pwl*f library in Python (Jekel and Venter, 2019) to obtain the two break

points, s_1 and s_2 , such that $s_1 < s_2$. Well-behaved segments have two clear break-points while the remaining segments don’t have clear break-points.

Fig. 8 shows the distribution of the minimum, maximum, s_1 , and s_2 speeds for all the valid segments for Nairobi, New York City, and São Paulo. We note that the median value of s_1 is about 20 kph across all three cities. This is an interesting observation, given that the other three quantities show significant differences across the three cities. In the next section, we hypothesize that s_1 value of a segment corresponds to an important phase-transition point in the traffic curve. This might be the reason why we get similar values of s_1 across all cities.

3.4. Estimating traffic curves

In practice, sudden jams happen over very short timescales, such as within 5 min (Jain et al., 2012), and we might not always have speed or density data at that fine granularity in order to be able to analyze them. In such situations, we cannot apply the formula from Eq. (1). Additionally, we might not have information about the actual vehicle density on the road. In such cases, under the framework of certain assumptions, we can estimate the traffic curve from the speed data alone, from which we can identify key threshold points for determining sudden jams.

Now we describe a methodology to compute the traffic curve, as in Fig. 4, from observed hourly speed data. We begin by dividing the traffic curve into three different regions based on our observations in mean-speed distributions, namely the free-flow region, the spiraling region, and the jam region. The free-flow region of the traffic curve represents a constant flow under equilibrium, which is only possible before the optimal buffer capacity (B_1^*) as shown in Fig. 4. The spiraling region of the traffic curve represents the phase after the optimal buffer capacity has been crossed and when the exit rate of the link rapidly drops. The spiraling region ultimately leads to the final region of the traffic curve, the jam region where entire traffic is brought to a standstill and moves forward at a very small constant speed. Fig. 9 provides an illustration of these regions. We make a few assumptions for simplistic modeling.

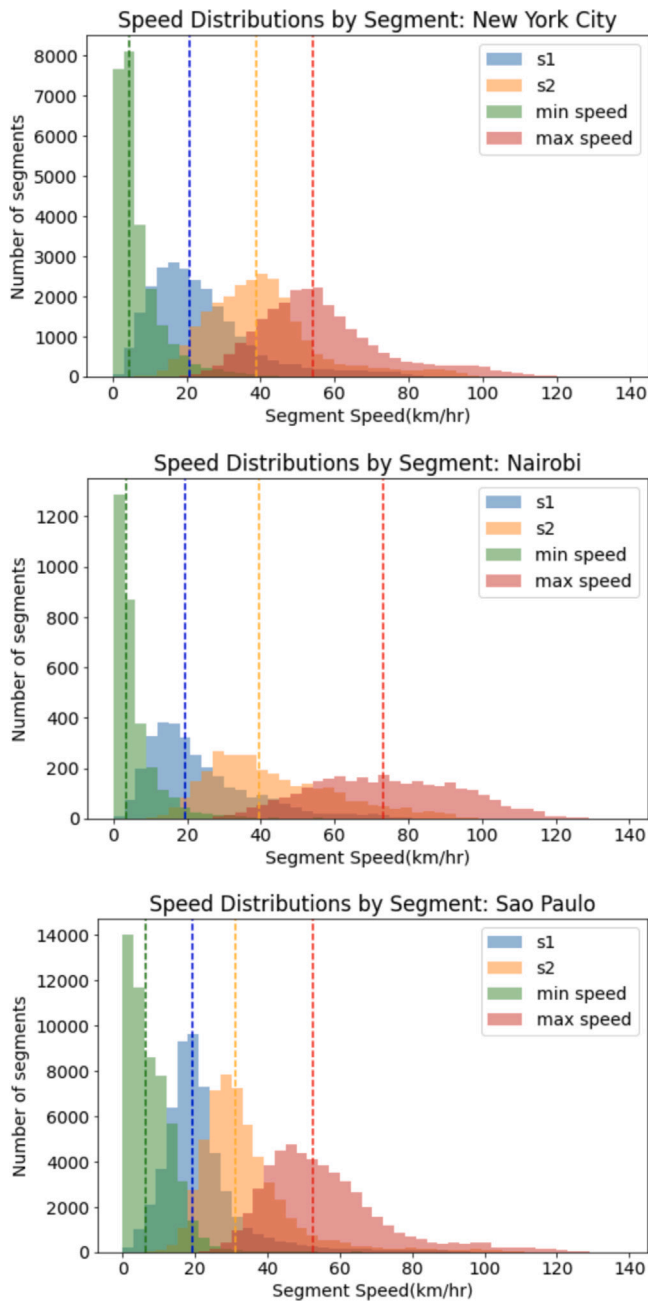


Fig. 8. Distribution of the maximum, minimum, s_1 and s_2 speeds across all segments with sufficient data points. The medians for each of the four quantities are shown at the top of the plot. We observe that the median value of s_1 corresponds to the value of 20 kph across all three cities.

- Drivers try to keep moving as fast as possible at all times, keeping with the pace of the rest of the traffic, while driving safely and not exceeding the speed limit. This is based on Newell's car-following traffic stream model (Newell, 2002), where every car follows the preceding car with a minimum space and time gap.
- All vehicles are modeled as rectangular blocks of fixed length and all roads are single lanes only, with only a single direction of travel. Overtaking dynamics are not considered. In terms of traffic queuing theory, we assume a FIFO (first in - first out) queue discipline, which is the most commonly used queuing model (May, 1990). This assumption is also implicit in the car-following traffic stream model.

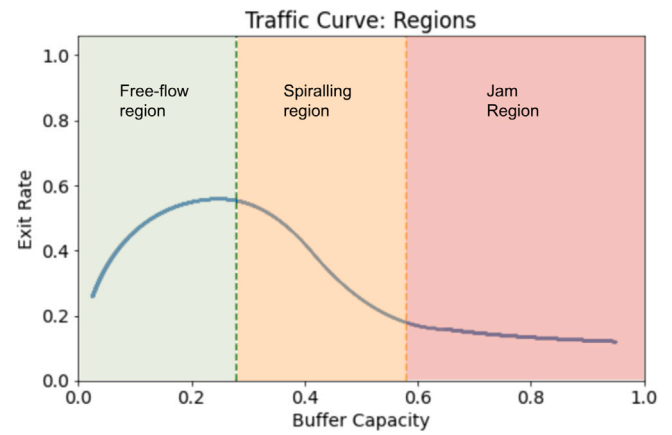


Fig. 9. Dividing the traffic curve into different regions based on traffic behavior. The free-flow region corresponds to equilibrium traffic where the link can maintain free-flow traffic before the optimal buffer capacity. The spiralling region corresponds to the rapid drop in exit rate with increasing buffer capacity, which ultimately leads to traffic collapse, leading to the jam region of the traffic curve.

First, we focus on the free-flow region of the traffic curve. Based on common knowledge in written driving tests (Driving Test Success), drivers are required to maintain a certain stopping distance with vehicles in front of them, which consists of two components, the thinking distance, and the braking distance. The thinking distance is based on the human reaction time and is linear with the speed of the vehicle while the braking distance is the distance it takes for the brakes to bring the vehicle to stop, which is quadratic with respect to the speed of the vehicle. Thus, the stopping distance can be expressed as:

$$d(s) = t \times s + t' \times s^2 \quad (2)$$

where t represents the human reaction time threshold while t' is the corresponding constant for the quadratic term, and s is the speed of the vehicle. It should be noted that the above relationship between inter-car distance and vehicle speed should be largely accurate for the low buffer-density phase of the traffic curve. Based on (Driving Test Success), the values for t and t' should be 0.675 and 0.076 respectively. Traffic mismanagement and chaotic driving conditions are likely to cause these values to be lower than the recommended values.

Now, let us assume an infinite road segment constrained under previous assumptions, filled with cars maintaining a constant speed s . In this case, the buffer density of the link would be:

$$B(s) = \frac{l_c}{l_c + d(s)} = \frac{l_c}{l_c + t \times s + t' \times s^2} \quad (3)$$

where d is the stopping distance and l_c is the length of each car. For simplicity, we assume all cars are of equal length $l_c = 4$ m. At the same time, we have the following relations between exit rate, buffer density and speed in the segment:

$$C(s) = B(s) \times \frac{s}{l_c} \implies C(s) = \frac{s}{l_c + d(s)} = \frac{s}{l_c + t \times s + t' \times s^2} \quad (4)$$

Eqs. (3) and (4) give us a parametric relationship between buffer capacity and exit rate, which is valid for the low-density phase of the traffic curve. Now, let us consider the other extreme phase of the traffic curve, the jam region.

The *jam density* is the upper limit on free-flowing traffic density, beyond which the road segment is said to be in the state of a jam (May, 1990). Although there is a bit of variation on this threshold density in literature (Knoop and Daamen, 2017; Wu, 2002), the theory according to May (1990) gives a value for jam density of 185 to 210 vehicles per kilometer per lane, which translates to between 74% and 84% lane occupancy, assuming an average car length of 4 m. We assume the threshold for a jam to be 66% occupancy of the road segment. This

corresponds to a spacing of half a car between every two cars. We believe that such a small distance between vehicles is enough to bring the traffic to a stand-still.

From the jam point onward, the relationship between buffer capacity and exit rate no longer follows the parametric Eqs. (3) and (4). For deriving a relationship between the two entities, we consider the case where the cars are packed in the road segment with inter-car spacing D and the traffic is at standstill. Say that after this particular road segment, we cross the choke point and the traffic resumes its free flow state. An example of such a situation could be a smaller road merging into an express highway. In this case, say that N cars are able to exit the road segment in time duration T . Then, the exit rate C_0 is given by $\frac{N}{T}$. Now, consider the case when the inter-car spacing is reduced by a small amount δD . Then, the buffer capacity of the new situation is given by:

$$B(\delta D) = \frac{l_c}{l_c + D - \delta D} \quad (5)$$

Also, due to the cars having lesser inter-car spacing, each car must wait an additional time δT before it can accelerate to the speed of the car in front of it. But, it should be noted that this additional time keeps getting accumulated for each car since the second car must wait δT for the first car before accelerating, while the third car must wait δT for the second car, and so on. Therefore, the time taken for N cars to cross the choke point will be $T + N \times \delta T$. Therefore, the exit rate, in this case, would be:

$$C(\delta T) = \frac{N}{T + N \times \delta T} \quad (6)$$

Now, we consider the lowest of the traffic corresponding to the jam exit rate C_0 as S . In that case, we have the following relation:

$$\delta D = S \times \delta T$$

Following the parametric forms in Eqs. (3) and (4), we have:

$$C_0 = \frac{S}{l_c + D} \implies \frac{N}{T} = \frac{S}{l_c + D} \implies N = \frac{S \times T}{l_c + D}$$

Then, substituting the above-obtained value of N in Eq. (6), we get:

$$C(\delta T) = \frac{S \times T}{(l_c + D) \times (T + \frac{ST\delta T}{l_c + D})} \implies C(\delta T) = \frac{S}{l_c + D + S \times \delta T}$$

Now, we use the relationship between δD and δT in the above equation. Thus, we have:

$$C(\delta D) = \frac{S}{l_c + D + \delta D} \quad (7)$$

Eqs. (5) and (7) give a parametric relationship between buffer capacity and exit rate for the high buffer density or jam-phase of the traffic curve. Thus, we have two different parametric forms of relationship between exit rate and buffer capacity in three different regions, namely the free-flow and jam regions, of the traffic curve. In the spiraling region lying in between the two regions, the dynamics of the exit rate and buffer capacity are non-trivially related and hard to deduce. As we saw in Section 3.3, the mean segment speed distributions show clear break points at s_1 and s_2 , where for speeds lower than s_1 , we observe rapidly decreasing values of mean segment speeds, showing the relationship between s_1 and traffic collapse. Thus, we hypothesize that the free-flow region of the traffic curve transitions to the spiraling region around the s_1 speed value for a given segment. This gives us the phase-transition point between the free-flow and the spiraling phases of the traffic curve. The spiraling phase would then smoothly join into the jam phase of the traffic curve. We assume that the jam phase starts at $B_0 = 0.66$ and the corresponding speed of traffic at this point (i.e. S) is 3.6 km/h (1 m/s). Then, we use parabolic curve fitting and smoothing to complete a continuous and differentiable transition between the two regions of the traffic curve and obtain the complete traffic curve as given in Fig. 10 for different segments.

3.5. Sudden jams in hourly resolution

In reality, the perception of a ‘‘jam’’ would depend on the particular road segment. On a highway, where speed limits may normally be upwards of 100 kph, drivers may perceive the state of driving at less than, say even 40 kph, to be in a state of a jam. Whereas on a city road, the limit is much lower.

We observe two inflection points in the speed distribution (s_1 and s_2), where there is a sudden change in the gradient. We also note that the median speed of the segment is close to $\frac{s_1 + s_2}{2}$. Another observation is that $s_2 \leq 3 \times s_1$ with high probability. Based on this, we heuristically define traffic jams (called slowdown jams) as instances of time when the value of mean segment speed falls below $\frac{s_1 + s_2}{4}$. Since we see a sudden drop in speeds below s_1 and $\frac{s_1 + s_2}{4} < s_1$ with high probability, we choose this as the threshold for our definition. Note that other values of the threshold can be chosen with similar justifications.

For understanding, if there is a correlation between sudden jams and slowdown jams, we generate the hourly resolution of a section of the NYCDOT dataset by averaging the speeds reported over a segment for each hour. Note that this method of generating hourly resolution is inaccurate because we don't have data about the density of cars in the given links. Also, we see that in hourly resolution data obtained by this process, most segments don't ‘‘behave well’’, in that they don't have clear breakpoints. Still, we proceed to apply the definition of slowdown jams on the hourly resolution of the NYCDOT data to compute the occurrences of slowdown jams. After this, for each occurrence of a slowdown jam, we take a look at the observations that make up the hourly average and check if we observe a sudden jam based on the definition given in Eq. (1). Taking the recent 2 million observations from the NYCDOT data, we find that in 67.38% of the cases where we observe a slowdown jam, we can see a sudden jam appearing in the corresponding hour. Note that this happens even for ‘‘ill-behaved’’ segments, hence, we can safely assume that the occurrence of slowdown jams, in reality, should be strongly correlated with sudden jams in reality.

4. Results

4.1. Sudden jams in NYCDOT dataset

When characterizing sudden jams, signals with average reporting rates of 90 seconds or less were used. Further, segments had to have a polyline component. These two conditions reduced the total set of 153 segments to 98. Finally, to be eligible for jam classification, the observation and prediction components of a given window cannot contain missing data.

Recall that a window consists of three components—observation, prediction, and target. A window is considered a sudden traffic jam if the average speed between the observation and target portions decreases by more than a particular rate. ‘‘Prediction window’’ is thus the time between the observation and target portions. Fig. 11 presents an overview of frequency using a fixed $\alpha = -0.002$, and window sizes varying from 1 to 10 min. The two window portions – observation and target – remained equal throughout. As such, they are denoted simply as ‘‘adjacent.’’ The frequency of sudden jams depends not only on the characteristics of the segment but on parameters to Eq. (1); in particular, the size of the window and the choice of α .

Fig. 11(a) outlines the inverse relationship between sudden jam frequency and respective window sizes on a sample segment. Prediction windows range from 0 to 10, where 0 signifies a comparison between adjacent points in time. The adjacent windows range from 1 to 10, as a lower bound of 0 would mean a comparison between non-existent windows. Darker cells denote higher occurrences, with each cell annotated with the number of occurrences. The number of sudden jams increases as the size of the adjacent window increases and as the size of

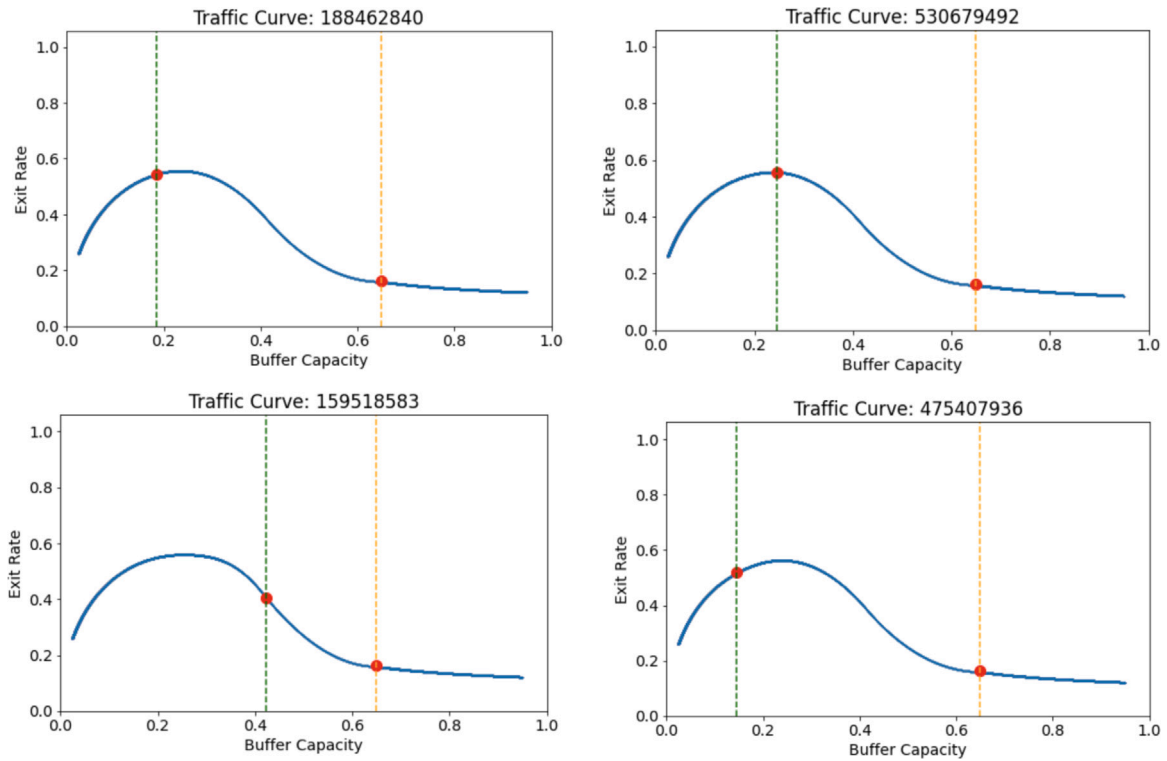
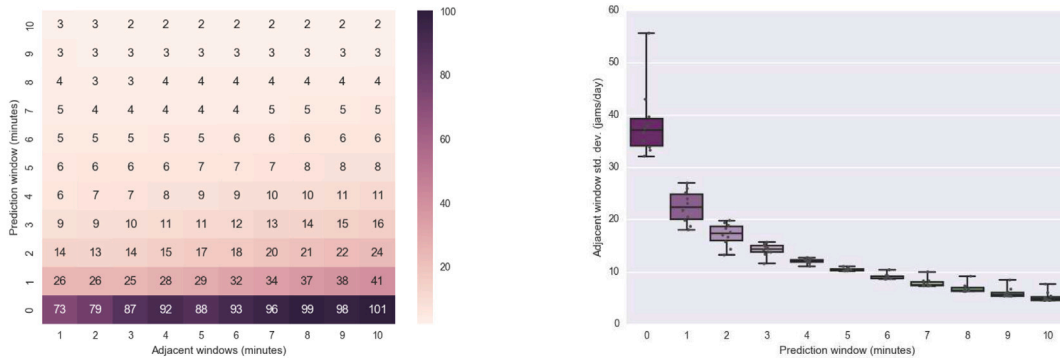


Fig. 10. Empirically derived traffic curves for four different road segments with their open street map way id given in figure titles. As can be seen, the first phase transition point lies close to the values of exit rate and buffer capacity given by s_1 according to Eqs. (3) and (4). The second transition point is fixed on the traffic curve based on our assumptions.



(a) Average number of jams, per day, per segment, across varying prediction and target windows. Observation and target windows were held equal, denoted here simply as "adjacent".
 (b) Average standard deviation of per day, per segment jam counts

Fig. 11. Sudden jam characterization across all segments in the network.

the prediction window decreases. The increase across prediction windows is exponential, while the increase across the adjacent windows is linear. With respect to the adjacent windows, this relationship is likely due to variability in average speed. Using observation and prediction windows of size one means that the model is taking into consideration "unsmoothed" values. In some cases, these observations may not be entirely representative of actual conditions; a result of measurement anomalies, for example. As window sizes increase, spurious values can augment average speed such that a sudden jam is harder to determine. With respect to prediction windows, the inverse relationship is largely a result of the manner in which sudden traffic is calculated: larger prediction windows require a larger decrease in speed between the observed and target windows to be classified as traffic events.

Fig. 11(b) presents the average standard deviation at varying adjacent window sizes across all segments considered. The large number

of traffic incidents shown in Fig. 11(a) for smaller adjacent window segments is dominated by a relatively small number of road segments. Measurements across mid to large windows, while less in number, are spread more evenly.

4.2. Slowdown jams in Uber movements dataset

We apply the $\frac{s_1 + s_2}{4}$ threshold on every one of the valid segments in all three cities. We observe the statistics for the slowdown jams in the three cities as shown in Table 1. From the table, it is clear that the average jam time per segment per day for Nairobi is outrageously higher than (almost 3x) the same for New York City and São Paulo.

We present the histogram distribution of the time of jams occurring in Fig. 12 and of the duration of jams in Fig. 13. From Fig. 12, we

Table 1

Jam statistics for slowdown jams across all three cities. We present the number of segments, the number of valid segments with sufficient data points, the number of slowdown jams observed, the total hours across all jams, and the mean amount of time spent in jams for each segment each day in hours.

City	Segments	Valid segments	Jams	Total jam hours	Mean jam time
Nairobi	4949	3063	3930954	4753618	1.89
New York	35602	25371	8400572	14033002	0.67
São Paulo	89121	53658	20091200	29340467	0.67

Table 2

Sample of 9 segments from the 6 junctions, showing total hours in jams, the number jams observed, the mean hours per day spent in jams, the mode hours (most commonly observed jam duration), and the mode hour of the day (most commonly observed time of day when jams occur).

Way ID	Total hours in jams	Number of jams	Mean hours in a jam	Mode hours in a jam	Mode hour of day
9931279	1453	1453	1.77	1.0	05:00
678371493	865	334	1.05	2.0	13:00
364279376	643	219	0.78	2.0	13:00
678371494	658	259	0.80	2.0	13:00
4724017	2763	2763	3.37	1.0	12:00
336067605	3303	3303	4.02	1.0	15:00
39573541	405	162	0.49	2.0	11:00
580233744	4092	4082	4.92	1.0	14:00
4742016	4794	4794	5.84	1.0	14:00

observe that the jam time patterns for São Paulo and New York City match each other largely, with both cities suffering congestion during late night hours while Nairobi presents an inverted case with most of the congestion occurring during afternoon hours. In Fig. 13, we observe that for all three cities, most jams last only one hour. This refers to the pattern of finding only one entry in the speed time series to be less than the threshold surrounded by entries that are larger than the threshold.

4.3. Investigating important junctions

Table 2 shows statistics for six representative junctions in the city of Nairobi, covering a variety of settings — T-junctions in the city, highway merges and roundabouts, and Figs. 14 to 15 show some sample snapshots for the six segments on various days. The faint lines show the input segments into the output segment(s) of interest, which are shown with thicker lines. The red color shows a jam, which is when the speed in a segment drops below the $\frac{s_1+s_2}{4}$ for that segment. The snapshots show how the average speeds in the segments vary over the course of the day.³ We note a large number of instances of prolonged congestion (i.e. jam) in the sink segment in junctions B, D, and E. In the roundabout junctions C and F, we observe jams at one or two output sink segments. We observe that in many cases, the jams persist for over 2 to 3 h. In the junctions B, D, and E particularly, which are all 2–1 merges, we observe that the average speed in the output segment is below the respective threshold continuously for more than several hours on multiple days.

4.4. Travel time impact

For an individual segment of fixed length, the travel time for the segment and the mean speed are expected to follow an inversely proportional relation. We validated this assumption for the NYCDOT dataset by computing the product of travel time and mean speed for all entries of particular segments, and computing the coefficient of variation of these products. The coefficient of variation equals the ratio of standard deviation to the mean of the data, and a value of less than one is considered to be a small variation in data. Given that we expect

³ For the sake of brevity, we only show handpicked snapshots for each of the junctions that best illustrate the different phases. We will be happy to share our analyzed data in greater detail as needed for the reproduction of our results.

the product to remain constant, our threshold for the coefficient of variation should be much smaller. In practice, we obtained a value of less than 0.1 for all road segments in the NYCDOT dataset except 2 outliers, with most segments having values in the range of ~0.01. Thus, we can assume for individual segments:

$$\text{travel time} \propto \frac{1}{\text{mean speed}}$$

Under jam conditions, the average segment speed can fall down to 5 km/h for a segment with an operational speed of 50 km/h, meaning that the travel time for jammed segments can see up to 10x increase. Depending on the source and destination as well as the commute path taken, the impact of jams on travel time will be variable. In the Uber movements dataset, we find the travel time data for cities of Nairobi and São Paulo between source and destination hex-clusters averaged by hour buckets per day from January 2020 to March 2020. Due to differences in resolution from speed data as well as lacking data on the commute path between the clusters, we have been unable to join the travel time data with the speed data to understand the impact of jams on travel time with high granularity. Still, we compared the maximum observed mean travel time on the segments to the average of the mean travel time. The results are shown in Fig. 16. We observe that for a majority of source–destination pairs, the travel time in a jam can be more than 50% of the average travel time. For a smaller fraction of cases, the jam can cause the travel time to increase by 3x or higher.

5. Discussion and conclusion

In this paper, we have presented the concept of sudden traffic jams, with a formal definition, the methodology to detect such jams and the corresponding theoretical foundation. It should be noted that sudden traffic jams are not a phenomenon emerging due to over-subscription of traffic flow but rather due to momentary bursts in the traffic. In our analysis, we found that highways and other free-flow segments are also susceptible to the phenomenon. Based on the comparisons between the three cities of New York, Nairobi, and São Paulo, we have determined that the city of Nairobi experiences unexpectedly higher severity of jams as compared to the other two cities.

Table 3 presents some key development indicators of the three cities including the population, percentage of households that own a car, number of segments in the Uber Movements dataset, and the mean jam

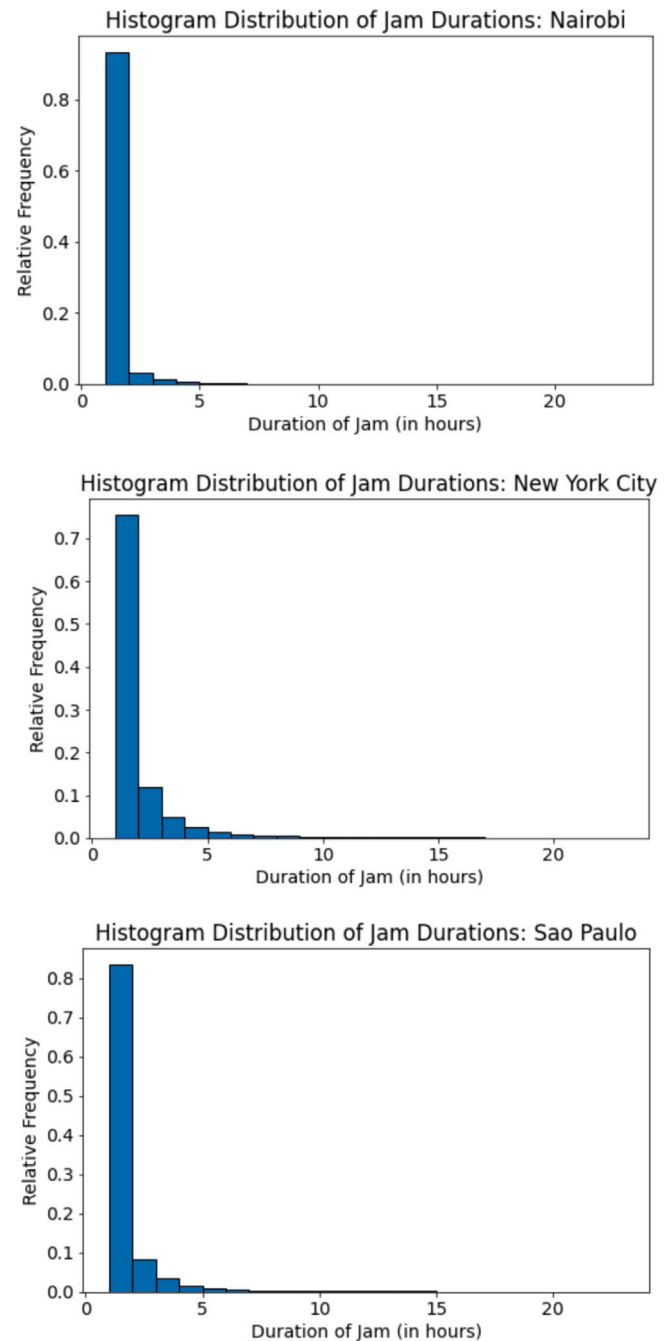
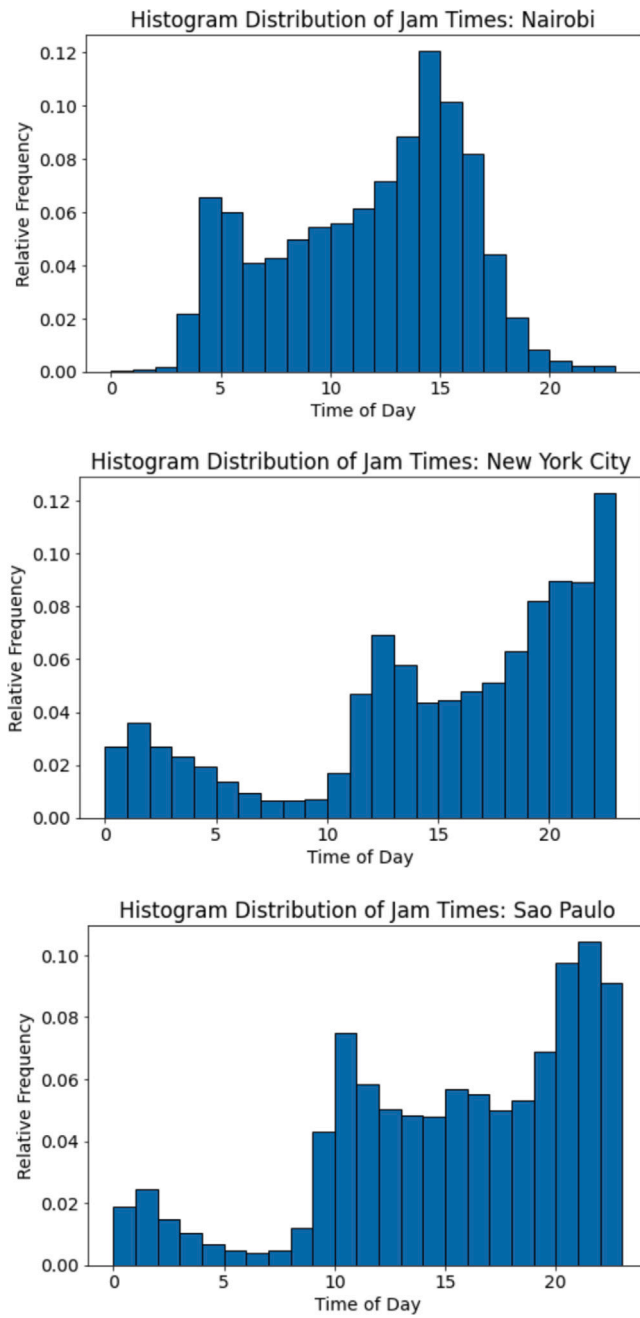


Fig. 12. Histogram of times when jams occur in all three cities. Both São Paulo and New York City show a similar pattern in terms of jam times, with many jams observed during late night hours while in Nairobi, we observe most jams around the afternoon.

Fig. 13. Histogram of the total duration of jams in all three cities. All cities, Nairobi, São Paulo, and New York City show a similar pattern in terms of jam duration, with most jams observed to last only a single hour. In New York City, we see the highest fraction of jams lasting more than one hour.

time per segment per day (see Table 1). We calculate the mean traffic load per segment in the city per segment by the simple formula given:

$$\text{Mean Traffic Load} = \frac{\text{Population}}{\text{Size of Household}} * \frac{\text{Household-Vehicle Percentage}}{100} * \frac{1}{\#\text{Segments}}$$

We assume an average household size of 4. We also assume that the result of the above formula is representative of the actual traffic load on the road segments in the cities.

We note that the ratio of mean time spent in jam per road segment for São Paulo, New York, and Nairobi is approximately 1:1:3. At the same time, we also see that the corresponding ratio of traffic load per road segment is approximately 1:2:3. Comparison between Nairobi

and São Paulo shows that approximately reducing the traffic load per segment by a factor of 3 also reduces the mean jam time by the same factor, as expected. On the other hand, consider the comparison between New York City and São Paulo, where São Paulo, despite having 2x lower mean traffic load, suffers on average, equal mean jam time. This shows that the traffic management in New York City might be better than São Paulo and Nairobi, leading to higher utilization of road capacity.

While the lack of road infrastructure seems to be a key reason for the traffic crisis in Nairobi, we believe that traffic mismanagement and chaotic driving conditions are also some of the factors contributing

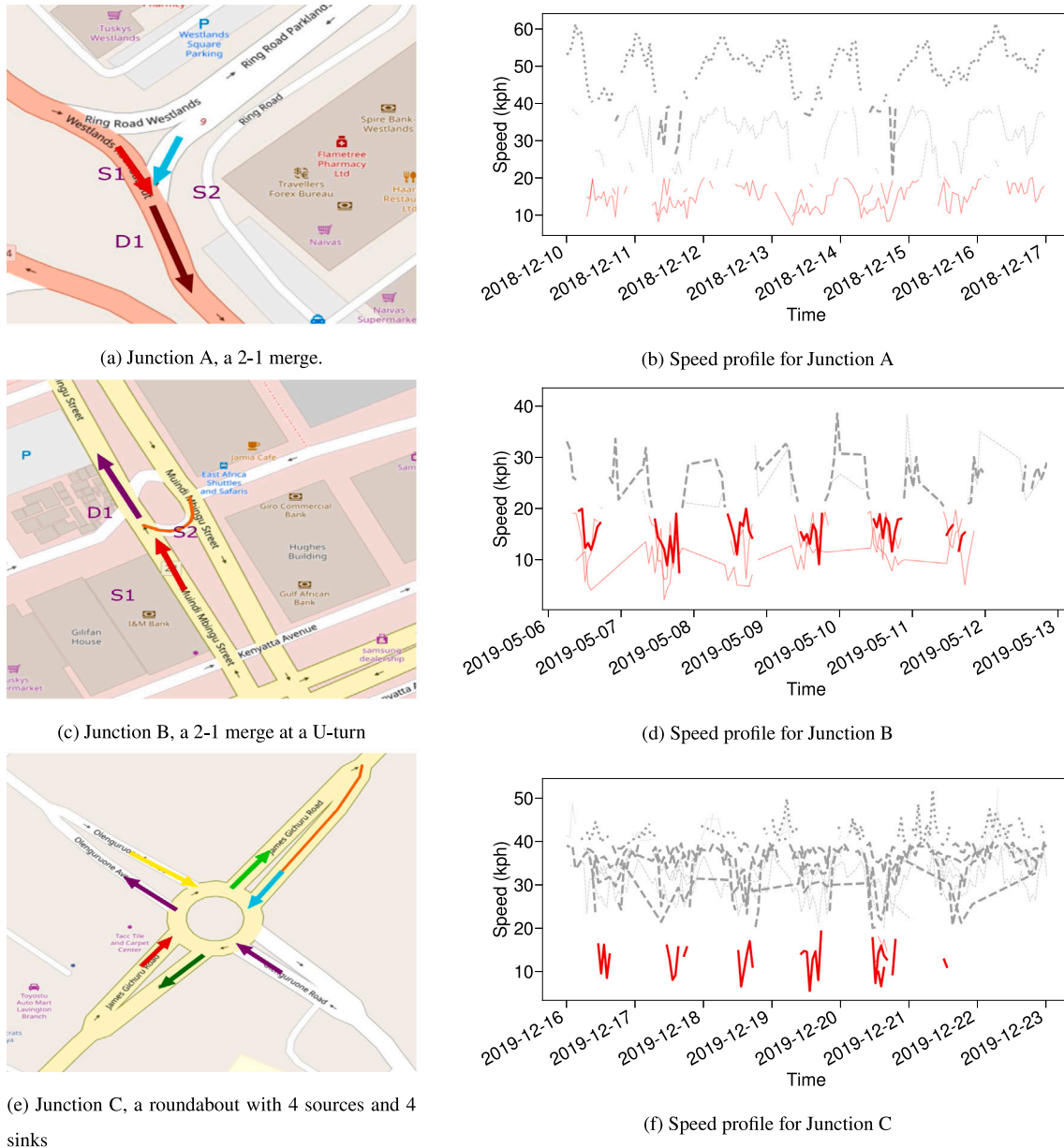


Fig. 14. Junctions A through C. On the left, we have maps with arrows showing the direction of traffic flow. On the right, we see the speeds observed in different segments. Faint lines represent in-flow traffic while thick lines represent out-flow traffic. Red is used to represent speeds below the threshold. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3

Development statistics all three cities where we present the number of segments observed in Uber Movements data, the population based on recent surveys, percentage of households owning motor vehicles (cars), the mean amount of time spent in jams for each segment each day in hours and the average number of vehicles per segment.

Source: NYCEDC, Wambu, EMBARQ Network.

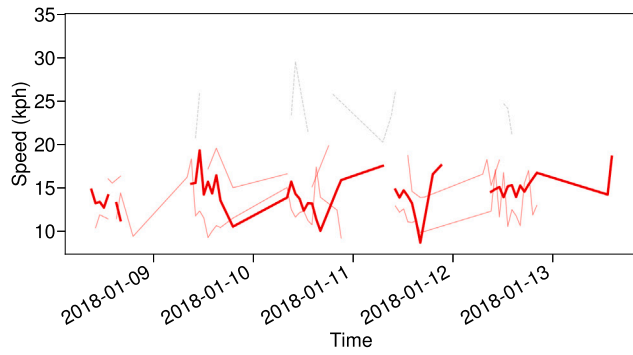
City	Segments	Population	Household-vehicle percentage	Mean jam time	Mean traffic load
Nairobi	4949	4.4 M	16.9%	1.89	37.56
New York	35602	8.4 M	45%	0.67	26.54
São Paulo	89121	12.33 M	40%	0.67	13.84

to this. As described in Section 3.1, the phenomenon of sudden jams emerges from the traffic collapse resulting from the decreasing link capacity in the spiraling region of the traffic curve. We now show how an aggressive driving culture can exacerbate the traffic congestion

problem. Consider the empirical formulation of traffic curve presented in Section 3.4. Based on our observations from speed-distributions shown in Section 3.3, we hypothesize that the traffic curve shifts from the free-flow phase to the spiraling phase at segment speeds around the



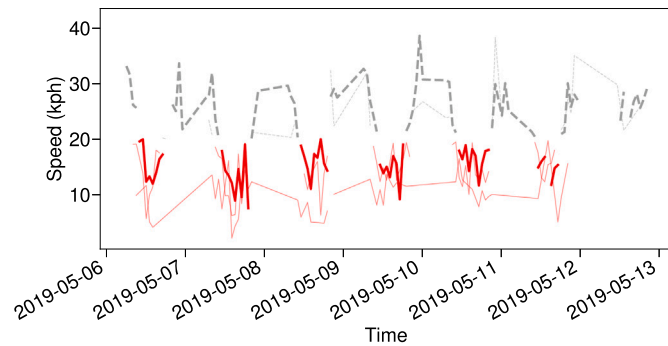
(a) Junction D, a 2-1 merge at a T-junction



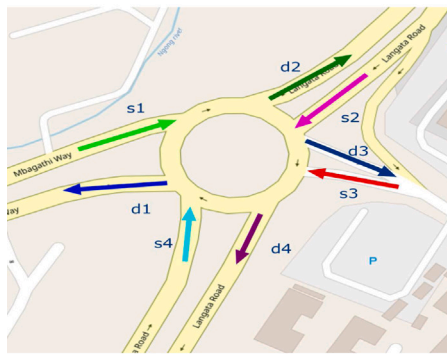
(b) Speed profile for Junction D



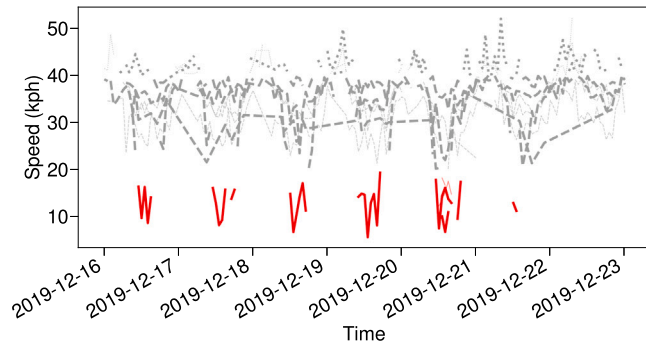
(c) Junction E, a 2-1 merge at another T-junction



(d) Speed profile for Junction B



(e) Junction F, another roundabout with 4 sources and 4 sinks



(f) Speed profile for Junction C

Fig. 15. Junctions D through F. Again, we have maps on the left and speed profiles on the right. Faint lines represent in-flow traffic while thick lines represent out-flow traffic. Red is used to represent speeds below the threshold. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

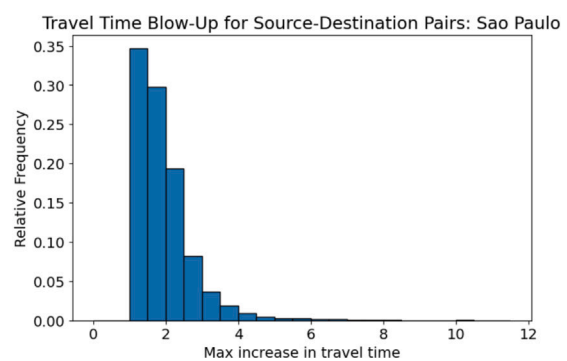
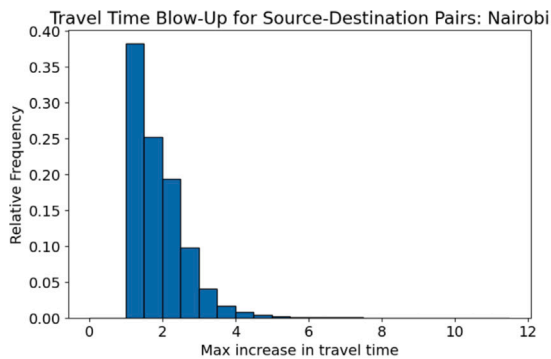


Fig. 16. Travel time blow up for different source–destination pairs in Nairobi and Sao Paulo. In more than 50% of such pairs, we find an increase in travel time of more than 50%.

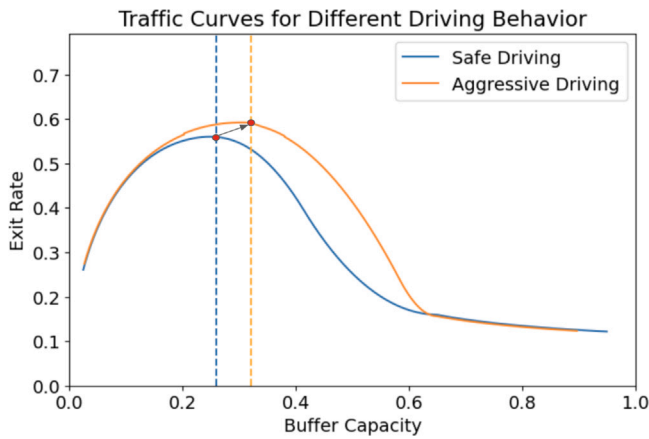


Fig. 17. Traffic curves for aggressive and safe driving behavior. It can be seen that the transition point from the free-flow to the spiraling region in the curve moves diagonally upwards, leading to a higher downward slope in the spiraling region for the aggressive curve.

value of s_1 , or the first break point in the speed distribution. The point on traffic curve corresponding to this can be found by substituting s with s_1 in Eqs. (3) and (4) as follows:

$$B(s_1) = \frac{l_c}{l_c + ts_1 + t's_1^2} \quad (8)$$

$$C(s_1) = \frac{s_1}{l_c + ts_1 + t's_1^2}$$

For aggressive driving behavior, we would consider smaller values of t and t' . This would result in a higher value for both $B(s_1)$ and $C(s_1)$, meaning that the point on the traffic curve will shift to the right and up. Note that the second transition point on the curve is fixed, and hence after curve fitting, we will obtain a higher downward slope in the second case (see Fig. 17). A higher slope in the spiraling region of the traffic curve can lead to more frequent and rapidly evolving jams, causing a larger duration of jam-time.

How do we mitigate these jams? Obviously, the development of the road infrastructure will help the problem to a large extent, as in the case of São Paulo, but it comes at a significant economic cost. Hence, we ask: is there a smarter way of traffic management that can increase the utilization of existing road infrastructure? Again, turning to the traffic curve from Section 3.1, the key idea is to be aware of the jam density and the threshold speed in a segment. Real-time instrumentation, such as the one in New York City, can inform drivers through navigation apps about impending congestion in a segment. Such knowledge can also be used in signaling at important points such as $n-1$ merges, where we observe the most instances of congestion resulting in prolonged jams. We believe designing such solutions will greatly benefit the developing cities around the world, which are disproportionately affected by the problem of traffic congestion.

Thus, to conclude, the key takeaways from our work are the collections of observations about traffic jams – they can appear in various scenarios, for various durations, and can potentially result in prolonged congestion of vehicles on the road, on many occasions lasting several hours at a stretch. Sudden jams that appear due to an increase in vehicle density beyond the threshold *jam density* are particular types of jams that last for long periods of time. We have provided a traffic curve formalism for understanding the phenomena of traffic collapse leading up to sudden jams, and a formula for sudden jams in terms of the drop in acceleration. For more sparse data such as in the Uber movements dataset, where we do not have minute-on-minute data to use the formula, we have proposed an alternative approach to estimating the traffic curve from the speed data and some basic assumptions. We compute an upper limit on the speed at the jam density in a road

segment from the speed distribution using two break points s_1 and s_2 in the cumulative distribution of speeds and call these jams as *slowdown jams*. Using the loop detector data from the New York City Department of Transportation, we show that slowdown jams and sudden jams have a high correlation. Applying the hourly definition to Uber movements speed data, we see that the city of Nairobi experiences unexpectedly high severity of traffic jams as compared to New York City or São Paulo, cities representative of the developed and developing economies. We attribute some part of this result to chaotic driving conditions and traffic mismanagement and propose that designing smart solutions for increasing road capacity utilization based on our work would reduce the economic budget for tackling the problem.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Shiva R Iyer reports financial support was provided by NYUWIRELESS Industry Affiliates. Lakshminarayanan Subramanian reports a relationship with Velai Inc that includes: equity or stocks. Lakshminarayanan Subramanian reports a relationship with Gaius Networks Inc that includes: board membership and equity or stocks. Lakshminarayanan Subramanian reports a relationship with Entrupy Inc that includes: board membership and equity or stocks. Lakshminarayanan Subramanian reports a relationship with World Bank Group that includes: consulting or advisory. Lakshminarayanan Subramanian reports a relationship with The Governance Lab that includes: consulting or advisory.

Data availability

The data used is available in public domain and the corresponding links have been cited in the manuscript.

Acknowledgments

The work done by the authors Ankit Bhardwaj, Shiva Iyer, Sriram Ramesh, Jerome White, and Lakshminarayanan Subramanian in this paper has been supported by funding in part from the NYU CTED,⁴ and the industrial affiliates in the NYUWIRELESS research group⁵ that funded Shiva Iyer in part as well as the air quality sensors used in the study. Shiva was also funded in part by an NSF Grant (award number OAC-2004572) titled “A Data-informed Framework for the Representation of Sub-grid Scale Gravity Waves to Improve Climate Prediction”. Shiva Iyer has also been supported by the Dean’s Dissertation Fellowship and Graduate Teaching Assistantship in the Graduate School of Arts and Sciences at New York University. Ankit Bhardwaj was funded by the New York University Henry M. MacCracken Fellowship. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of New York University or Wadhvani AI.

References

- Badger, E., 2013. How traffic congestion affects economic growth. <https://www.bloomberg.com/news/articles/2013-10-22/how-traffic-congestion-affects-economic-growth>.
- Driving Test Success, Stopping distances and the Theory Test, <https://www.drivingtestsuccess.com/blog/stopping-distances-and-theory-test>.
- EMBARQ Network, Study: São Paulo Commuters Lose One Month per Year Sitting in Traffic, <https://www.smartcitiesdive.com/ex/sustainablecitiescollective/study-sao-paulo-commuters-lose-one-month-year-sitting-traffic/278306/>.
- Friedman, J., 2020. 10 Cities with the worst traffic in the world. <https://www.usnews.com/news/cities/slideshows/cities-with-the-worst-traffic-in-the-world?slide=12>.

⁴ <https://www.nyucted.org/>

⁵ <https://www.nyuwireless.com>

- Harbluk, J.L., Noy, Y.I., Trbovich, P.L., Eizenman, M., 2007. An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accid. Anal. Prev.* 39 (2), 372–379. <http://dx.doi.org/10.1016/j.aap.2006.08.013>.
- Hodge, K., Guide to driving in Kenya, <https://www.rhinocarhire.com/Drive-Smart-Blog/Drive-Smart-Kenya.aspx>.
- Hu, H., Li, G., Bao, Z., Cui, Y., Feng, J., 2016. Crowdsourcing-based real-time urban traffic speed estimation: From trends to speeds. In: 2016 IEEE 32nd International Conference on Data Engineering (ICDE). pp. 883–894. <http://dx.doi.org/10.1109/ICDE.2016.7498298>.
- Iyer, S.R., An, U., Subramanian, L., 2020. Forecasting Sparse Traffic Congestion Patterns Using Message-Passing RNNs. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 3772–3776. <http://dx.doi.org/10.1109/ICASSP40776.2020.9052963>.
- Jain, V., Sharma, A., Subramanian, L., 2012. Road traffic congestion in the developing world. In: Proceedings of the 2nd ACM Symposium on Computing for Development. In: ACM DEV '12, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/2160601.2160616>.
- Jekel, C.F., Venter, G., 2019. `pwlf`: a python library for fitting 1D continuous piecewise linear functions. URL https://github.com/cjekel/piecewise_linear_fit.py.
- Kerner, B.S., Rehborn, H., 1996. Experimental features and characteristics of traffic jams. *Phys. Rev. E* 53 (2), R1297–R1300. <http://dx.doi.org/10.1103/PhysRevE.53.R1297> Publisher: American Physical Society.
- Knoop, V.L., Daamen, W., 2017. Automatic fitting procedure for the fundamental diagram. *Transp. B Transp. Dyn.* 5 (2), 129–144. <http://dx.doi.org/10.1080/21680566.2016.1256239>.
- Knorr, F., Baselt, D., Schreckenberg, M., Mauve, M., 2012. Reducing Traffic Jams via VANETS. *IEEE Trans. Veh. Technol.* 61 (8), 3490–3498. <http://dx.doi.org/10.1109/TVT.2012.2209690>, Conference Name: IEEE Transactions on Vehicular Technology.
- Lozano, A., Manfredi, G., Nieddu, L., 2009. An algorithm for the recognition of levels of congestion in road traffic problems. *Math. Comput. Simulation* 79 (6), 1926–1934. <http://dx.doi.org/10.1016/j.matcom.2007.06.008>.
- May, A., 1990. *Traffic Flow Fundamentals*. Prentice Hall, URL <https://books.google.com/books?id=JYJPAAMAAMAJ>.
- Nesbitt, D., Dara-Abrams, D., 2017. OSMLR traffic segments for the entire planet by MapZen. <https://www.mapzen.com/blog/osmlr-2nd-technical-preview/>.
- New York City Department of Transport, Data Feeds, <https://data.cityofnewyork.us/Transportation/DOT-Traffic-Speeds-NBE/i4gi-tjb9>.
- Newell, G.F., 2002. A simplified car-following theory: a lower order model. *Transp. Res. B* 36, 195–205.
- NYCEDC, New Yorkers and Their Cars, <https://edc.nyc/article/new-yorkers-and-their-cars>.
- Oh, S., Byon, Y.-J., Jang, K., Yeo, H., 2015. Short-term travel-time prediction on highway: A review of the data-driven approach. *Transp. Rev.* 35 (1), 4–32. <http://dx.doi.org/10.1080/01441647.2014.992496>.
- OpenStreetMap contributors, 2017. Planet dump retrieved from <https://planet.osm.org>. <https://www.openstreetmap.org>.
- Orosz, G., Wilson, R.E., Stépán, G., 2010. Traffic jams: dynamics and control. *Phil. Trans. R. Soc. A* 368 (1928), 4455–4479. <http://dx.doi.org/10.1098/rsta.2010.0205>, URL <https://royalsocietypublishing-org.proxy.library.nyu.edu/doi/full/10.1098/rsta.2010.0205> Publisher: Royal Society.
- Pattara-Atikom, W., Pongpaibool, P., Thajchayapong, S., 2006. Estimating road traffic congestion using vehicle velocity. In: International Conference on ITS Telecommunications. IEEE, pp. 1001–1004.
- Porikli, F., Li, X., 2004. Traffic congestion estimation using HMM models without vehicle tracking. In: Intelligent Vehicles Symposium. IEEE, pp. 188–193. <http://dx.doi.org/10.1109/IVS.2004.1336379>.
- Roy, S., Sen, R., Kulkarni, S., Kulkarni, P., Raman, B., Singh, L.K., 2011. Wireless across road: RF based road traffic congestion detection. In: 2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011). pp. 1–6. <http://dx.doi.org/10.1109/COMSNETS.2011.5716525>.
- Sen, R., Cross, A., Vashistha, A., Padmanabhan, V.N., Cutrell, E., Thies, W., 2013. Accurate speed and density measurement for road traffic in India. In: Proceedings of the 3rd ACM Symposium on Computing for Development. In: ACM DEV '13, Association for Computing Machinery, New York, NY, USA, pp. 1–10. <http://dx.doi.org/10.1145/2442882.2442901>.
- Sharma, M., Dikshit, O., 2016. *Comprehensive Study on Air Pollution and Green House Gases (GHGs) in Delhi*. Technical Report, IIT Kanpur.
- Simons-Morton, B.G., Ouimet, M.C., Wang, J., Klauer, S.G., Lee, S.E., Dingus, T.A., 2009. Hard braking events among novice teenage drivers by passenger characteristics. In: International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, vol. 2009. pp. 236–242.
- Stathopoulos, A., Karlaftis, M.G., 2002. Modeling Duration of Urban Traffic Congestion. *J. Transp. Eng.* 128 (6), 587–590. [http://dx.doi.org/10.1061/\(ASCE\)0733-947X\(2002\)128:6\(587\)](http://dx.doi.org/10.1061/(ASCE)0733-947X(2002)128:6(587)), URL <http://ascelibrary.org/doi/10.1061/%28ASCE%290733-947X%282002%29128%3A6%28587%29>.
- Texas Transportation Institute, 2009 Urban Mobility Report, <http://mobility.tamu.edu/ums/>.
- TomTom International B.V., Full Ranking 2020 https://www.tomtom.com/en_gb/traffic-index/ranking/.
- Treiterer, J., Myers, J.A., 1974. THE hysteresis phenomenon IN TRAFFIC FLOW. *Transp. Traffic Theory Proc.*
- Treiterer, J., Taylor, J., 1966. TRAFFIC FLOW investigations BY photogrammetric techniques. *Highway Res. Rec.*
- Uber Technologies, Uber Movement Speeds, <https://movement.uber.com/>.
- Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2014. Short-term traffic forecasting: Where we are and where we're going. *Transp. Res. C* 43, Part 1, 3–19. <http://dx.doi.org/10.1016/j.trc.2014.01.005>, Special Issue on Short-term Traffic Flow Forecasting.
- Wambu, W., Census report: This is where Kenya's wealthiest live, <https://www.standardmedia.co.ke/entertainment/news/article/2001361612/census-report-this-is-where-kenyas-wealthiest-live>.
- Wu, N., 2002. A new approach for modeling of fundamental diagrams. *Transp. Res. A Policy Pract.* 36 (10), 867–884. [http://dx.doi.org/10.1016/S0965-8564\(01\)00043-X](http://dx.doi.org/10.1016/S0965-8564(01)00043-X), URL <https://www.sciencedirect.com/science/article/pii/S096585640100043X>.
- Wu, H., 2018. Comparing google maps and uber movement travel time data. *Transp. Findings* <http://dx.doi.org/10.32866/5115>.
- Yang, M.H., Luong, T.T., Recker, W., 2014. Extracting traffic patterns from loop detector data using multiple changepoints detection. In: Transportation Research Board 93rd Annual Meeting. (14–1366).
- Yoon, J., Noble, B., Liu, M., 2007. Surface street traffic estimation. In: International Conference on Mobile Systems, Applications and Services. ACM, pp. 220–232.
- Zhao, J., Li, D., Sanhedrai, H., Cohen, R., Havlin, S., 2016. Spatio-temporal propagation of cascading overload failures in spatially embedded networks. *Nature Commun.* 7 (1), 10094. <http://dx.doi.org/10.1038/ncomms10094>, URL <https://www.nature.com/articles/ncomms10094>.