

Congestion Propagation Based Bottleneck Identification in Urban Road Networks

Changle Li , Senior Member, IEEE, Wenwei Yue , Student Member, IEEE, Guoqiang Mao , Fellow, IEEE, and Zhigang Xu , Member, IEEE

Abstract—Due to the rapid development of vehicular transportation and urbanization, traffic congestion has been increasing and becomes a serious problem in almost all major cities worldwide. Many instances of traffic congestion can be traced to their root causes, the so-called traffic bottlenecks, where relief of traffic congestion at bottlenecks can bring network-wide improvement. Therefore, it is important to identify the locations of bottlenecks and very often the most effective way to improve traffic flow and relieve traffic congestion is to improve traffic situations at bottlenecks. In this article, we first propose a novel definition of traffic bottleneck taking into account both the congestion level cost of a road segment itself and the contagion cost that the congestion may propagate to other road segments. Then, an algorithm is presented to identify congested road segments and construct congestion propagation graphs to model congestion propagation in urban road networks. Using the graphs, maximal spanning trees are constructed that allow an easy identification of the causal relationship between congestion at different road segments. Moreover, using Markov analysis to determine the probabilities of congestion propagation from one road segment to another road segment, we can calculate the aforementioned congestion cost and identify bottlenecks in the road network. Finally, simulation studies using SUMO confirm that traffic relief at the bottlenecks identified using the proposed technique can bring more effective network-wide improvement. Furthermore, when considering the impact of congestion propagation, the most congested road segments are not necessarily bottlenecks in the road network. The proposed approach can better capture the features of urban bottlenecks and lead to a more effective way to identify bottlenecks for traffic improvement. Experiments are further conducted using data collected from inductive loop detectors in Taipei road network and some road segments are identified as bottlenecks using the proposed method.

Index Terms—Bottleneck identification, congestion propagation, bottleneck validation, graph theory, maximal spanning tree, SUMO.

Manuscript received June 25, 2019; revised September 10, 2019 and January 4, 2020; accepted February 5, 2020. Date of publication March 20, 2020; date of current version May 14, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant U1801266, in part by the National Key R&D Program of China under Grant 2019YFB1600100, in part by the Key R&D Program of Shaanxi under Grants 2018ZDXM-GY-038 and 2018ZDCXL-GY-04-02, in part by the Youth Innovation Team of Shaanxi Universities, and in part by the Science and Technology Projects of Xi'an, China under Grant 201809170CX11JC12. The presented work is completed at Xidian University. The review of this paper was coordinated by Dr. Z. Liu. (Corresponding author: Changle Li.)

Changle Li and Wenwei Yue are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, Shaanxi 710071, China (e-mail: clli@mail.xidian.edu.cn; wenwei.yue@stu.xidian.edu.cn).

Guoqiang Mao is with the Xidian University, Xi'an 710071, China, and also with the University of Technology of Sydney, Ultimo, NSW 2007, Australia (e-mail: gqmao@xidian.edu.cn).

Zhigang Xu is with the School of Information Engineering, Chang'an University, Xi'an, Shaanxi 710064, China (e-mail: xuzhigang@chd.edu.cn).

Digital Object Identifier 10.1109/TVT.2020.2973404

I. INTRODUCTION

TRAFFIC congestion has become a serious problem in almost all modern metropolitan cities due to increased use of vehicular transportation, urbanization and population increases. Congestion reduces the efficiency of transportation infrastructure and increases travel time, air pollution and fuel consumption, which in turn result in various social and economical problems [1]–[5].

As a major contributor to congestion, traffic bottlenecks account for 40% traffic congestion [6]. Therefore, locating traffic bottlenecks to identify the root causes of congestion is important and provides effective and cost-efficient means for traffic improvement. In addition to increasing the capacity of bottleneck by widening the road, advanced traffic control strategies such as traffic light control and vehicle rerouting can be implemented to relieve congestion at traffic bottlenecks [7]–[13].

Most works in the literature on bottleneck identification focused on freeway [14]–[16]. However, bottleneck identification in urban road networks is much more challenging. First, road topology is more intricate in urban networks. Consequently, vehicle travel pattern and congestion propagation pattern are more difficult to be estimated. Second, there is more traffic in urban roadways, which leads to more unexpected traffic conditions in road networks. Third, other factors such as traffic signals and social events have more significant impact on urban roadways than on freeways. Recently, urban bottleneck identification has received significant attention. In [24], Ma *et al.* defined a parameters Im based on traffic impedance C_{rs} and network effectiveness E . They compared the parameter Im before and after a particular road segment failure (in congestion) and regarded the road segment with more difference of parameter Im as a bottleneck. Ye *et al.* [17] used a critical index v/c based on the ratio of traffic flow and road capacity of a road segment to identify whether a road segment is a bottleneck or not. Lee *et al.* [25] developed a three-phrase spatio-temporal bottleneck mining model to identify bottlenecks in urban road networks and considered that bottlenecks most likely existed in the spatial cross section of two congestion propagation patterns.

Intuitively, the notion of a bottleneck implies that removal of the bottleneck should bring network wide traffic improvement, not just the traffic situation at the bottleneck location. As the most significant road segments causing congestion, when congestion occurs at bottlenecks, it is more likely to propagate to the other road segments and cause a large-scale of congestion in road networks. Thus, if we can identify and improve the traffic

conditions at bottlenecks, congestion will not only be mitigated at bottlenecks, traffic congestion in the entire road network will also be alleviated. Following the intuitive arguments above, to gauge bottlenecks properly, two main features should be considered: 1) the congestion level of a bottleneck itself; 2) the consequences of congestion propagation to other road segments. However, most of existing works on bottleneck identification only take congestion level cost of a bottleneck itself into consideration and neglect the congestion propagation effects, which may lead to erroneous bottleneck identification.

To fill the gap, in this paper, we first propose a novel bottleneck definition considering both congestion costs on road segments themselves and congestion propagation costs to other road segments. Then, a graph-theoretic method is proposed to model congestion propagation in a road network. Furthermore, with the combination of graph theory and Markov analysis, this paper quantifies the congestion costs of all road segments and identifies bottlenecks. Finally, both simulations using SUMO [27], [28] and experiments using inductive loop detector data of Taipei urban areas are conducted to validate our proposed bottleneck identification method and provide an application example over a real urban road network, respectively. More specifically, the following contributions are made in this paper:

- An intuitive bottleneck definition is proposed considering both the congestion level cost of a road segment itself and the congestion contagion cost, which can better capture the impact of bottlenecks on the road network and the causal relationship between congestion at different road segments. The proposed metric provides a more rigorous way to identify traffic bottlenecks;
- A novel technique is proposed, based on a combined use of graphical models, maximal spanning trees and Markov analysis, to model and analyze congestion propagation in urban road networks, which presents an effective approach to quantify congestion propagation processes and congestion costs of all road segments in road networks;
- Simulations are conducted using SUMO which demonstrates that compared with those techniques in the literature only considering congestion on road segments themselves for bottleneck identification, the proposed method can capture the features of urban bottlenecks and is more effective in identifying bottlenecks;
- Using the inductive loop detector data, the proposed technique is also applied to identify bottlenecks in urban areas of Taipei and shows that the most congested road segments are not necessarily bottlenecks.

The rest of paper is organized as follows. Section II presents an overview of the existing congestion propagation and bottleneck identification methods in urban areas. Section III presents the proposed bottleneck identification technique. Simulations based on SUMO and discussions are conducted in Section IV. Experimental results are presented in Section V. Finally, Section VI concludes the paper.

II. RELATED WORK

Traffic congestion in urban networks is a long-standing and even increasing problem in modern society. Many advanced

traffic control strategies, such as traffic signal control and vehicle rerouting, have been introduced to mitigate congestion and improve the network performance in urban traffic control systems. For signal control, Split, Cycle and Offset Optimization Technique (SCOOT) [29] and Sydney Cooperative Adaptive Traffic System (SCATS) [30] have been applied in many cities around the world extensively to achieve reduced vehicle delays and mitigate road congestion. For route guidance, the rerouting methods have also been widely exploited to manage traffic networks in a more organized way and relieve congestion in road networks. Pan *et al.* [31] presented proactive vehicle rerouting strategies, which computed tailored rerouting for drivers when congestion was predicted on their routes. Simulation results demonstrated that the route choice strategies were effective in mitigating congestion and improving traffic efficiency. In [32] and [33], Cao *et al.* developed several route guidance strategies based on a multi-agent framework, where each local agent collected vehicle intentions and then provided route guidance for these vehicle by solving a route assignment problem. Simulations were conducted based on traffic networks in Singapore and New York using SUMO to validate the effectiveness of proposed strategies in improving average probability of arriving on time and total travel time. Moreover, there are also some works focusing on the combination of signal control and route guidance. Xiao and Lo [34] formulated a traffic system considering both day-to-day route choice and signal control, which was crucial for developing an optimal control strategy to increase traffic fluency and even relieve possible congestion in traffic networks. Researchers in [35] proposed a pheromone-based traffic management system for traffic congestion mitigation, which unified both dynamic route choice and signal control. Simulation results showed that the proposed system outperformed other approaches that only considered vehicle rerouting or signal control in terms of road congestion levels, travel delays, air pollution and fuel consumption.

For the entire traffic network, it is not reasonable to apply traffic control strategies on all intersections or road segments to mitigate congestion, which not only leads to more computational complexity, but also causes more extra installation and maintenance costs. Therefore, it is necessary to locate the most critical road segments (bottlenecks) in road networks and when congestion on the identified bottlenecks can be relieved by applying these advanced traffic control strategies, traffic conditions will be improved in the whole traffic network effectively and efficiently. Most works on bottleneck identification are based on the congestion level of a road segment itself, which evaluates the congestion according to average travel time, travel speed and so on, to identify the most congested road segments as bottlenecks in urban road networks. Ye *et al.* [17] employed a route choice model to simulate traffic flow and identify bottlenecks in urban areas. They first determined an index v/c based on the ratio of the traffic flow and the road capacity of a road segment. If the real-time v/c is higher than a critical value, the road segment was considered to be a bottleneck. The critical threshold in this paper was considered as a fixed value 0.539, which was determined empirically. However, this definition is better suited to describe the congestion level of road segments and is really not a good metric for bottleneck as it does not capture the important aspect

of bottleneck that congestion at a bottleneck may spread to other road segments. A similar definition was employed in [18], where Long *et al.* took the average journey speed of a road segment as a key parameter to determine whether the road segment can be considered as a bottleneck. Specifically, if the average travel speed of a road segment is larger than 20 km/h, the road segment can be seen as a bottleneck. In [19], Gong and Wang identified the congestion of each road segment based on the road occupancy on each road segment itself, where 30% of the average occupancy was considered as the threshold to differentiate the “uncongested state” and “congested state”. Moreover, they analyzed the temporal relations among the detected road segments and regarded the road segment where congestion occurred first as a bottleneck. However, the first congested road segment is not necessarily the root cause of congestion in urban areas. In summary, these bottleneck identification methods were only based on the congestion level of road segment itself and did not consider the congestion propagation effects.

On the other hand, as a novel insight to analyze causal relationship between congestion on different road segments, congestion propagation in urban areas has received significant attentions recently [20]–[25]. In [20], Nguyen, Liu and Chen introduced an algorithm to construct causality trees based on congestion propagation, which demonstrated the congestion propagation pattern and estimated their propagation probabilities based on temporal and spatial information of congestion. Then, they found the frequency sub-structures in these causality trees to discover frequent patterns of congestion propagation in the road network, which can reveal not only the recurring interactions among spatio-temporal congestion, but also potential bottlenecks in road network. Wang *et al.* [21] proposed a three-phase framework to study congestion correlation between road segments using GPS trajectories of taxis. They extracted various features on each pair of road segments and analyzed the important features which led to congestion correlation between two road segments in urban areas. They found the important features that could lead to a high/low congestion correlation, such as time of the day, betweenness and closeness of each road segment. In [22], Tao *et al.* analyzed the congestion relationship between road segments and their simulation results suggested that the congestion of a road segment was affected by the road network structure and congestion of its adjacent road segments. The same conclusion was drawn in [23], where Li, Liu and Zou utilized a coordination game model to analyze the critical condition that could lead to congestion propagation. They found that the influence of congested road segments on other adjacent road segments is a key contributor to traffic congestion propagation in urban environment, and when the influence reaches to a certain critical threshold (determined by the road network topology), a massive traffic congestion will be generated.

Based on the aforementioned researches, some urban bottleneck identification methods have been proposed taking congestion propagation into account directly or indirectly. In [24], Ma *et al.* took traffic impedance C_{rs} and network effectiveness E into consideration and combined the two parameters together as a new parameter Im based on a certain weight. They regarded the road segments with significant difference in the parameter Im before and after a particular road segment failure (in congestion)

as bottlenecks. However, because the traffic impedance and network effectiveness are difficult to calculate, the performance of their proposed method is yet to be comprehensively validated. Moreover, this method did not explicitly consider the congestion propagation relationships between road segments. Lee *et al.* in [25] developed a three-phase spatio-temporal traffic bottleneck mining model to identify bottlenecks and considered that bottlenecks most likely existed in the spatial cross area of two congestion propagation patterns. A congestion propagation pattern depicts the congestion propagation relationship between two congested areas. Using data collected from a taxi dispatching system, the experimental results showed the effectiveness of the proposed method in congestion prediction and bottleneck identification. Reference [25] was the first paper that identify urban bottlenecks explicitly based on the congestion propagation feature and spatio-temporal property of urban bottleneck. However, the bottleneck identification method in [25] defined urban bottlenecks as the spatial cross area of two congestion propagation patterns, which failed to quantify the congestion cost to give a rigorous and intuitive identification of bottlenecks. Moreover, [25] considered only congestion propagation of two congested areas and could not extend to the whole urban road network. Our previous work [26] also studied a bottleneck identification method considering congestion propagation effects, however, the definition and calculation of congestion costs were not well-conducted and the validation of identified bottlenecks should also be further investigated.

In order to overcome the shortcomings of the aforementioned works, in this paper, we propose a novel congestion bottleneck definition taking into account the two identifying features of bottlenecks: congestion level cost of a road segment itself and contagion cost that the congestion may propagate to other road segments. Novel graph theoretical technique and Markov analysis are employed to analyze and quantitatively determine the two costs. Both simulations using SUMO and experiments using inductive loop detector data are presented to demonstrate the effectiveness of our proposed method in urban bottleneck identification.

III. BOTTLENECK IDENTIFICATION TECHNIQUE

In this section, a bottleneck identification technique is presented and a roadmap of the technique is given in Fig. 1. Firstly, we introduce a metric about road congestion (Fig. 1(a)) and a definition of congestion correlation between two road segments (Fig. 1(b)), which indicates the causal relationship between congestion at different road segments. Then, based on the correlated congestion, we build a graphical model to represent the congestion correlations and analyze the congestion propagation pattern in the network (Fig. 1(c) (d)). Finally, considering both congestion of a road segment itself and congestion propagation effects, we quantify congestion costs of all road segments and identify traffic bottlenecks (Fig. 1(e)).

A. Congestion and Congestion Correlation

Traffic congestion has become one of the most major and costly problems in many cities, which always leads to an increase in travel time and a decrease in velocity. Generally,

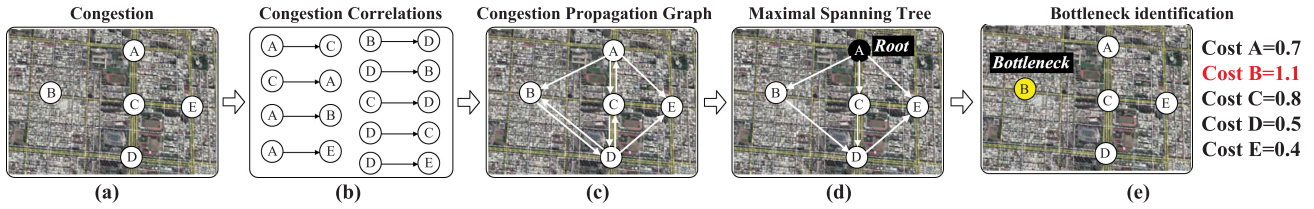


Fig. 1. Roadmap of bottleneck identification technique.

traffic congestion can be categorized as recurrent congestion and non-recurrent congestion and their identification and detection methods are also different. On the one hand, researchers often identify recurrent congestion by setting up a critical threshold for a variety of metrics, such as travel time, speed and road occupancy. Most of the critical thresholds are seen as a fixed value [41], [42] and when the monitored real-time traffic metric on a road segment is higher or lower than a pre-designated threshold, the traffic condition of this road segment can be seen as congestion. For example, following a document of the Ministry of Public Security of China [43], if the average speed of vehicles on a road segment is less than 20 km/h, this road segment can be regarded to be congested. However, the fixed threshold is usually determined empirically and does not take the characteristics of individual road segment into consideration, such as road lengths, number of lanes and speed limits. In this case, some works defined recurrent congestion considering different properties of each road segment [20], [44]. For instance, Nguyen *et al.* [20] suggested that a segment was considered as congested at a specific time if the average travel time is longer than 80% of its time distribution. On the other hand, non-recurrent congestion in a road network is mainly caused by incidents, workzones, special events and extreme weather [45]. Comparatively, the identification of non-recurrent congestion is much difficult, which is often dealt as a pattern recognition problem and many classifiers are utilized to determine the locations and severities of non-recurrent congestion [46], [47]. In this paper, to identify the long-term traffic bottlenecks in urban traffic networks, we mainly concern about the identification of recurrent congestion.

As an important metric to evaluate road conditions, researches about vehicle travel speed have been conducted along several directions, such as travel speed forecast [36]–[40] and traffic congestion detection [41], [42]. Existing studies about travel speed forecasting have shown their capabilities in predicting future traffic states, such as autoregressive integrated moving average (ARIMA) [36], Convolutional Neural Networks (CNN) [37] and Graph Convolutional Networks (GCN) [38]–[40]. These methods have the potential to be applied in congestion detection by using the critical threshold for average travel speed to classify the real-time congestion status of road segments. Therefore, in this paper, to identify traffic congestion, the evaluation is done based on the average travel speed of each road segment. More specifically, a road segment is considered as congested at a specific time if its travel speed is lower than the $n\%$ of its average travel speed, where n is varies between 10 and 90 in our experiments. Fig. 2 shows the travel speed on a road segment over 12 hours, where each point represents the travel speed at

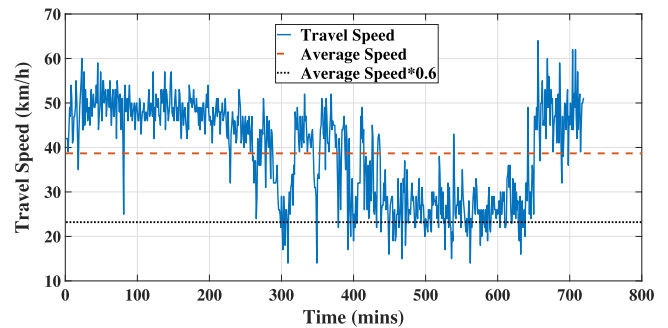


Fig. 2. The average travel speed of a road segment within 12 hours.

each minute. The average travel speed (38.67 km/h) and 60% of the average travel speed (23.2 km/h) on the road segment are presented by red dotted line and black dotted line, respectively. This road segment can be marked as congested where the average travel speed is lower than 23.2 km/h.

Furthermore, when congestion occurs on a road segment, it may potentially affect the traffic flows of surrounding road segments and lead to more congested road segments. Therefore, it is essential to analyze and uncover the correlation between congestion on different road segments. To achieve this, we propose a congestion correlation definition based on the spatial-temporal relationship of two road segments, as follows:

Definition 1. (Congestion correlation between two road segments): Congestion on a road segment A is correlated with congestion on road segment B , if the following requirements are satisfied.

- Spatial threshold: the shortest path distance between congestion on road segments A and B is less than a pre-designated spatial threshold.
- Congestion propagation speed interval: according to the shortest path distance and time difference between congestion occurring at road segments A and B , the congestion propagation speed between the two road segments should be within a pre-designated (and empirically set) congestion propagation speed interval.

In this paper, we consider that congestion at two different road segments are correlated *only if* both the spatial threshold and the congestion propagation speed interval are met. Compared with existing works [21], [25], the congestion correlation definition in this paper has two merits: firstly, determining the spatial threshold T_s based on the shortest path distance can indicate the congestion propagation path and the propagation direction in the actual traffic network; secondly, the congestion propagation speed interval can better capture the spatial-temporal

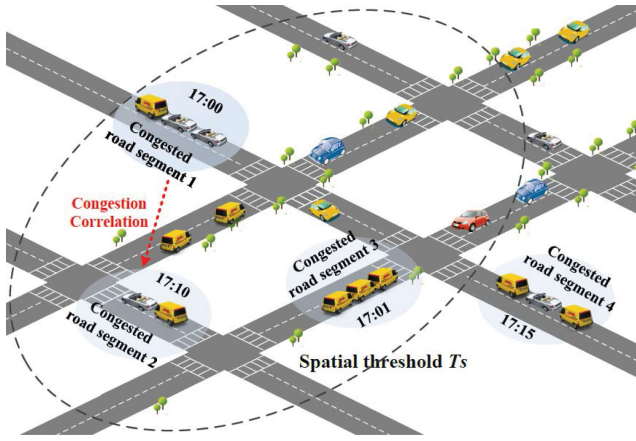


Fig. 3. An illustration of the congestion correlation between two road segments.

relationship of congestion propagation. The two thresholds are set empirically and may be different for different cities.

An example is given in Fig. 3. Congestion occurs on road segment 1 at 17:00 and we need to investigate the correlated congested road segments for road segment 1 based on our proposed congestion correlation definition. According to the spatial threshold T_s , we first determine the shortest path distance from congested road segment 1 to its upstream road segments in the road network, as shown in Fig. 3. In this way, because the shortest path distance between congested road segment 1 and congested road segment 4 is larger than the spatial threshold T_s , congestion on road segment 4 is not considered to be correlated with congestion on road segment 1. Moreover, in terms of congested road segments 2 and 3, their shortest path distances from congested road segment 1 are both less than the spatial threshold T_s . However, congestions on road segments 1 and 3 occur at almost the same time despite that the shortest path distance between them is relatively large. Therefore, when congestion propagation speed is taken into account, it is unlikely that congestions on road segments 1 and 3 are correlated. Hence considering both the spatial threshold and the congestion propagation speed interval, only the congestion on road segment 2 is considered to be correlated with the congestion on road segment 1 and we can further obtain the causal relationship “congested road segment 1 \rightarrow congested road segment 2”.

B. Congestion Propagation Graph and Maximal Spanning Tree

In this subsection, according to the aforementioned congestion correlation definition, we first connect these correlated segments together based on their spatial relationship to construct congestion propagation graphs (CPG). Then, using the constructed congestion propagation graphs, we employ a maximal spanning tree algorithm to obtain a set of trees in the graphs, where each tree includes as many edges in the congestion propagation graphs as possible, to capture the causal relation among congestion at different road segments.

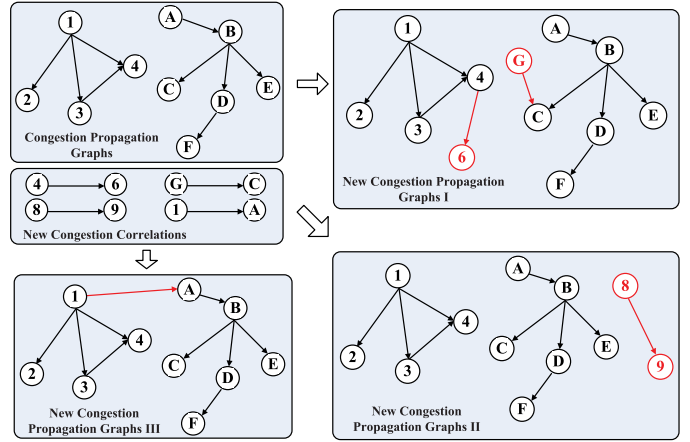


Fig. 4. An illustration of the construction of congestion propagation graphs.

More specifically, using the procedure described in the previous subsection, we can obtain a set of congestion correlated road segments and each correlation relationship can be seen as a directed edge, which indicates that congestion propagates from the road segment corresponding to the start vertex of the edge to the second road segment corresponding to the end vertex of the edge. Then, using Algorithm 1, we can connect the obtained congestion correlations (i.e., these directed edges) together to construct a set of disjoint directed graphs. An example is shown in Fig. 4. Assuming that we have constructed two disjoint congestion propagation graphs, we need to add the other four new correlations 4 \rightarrow 6, $G \rightarrow C$, 8 \rightarrow 9 and 1 \rightarrow A into the graphs. As depicted in the congestion propagation graph I of Fig. 4, if either road segment in a correlation relation already exists in the current graphs, such as correlation 4 \rightarrow 6 and $G \rightarrow C$, we can connect the correlation to the corresponding graph (lines 7–8). If none of the two road segments in a correlation relation are in the existing graphs, such as correlation 8 \rightarrow 9, then this edge (and the associated vertices) should form the first edge of a new graph, as shown in the congestion propagation graph II of Fig. 4 (lines 9–11). Moreover, if one road segment in a correlation is in a graph and another road segment in a correlation is in another graph, such as correlation 1 \rightarrow A, then we can join the two graphs together and form one graph, as shown in the congestion propagation graphs III (lines 12–13). However, if two road segments in a correlation are both in a same graph, then we can delete this correlation (lines 15). In this way, we can construct several disjoint congestion propagation graphs using the aforementioned correlation relationship.

Then, using the constructed congestion propagation graphs, we can construct maximal spanning trees from the graphs, which is required to compute the congestion cost of each road segment in the trees. The following gives a formal definition of the maximal spanning trees considered in this paper.

Definition 2. (Maximal spanning tree): A maximal spanning tree is a tree with a maximal set of directed edges (i.e., correlations) such that there is a unique (directed) path from the root of the tree (i.e., a road segment) to any other vertex (i.e., the end point of an edge) of the tree.

Algorithm 1: Constructing Congestion Propagation Graph (CPG).

```

1: Input: a set of congestion correlations and there are
   two road segments (road segment A and road segment
   B) in a correlation indicating congestion propagates
   from road segment A to road segment B.
2: Output: a set of congestion propagation graphs.
3:  $Graph_1 \leftarrow$  an empty set;
4:  $CPG \leftarrow$  an empty set;
5:  $CPG \leftarrow CPG \cup Graph_1$ ;
6: for Each  $correlation_i$  ( $i \in (1, \dots, N)$ ) do
7:   if ((road A  $\in$  one of graphs in  $CPG$  (e.g.  $Graph_m$ ))
      && (road B  $\notin$  all graphs in  $CPG$ )) || ((road B  $\in$  one
      of graphs in  $CPG$ ) && (road A  $\notin$  all graphs in
       $CPG$ )) then
8:      $Graph_m \leftarrow Graph_m \cup correlation_i$ ;
9:   else if (road A  $\notin$  all graphs in  $CPG$ ) && (road B  $\notin$ 
      all graphs in  $CPG$ ) then
10:     $Graph_{new} \leftarrow correlation_i$ ;
11:     $CPG \leftarrow CPG \cup Graph_{new}$ ;
12:   else if (road A  $\in$  the  $m$ th graph  $Graph_m$  in  $CPG$ )
      && (road B  $\in$  the  $n$ th graph  $Graph_n$  in  $CPG$ ) then
13:     $Graph_m \leftarrow Graph_m \cup Graph_n$ ;
14:   else
15:     Delete  $correlation_i$ ;
16:   end if
17: end for
18: return  $CPG$ ;

```

Following Definition 2, in order to quantify the congestion propagation effects caused by one road segment and calculate its congestion propagation cost, we present an algorithm based on Breadth First Search (BFS) as shown in Algorithm 2 and consider each road segment as a root node to construct a maximal spanning tree respectively from congestion propagation graphs. An example is given to illustrate the construction of maximal spanning trees. As depicted in Fig. 5, a congestion propagation graph is obtained based on our proposed method above, which consists of 5 vertices (road segments) and 9 directed edges (correlations). Regarding road segments A , B , C , D and E as the root of a tree respectively, we can get 5 different maximal spanning trees (because congestion on road segment E does not propagate to the other road segments, the fifth tree only consists of a root node, i.e., road segment E) and each of them indicates the congestion propagation path and influence areas when congestion occurs on the root road segment.

C. Bottleneck Identification

Bottlenecks are a result of specific physical conditions, such as road network geometries, roadway operation strategies and traffic demand fluctuations, which are often the most vulnerable points in a road network [48]. Therefore, it is essential to identify urban bottlenecks for locating the root cause of congestion in a road network. In this subsection, according to the existing

Algorithm 2: Constructing Maximal Spanning Trees in a Graph.

```

1: Input: a set of congestion propagation graphs.
2: Output: a set of maximal spanning trees.
3: for Each  $graph_n$  ( $n \in (1, \dots, N)$ ) do
4:   Compress  $graph_n$  to adjacency matrix  $A_n$ ;
5:   for Each  $node_m$  ( $m \in (1, \dots, M)$ ) do
6:      $head = 1$ ;
7:      $tail = 1$ ;
8:      $queue(head) = m$ ;
9:      $head = head + 1$ ;
10:     $flag = m$ ;
11:     $tree \leftarrow$  an empty set;
12:    while  $tail \neq head$  do
13:       $i = queue(tail)$ ;
14:      for  $j = 1$  to the dimension of  $A_n$  do
15:        if  $A_n(i, j) == 1$  &&  $j \notin flag$  then
16:           $queue(head) = j$ ;
17:           $head = head + 1$ ;
18:           $flag \leftarrow flag \cup j$ ;
19:           $tree \leftarrow tree \cup [i, j]$ ;
20:        end if
21:      end for
22:       $tail = tail + 1$ ;
23:    end while
24:    return  $tree$ ;
25:  end for
26: end for

```

definitions about congestion costs [24], [49], [50], we first propose a novel congestion cost definition and the corresponding bottleneck definition. Then, we elaborate the procedures to identify bottlenecks by calculating both the congestion level cost of a road segment itself and the contagion cost that the congestion may propagate to other road segments.

To better identify long-term bottlenecks and quantify their negative effects to the entire road network precisely, we should not only calculate the congestion levels on road segments, but also analyze the congestion propagation influences of these roads to their neighboring road segments. For the quantification of congestion level on each road segment, in order to describe the long-term traffic condition of a road segment, in this paper, we calculate the average road occupancy with consideration of road importance to measure the congestion level cost of a road segment itself. Moreover, due to the difficulties in congestion propagation process estimation and quantification, in this paper, we utilize Markov analysis to build the bridge between congestion on two different road segments by the congestion propagation probabilities. Then, combining with the maximal spanning trees, the congestion propagation costs of all road segments in a road network can be obtained and quantified. Specifically, the definition of congestion cost and urban bottlenecks are demonstrated in Definition 3 and Definition 4, respectively.

Definition 3. (Congestion cost): Traffic congestion cost of a road segment indicates the quantification of congestion on

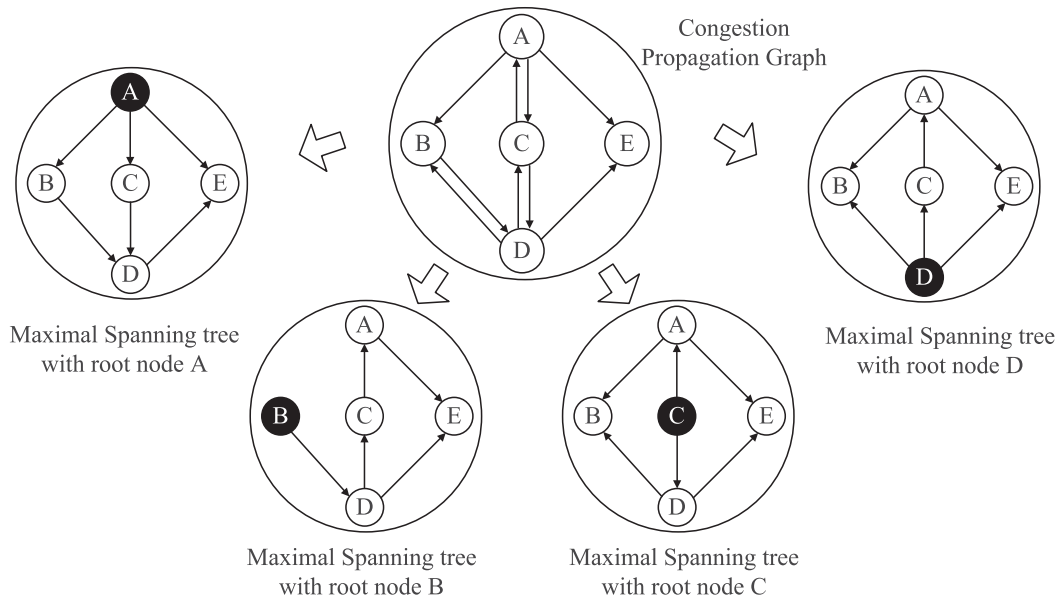


Fig. 5. An illustration of the construction of maximal spanning trees.

this road segment that causes the negative influence on the whole road network, which can be expressed as the sum of the congestion level cost of the road segment itself and the congestion propagation cost that the congestion may propagate to other road segments. The first item is calculated based on the normalized average road occupancy and the importance of the road segment and the second item is determined using the congestion propagation probabilities from this road segment to the other road segments and the congestion costs of the involved road segments based on the obtained maximal spanning trees.

Definition 4. (Bottlenecks in urban areas): Urban traffic bottlenecks indicate that the most significant road segments that cause more congestion costs in urban areas, which can be determined when congestion costs of road segments exceed a pre-designated threshold.

Firstly, we can calculate the congestion costs of road segments themselves in the road network. By using the inductive loop detector data, it is convenient to collect road occupancy data to quantify the real-time road condition on each road segment, which is defined as the percentage of a road segment occupied by vehicles [51]. Therefore, in this paper, we utilize the average road occupancy to indicate congestion levels of road segments. However, the same level of congestion on different road segments may have different impact on a road network, so we normalize the average traffic flow on each road segment by the maximum flow as the weights of their road congestion levels. For road network with N road segments, let the normalized average traffic flows on all road segments be $[X_1, X_2, \dots, X_N]$ and the average occupancy of all road segments be $[Y_1, Y_2, \dots, Y_N]$. The weighted congestion levels of road segments can be presented as

$$[X_1 * Y_1, X_2 * Y_2, \dots, X_N * Y_N]. \quad (1)$$

Furthermore, we utilize the normalized congestion levels of road segments to indicate congestion costs of road segments

Algorithm 3: Calculating the Total Congestion Cost of Root Node in a Maximal Spanning Tree.

- 1: **Input:** a maximal spanning tree; congestion levels of all nodes in the tree; and congestion propagation probabilities of all vertexes in the tree.
 - 2: **Output:** the total congestion cost of root node.
 - 3: **while** there are n ($n \geq 2$) nodes in the tree **do**
 - 4: Find the leaf nodes in the tree.
 - 5: Calculate congestion contagion costs from leaf nodes to their parent nodes.
 - 6: Add the congestion contagion costs to the corresponding parent nodes.
 - 7: Delete the leaf nodes and the connected edges to form a new tree.
 - 8: **end while**
 - 9: **return** the total congestion cost of root node.
-

themselves, which is

$$[W_1, W_2, \dots, W_N], \quad (2)$$

where $W_i = \frac{X_i * Y_i}{\max\{X_1 * Y_1, X_2 * Y_2, \dots, X_N * Y_N\}}$.

Next, utilizing the obtained maximal spanning trees, we can calculate the congestion contagion costs of all the road segments in the trees. This is done by the first determining the congestion propagation probabilities between two road segments connected by an edge, such as $A \rightarrow B, A \rightarrow C, A \rightarrow D$ and $D \rightarrow F$ in Fig. 7. In this paper, we utilize Markov analysis to determine the probabilities of congestion propagation from one road segment to another road segment [52], which provides an intuitive way to better capture the causal relationships between congestion on different road segments and quantify congestion propagation probabilities. Specifically, suppose that $A_t = 1$ denotes the event that road segment A is congested at time t and

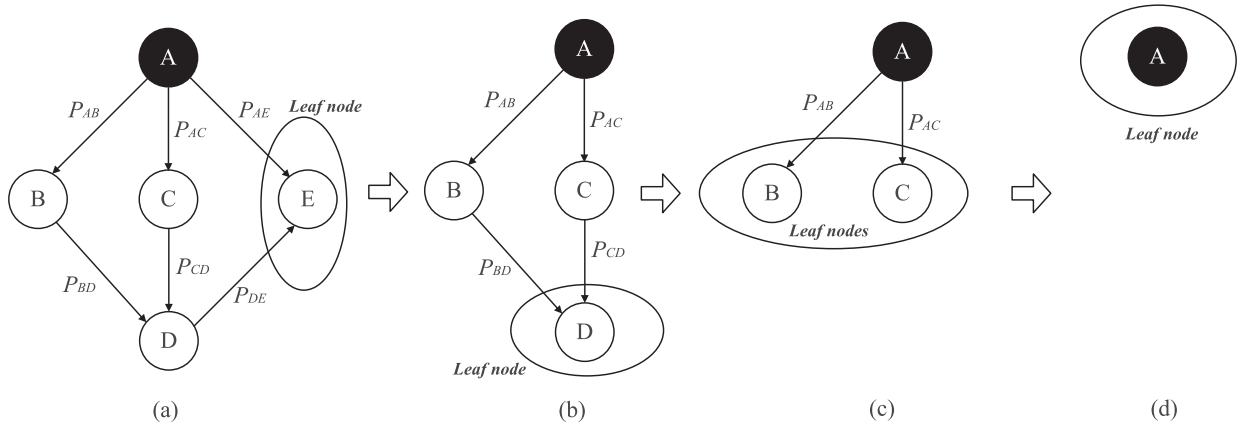


Fig. 6. Congestion cost calculation of root node in a tree.

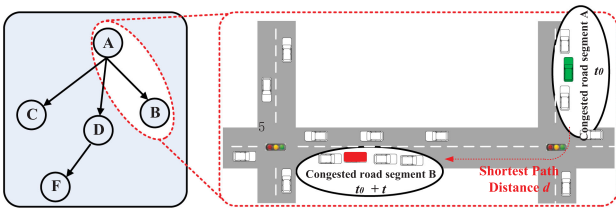


Fig. 7. An illustration of the determination of the congestion propagation probability.

$B_t = 1$ denotes the event that road segment B is congested at time t , the congestion propagation probability P_{AB} from road segment A to road segment B ($A \rightarrow B$) can be calculated according to a conditional probability in (3), which indicates the probability that road segment B is congested at time $t_0 + \tau$ ($B_{t_0+\tau} = 1$) given that congestion occurs on road segment A at time t_0 ($A_{t_0} = 1$).

$$P_{AB} = P(B_{t_0+\tau} = 1 | A_{t_0} = 1), \quad (3)$$

where τ indicates a time instant and should fulfill the condition in (4).

$$\frac{Distance(A, B)}{Speed_{\max}} \leq \tau \leq \frac{Distance(A, B)}{Speed_{\min}}, \quad (4)$$

where $Distance(A, B)$ denotes the shortest path distance $Distance(A, B)$ between road segments A and B , $Speed_{\max}$ and $Speed_{\min}$ are determined according to the upper and lower bounds of congestion propagation speed interval in Definition 1, respectively.

Based on the obtained congestion propagation probabilities of all congestion correlations and the topologies of maximal spanning trees with different root nodes, we propose an algorithm to calculate the congestion costs of all road segments in urban traffic network, as shown in Algorithm 3. For easy implementation, we calculate the congestion cost of root node in each tree starting from the leaf nodes (outdegree = 0) recursively to the parent nodes of these leaf nodes and so on, till reaching the root of the trees. An example is shown in Fig. 6, and there are five nodes with congestion costs $[W_A, W_B, W_C, W_D, W_E]$ and six directed edges with the corresponding propagation probabilities.

As we can see that, node A is the root of the tree and congestion at road segment A propagates to road segments B, C, D and E gradually. To calculate the congestion cost of node A , we start from the leaf node E and the corresponding directed edges AE and DE . Using the normalized road congestion cost of node E , W_E and congestion propagation probabilities of edge AE and DE , P_{AE} and P_{DE} , we can first get the congestion contagion costs caused by congestion propagation from road segment A to road segment E and road segment D to road segment E , which is $P_{AE}W_E$ and $P_{DE}W_E$, respectively. Then, we delete the node E and edges AE and DE , thus the tree in Fig. 6(a) can be presented as Fig. 6(b). Because node D is the leaf node in this new tree, considering both normalized congestion level W_D and congestion contagion costs propagating to the other nodes, $P_{DE}W_E$, the total congestion cost of node D can be written as $W_D + P_{DE}W_E$. In this way, we can recursively obtain the congestion costs of node B and node C and eventually get the total congestion cost of the root node road segment A , as follows:

- Road segment B : $W_B + P_{BD}(W_D + P_{DE}W_E)$;
- Road segment C : $W_C + P_{CD}(W_D + P_{DE}W_E)$;
- Road segment A : $W_A + P_{AB}[W_B + P_{BD}(W_D + P_{DE}W_E)] + P_{AC}[W_C + P_{CD}(W_D + P_{DE}W_E)] + P_{AE}W_E$.

Hence considering both the congestion level cost of road segment A itself and its congestion propagation effects, we can rigorously quantify the total congestion cost caused by road segment A to the whole traffic network using our proposed graph-based approach. Similarly, for the other road segments, we can also calculate their total congestion costs according to the maximal spanning trees with different root nodes. Further, following Definition 4, we regard the road segments with higher congestion costs as the bottlenecks in the whole urban road network.

IV. SIMULATION AND DISCUSSION

Because of the difficulty in validating our proposed bottleneck identification method in actual road networks, in this section, using a traffic simulator SUMO, we first identify bottlenecks in a simplified network based on the City of Sioux Falls, South Dakota, USA. Then, we increase the number of lanes at each

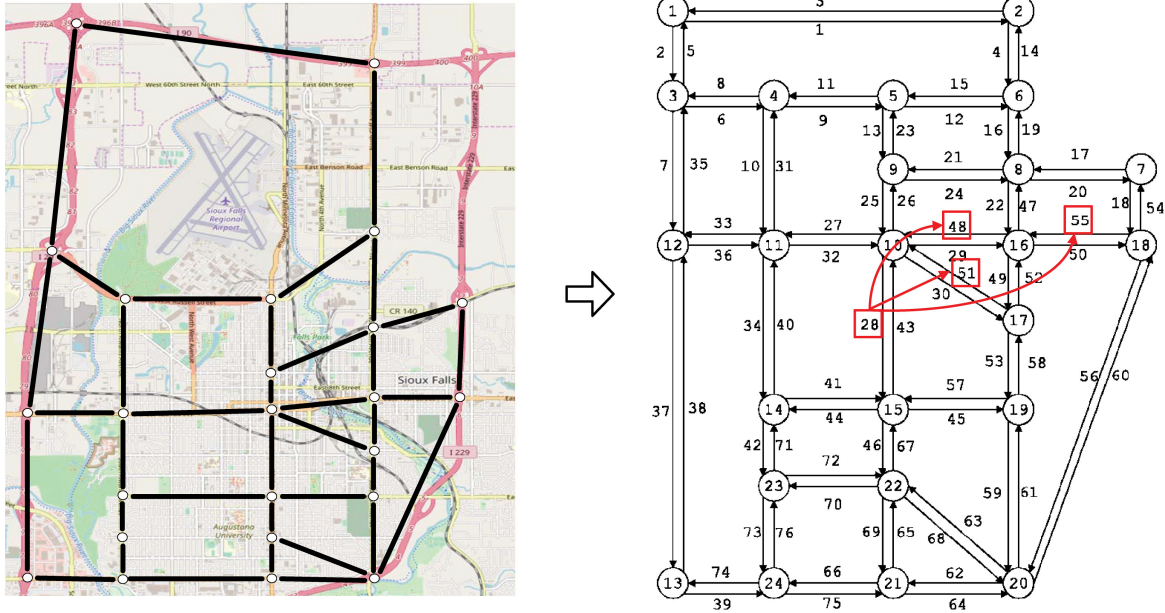


Fig. 8. Sioux falls network.

road segment and compare the average traffic speed in the total network before and after these increases to validate the effectiveness of our proposed graph-theoretic approach. Moreover, based on different vehicle arrival rates, we compare the average traffic speed under the real road network, road network with an increased number of lanes on the identified bottlenecks using existing method in the literature and network with increased number of lanes on the identified bottlenecks using our proposed method to validate the effectiveness of the proposed graph-theoretic bottleneck identification approach.

A. Sioux Falls Network

Sioux Falls network is very popular within the transport research community and has been used as a benchmark and test scenario in many publications [53]–[55]. As shown in Fig. 8, using a map of the City of Sioux Falls, 24 nodes of the network are matched to the major intersections of the city and 76 directed edges of the network roughly are matched to the major arterial roads of the city. The lengths of all road segments are set equal to the Euclidian distances between the respective two intersections in real road network and each road segment is set with 2 lanes, which is also a good simulation of the real conditions [53].

B. Bottleneck Identification Based on Our Proposed Method

In this subsection, we first set the average vehicle arrival rate of Sioux Falls network to be 7200 vehicles per hour. Then, using our proposed graph-theoretic approach, we identify bottlenecks under different congestion thresholds respectively and evaluate the travel speed improvement of the road network after increasing the number of lanes on each identified bottleneck. These results can be utilized to analyze the effectiveness of our proposed bottleneck identification method under different congestion thresholds and determine the most appropriate threshold

TABLE I
THE EVALUATION OF TRAVEL SPEED IMPROVEMENT UNDER
DIFFERENT CONGESTION THRESHOLDS

Percentage (%)	Number of trees	Percentage of speed improvement (%)	
		Level	Propagation
10	27	0.7014	0.7565
20	31	0.7014	0.8013
30	39	0.7014	0.8261
40	47	0.5815	0.8612
50	57	0.4821	0.9018
60	65	0.4609	0.9437
70	74	0.4475	0.8423
80	74	0.4475	0.6994
90	74	0.4475	0.6473

for classification of road segments' congestion status. Finally, based on the threshold, we represent the congestion costs on road segments themselves and congestion propagation costs, respectively, and identify bottlenecks in Sioux Falls network.

As illustrated in Table I, we investigate the travel speed improvement with the congestion identification thresholds of road segments, which varies from 10% of their average travel speeds to 90% of their average travel speeds. It can be noted that as the percentage increases, the number of congestion propagation trees follows an upward trend until congestion occurs on almost all road segments in Sioux Falls network. This trend is expected, because more and more traffic statuses are regarded as congestion with an increase of the congestion threshold. More importantly, according to different congestion thresholds, we utilize the existing method (based on congestion level only) and our proposed method (with combination of congestion level and congestion propagation) to identify bottlenecks in Sioux Falls network, respectively. We can see that after increasing the number of lanes at the identified bottlenecks, the travel speed

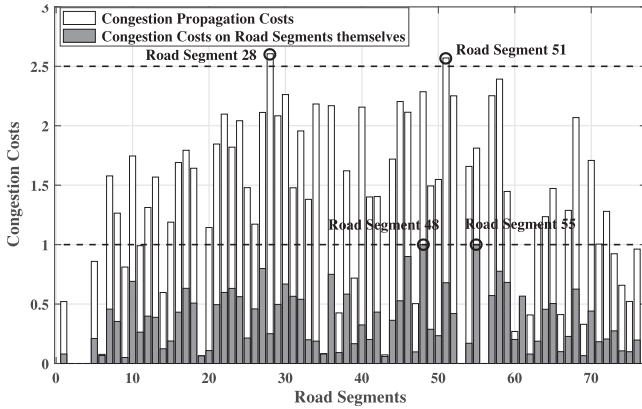


Fig. 9. Congestion costs of road segments and identified bottlenecks in Sioux Falls network.

improvement follows a declining trend with an increase of the congestion threshold for the existing method. The reason lies in that the existing method about bottleneck identification is better suited to describe the congestion level of road segments. At the same time, the smaller the congestion threshold is, the worse road conditions of the identified bottlenecks are. However, when the threshold is determined as 60%, our proposed bottleneck identification method can achieve the highest improvement of the average travel speed in Sioux Falls network. Thus, in this paper, we choose 60% of the average travel speeds as the metric to classify the real-time congestion status of road segments.

Using the threshold, we display the congestion costs on road segments themselves and congestion propagation costs respectively. As shown in Fig. 9, the horizontal axis indicates each road segment in Sioux Falls network and the vertical axis demonstrates the congestion costs of all road segments. The grey bars indicate congestion costs of road segments themselves, which are obtained according to the normalized congestion levels on each road segments (in Subsection III-C) and the white bars describe the congestion propagation costs which are calculated based on the aforementioned congestion propagation graphs and maximal spanning trees (in Subsection III-B). The sum of congestion level costs and congestion propagation costs are considered as the total congestion costs of road segments in the road network. We can see that based on the existing bottleneck identification method (considering congestion levels of road segments only), road segments 48 and 55 will be identified as bottlenecks in Sioux Falls network. However, when we also take congestion propagation effects into consideration to quantify the congestion costs for all road segments in the road network, road segments 28 and 51 are more likely to be regarded as bottlenecks, as illustrated in Fig. 9. Especially, the congestion level on road segment 28 is not very high compared with the other road segments (e.g., road segments 48, 51 and 55). However, as shown in Fig. 8, congestion on road segment 28 tends to propagate to road segments 48, 51 and 55 due to the spatial connections between these road segments. In this case, considering both congestion level and congestion propagation, the congestion cost of road segment 28 is much higher such that this road segment is more likely to be a bottleneck in Sioux Falls

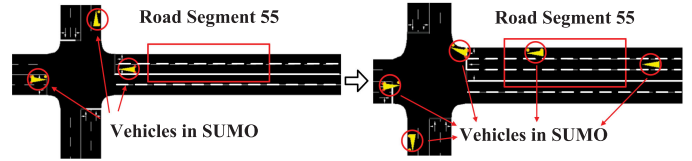


Fig. 10. Increasing the number of lanes on each road segment.

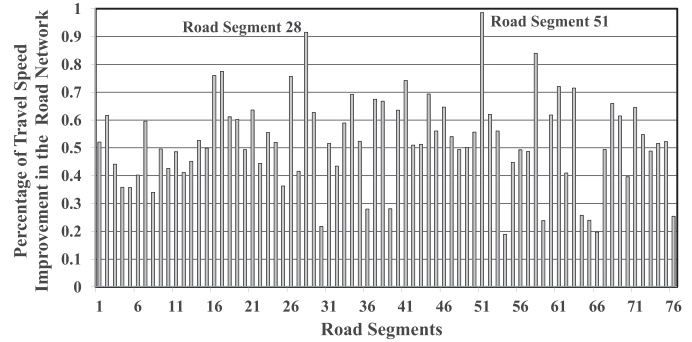


Fig. 11. Percentage of travel speed improvement in road network after increasing the number of lanes on each road segment.

network. These results demonstrate that the most congested road segments do not always incur more congestion costs in the whole traffic network and identifying bottlenecks in urban areas only according to the congestion levels of road segments is not always effective. Moreover, by quantifying congestion costs for all road segments in a road network, more than one bottleneck can be identified in a spatio-temporal congestion propagation area.

C. Bottleneck Verification

As mentioned in Subsection III-C, an intuitive definition of bottlenecks is that traffic improvement at bottlenecks brings the most significant network-wide traffic improvement. Therefore, in this subsection, in order to validate the effectiveness of our proposed method in identifying bottlenecks, we increase the number of lanes on each road segment on SUMO and compare the percentage of travel speed improvement in the road network before and after each increase respectively. An example is given in Fig. 10, where the number of lanes on road segment 55 is increased from 2 to 3 and we can analyze the importance of congestion on road segment 55 to the traffic conditions in the entire Sioux Falls network by comparing the average network travel speed under the two scenarios respectively.

As indicated in Fig. 11, after increasing the number of lanes on each road segment, the average travel speed of road network will be improved. Especially, there are significant improvements of the average travel speed in the road network after increasing the number of lanes on road segments 28 and 51, which are 91.4% and 98.3% respectively. Thus, road segments 28 and 51 have greater impacts on traffic conditions of the entire road network and according to our proposed bottleneck definition in Definition 4, the two road segments can be considered as bottlenecks in Sioux Falls network. The result suggests that our proposed bottleneck identification method considering congestion levels and congestion propagation costs simultaneously

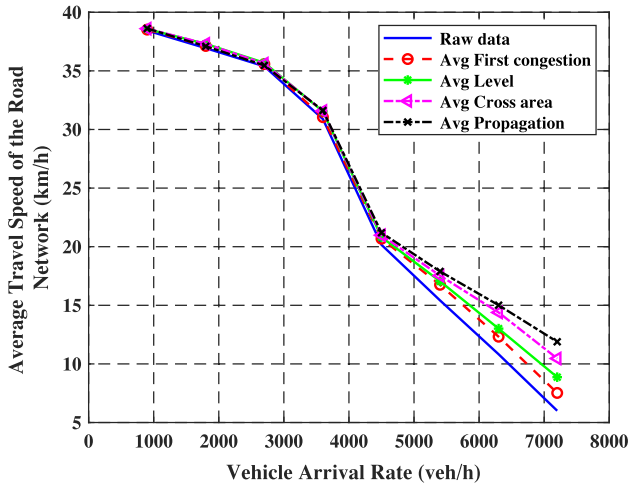


Fig. 12. Average travel speed of the road network after increasing the number of lanes on identified bottlenecks based on the existing methods and our proposed method.

can better capture the features of urban bottlenecks and have a superior performance in identifying bottlenecks for urban traffic networks.

D. Comparison With the Existing Bottleneck Identification Method

In this subsection, we first identify bottlenecks by using our proposed graph-theoretic method, the first congestion based method [19], congestion level based method and spatial cross area based method [25], respectively. Specifically, for our proposed method, as shown in Fig. 8, road segments 28 and 51 are regarded as bottlenecks in Sioux Falls network. For the first congestion based method, the road segment where congestion occurs first is more likely to be considered as a bottleneck, thus road segments 65 and 72 are seen as bottlenecks in the road network. For the congestion level based method, the road segments with higher congestion level are considered as bottleneck, as illustrated in Fig. 8, road segments 48 and 55 are regarded as bottlenecks. Finally, according to the spatial cross area based method, the road segments located at the spatial cross area of two congestion propagation patterns are more likely to be bottlenecks, in this case, road segments 51 and 63 are regarded as bottlenecks in Sioux Falls network.

Then, we increase the number of lanes on each identified bottleneck and compare the average travel speed of the road network based on different bottleneck identification methods under different vehicle arrival rates. As shown in Fig. 12, the horizontal axis indicates the vehicle arrival rate of the entire road network and the vertical axis describes the average network travel speed. We can see that when the vehicle arrival rate is small, after increasing the number of lanes on each identified bottleneck based on the existing methods and our proposed method, the average travel speed of the road network has little improvements. This is expected because the capacity of the road network has not been saturated and increasing the number of lanes on bottlenecks can hardly improve the average travel speed

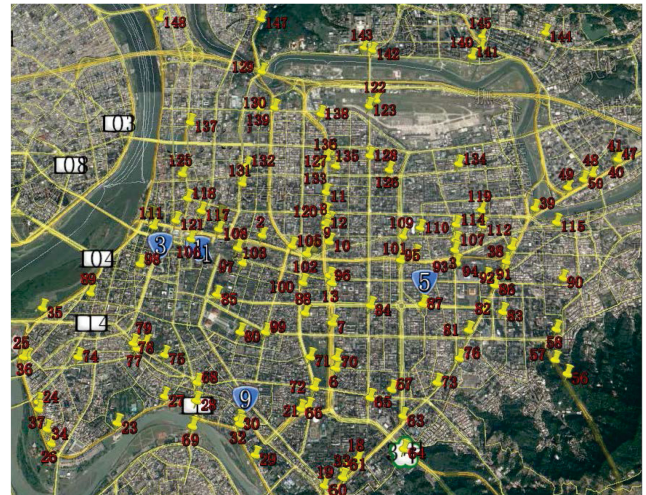


Fig. 13. An illustration of the locations of inductive loop detectors in Taipei.

of the road network. However, with the increase of vehicle arrival rates, congestion starts to occur in the road network, and in this occasion, when we increase the number of lanes on each bottleneck identified by both the existing methods and our proposed method, the average travel speed of the road network will be improved obviously. Furthermore, when vehicle arrival rates are large enough, increasing the number of lanes on bottlenecks identified by our proposed method will bring more improvement on network travel speed than increasing the number of lanes on bottlenecks identified by the existing methods. Especially, when the vehicle arrival rate is 7200 veh/h, the average travel speed can be improved by 25.3% using the first congestion based method, by 47.7% using the congestion level based method and by 74.3% using the spatial cross area based method, while our proposed method can provide 95.2% travel speed improvement for the road network, which indicates that our proposed bottleneck identification approach considering congestion propagation can provide a more effective and rigorous way to identify bottlenecks in road networks.

V. EXPERIMENTS AND DISCUSSION

In this section, we carry out experiments utilizing data collected from loop detective sensors in the traffic network of Taipei, Taiwan and use the proposed technique to identify bottlenecks in Taipei.

A. Data

An inductive loop detectors data set from the urban traffic network of Taipei, Taiwan is used for this research. The detector data were collected from 1 April, 2013 to 30 April, 2013 and we choose the weekday data to implement our experiments. In this data set, there are 153 detectors in the urban areas, as shown in Fig. 13 and the average speed, occupancy and flow data of all the detectors over 1-minute intervals for 24 hours a day were available. In this paper, we choose the average speed data to determine whether the road segments are congested

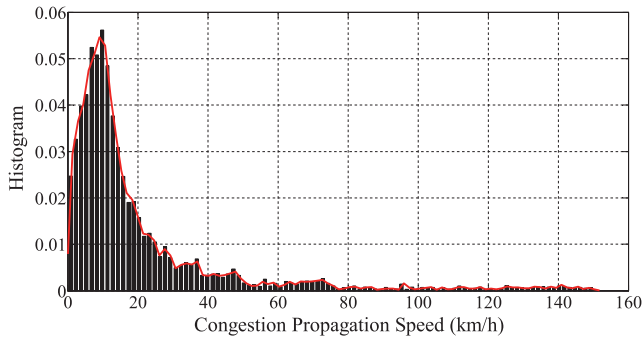


Fig. 14. An illustration of the histogram of the congestion propagation speed.

and implement our bottleneck identification using the proposed technique.

B. Experiments on Congestion Correlations

Utilizing the classification of congestion and non-congestion introduced earlier in the paper, we first obtain a set of congested road segments based on the inductive loop detectors data in Taipei. Then, we need to enforce the spatial threshold T_s and congestion propagation speed interval to get a set of congestion correlations based on the proposed Definition 2.

Term two road segments connected directly as first-order spatial neighbours. Then, it naturally follows that the second-order spatial neighbours of a road segment are the first-order neighbours of their first-order neighbours (excluding itself) and so on [56]–[58]. According to the simulation results in Section IV and the analytical result in [57], we find that the presence of spatio-temporal correlation between road segments extends to spatial order three but the strength of the correlation is diminished significantly beyond three orders. Therefore, in the experiments, we set the spatial threshold T_s as 2 km according to the actual average road length of traffic network in Taipei. In this way, we can connect two congested road segments as congestion correlated road segments. However, because of incidents in urban areas, such as traffic accidents and road construction, there might be some incidental congestion correlations which occur only few times and if we take these congestion correlations into consideration, it will lead to erroneous bottleneck identification. Therefore, in this paper, we need to delete these rarely happened congestion correlations and obtain a set of preliminary congestion correlations.

Moreover, we also need to determine the congestion propagation speed interval to pick out the realistic congestion correlations. Therefore, we calculate the congestion propagation speed of all the preliminary congestion correlations which is obtained only based on the spatial threshold T_s using their shortest path distances and the corresponding congestion time. Then, we can obtain the histogram of the speed of the preliminary congestion correlations, as shown in Fig. 14. However, if the shortest path distance between two road segments in a correlation is comparatively large (smaller than T_s) and congestion in the two road segments occur almost at the same time, then the congestion propagation speed will be quite large. Similarly, if the shortest

path distance between two road segments in a correlation is small and congestion in the two road segments occurs successively during a long time period, then the congestion propagation speed will be quite small. Both scenarios suggest a possible non-casual relationship. Therefore, in order to eliminate the impact of the aforementioned extreme congestion propagation speed, we choose a 80% confidence interval to determine the congestion propagation speed interval where the area in left tail is 15% and the area in right tail is 5%, according to the existing studies [21], [25] about congestion propagation speed and the actual road traffic network of Taipei. After deleting the preliminary congestion correlations which are not within the congestion propagation speed interval, we can get a set of congestion correlations and all of them can be seen as directed edges to construct congestion propagation graphs.

C. Experiments on Congestion Propagation Graphs and Maximal Spanning Tree

In this subsection, using the obtained congestion correlations, we connect them together based on our proposed method in Subsection III-B to construct congestion propagation graphs and map them onto the urban road network of Taipei, as shown in Fig. 15. We can see that there are 5 connected congestion propagation graphs in Fig. 15 and all the graphs are marked by different colors. The largest congestion propagation graph is located in the west of urban areas (the red solid line) and includes the most number of congestion correlations. This suggests that if congestion occurs in this area, the congestion tends to propagate to the more road segments and bottlenecks are more likely located among these road segments.

Then, utilizing Algorithm 2, we can build the maximal spanning trees from these congestion propagation graphs by regarding each road segment of these graphs as the root node of a congestion propagation tree respectively. For convenience, we present one maximal spanning tree from each congestion propagation graph. As shown in Fig. 16, road segment 27 is the root of one of the maximal spanning trees in CPG 1, road segment 18 is the root of one of the maximal spanning trees in CPG 2, road segment 63 is the root of one of the maximal spanning trees in CPG 3, road segment 91 is the root of one of the maximal spanning trees in CPG 4, and road segment 39 is the root of one of the maximal spanning trees in CPG 5. Particularly, the maximal spanning tree with root road segment 27 consists of 11 edges and congestion from road segment 27 can almost propagate to the whole west urban area of Taipei.

D. Experiments on Bottleneck Identification

In this subsection, according to the obtained maximal spanning trees, we calculate congestion costs of all road segments and then identify bottlenecks in urban traffic network of Taipei based on our proposed bottleneck identification approach. As shown in Fig. 17, the horizontal axis indicates each road segment in Taipei and the vertical axis describes congestion costs of all the 153 road segments. Moreover, the congestion propagation costs of road segments and the congestion costs of road segments themselves are marked by the white and grey bars respectively, and the sum

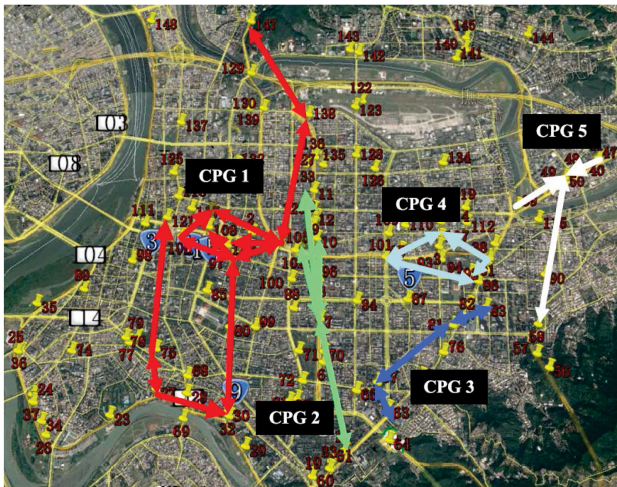


Fig. 15. An illustration of congestion propagation graphs in Taipei.

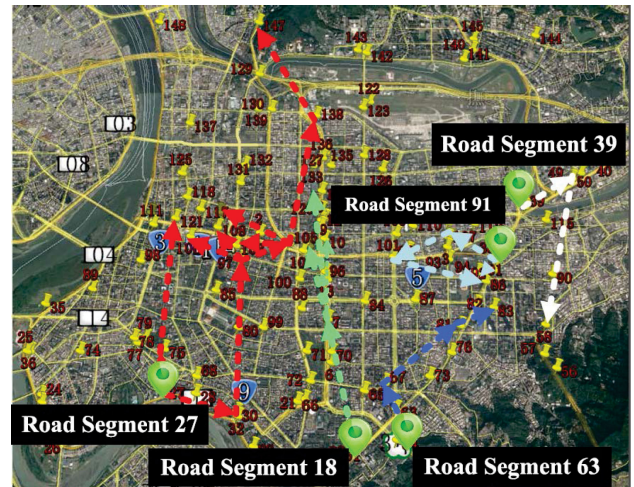


Fig. 16. An illustration of maximal spanning trees in Taipei.

of them can be seen as the total congestion cost of each road segment in Taipei road network. We can see that the congestion costs of some road segments are mainly caused by congestion costs on road segments themselves, such as road segments 30, 42, 50 and 121. While congestion on some road segments trends to propagate and leads to congestion on the other road segments, so the congestion costs of these road segments are mainly caused by congestion propagation costs, such as road segments 48 and 125. In this case, when we utilize the existing method that considers congestion levels on road segments only to identify bottlenecks, road segments 30, 42, 50 and 121 will be regarded as bottlenecks. However, as illustrated in Fig. 17, although the congestion levels on some road segments (e.g., road segments 48 and 125) are not as high as some other road segments (e.g., road segments 30, 42 and 50), their congestion propagation effects to the whole traffic network are significant and mitigating congestion on these road segments can lead to a network-wide traffic improvement. Thus, these road segments should also be considered as bottlenecks in the road network. Compared with existing bottleneck identification methods, the proposed urban bottleneck definition takes both congestion levels and congestion propagation costs into account, which provides a more intuitive and also effective notion for urban bottleneck identification. We can see that in Fig. 17, road segments 48, 50, 113, 121 and 125 are five road segments that incur the most congestion costs in the urban road network of Taipei and these road segments can be considered as bottlenecks in the road network of Taipei. In summary, defining bottlenecks in urban areas only based on congestion costs of road segments themselves will lead to the inaccuracy and ineffectiveness in identifying urban bottlenecks because the most congested road segments are not necessarily bottlenecks in urban traffic networks.

Moreover, according to the congestion costs of all 153 road segments, we divide them into five categories and map them onto the traffic network in Taipei. As shown in Fig. 18, road segments are marked by red labels when their congestion costs are greater than 1; marked by purple labels when their congestion costs are greater than 0.8 and less than 1; marked by yellow labels

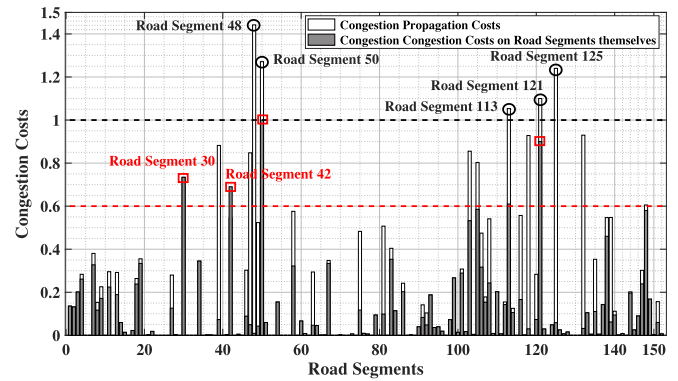


Fig. 17. Congestion costs of road segments in Taipei.



Fig. 18. Bottlenecks in Taipei.

when their congestion costs are greater than 0.6 and less than 0.8; marked by green labels when their congestion costs are greater than 0.4 and less than 0.6; marked by white labels if their congestion costs are less than 0.4. Specifically, the road segments marked by red labels (road segments 48, 50, 113, 121 and 125) are more likely to be bottlenecks in the road network of Taipei and if the traffic conditions on these road segments can

be improved, congestion in the entire urban traffic network can be mitigated significantly.

VI. CONCLUSION

In this paper, in order to identify bottlenecks in urban traffic network, we proposed a novel urban bottleneck definition, which calculates congestion costs of road segments to identify bottlenecks in urban areas taking into account both road congestion level cost and congestion contagion cost. First, we obtained a set of congestion correlations which connect congestion in two road segments according to the shortest path distance between the two road segments and the corresponding congestion propagation speed. Then, we proposed an algorithm to connect these correlations together and obtained congestion propagation graphs. We also presented an algorithm to build maximal spanning trees in the congestion propagation graphs. After that, we calculated the road congestion level costs themselves according to the normalized average road occupancy and importance of road segments and then obtained the congestion contagion costs of road segments based on the maximal spanning trees. Moreover, using road congestion level costs themselves and congestion contagion costs, we calculated congestion costs of road segments and identified bottlenecks in urban areas. Finally, we validated that the proposed bottleneck identification technique on SUMO and the results indicated that our proposed method could provide a more effective and rigorous way in identifying urban bottlenecks. We further utilized our proposed method to identify the bottlenecks in urban areas of Taipei based on inductive loop detector data and the experiment results showed that the most congested road segments are not necessarily bottlenecks in the road network, which suggests the effectiveness of our proposed method for urban bottleneck identification.

To the best of our knowledge, this is the first work to quantify the congestion costs of road segments and identify bottlenecks using both congestion costs on road segments themselves and congestion propagation costs. The simulation and experiment results derived in this paper can be utilized to provide proper guidance for road capacity improvement and congestion mitigation in urban traffic networks. In the future, combining the traffic data with more details about road characteristics, our proposed method can better analyze congestion propagation and achieve finer-grained bottleneck identification in urban traffic networks. Moreover, the notion of congestion propagation in this paper can also be utilized to predict congestion in urban areas by using deep learning, such as CNN and GCN, which will also play an important role to further improve traffic performance in road networks.

REFERENCES

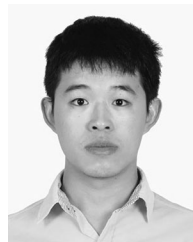
- [1] A. Bazzan, D. de Oliveira, F. Klügl, and K. Nagel, "To adapt or not to adapt—Consequences of adapting driver and traffic light agents," *Adaptive Agents and Multi-Agent Systems III. Adaptation and Multi-Agent Learning*, New York, NY, USA: Springer, pp. 1–14, 2008.
- [2] Y. Zhang, C. Li, T. Luan, Y. Fu, W. Shi, and L. Zhu, "A mobility-aware vehicular caching scheme in content centric networks: Model and optimization," *IEEE Trans. Veh. Tech.*, vol. 68, no. 4, pp. 3100–3112, Apr. 2019.
- [3] L. Morgan, "The effects of traffic congestion," Sep. 2014. [Online]. Available: <http://traveltips.usatoday.com/effects-traffic-congestion-61043.html>
- [4] A. Ahmad, R. Arshad, S. Mahmud, G. Khan, and H. Al-Raweshidy, "Earliest-deadline-based scheduling to reduce urban traffic congestion," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1510–1526, Aug. 2014.
- [5] D. Schrank, B. Eisele, T. Lomax, and J. Bak, "Urban mobility scorecard," *Texas A&M Transp. Inst.*, Accessed on: Sept. 2015.
- [6] N. Mineta, "National strategy to reduce congestion on America's Transportation Network," U.S. Dept. Transp., Washington, 2006.
- [7] L. Qi, M. Zhou, and W. Luan, "A two-level traffic light control strategy for preventing incident-based urban traffic congestion," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 13–24, Jan. 2018.
- [8] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialo, and Y. Wang, "Review of road traffic control strategies," *Proc. IEEE*, vol. 91, no. 12, pp. 2043–2067, Dec. 2003.
- [9] Y. Fu, C. Li, T. Luan, Y. Zhang, and F. Yu, "Graded warning for rear-end collision: An artificial intelligence aided algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 2, pp. 565–579, Feb. 2020.
- [10] A. Kouvelas, K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "A hybrid strategy for real-time traffic signal control of urban road networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 884–894, Sep. 2011.
- [11] J. Haddad and N. Geroliminis, "On the stability of traffic perimeter control in two-region urban cities," *Transp. Res. B*, vol. 46, no. 9, pp. 1159–1176, 2012.
- [12] N. Geroliminis, J. Haddad, and M. Ramezani, "Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 348–359, Mar. 2013.
- [13] G. Kim, Y. Ong, T. Cheong, and P. Tan, "Solving the dynamic vehicle routing problem under traffic congestion," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2367–2380, Aug. 2016.
- [14] B. Kerner, "Control of spatiotemporal congested traffic patterns at highway bottlenecks," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 2, pp. 308–320, Jun. 2007.
- [15] B. Coifman and S. Kim, "Extended bottlenecks, the fundamental relationship, and capacity drop on freeways," *Transp. Res. A*, vol. 45, no. 9, pp. 980–991, 2011.
- [16] L. Zhang and D. Levinson, "Ramp metering and freeway bottleneck capacity," *Transp. Res. A*, vol. 44, no. 4, pp. 218–235, 2010.
- [17] X. Ye, S. Deng, W. Yang, Z. Bao, and J. Chen, "Evaluating the impacts of travel information on urban traffic congestion propagation and bottleneck identification," in *Proc. Cota Int. Conf. Transp. Professionals*, 2014, pp. 1890–1901.
- [18] J. Long, Z. Gao, H. Ren, and A. Lian, "Urban traffic congestion propagation and bottleneck identification," *Sci. China*, vol. 51, no. 7, pp. 948–964, 2008.
- [19] J. Gong and W. Yang, "The traffic bottleneck analysis on urban expressway under information condition," in *Proc. Int. Conf. Power Electron. Intell. Transp. Syst.*, 2010, pp. 400–403.
- [20] H. Nguyen, W. Liu, and F. Chen, "Discovering congestion propagation patterns in spatio-temporal traffic data," *IEEE Trans. Big Data*, vol. 3, no. 2, pp. 169–180, Jun. 2017.
- [21] Y. Wang, J. Cao, W. Li, and T. Gu, "Mining traffic congestion correlation between road segments on GPS trajectories," in *Proc. IEEE Int. Conf. Smart Comput.*, 2016, pp. 1–8.
- [22] R. Tao, Y. Xi, and D. Li, "Simulation analysis on urban traffic congestion propagation based on complex network," in *Proc. IEEE Int. Conf. Service Operations Logistics, Informat.*, 2016, pp. 217–222.
- [23] Y. Li, Y. Liu, and K. Zou, "Research on the critical value of traffic congestion propagation based on coordination game," *Procedia Eng.*, vol. 137, no. 1, pp. 754–761, 2016.
- [24] J. Ma, C. Li, Z. Liu, and Y. Duan, "On traffic bottleneck in green ITS navigation: An identification method," in *Proc. IEEE Veh. Technol. Conf.*, 2016, pp. 1–5.
- [25] W. Lee, S. Tseng, J. Shieh, and H. Chen, "Discovering traffic bottlenecks in an urban network by spatiotemporal data mining on location-based services," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1047–1056, Dec. 2011.
- [26] W. Yue, C. Li, and G. Mao, "Urban traffic bottleneck identification based on congestion propagation," in *Proc. IEEE Int. Conf. Commun.*, 2018, pp. 1–6.
- [27] 2006. [Online]. Available: https://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/
- [28] 2011. [Online]. Available: http://www.sumo.dlr.de/userdoc/Sumo_at_a_Glance.html

- [29] P. Hunt, D. Robertson, R. Bretherton, and R. Winton, "SCOOT: A traffic responsive method of coordinating signals," *Transport Road Res. Lab.*, Tech. Rep. TRRL-LR-1014, 1981.
- [30] P. Lowrie, "The Sydney coordinated adaptive traffic system," in *Proc. Int. Conf. Road Traffic Signal.*, 1982, pp. 67–70.
- [31] J. Pan, I. Popa, K. Zeitouni, and C. Borcea, "Proactive vehicular traffic rerouting for lower travel time," *IEEE Trans. Veh. Tech.*, vol. 62, no. 8, pp. 3551–3568, Oct. 2013.
- [32] Z. Cao, H. Guo, and J. Zhang, "A multiagent-based approach for vehicle routing by considering both arriving on time and total travel time," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 3, pp. 1–21, 2018.
- [33] Z. Cao, H. Guo, and J. Zhang and U. Fastenrath, "Multiagent-based route guidance for increasing the chance of arrival on time," in *Proc. 14th AAAI Conf. Artif. Intell.*, 2016, pp. 3814–3820.
- [34] L. Xiao and H. Lo, "Combined route choice and adaptive traffic control in a day-to-day dynamical system," *Netw. Spatial Econ.*, vol. 15, no. 3, pp. 697–717, 2015.
- [35] Z. Cao, S. Jiang, J. Zhang, and H. Guo, "A unified framework for vehicle rerouting and traffic light control to reduce traffic congestion," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1958–1973, Jul. 2017.
- [36] W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing, "Discovering spatio-temporal causal interactions in traffic data streams," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 1010–1018.
- [37] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, pp. 818–833, 2017.
- [38] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*.
- [39] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," in *Proc. Int. Conf. Learn. Representations*, 2014, pp. 1–14.
- [40] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: data-driven traffic forecasting," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–16.
- [41] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 694–711, May 2006.
- [42] Y. Ki and D. Baik, "Model for accurate speed measurement using double-loop detectors," *IEEE Trans. Veh. Technol.*, vol. 55, no. 4, pp. 1094–1101, Jul. 2006.
- [43] H. Sun, J. Wu, D. Ms, and J. Long, "Spatial distribution complexities of traffic congestion and bottlenecks in different network topologies," *Appl. Math. Modell.*, vol. 38, no. 2, pp. 496–505, 2014.
- [44] G. Marfia and M. Rocchetti, "Vehicular congestion detection and short-term forecasting: A new model with results," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 2936–2948, Sep. 2011.
- [45] A. Chow, A. Santacreu, I. Tsapakis, G. Tanasaranond, and T. Cheng, "Empirical assessment of urban traffic congestion," *J. Adv. Transp.*, vol. 48, no. 8, pp. 1000–1016, 2014.
- [46] E. Gakis, D. Kehagias, and D. Tzovaras, "Mining traffic data for road incidents detection," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2014, pp. 1–6.
- [47] J. Ren, Y. Chen, L. Xin, J. Shi, B. Li, and Y. Liu, "Detecting and positioning of traffic incidents via video-based analysis of traffic states in a road segment," *IET Intel. Transport Syst.*, vol. 10, no. 6, pp. 428–437, 2016.
- [48] D. Hale, R. Jagannathan, M. Xyntarakis, P. Su, X. Jiang, J. Ma, J. Hu, and C. Krause, "Traffic Bottlenecks: Identification and Solutions," *Federal Highway Admin. (FHWA)*, U.S. Dept. Transp., Washington, D.C., USA, Tech. Rep. FHWA-HRT-16-036, 2016.
- [49] M. Ali, M. Adnan, S. Noman, and S. Baqueri, "Estimation of traffic congestion cost: A case study of a major arterial in Karachi," *Procedia Eng.*, vol. 77, no. 2014, pp. 37–44, 2014.
- [50] BTRE, "Estimating urban traffic and congestion cost trends for Australian cities," work paper 71, BTRE, Bureau Transport Regional Econ., Canberra ACT, 2007.
- [51] S. Wang, S. Djahel, Z. Zhang, and J. McManis, "Next road rerouting: A multiagent system for mitigating unexpected urban traffic congestion," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2888–2899, Oct. 2016.
- [52] E. Wang, Y. Yang, J. Wu, W. Liu, and X. Wang, "An efficient prediction-based user recruitment for mobile crowdsensing," *IEEE Trans. Mob. Comput.*, vol. 17, no. 1, pp. 16–28, Jan. 2018.
- [53] A. Chakirov and P. Fourie, "Enriched sioux falls scenario with dynamic and disaggregate demand," Tech. Rep., Future Cities Laboratory, Singapore ETH Centre (SEC), Singapore, 2014.
- [54] L. J. LeBlanc, E. K. Morlok, and W. P. Pierskalla, "An efficient approach to solving the road network equilibrium traffic assignment problem," *Transp. Res.*, vol. 9, no. 5, pp. 309–318, 1975.
- [55] H. Bar-Gera, F. Hellman, and M. Patriksson, "Computational precision of traffic equilibria sensitivities in automatic network design and road pricing," *Procedia - Social Behav. Sci.*, vol. 80, no. 7, pp. 41–60, 2013.
- [56] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transp. Res. C*, vol. 19, no. 4, pp. 606–616, 2011.
- [57] T. Cheng, J. Haworth, and J. Wang, "Spatio-temporal autocorrelation of road network data," *J. Geog. Sys.*, vol. 14, no. 4, pp. 389–413, 2011.
- [58] M. Asif *et al.*, "Spatiotemporal patterns in large-scale traffic speed prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 794–804, Apr. 2014.



ad hoc networks, and wireless sensor networks.

Changle Li (Senior Member, IEEE) received the Ph.D. degree in communication and information system from Xidian University, Xi'an, China, in 2005. He conducted his Postdoctoral Research in Canada and the National Institute of Information and Communications Technology, Japan, respectively. He was a Visiting Scholar with the University of Technology Sydney and is currently a Professor with the State Key Laboratory of Integrated Services Networks, Xidian University. His research interests include intelligent transportation systems, vehicular networks, mobile



Wenwei Yue (Student Member, IEEE) received the B.E. degree in information engineering, in 2015, from Xidian University, Xian, China, where he is currently working toward the Ph.D. degree in communication and information systems. From February 2017 to February 2019, he was a visiting Ph.D. student with the School of Computing and Communications, University of Technology Sydney. His current research interests include intelligent transportation systems, vehicular networks, and Big Data.



Guoqiang Mao (Fellow IEEE) has authored and coauthored more than 200 papers in international conferences and journals, which have been cited more than 8500 times. His research interests include intelligent transport systems, applied graph theory and its applications in telecommunications, Internet of Things, wireless sensor networks, wireless localization techniques, network modeling, and performance analysis. He has been an editor for the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (since 2018), IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2014–2019), IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (since 2010) and received "Top Editor" award for outstanding contributions to the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY in 2011, 2014, and 2015. He was a Co-Chair of IEEE Intelligent Transport Systems Society Technical Committee on Communication Networks. He has served as a Chair, Co-Chair, and TPC Member in a number of international conferences. He is a Fellow of IET.



Zhigang Xu (Member, IEEE) received the B.S. degree in automation, and the M.S. and Ph.D. degrees in traffic information engineering and control from Chang'an University, Xi'an, China, in 2002, 2005, and 2012, respectively. He is currently a Professor with the School of Information Engineering, and the Director of the Lab of Traffic information Sensing and Control at Chang'an University. He was with the University of California, Davis as a Visiting Scholar, in 2015. His research interests include connected and autonomous vehicles, traffic flow analysis, transportation optimization, and intelligent transportation systems. He is the Chair of CAVs Committee of World Transportation Convention and a Member of IEEE Intelligent Transportation Systems Society and American Society of Civil Engineers.