

# Empirical observations of congestion propagation and dynamic partitioning with probe data for large scale systems

## **Yuxuan Ji**

*Urban Transport Systems Laboratory  
School of Computer and Communication Sciences  
École Polytechnique Fédérale de Lausanne  
EPFL ENAC INTER LUTS, GC C2 390, 1015-Lausanne, Switzerland  
Phone: +41(21)693-5397  
E-mail: [yuxuan.ji@epfl.ch](mailto:yuxuan.ji@epfl.ch)*

## **Jun Luo**

*Huawei Noah's Ark Laboratory, Hong Kong, China and  
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China  
E-mail: [jun.luo@siat.ac.cn](mailto:jun.luo@siat.ac.cn)*

## **Nikolas Geroliminis\***

*Urban Transport Systems Laboratory  
School of Architecture, Civil and Environmental Engineering  
École Polytechnique Fédérale de Lausanne  
EPFL ENAC INTER LUTS, GC C2 389, 1015-Lausanne, Switzerland  
Phone/Fax: +41(21)693-2481 / +41(21)693-2481  
E-mail: [nikolas.geroliminis@epfl.ch](mailto:nikolas.geroliminis@epfl.ch)*

\*Corresponding author

Cite us:

Y. Ji, J. Luo and N. Geroliminis. Empirical Observations of Congestion Propagation and Dynamic Partitioning with Probe Data for Large-Scale Systems, in Transportation Research Record, vol. 2422, num. 2, p. 1-11, 2014.

This work obtained the 2014 Greenshields Prize

## ABSTRACT

Research on congestion propagation in large urban city networks has been mainly based on microsimulations of link-level traffic dynamics. However, both the unpredictability of travel behaviors and high complexity of accurate physical modeling remain challenging and simulation results may be far time consuming and not realistic. The main purpose of this paper is to reveal the hidden information during the process of congestion formation by exploring empirical data from large-scale urban networks. Specifically, we aim at studying the spatiotemporal relation of congested links, observing congestion propagation from an macroscopic perspective, and finally identifying critical congestion regimes to aid the design of peripheral control strategies. To achieve these goals, we use maximum connected component of congested links to capture congestion propagation in the city by utilizing a large dataset of 20,000 taxis with GPS data from Shenzhen of China. We present the empirical macroscopic fundamental diagrams of congested regions observed in propagation and quantify the critical congestion regimes. The findings show that the proposed methodology can effectively distinguish the congestion pockets from the rest of the network and efficiently track the congestion evolution in linear time  $O(n)$ .

## INTRODUCTION

Realistic modeling and efficient control/management of transportation systems remain a big challenge, due to the high unpredictability of choices of travelers (in terms of route, time and mode of travel), the uncertainty in their reactions to the control and the spatiotemporal propagation of congestion, the lack of coordinated actions coupled with the limited infrastructure available. While there is a strong understanding and vast literature of congestion dynamics, control and spreading in one-dimensional traffic systems with a single mode of traffic, most of the analysis at the network level is based on simplistic models or simulation, (i) which are difficult to calibrate and (ii) require a large number of input variables/parameters, that might not be observable with the current available data. There is a weak explanation on why traffic significantly varies from one day to another even if demand profiles are similar. While cascading phenomena are present in many types of physical or social systems (financing, human interactions) city traffic has interesting irregularities that should be studied. It is not also clear how one could predict spatiotemporally the development and propagation of congestion and how people adapt to their route choices under different congested conditions with small errors. The severity of these effects will be even stronger in case of non-recurrent events (e.g. accidents, road constructions), which can affect the resilience and productivity of transportation systems. How all of the above are related to the network topology and how perturbations in demand profile and network characteristics affect choices of people (in terms of route, departure time and mode) is a challenging question.

Cities are hierarchical self-organizing complex systems, but their form and structure are often far beyond of what we model as ideal cities. Nevertheless, cities are becoming smarter in ways that enable us to monitor, understand, analyze and plan to improve traffic congestion in real time. This requires a fundamental change in the way we need to model these complex physical systems and their interactions. The problem is challenging because transport in cities is governed by non-linear interactions with components that influence the system and vice versa. These components also adapt their behavior and decisions when the operators intervene in the system. At the same time congestion is not local and it spreads in time and space in a non-trivial way. All these combined with imperfect information for the users create limits both in the predictability and the control of the system.

The main objective of this work is to understand how the dynamics of congestion development and spreading can be observed and modeled in a way to allow real-time control by utilizing a very detailed dataset from 20,000 taxis in Shenzhen, China. We try to reveal the hidden information during the process of congestion formation. Specifically, we aim at exploring the spatiotemporal relation of congested links, observing and capturing congestion propagation from a macroscopic perspective, and finally identifying and quantifying critical congestion regimes to aid the design of hierarchical control strategies to improve mobility. While there is strong effort in the last 30 years for traffic flow models in one-dimensional systems (see (1) for an overview), literature in congestion spreading in large-scale networks mainly focuses on graphical representations of congestion without a metric or a dynamic model. This is due to the high complexity of the system. We plan to connect the formation of congestion by investigating the evolution of the maximum interconnected component (MIC) of congested roads in different regions of a city.

Literature in network level dynamics and congestion propagation is limited. It was recently observed from empirical data in downtown Yokohama (2) that by spatially aggregating the highly scattered plots of flow vs. density from individual detectors, the scatter almost disappeared and a well-defined Macroscopic Fundamental Diagram (MFD) exists between space-mean flow and density. The idea of an MFD with a critical accumulation belongs to Godfrey (3) and similar approaches were introduced later in (4), (5) and (6). The empirical verification of its existence with dynamic features is recent (2). This property is important for modeling purposes, as details in individual links are not needed to describe the congestion level of cities and its dynamics. It can also be utilized to introduce simple control strategies to improve mobility in homogeneous city centers building on the concept of an MFD, like in (7), (8) and (9). The main logic of the strategies is that they try to decrease the inflow in regions with points in the decreased part of an MFD. This work showed that efficient active traffic management schemes could be developed without a detailed knowledge of origin-destination matrices.

Despite these recent findings for the existence of MFDs with low scatter, these curves should not be a universal law. In particular, networks with an uneven and inconsistent distribution of congestion may exhibit traffic states that are well below the upper bound of an MFD and much too scattered to line along an MFD. Recent findings from empirical and simulated data (10), (11) have identified the spatial distribution of vehicle density in the network as one of the key components that affect the scatter of an MFD and its shape. They observed well-defined relations between flow and density when link vehicle density variance is about constant. They also observed low scatter relationships between network flow and variance of link density for a given network. In other words, the average network flow is consistently higher when link density variance is low for the same network density, but higher densities can create points below an MFD when they are heterogeneously distributed. The properties of a well-defined MFD, stability and observability are analyzed through many other simulation studies and experimental tests in (12)-(21) and others.

These findings are of great importance because the concept of an MFD can be applied for heterogeneously loaded cities with multiple centers of congestion, if these cities can be partitioned in a small number of homogeneous clusters. Recent work created clustering algorithms for heterogeneous transportation networks. The objectives of partitioning are to obtain (i) small variance of link densities within a cluster, which increases the network flow for the same average density and (ii) spatial compactness of each cluster which makes feasible the application of perimeter control strategies (18). The proposed mechanism can produce a partitioning with a desired number of clusters that has both small link density variances and spatially compact shapes, which are validated by both density variance metrics and spatial compactness metrics. Furthermore, by comparing with some other clustering algorithms this mechanism demonstrated superiority of both effectiveness and robustness in partitioning a real urban transportation network. The spatiotemporal propagation of congestion is highly related with the type of partitioning and the application of the appropriate control strategy. Urban systems experience highly dynamic behavior and different traffic patterns may arise for different times of day (think of morning-evening commuting patterns or stochastic variations of traffic flow). In these cases, very likely one needs to identify different optimal sets of clusters depending on these patterns.

The work in (18) has laid a solid foundation for static partitioning in transportation networks and the current paper investigates the traffic propagation by exploring the spatial and temporal features of congestion and their correlations. If dynamic partitioning could describe the spatiotemporal characteristics of congestion, then perimeter control strategies for multiple regions (see for example (8), (9)) can be extended to capture dynamic boundaries. The remainder of this paper is organized as follows. Section 2 estimates the dynamic traffic conditions of Shenzhen city for a typical day in China. Static partitioning algorithms proposed in (18) are also applied to the new network for the traffic condition at a specific time period. Section 3 demonstrates the methodology proposed to observe and capture congestion propagation. Section 4 presents the empirical macroscopic fundamental diagrams of the congested regions observed in propagation and Section 5 extends and discusses the MFD for a whole day. Finally, Section 6 concludes our work and discusses potential issues to be studied in the future.

## **DATA PROCESSING AND ANALYSIS**

The data contains the network structure and daily GPS data of taxis for one month in the city of Shenzhen in China. Shenzhen is a major city in the south of Southern China's Guangdong Province, situated immediately north of Hong Kong. The area became one of the most successful Special Economic Zones (SEZs) in China established in 1979, before which it was only a small village. The rapid foreign investment created one of the fastest-growing cities in the world with a population close to 11 million and, as expected large congestion problems both in the urban and freeway system of the city. There are 11125 nodes connected by 12305 links. The number of taxis is around 20 thousand and varies from day to day. Each day, the taxis make records of current time, GPS locations in form of coordinates and passenger information. The time of recording is random and the length of interval between two records ranges from 1 second to several hours. In total there are around 30 to 50 million records per day.

Data processing is the first and critical step of data analysis. It plays an essential role by extracting useful information and facilitating the work of later stages in observing the underlying patterns and deriving insightful results. Thus the final messages inferred and phenomena revealed are directly related to the quality of data and the effectiveness of the processing methods. In this section, we describe the raw data and the data processing methods used to obtain the hidden information. In addition, the static partitioning methodology proposed by (18) is applied to the new network. We show that the algorithms work well as expected in segmenting a network into regions with different congestion levels for a static time period. However, it is also unveiled that they are inefficient in studying the propagation of congestion in a dynamic network where traffic condition evolves as time passes.

For the purpose of studying dynamic traffic conditions, we estimate the link speeds in specified regions. The speed of each link is estimated as its average speed during a 15 minute interval based on the following formula:

$$v_l^t = \sum_{i \in P_{l,t}} s_i / \sum_{i \in P_{l,t}} h_i \quad (1)$$

$v_l^t$  denotes the average speed of link  $l$  during the time period  $[t, t+15]$ .  $i$  is the identity of a taxi and  $P_{l,t}$  is the set of taxis that pass link  $l$  during time period  $[t, t+15]$ .  $s_i$  and  $h_i$  are respectively the total distance traveled and time spent by taxi  $i$  on link  $l$  during time period  $[t, t+15]$ . The estimation is repeated every 5 minutes throughout the day.

The quality of the estimation is improved by several means. Firstly, in order to reflect the demand more appropriately, only  $s_i$  and  $h_i$  with passengers inside taxi  $i$  are considered, which implies that the estimated link speed is calculated based on passenger travel distance and time. Secondly,  $s_i$  and  $h_i$  are obtained based on the passenger trips less than 5 minutes. Longer trips are cut into minimum number of small trip segments, each of which has a time duration less than 5 minutes.  $s_i$  is calculated as the shortest path between the beginning and ending positions of each small trip in the network. Since full vehicle trajectory is not always available, the segmented small trips can better approximate the real distance traveled than a long trip based on the shortest path. Thirdly, with either  $s_i = 0$  or  $h_i = 0$ , the pair  $(s_i, h_i)$  is not used for the sake of possible hardware faults in clock timing or GPS positioning. Finally, the GPS position is mapped to the closest link in the network and a record is considered only when its mapping error is less than 50 meters.

Figure 1 shows the trajectories of 3 different taxis with identities X, Y and Z (anonymous ID name for data privacy reason) during the whole day of 01.09.2011. The network structure is shown in green. The whole map can be observed in Figure 1.(c), while Figure 1.(a) and Figure 1.(b) respectively zoom in the east and central parts of the city. The blue lines are the trajectories and the small red points are the GPS positions of taxis. It is shown that the trajectories of the taxis can be clearly traced with GPS positions. However it is also observed that the trajectories are not always continuous, due to either the failure of mapping (i.e., error above 50 meters) or the intermittency of GPS recording. We can also see that the activities of some taxis are restricted in a certain area while others can travel through the whole city. This observation is consequent to the local policy that there are two kinds of taxis: red and green. The red ones are allowed to travel in both inner zone and outer zone of the city, while the green ones are only allowed to travel the outer zone.

(place Figure 1 approximately here)

Based on the above method of estimating link speeds, the network condition at 8:30am is obtained and shown in Figure 2.1, with white link indicating higher speed and black ones for lower speed. It is consistent with the reality that the upper and lower regions of Shenzhen are more congested than the middle part, due to the fact that there are more main roads in middle part. In this data we apply the

static partitioning algorithm developed in (18) for transportation networks and the Normalized Cut algorithm (19), which is a state-of-the art clustering algorithm that reference (18) has extended/adjusted for traffic data. Figure 2.2 shows the partitioning of Normalized Cut in (19) and Figure 2.3 shows the algorithm developed in (18) by the same authors as in this paper. First from the visual perspective, our static partitioning can better differentiate and segment regions with different traffic conditions. Secondly, two main metrics evaluating the quality of partitioning is shown in Table 1. “Silhouette” measures the ratio of intra-cluster dissimilarity over inter-cluster dissimilarity, while the “Mean difference” from another aspect tells the difference among the clusters obtained. A good quality cluster has small “Silhouette” value and a large “Mean difference” value as the objective is that different clusters have significantly different value of mean speed (to highlight different levels of congestion). Thus it is shown that our static partitioning algorithm outperforms against the classic algorithm of Normalized Cut in the new network analyzed. More details and discussion on static partitioning can be found in (18).

(place Figure 2 approximately here)

(place Table 1 approximately here)

Regarding dynamic traffic conditions, we focus on studying the upper side of Shenzhen (the blue cluster in Figure 2.3) where at 8:30am the congestion has formed to some degree, seen from Figure 2.1. We estimate and show the average speed of links in the region on Thursday 01.09.2011 from 1am to 11pm in Figure 3.1. It can be seen clearly that the sharpest decline of the speed occurs during the time period from 6am to 8am, which is the typical morning rush period in China. The blue curve in Figure 3.2 shows the time series of the total number of taxis in the upper region of Shenzhen and an increasing part is observed during the morning rush hour as expected. Similarly as speed, the number of taxis is also estimated every 5 minutes and for each time interval of 15 minutes. The estimation is based on the assumption that a taxi stays in the specified region for the whole interval as long as one GPS record shows the existence of the taxi in the region during the time interval. This assumption is reasonable thanks to the large extent of the region. We also estimate for each taxi the percentage of time spent as full and empty passenger during the time interval in the region. Due to external faults or GPS positioning errors as mentioned earlier, the complete trajectory of a taxi can not always be obtained. Thus for a taxi, the sum of the percentage of full time and empty time for a time interval is less than or equal to 1. The red curve in Figure 3.2 is the number of full taxis estimated by adding the percentage of full time of all taxis during the time interval in the upper region. The green one is the number of empty taxis. It is shown that the trend of the number of full and empty taxis can reflect the passenger demand while the sum of the two is less than the real total number of taxis (the blue curve) for the reason explained above. However, we will utilize the ratio of the number of full over empty taxis to estimate the total number of taxis in several smaller subregions in Section 4.

(place Figure 3 approximately here)

As discussed in (18), tracking congestion propagation and dynamic partitioning have been a difficult task. Boundary adjustment algorithms based on an initial static partitioning have been proposed by (18) to observe the shrinking and growing of congestion. Conceptually speaking, after a partitioning is obtained that maximizes the objective of intra-similarity and inter-dissimilarity among regions, a boundary adjustment is performed to locally improve the estimation. The reasoning behind this is that the first steps of the algorithm might not capture local congestion phenomena. However this process can not be properly or easily implemented in reality for mainly the following reasons. First of all, the results obtained will be strongly dependent on the initial states. Choosing an appropriate time period for initial partitioning is difficult, especially at the beginning of congestion formation where the network is usually not partitionable. In addition, a boundary adjustment algorithm will not be able to

observe the development of new pockets of congestion. Finally, the computational complexity of boundary adjustment algorithms is high and thus can be hardly developed in a real system. Another straightforward approach is to repeatedly apply the static partitioning method at each time period independently.

We take a closer look at the link speed distribution of the upper region for a 15min interval starting at 6:00am, 7:00am, 7:30am and 8:00am, in both histograms (first column) and grayscale contour plots (second column - dark colors show lower speeds) in Figure 4. Both show that the network condition is decreasing, with particularly the histogram of link speeds gradually shifting towards the left side (i.e., lower speeds). However, it can not be observed clearly from the grayscale graphs how the congestion grows to the final state. The results of partitioning corresponding to each time period based on (18) are shown in the third column in Figure 4. Unfortunately, the graph sequence obtained by static partitioning at each time step can not tell us much on congestion propagation. There are several reasons. First of all, the static partitioning algorithms aim at clustering the most similar links instead of congested links. Thus it can not reflect the physical logic behind the formation of congestion. For instance, all the congested links will probably be separated into different groups due to the fact that they have different congestion levels. Second, there is no time correlation among the static partitioning, implying that the current partitioning is unconditional on the previous one. Thus a continuous and coherent congestion propagation can be hardly observed based on the time series profiles of link speeds. These aspects on the other hand provide inspirations on developing new methodologies that take into consideration both the physical and spatio-temporal properties and correlations of congestion.

(place Figure 4 approximately here)

## PROPAGATION MODELING AND OBSERVATION

The model to observe the congestion propagation is proposed as follows. Firstly, the network  $G$  is built as a graph  $G = (V, E)$  where  $V$  and  $E$  are the sets of nodes and edges respectively. Each node  $u \in V$  represents a link in the network, with a value assigned by the estimated speed. Each edge  $e(u, v) \in E$  indicates that two links  $u$  and  $v$  are directly connected if they are spatially adjacent in the network (i.e., two links having a common intersection). Secondly, a threshold  $r$  which is around 1/3 of the free flow speed of each link is set to differentiate between congested and uncongested links, which is consistent with data from Yokohama (2). Thus the links fall into two groups: congested links in  $C$  with speed less than or equal to  $r$  and uncongested ones in  $U$  above  $r$ . Here we do a step of spatial smoothing to better observe congestion propagation. Specifically, after searching for all the links in  $U$  that have more neighbor links (i.e., links that are spatially adjacent) that are in  $C$  than in  $U$ , we move them into group  $C$  from  $U$ . Simply speaking, an uncongested link is remarked as a congested one if it has more congested neighbors than uncongested ones. The spatial smoothing operation can also be modified based on specific needs. It can be either omitted or the “neighbor links” can be generalized into “neighbor links of degree  $d$ ”. In our case, we define the neighbor links as two links having a distance 1 (distance calculated as the minimum number of edges passed between two nodes, namely shortest path) in graph  $G = (V, E)$ . Thus “neighbor links of degree  $d$ ” can be redefined as two links having a shortest path of length  $d$ . We expect that this smoothing will not alter the physical properties of the system (as only less than 3% of the links change from  $U$  to  $C$ ), but it will disregard location errors or travel time stochasticity due to small sample, e.g. only one taxi in a link during 15min. Finally, we find the maximum connected components (MCCs) of the congested links in  $C$ . The algorithms for locating the connected components use either breadth-first search or depth-first search in a graph. The computational complexity of locating all the connected components of congested links in the network is thus  $O(n)$  in linear time. Denote by  $MCC_i(t)$  the  $i$ -th largest component during time  $t$ , where  $t$  varies between 1 and 24 (24 time intervals between 6:00am and 8:00am).

Connected component is a concept in graph theory. A connected component of the congested links is defined to satisfy the following four conditions.

1. The connected component  $G' = (V', E')$  is a subgraph of the network  $G = (V, E)$ , where  $V' \subseteq V, E' \subseteq E$ .
2.  $V' \subseteq C$ , namely the connected component is composed of only congested links.
3.  $G'$  is a connected graph, which means that there is always a path between any pair of nodes  $u, v \in V'$ .
4. No additional node can be added into  $G'$  without breaking the conditions 1, 2 and 3.

The largest six connected components  $MCC_i(t)$ ,  $1 \leq i \leq 6$  and the upper region network are shown in different colors in the order of red, green, blue, cyan, magenta and black in Figure 5. The network is in yellow. Figure 5 demonstrates the evolution of the  $MCC_i(t)$  as time passes during the rush hour, with 4 snapshots at representative time periods. These periods are (left-right, up-down): 6:30am, 7:00am, 7:15am, 7:40am. We pick up these time periods from 6-8am to show the most typical phenomena of congestion growing during morning rush hour. Note that congestion shrinking may also happen for a short period of time but is not shown here due to space limit. This also happens during 8am-3pm when demand is lower. It is clear from Figure 5 that  $MCC_1$  (red) grows by “eating” the other small components as time passes and gradually dominates the network. Another interesting observation is that the first 3 largest components of congestion  $MCC_i(t)$ ,  $1 \leq i \leq 3$  include more than 90% of the congested links of family  $C$ . This means that in congested city centers a few pockets of congestion can mainly describe the development and evolution of traffic conditions. This result is not surprising because in urban networks there is significant spatial correlation between adjacent links. Congestion is not a spatially random process, but it grows closely to areas/links that are already congested.

Figure 6.1 shows the time series of the ratio of the speed of different parts of the network over the maximum speed. The red curve is for the maximum component  $MCC_1$  (the red cluster in Figure 5), while the green one is for the rest of the network excluding  $MCC_1$ . The blue curve is for the whole network. The three speed evolution profiles show that the maximum component can always clearly differentiate with the rest of the network with respect to traffic conditions. Figure 6.2 shows the evolution of the size of  $MCC_1$  (red cluster in Figure 5) by red curve,  $MCC_2$  (green cluster) by green curve and  $MCC_3$  (blue cluster) by blue. It is obvious that the sizes of all the other components in the network are ignorable compared to  $MCC_1$ . Due to the fact that the evolution of the size of  $MCC_1$  to a large extent dominates the other components in the network, it is sufficient to use only  $MCC_1$  (the red cluster) to describe the congestion propagation. Therefore, the maximum component effectively captures and efficiently highlights the evolution of the congested area in all three dimensional senses: time, space and speed. Figure 6.3 demonstrates a monotonically increasing relation between the size of  $MCC_1$  and the size of  $C$  (i.e., number of congested links in the whole upper region network). We observe that when the number of congested links reaches around 75% of the total number of links in the network, the size of  $MCC_1$  reaches half of the network. Currently we are investigating the effect of different network structures on this relation.

(place Figure 5 approximately here)

(place Figure 6 approximately here)

## MACROSCOPIC FUNDAMENTAL DIAGRAM

Macroscopic fundamental diagram (MFD) describes the capabilities of a network by associating the performance (in terms of flow or vehicle kilometers traveled namely VKT) with the number of vehicles in the network (2). Previous research showed that for homogeneously congested regions this relation is not very sensitive to detailed demand profile and can be used to predict traffic conditions and design control strategies. We study the MFD of each  $MCC_1$  (red component) in Figure 5 by estimating its speed (average speed of all the links in  $MCC_1$ ) and total number of taxis. While the speed of full taxis can be considered a good proxy for the speed of the normal vehicles in the network, it is not straightforward how the density of taxis can be directly associated with the density of cars. Previous research (2) utilized taxi and loop detector data to estimate the density, but detector data is



not available in this study. The speed estimation process is explained in Section 2 and the total number of taxis is estimated by the following formula:

$$n_j = n_u \times (l_j/l_u) \times ((n_j^f/n_j^e)/(n_u^f/n_u^e)) \quad (2)$$

where  $j$  denotes the subgraph of the network for  $MCC_1(t)$  and  $u$  denotes the whole network of the upper Shenzhen region (shown in Figure 5). Other notations have the following meanings:  $n_j$ , the total number of taxis in  $j$ ;  $n_u$ , the total number of taxis in  $u$  (i.e., the blue curve estimated in Figure 3.2);  $l_j$  and  $l_u$ , the number of links in  $j$  and  $u$ ;  $n_u^f$  and  $n_u^e$ , the number of full and empty taxis in  $u$  (i.e., the red and green curves estimated in Figure 3.2);  $n_j^f$  and  $n_j^e$ , the number of full and empty taxis in  $j$  estimated by the same method for  $u$  explained in Section 2. The above approximation is based on two assumptions. Firstly, the number of taxis is proportional to the size of the region. Secondly, it is also proportional to the ratio of full over empty taxis. Note that  $n_j$  is estimated by Equation 2 instead of the same way as  $n_u$ , because the method for estimating  $n_u$  is appropriate only when  $u$  is a large region. Besides, the ratio  $n_j^f/n_j^e$  is justified by most of the GPS positioning that is accurate, though  $n_j^f + n_j^e \leq n_j$  due to the existence of GPS errors or external faults explained in Section 2. The physical reasoning is that as the ratio of full taxis increases in a region of a city, this implies that this region also attracts a large number of cars. Analysis of similar taxi data combined with loop detectors in Yokohama (2) supports such an approximation.

Figure 7.1 shows the average flow vs. density (i.e., MFD) for several different  $MCC_1$  during 1-8am. Based on these MFDs, we calculate their critical time, critical density, maximum flow and critical speed. Critical density is estimated as the number of taxis in the region divided by the length of the region  $l$  (number of links) when flow is maximum. Critical time is the first time density exceeds the critical threshold, while maximum flow is also divided by the length of the region. These dimensionless transformations are necessary to compare regions of different sizes. Figure 7.1 shows that all the  $MCC_1$ s reach the similar critical density, flow and speed, which defines the same congestion state for all the regions. Thus congestion propagation could be captured and quantified. It also highlights that the MFD for the greater region of Shenzhen has similar properties in different parts of the urban network. Note that the critical speed is slightly higher for the outer regions, which might be due to the topology of the network (longer links and faster roads). It suggests that the small region first reaches the critical state of congestion and later propagates and becomes a larger region. It is clear that they experience different degrees of congestion regime. The MFD of the large region has a longer period of increasing flow than small regions before it starts decreasing, which also implies the congestion propagates from small regions to large ones.

Figure 7.2 shows the critical time for each  $MCC_1(t)$ ,  $6 \leq t \leq 24$ . The results for  $MCC_1(t)$ ,  $1 \leq t \leq 5$  are ignored because the  $MCC_1$  is too small at these time periods. It reveals that there are around 3 critical time periods that characterize congestion propagation. Regions  $MCC_1(6)$  and  $MCC_1(7)$  reach congestion at around 6:10am,  $MCC_1(12)$  and  $MCC_1(13)$  at 6:20am while the  $MCC_1(t)$ ,  $15 \leq t \leq 24$  at 6:40am. The critical time of the regions are not always increasing as some  $MCC_1$  shrinks from the previous one such as  $MCC_1(10)$ . Furthermore, due to the shrinking phenomena, the size of  $MCC_1(t)$  around  $t = 10, 11$  is too small and thus the critical values are not very stable or significant at that time. The fact that some regions have the same critical time (i.e.,  $MCC_1(15)$  and  $MCC_1(24)$ ) does not mean that they are exactly the same area with same size. It implies that a small number of uncongested links at  $MCC_1(15)$  become congested (to help form larger  $MCC_1$ ) later at  $MCC_1(24)$  without changing much the overall traffic conditions. Note also that the critical time that a region reaches congestion is earlier than when this region creates a large connected component (compare for example the times in Figure 5, with the times in Figure 7.2). We further investigated this and we identified that a region becomes congested (decreasing point in the MFD) when about 60-70% of links are congested. During this time, these congested links do not form a large connected component, but this occurs later when 90-95% of links are congested. Finally we can summarize that the congestion propagates from the red

cluster in Figure 5.1 to the one in Figure 5.4 in around half an hour from 6:10am to 6:40am, following the patterns observed in Figure 5.

(place Figure 7 approximately here)

## EXTENSIONS

We show the MFD for region  $MCC_1(24)$  by integrating all the estimation from 1am to 3pm in Figure 8. The green part is from 1am-6am and it shows that the output of the network is increasing in early morning. The red part is from 6am-8am where the turning point of the MFD occurs and is already studied in detail in the previous Sections. The blue part is from 8am-3pm where we can see that the MFD is trying to climb back to the maximum point. This is consistent with Figure 3.1 which shows that the speed is increasing from 8am to 1pm, though not yet reaching the maximum speed before it starts decreasing again after 1pm.

The MFD for the region is shown by utilizing the total number of taxis, based on the assumption that the number of taxis is constantly proportional to the total number of vehicles. However, this assumption may not hold through 24 hours. As seen from Figure 3, though the speed keeps decreasing after 3pm, both the number of full and total taxis keep almost constant. This implies that the proportion of other vehicles especially private cars is increasing and thus contributing more to the traffics in the evening hours. Thus the MFD based solely on taxis is not expected for a whole day unless further information is available to estimate the total number of vehicles in the city, e.g. from loop detector data.

(place Figure 8 approximately here)

## DISCUSSION

The work in this paper has laid a solid foundation for understanding with sparse probe vehicle data how congestion is developed and propagates in large urban systems. By identifying interconnected congested links, the critical pockets of congestion have been identified and Macroscopic Fundamental Diagrams were observed with low scatter. An important contribution is that we connect the formation of congestion by investigating the evolution of the maximum connected component of congested roads in different regions of a city. To estimate density of cars more accurately, taxi GPS info is not enough and has to be associated with loop detector data. In this case an approach similar to (2) can be applied. Nevertheless, our approximation of Equation 2 gives realistic results for the space and time location of congestion (as expressed with decreasing parts of MFD). We also observe that the ratio of full and empty taxis varies smoothly with time and our approximation is robust.

Ongoing work continues to study the traffic propagation by exploring the spatial and temporal features of congestion and their correlations for different city structures by utilizing different sensor data. We are also quantifying how network variables for robustness and redundancy are related with the network capacity and evolution of MCC for different demand profiles (e.g. network connectivity, betweenness centrality, road density, circuitry and others). While there is a vast literature in topology and robustness of graphs, these variables have not been associated with traffic congestion (distance shortest path is used). We model the formation of congestion by investigating the evolution of the MCC of congested roads in different regions of a city. A probabilistic model of propagation (similar to disease spreading) is being developed if one considers that a road close to congested ones has higher probability to also become congested compared to a road far from congestion. Based on these findings, we are currently working on designing control strategies for the heterogeneous network with different levels of congestion and dynamic partitioning. Another research priority is to investigate partitioning and perimeter control strategies for networks with very heterogeneous topology or hierarchical structure (e.g., mixed networks of grid arterials and freeway systems).

## ACKNOWLEDGMENT

The authors are grateful to Prof. Jun Luo from SIAT in Shenzhen for the continuous data support and for hosting Yuxuan Ji at SIAT in 2012.

## References

- (1) Helbing, D. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, 73, 2001, pp. 1067-1141.
- (2) Geroliminis, N., and C. Daganzo. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. *Transportation Research Part B*, Vol. 42, No. 9, 2008, pp. 759-770.
- (3) Godfrey, J. The mechanism of a road network. *Traffic Engineering and Control*, Vol. 11, No. 7, 1969, pp.323-327.
- (4) Herman, R. and I. Prigogine. A two-fluid approach to town traffic. *Science*, Vol. 204, No. 4389, 1979, pp. 148-151.
- (5) Mahmassani, H., J. Williams and R. Herman. Performance of urban traffic networks. In *10<sup>th</sup> International Symposium on Transportation and Traffic Theory*, Amsterdam, the Netherlands, 1987.
- (6) Daganzo, C. Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B*, Vol. 41, No. 1, 2007, pp. 49-62.
- (7) Keyvan-Ekbatani, M., A. Kouvelas, I. Papamichail and M. Papageorgiou. Exploiting the fundamental diagram of urban network for feedback-based gating. *Transport Research Part B*, Vol. 46, No.10, 2012, pp. 1393-1403.
- (8) Aboudolas, K., and N. Geroliminis. Perimeter and boundary control in multi-reservoir heterogeneous networks. *Transportation Research Part B*, Vol. 55, 2013, pp. 265-281.
- (9) Geroliminis, N., J. Haddad and M. Ramezani. Optimal Perimeter Control for Two Urban Regions with Macroscopic Fundamental Diagrams: A Model Predictive Approach. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No.1, 2012, pp. 348-359.
- (10) Geroliminis, N., and J. Sun. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B*, Vol. 45, No. 3, 2011, pp. 605-617.
- (11) Mazloumian, A., N. Geroliminis and D. Helbing. The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of Royal Society A*, Vol. 368, No. 1928, 2010, pp. 4627-4648.
- (12) Gayah V.V. and V.V. Dixit. Using mobile probe data and the macroscopic fundamental diagram to estimate network densities. *Transportation Research Record*, Vol. 2390, 2013, pp. 76-86.
- (13) Buisson, C., and C. Ladier. Exploring the impact of homogeneity of traffic measurements on the existence of Macroscopic Fundamental Diagrams. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2124, 2009, pp. 127-136.
- (14) Ji, Y., W. Daamen, S. Hoogendoorn, S. Hoogendoorn-Lanser, and X. Qian. Macroscopic fundamental diagram: investigating its shape using simulation data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2161, 2010, pp. 42-48.
- (15) Mahmassani, H., M. Saberi, and A. Zockaie. Urban network gridlock: Theory, characteristics, and dynamics. *Transportation Research Part C*, Vol. 36, 2013, pp. 480-497.
- (16) Knoop, V., S. Hoogendoorn and H. van Lint. The impact of traffic dynamics on the Macroscopic Fundamental Diagram. In *92<sup>nd</sup> Annual Meeting of the Transportation Research Board*, Washington D.C., USA, 2013.
- (17) Gayah, V., and C. Daganzo. Clockwise hysteresis loops in the Macroscopic Fundamental Diagram: An effect of network instability. *Transportation Research Part B*, Vol. 45, No. 4, 2011, pp. 643-655.
- (18) Ji, Y., and N. Geroliminis. On the spatial partitioning of urban transportation network. *Transportation Research Part B*, Vol. 46, No. 10, 2012, pp. 1639-1656.
- (19) Shi, J., and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.22, No.8, 2000, pp.888-905.
- (20) Leclercq L., N. Chiabaut and B. Trinquier. Macroscopic Fundamental Diagrams: A cross-comparison of estimation methods. *Transportation Research Part B*, Vol. 62, 2014, pp. 1-12.
- (21) Leclercq L. and N. Geroliminis. Estimating MFDs in simple networks with route choice. *Transportation Research Part B*, Vol. 57, 2013, pp. 468-484.

Figure 1 Taxi trajectories (←North) .....	13
Figure 2 Static partitioning.....	14
Figure 3 Evolution of speed and number of taxis in upper region of Shenzhen .....	16
Figure 4 Link speed distribution and independent static partitioning at different time periods .....	18
Figure 5 Congestion propagation observed during the morning rush hour.....	18
Figure 6 Evolution of speed and size of components.....	19
Figure 7 Macroscopic fundamental diagrams and critical time .....	20
Figure 8 MFD for MCC1(24) estimated during 1am-3pm .....	21
Table 1 Metrics evaluating partitioning .....	15

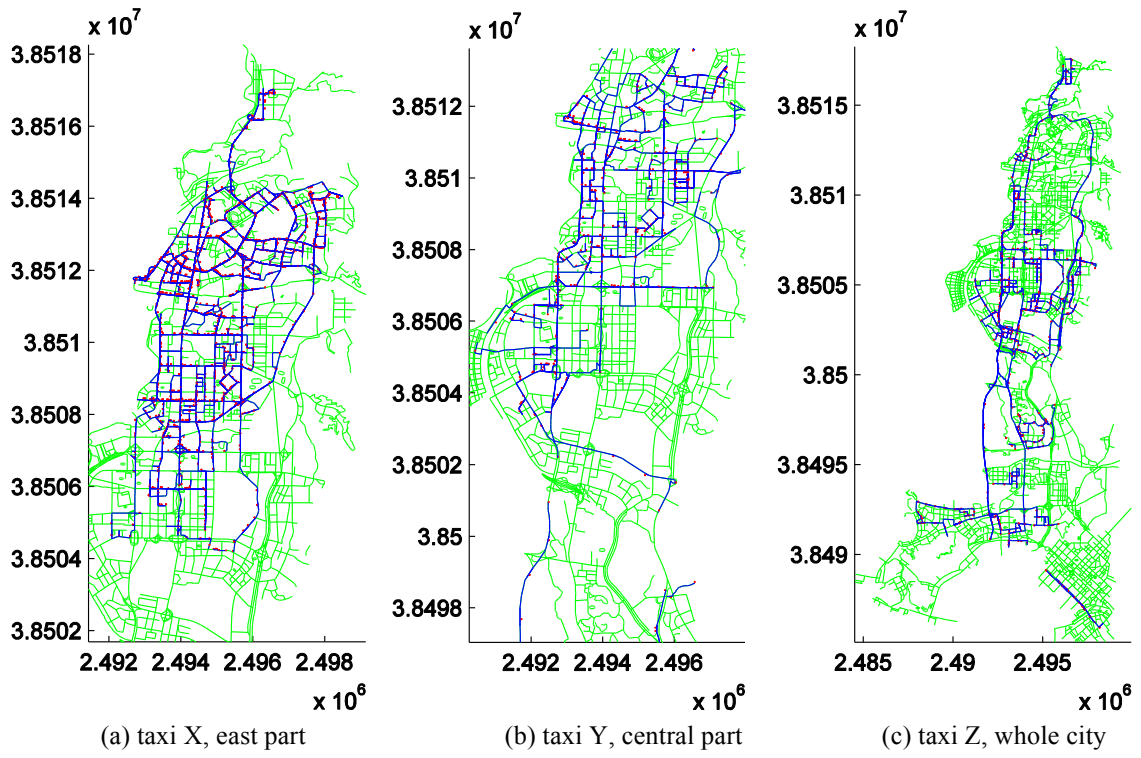


Figure 1 Taxi trajectories (←North)

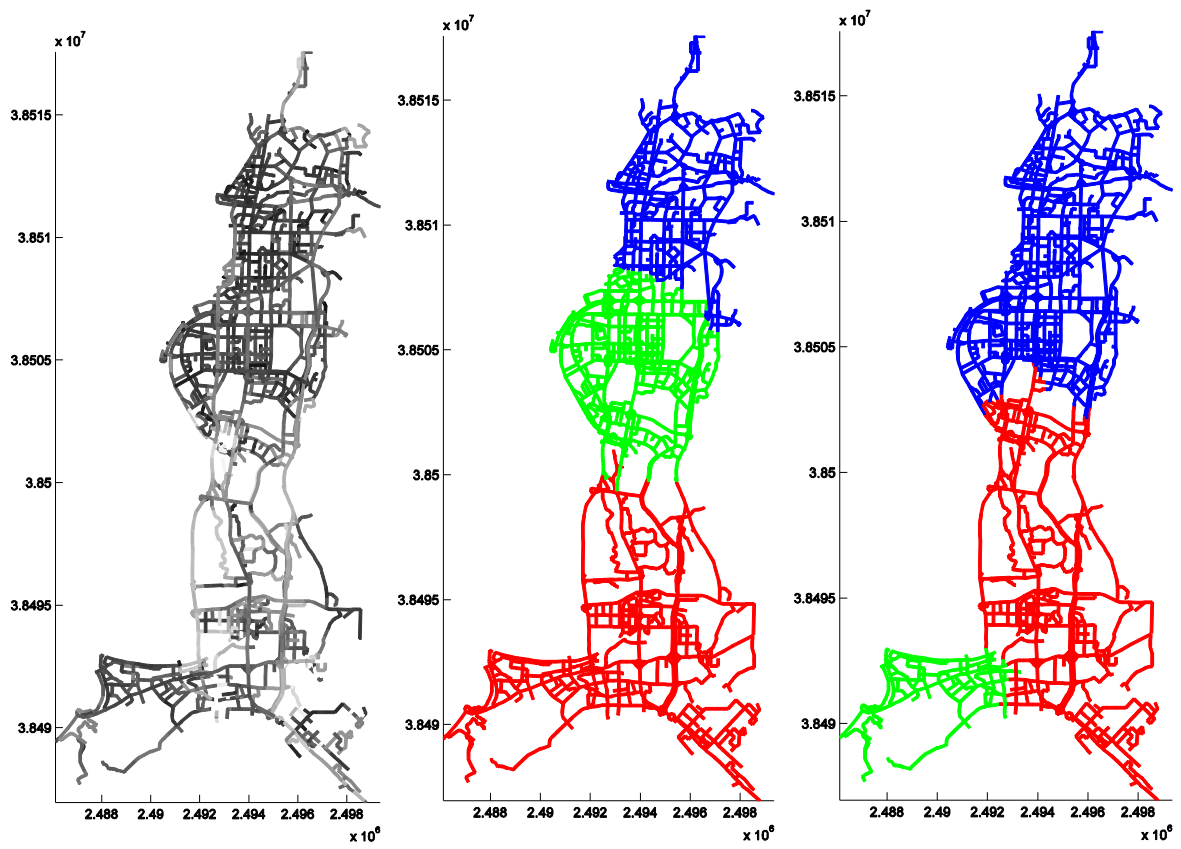


Figure 2 Static partitioning

Table 1 Metrics evaluating partitioning

Metrics	Normalized Cut	Our static partitioning
Silhouette	0.9745	0.6306
Mean difference (m/s)	2.9650	5.5992

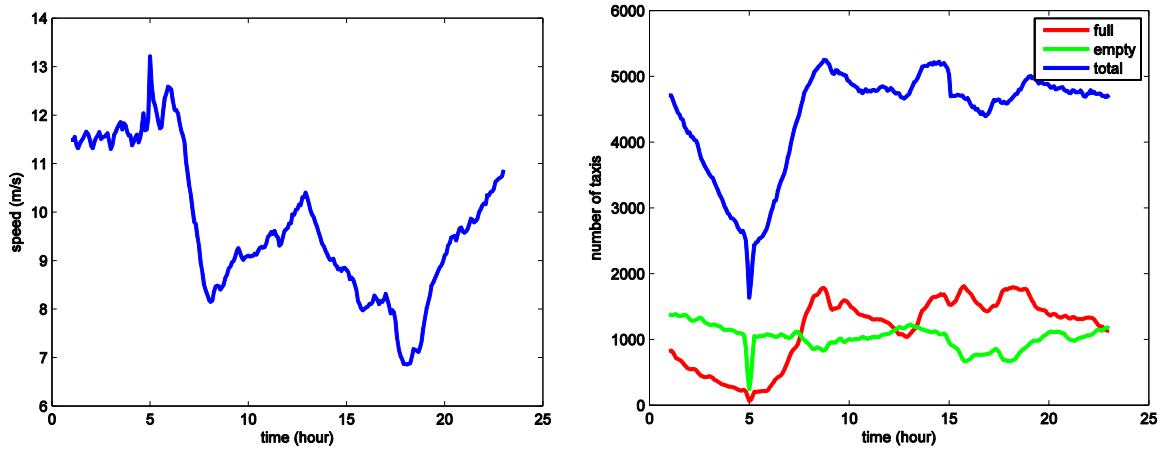


Figure 3 Evolution of speed and number of taxis in upper region of Shenzhen



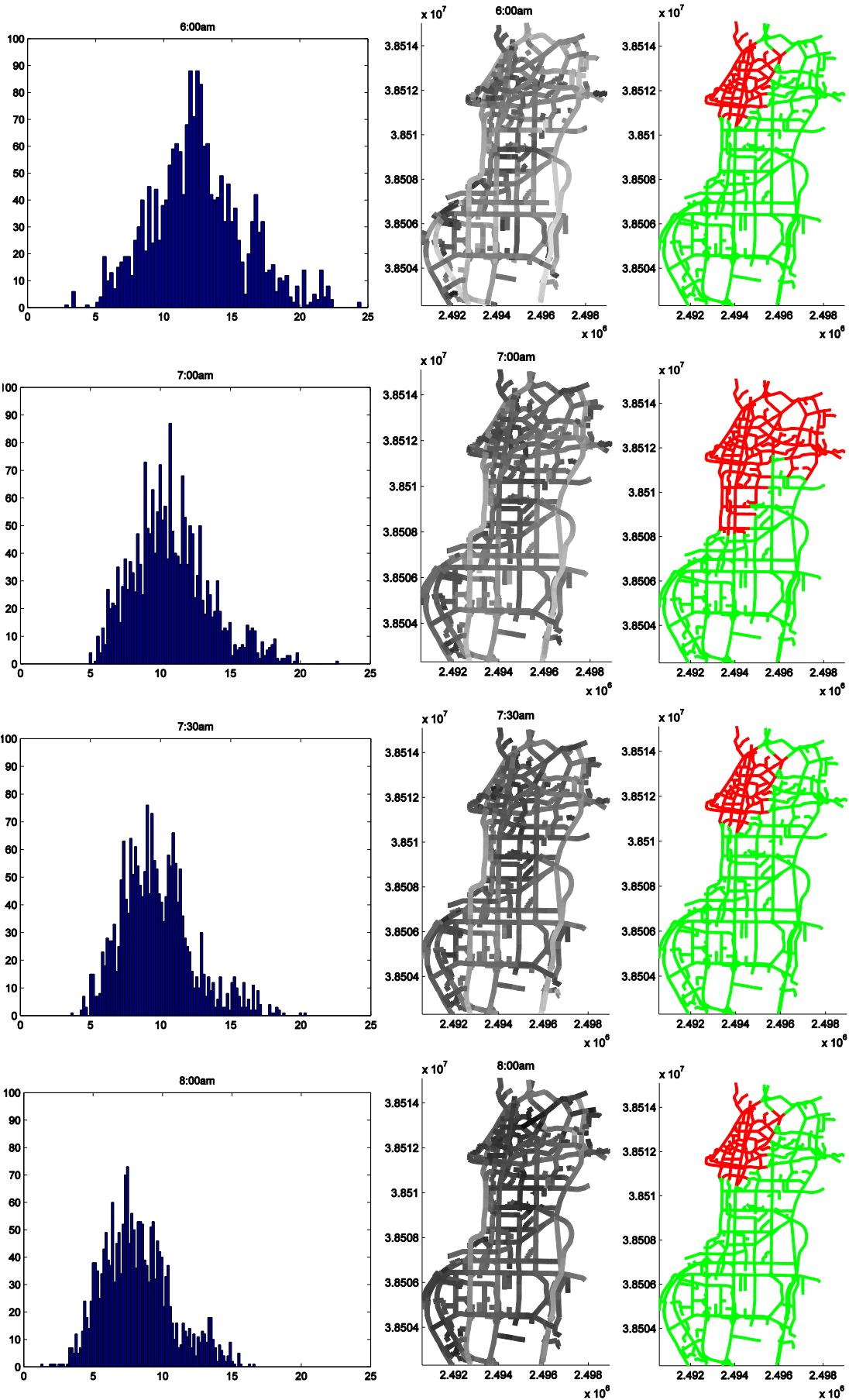


Figure 4 Link speed distribution and independent static partitioning at different time periods

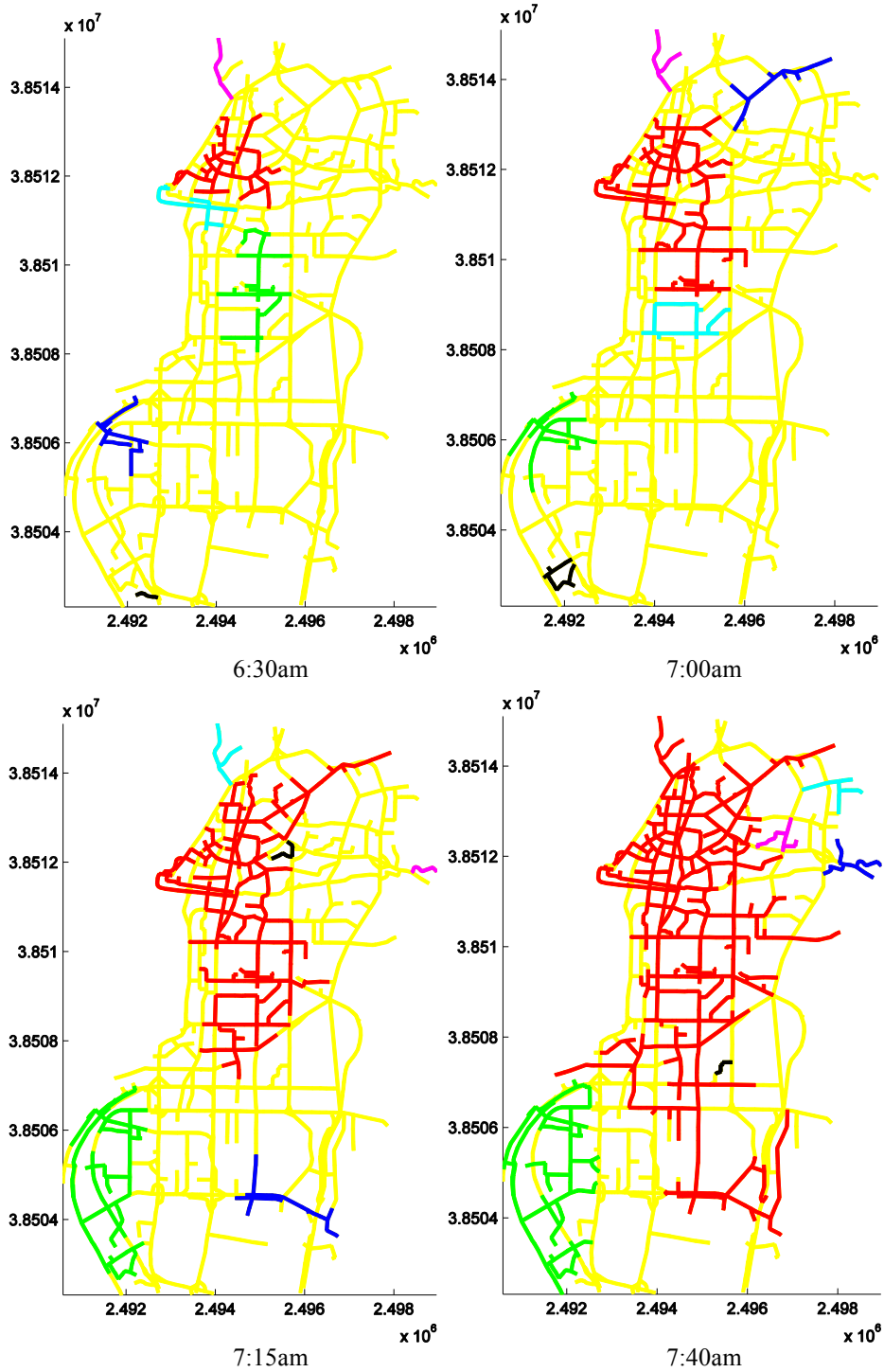


Figure 5 Congestion propagation observed during the morning rush hour

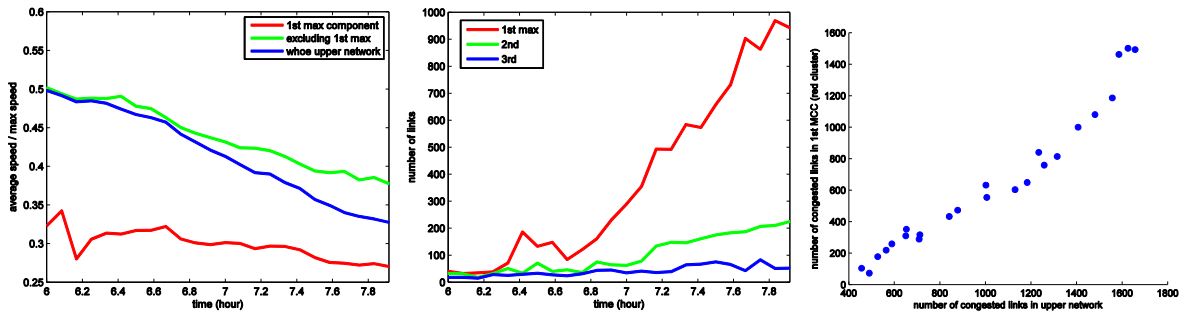


Figure 6 Evolution of speed and size of components

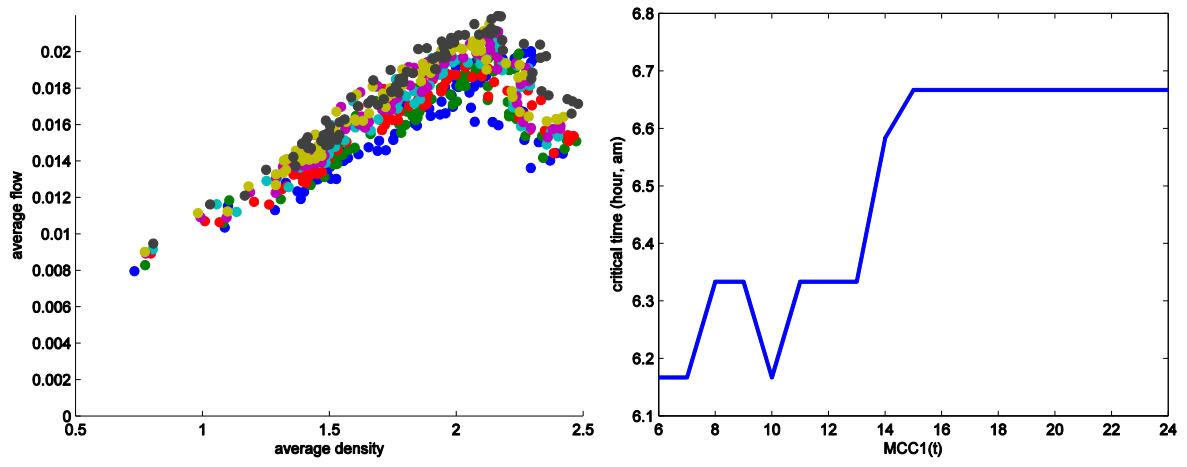


Figure 7 Macroscopic fundamental diagrams and critical time

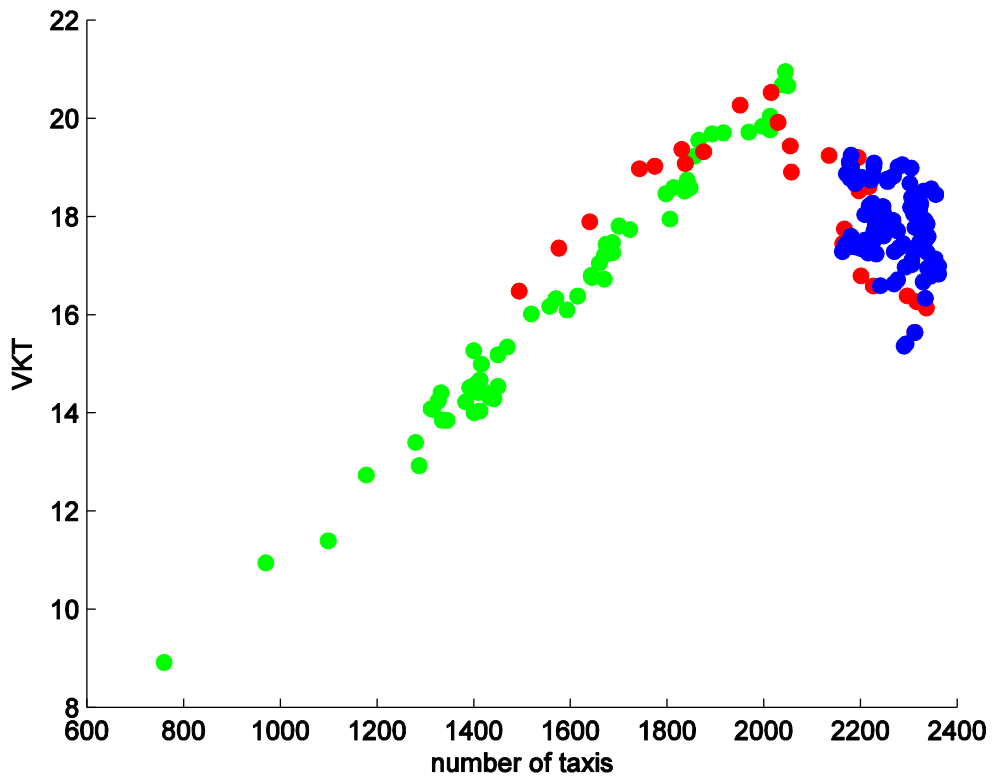


Figure 8 MFD for  $MCC_1(24)$  estimated during 1am-3pm