What Is the Root Cause of Congestion in Urban Traffic Networks: Road Infrastructure or Signal Control?

Wenwei Yue, *Member, IEEE*, Changle Li, *Senior Member, IEEE*, Yue Chen, Peibo Duan, *Member, IEEE*, and Guoqiang Mao, *Fellow, IEEE*

***Abstract*— Identifying the root cause of congestion and taking appropriate strategies to improve traffic network performance are important goals of Advanced Traffic Management Systems (ATMS). On many occasions, the causes of congestion are not necessarily attributable to road infrastructures themselves. Instead, signal control strategies at intersections are very often the major contributors of congestion. In lieu of this, in this paper, a root cause identification method is developed with consideration of the impact from both road infrastructure and traffic signal control. Firstly, we differentiate congestion effects between road segments and intersections to attribute the causes of congestion to road infrastructure and signal control respectively. Then, we construct causal congestion trees to model congestion propagation and quantify congestion costs for each road segment and intersection in the whole road network. A Markov model is utilized to capture congestion spatio-temporal correlation among multiple road segments and intersections simultaneously, with which the most critical root cause can be located. Furthermore, a gradient boosting decision tree based method is presented to predict the root cause of congestion according to traffic flows, signal control strategies and road topology in traffic networks. Finally, simulations based on Simulation of Urban Mobility (SUMO) validate the effectiveness of our proposed method in identifying and predicting the congestion root cause. Experiments are further conducted using inductive loop detector data to identify the root cause for the road network of Taipei.**

***Index Terms*— Congestion root cause identification, road infrastructure, signal control, congestion propagation, gradient boosting decision tree, SUMO.**

I. Introduction

TRAFFIC congestion is an important problem globally and is getting worse in many countries due to growing urbanization [1]. First, congestion causes excessive travel delay, which not only causes economical losses due to reduced productivity but also results in much more serious consequences, *e.g.*, the high-priority vehicles such as ambulances are stuck in congestion [2]. Second, traffic congestion has a negative effect on the environment. The Texas Transportation Institution (TTI) reported that congestion caused 3.1 billion gallons of extra fuel consumption in 2014 [3], which led to more CO_2 emissions and serious greenhouse effects. Last but not the least, congestion costs are rapidly increasing in recent years. According to the urban transportation scorecard [3], from 471 U.S. urban areas, the congestion "invoice" for the cost of extra time and fuel was $160 billion in 2014 and is expected to reach $192 billion in 2020.

To relieve traffic congestion, a number of congestion mitigation strategies were developed by worldwide transportation authorities [4]–[10]. Among them, road infrastructure construction and traffic signal control are seen as the most extensively utilized approaches [11], [12]. However, two problems should be addressed first before implementing these congestion mitigation strategies: 1) what are the major contributors of congestion in a traffic network? 2) which strategy is more effective to mitigate congestion, by enlarging road infrastructures [13] to increase network capacity or using an improved signal control to enhance the travel fluency? The two problems naturally lead to two research directions to identify the root cause of congestion in a traffic network. On one hand, according to the U.S. Department of Transportation, as the most vulnerable roads for congestion, traffic bottlenecks account for 40% congestion [14], [15] and intuitively, the notion of long-term bottlenecks indicates that the removal of these bottlenecks can improve traffic conditions at bottleneck locations and even in the entire road networks [1], [15]. Therefore, locating long-term bottlenecks and implementing congestion mitigation strategies at these bottlenecks are regarded as a more efficient way to bring network-wide traffic improvement. On the other hand, each road in urban road networks can be divided into two parts: a road segment and an intersection. When congestion regularly occurs at a particular road segment, road infrastructure construction tends to be the more effective way to relieve congestion intuitively. On the contrary, when congestion occurs more often near a particular intersection, however, it is more beneficial to employ advanced signal control to improve traffic situations. Therefore, if the cause

Manuscript received July 1, 2020; revised March 13, 2021; accepted May 20, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB1600100, in part by the National Natural Science Foundation of China under Grant U1801266, and in part by the Youth Innovation Team of Shaanxi Universities. The Associate Editor for this article was S. Sacone. *(Corresponding author: Changle Li.)*

Wenwei Yue, Changle Li, Yue Chen, and Guoqiang Mao are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China, and also with the Research Institute of Smart Transportation, Xidian University, Xi'an 710071, China (e-mail: wwyue@xidian.edu.cn; clli@mail.xidian.edu.cn; chenyue_albert@stu.xidian.edu.cn; gqmao@xidian.edu.cn).

Peibo Duan is with the School of Software, Northeastern University, Shenyang 110819, China (e-mail: duanpeibo@swc.neu.edu.cn).

Digital Object Identifier 10.1109/TITS.2021.3085021

of a congestion can be pinpointed to a road segment or an intersection, more effective traffic congestion mitigation can be implemented.

As a promising way to mitigate traffic congestion, a number of studies have been investigated in identifying root causes of congestion in road networks [1], [16]–[18]. Gong and Yang [16] analyzed the major contributor of congestion on urban expressways by using the spatio-temporal occupancy scatter graph and regarded the place where congestion first occurred as the root cause of congestion in a road network. In [17], Lee *et al.* developed a spatio-temporal traffic bottleneck mining model based on spatio-temporal traffic patterns[1] to discover the root causes of congestion in urban road networks. Ma *et al.* [18] proposed a simulation based bottleneck identification method by using VISSIM simulator, where they compared the total travel time in a road network under the same traffic pattern before and after a particular road is congested and considered the roads whose congestion leads to more travel time difference as the root causes of traffic congestion. Our previous work [1] presented a bottleneck identification method considering both congestion costs on roads themselves and congestion propagation costs that congestion may propagate to the other roads and regarded the roads with higher congestion costs as the root causes of congestion in road networks. To the best of our knowledge, most existing works on congestion cause identification strive to locate roads contributing most to congestion. They provide effective solutions to the first question of the aforementioned problems. However, there are few researches targeting at the second question, which arguably is at least equally important as the first question.

To fill the gap, in this paper, we design a congestion root cause identification method, which quantifies the level of congestion at each road segment and intersection, and can provide more intuitive control strategies to relieve traffic congestion in road networks. Furthermore, with the gradient boosting decision trees (GBDT), we provide an approach to predict root causes of congestion according to traffic flows, road topology and signal control strategies in a road network. Finally, simulations based on Simulation of Urban Mobility (SUMO) are conducted to illustrate the effectiveness of our proposed method. Experiments using inductive loop detector data are further carried out to identify the root cause in the road network of Taipei. More specifically, contributions of this paper are presented as follows:

- Unlike most of existing techniques attributing causes of congestion at the most congested roads, the proposed method differentiates congestion effects of road segments and intersections, respectively which assists to better identify the root causes of congestion (*e.g.*, road infrastructures and traffic signal) and suggests a specific congestion mitigation strategy to enhance traffic condition in road networks;
- Based on a combined use of graphical models and fuzzy logic, a root cause identification method is proposed with consideration of congestion propagation among road segments and intersections in a road network, which can not only locate the main contributors of network congestion, but also quantify congestion costs of all road segments and intersections to better capture their congestion impacts to the whole road network;
- By using a Markov model, we identify the most critical congestion root cause among multiple correlated road segments and intersections, which provide an effective method to analyze and quantify the congestion relationships among multiple road segments and intersections simultaneously;
- A gradient boosting decision tree based method is utilized to predict root causes of congestion in road networks, which provides a rigorous approach to predict the possible root cause of congestion utilizing several easily-measured factors, such as traffic flows, signal control strategies and road topology.

The rest of this paper is organized as follows. Section II reviews the related works. Section III presents the proposed congestion root cause identification method. In Section IV, a gradient boosting decision trees based approach is developed to predict root causes of congestion according to traffic flows, signal control strategies and road topology. Simulations based on SUMO and experiments using inductive loop detector data are conducted in Section V. Finally, Section VI concludes the paper.

II. Related Works

Due to the rapid expansion of urbanization, almost all metropolitan cities in the world are suffering from an unprecedented increase in road congestion, which motivates the development of many control strategies (*e.g.*, route guidance, signal control and road infrastructure construction) to relieve the negative effects of congestion. For route guidance strategies, Claes *et al.* [4] presented an anticipatory method for vehicle route guidance that predicted road conditions in the near future and provided vehicles with the predictive rerouting information to avoid potential traffic congestion. Pan *et al.* [5] proposed a proactive control strategy to provide rerouting guidance for travelers when signs of congestion were observed. Simulation results based on SUMO illustrated the effectiveness of the proposed method in reducing vehicle travel times. However, the main purpose of route guidance strategies is to enable certain travelers to avoid the current road congestion and obtain more reliable and efficient routes from their origins to destinations, which cannot solve the congestion problem fundamentally and bring a network-wide traffic improvement. Moreover, there still exist many challenges in implementing route guidance strategies for road administrators, such as the real-time traffic information detection and dissemination and the compliance rate of drivers, which still need further investigation and evaluation. Therefore, as more intrinsic and controllable strategies, signal control and road infrastructure construction are more adopted [20]. A number of traffic responsive signal control strategies, such as Split, Cycle and Offset Optimization Technique (SCOOT) [6] and Sydney

[1]A spatio-temporal traffic pattern is a distribution of traffic flows in space and time [19].

Cooperative Adaptive Traffic System (SCATS) [7], were extensively applied in many cities to automatically respond to the real-time traffic conditions to enhance traffic efficiency and relieve road congestion. Aboudolas *et al.* [21] developed store-and-forward model based signal control methods in large-scale urban road networks to minimize the average queue lengths under saturated traffic conditions. Kouvelas *et al.* [22] employed a hybrid traffic signal control strategy that utilized a real-time Webster strategy for under saturated traffic conditions and adopted Traffic-Responsive Urban Control (TUC) when traffic conditions were close to saturation. Simulation results illustrated the capabilities of the hybrid strategy. Moreover, many road infrastructure construction projects were also developed by worldwide transportation agencies. The aims of scheme "A6 Clapham Bypass" were to "relieve congestion", "improve road safety" and "provide the opportunity for environmental improvement" in Clapham [8]. Hymel [9] investigated the impact of traffic congestion on employment growth in large U.S. metropolitan areas. The results indicated that high level congestion inhibited employment growth and increasing the efficiency of public infrastructure can spur economies. In order to improve traffic flows and decrease congestion, investment in road infrastructure increased in Great Britain by about 55% between 2001 and 2008 and in 2008-2009 over $16000 million of public expenditure was spent on transportation infrastructure and the major road network was extended by about 175 kilometers [10].

Obviously, it is impractical and uneconomical for transportation agencies to employ these congestion control strategies on all roads in a traffic network. Therefore, it is essential to locate the major congestion contributors and, in other words, the most vulnerable points in a road network [23]. For non-recurrent traffic congestion, Anbaroglu *et al.* [24] developed a non-recurrent congestion detection method using spatio-temporal clustering. Kwon *et al.* [25] utilized statistical regression to analyze the components of non-recurrent congestion. Chow *et al.* [26] employed an empirical assessment of urban traffic congestion in Central London and attributed congestion to different causes (*e.g.*, recurrent congestion, accidents, roadwork, special events, and strikes) using linear regression. On the other hand, recurrent congestion can reflect the long-term traffic conditions in the road network more effectively. Thus for recurrent traffic congestion, Ye *et al.* [27] defined a critical threshold v/c based on the ratio of travel speed and road capacity and when the real-time traffic information v/c on a road is higher than a pre-designated threshold, these roads were considered as bottlenecks in a road network. Gong and Yang [16] analyzed the bottlenecks in urban expressways and considered the roads where congestion occurred first as root causes of congestion in a road network. The proposed identification method was evaluated effectively based on detector data in the road network of Shanghai. Lee *et al.* [17] developed a three-phase spatio-temporal traffic bottleneck mining (STBM) model to identify the root cause of congestion in an urban network. Traffic information was collected in the first phase. The second phase defined and identified several traffic congestion related patterns according to the different spatio-temporal traffic flow distributions at road segments or intersections. The traffic bottleneck mining heuristics were proposed in Phase III to discover traffic bottlenecks in an urban network. Experiment results by using a taxi dispatching system confirmed that the discovered spatio-temporal bottlenecks matched the travelers' experience. Ma *et al.* [18] identified and evaluated bottlenecks in urban areas considering both vehicle travel costs and network effectiveness. They compared the total travel time of all vehicles under the same traffic flow before and after a particular road was in congestion by using VISSIM traffic simulator and regarded the roads with more travel time differences as congestion bottlenecks. Simulation results based on the road network in Nanjing illustrated the effectiveness of the proposed method in identifying urban traffic bottlenecks. A congestion propagation based bottleneck identification method was proposed in our previous work [1]. We first proposed a bottleneck definition with consideration of both congestion on roads themselves and congestion propagation effects to the other roads. Then, a combination of graphical models, maximal spanning trees and Markov model were utilized to identify traffic bottlenecks and quantify congestion effects of the identified bottlenecks to the entire network. Finally, simulation based on SUMO indicated the superiority of the proposed approach in bottleneck identification compared to existing bottleneck identification methods and experiments were also conducted to identify bottlenecks in the urban network of Taipei by using inductive loop detector data.

In summary, existing bottleneck identification methods for recurrent traffic congestion above can achieve a vigorous performance in locating the most vulnerable roads for congestion in a road network and relief of congestion at the identified bottlenecks can bring a network-wide traffic improvement. However, there are few researches focusing on the causes of congestion at the identified bottlenecks. Intuitively, improving signal control policy is considered as the more effective strategy to mitigate traffic congestion when congestion is more likely to occur at an intersection; and when congestion more often happens on a certain road segment, road infrastructure construction tends to be the dominant way to improve traffic flow and relieve road congestion. Therefore, if congestion effects between intersections and road segments can be differentiated, we may identify solutions that do not require infrastructure construction to increase road capacity, but employing an improved signal control to solve the problem. In this case, more specific root causes for congestion can be captured and more explicit control strategies can be further provided for road administrators to relieve traffic congestion effectively and efficiently.

III. Root Cause Identification

In this section, as shown in Fig. 1, we first introduce both the definitions of congestion on road segments and at intersections to differentiate the congestion effects on road segments and at intersections. Then, in order to analyze the congestion propagation process in a road network, we present the definition of correlation between congestion on road segments and at intersections. After that, we combine a graph-theoretic method

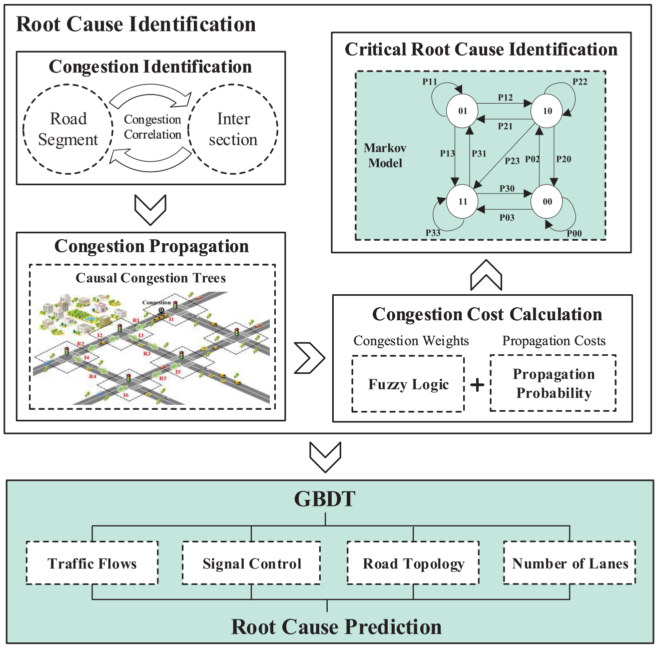

Fig. 1. Roadmap of root cause identification and prediction.

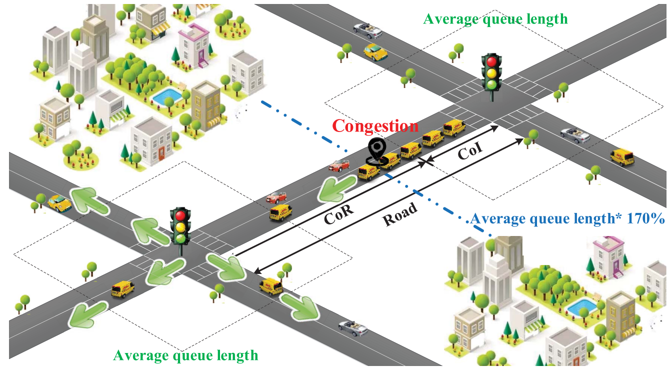

Fig. 2. An illustration of congestion propagation.

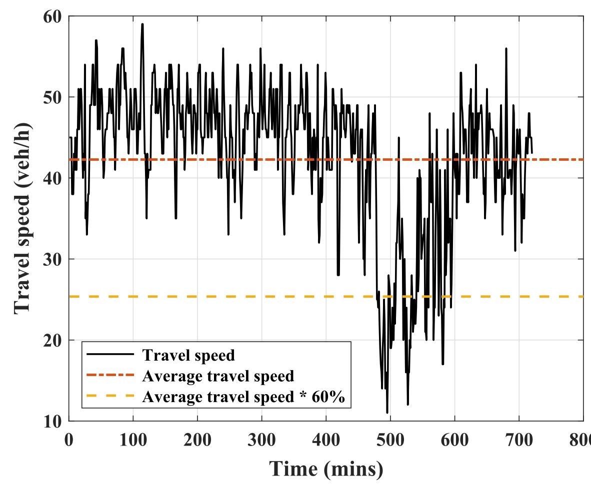

Fig. 3. An illustration of congestion on a road segment.

and fuzzy logic to quantify congestion costs of both road segments and intersections. Furthermore, a Markov model is utilized to identify the most critical root cause of congestion in a road network.

A. Congestion and Congestion Propagation

Real-time vehicle travel speed and road occupancy are often utilized to identify traffic congestion on road segments [28]–[30]; while the queue length is usually used to determine traffic conditions at intersections and evaluate the effectiveness of signal control strategies [31]–[33]. Moreover, to differentiate congestion effects on road segments and at intersections, and investigate congestion propagation in urban road networks, definitions about congestion at road segments and intersections should be clarified respectively. Most of existing works defined congestion on a road by using a fixed threshold for traffic information, such as average travel speed, road occupancy and traffic flow [28], [29], and when the real-time traffic information is higher or lower than the pre-designated threshold, the road is considered to be congested. For example, as reported by the Ministry of Public Security of China [30], when the travel speed on a road is less than 20 km/h, the road can be seen as congested. Li and Dai [34] determined the queue length threshold using video camera detectors at intersections, to identify traffic congestion. However, because different road segments and intersections have different characteristics (*e.g.*, road lengths, number of lanes and signal control strategies), a pre-designated threshold may not suit different road networks and even different road segments and intersections in the same road network. In this case, a number of researches identified traffic congestion by considering individual road characters. For instance, Nguyen *et al.* [35] suggested that a road segment is congested when the travel time on a road segment is above 80% of its time distribution. With consideration of the signal control strategy at each intersection, Luo *et al.* [36] divided traffic conditions into three levels (unsaturated, slightly saturated and over-saturated) according to average queue length during each cycle. In this paper, as illustrated in Fig. 2, we first give the congestion definitions of road segments (CoR) and intersections (CoI) adopting existing conventions, respectively.

Definition 1 (CoR): A road segment is considered to be congested if the real-time traffic speed is lower than the r% of its average travel speed.

According to [1], experiment results indicate the best average travel speed evaluation for congestion identification on road segments when $r = 60$, thus we utilize 60% of average travel speeds as the metric to identify congestion on each road segment. As shown in Fig. 3, the black line denotes the travel speed on a road segment over 1-minute intervals for 12 hours. The red dotted line represents the average speed of the road segment (42.27 km/h) and the yellow dotted line indicates the 60% of average travel speed (25.37 km/h). In the case, according to Definition 1, when the travel speed on this road segment is lower than 25.37 km/h at a specific time, the road segment is considered to be congested.

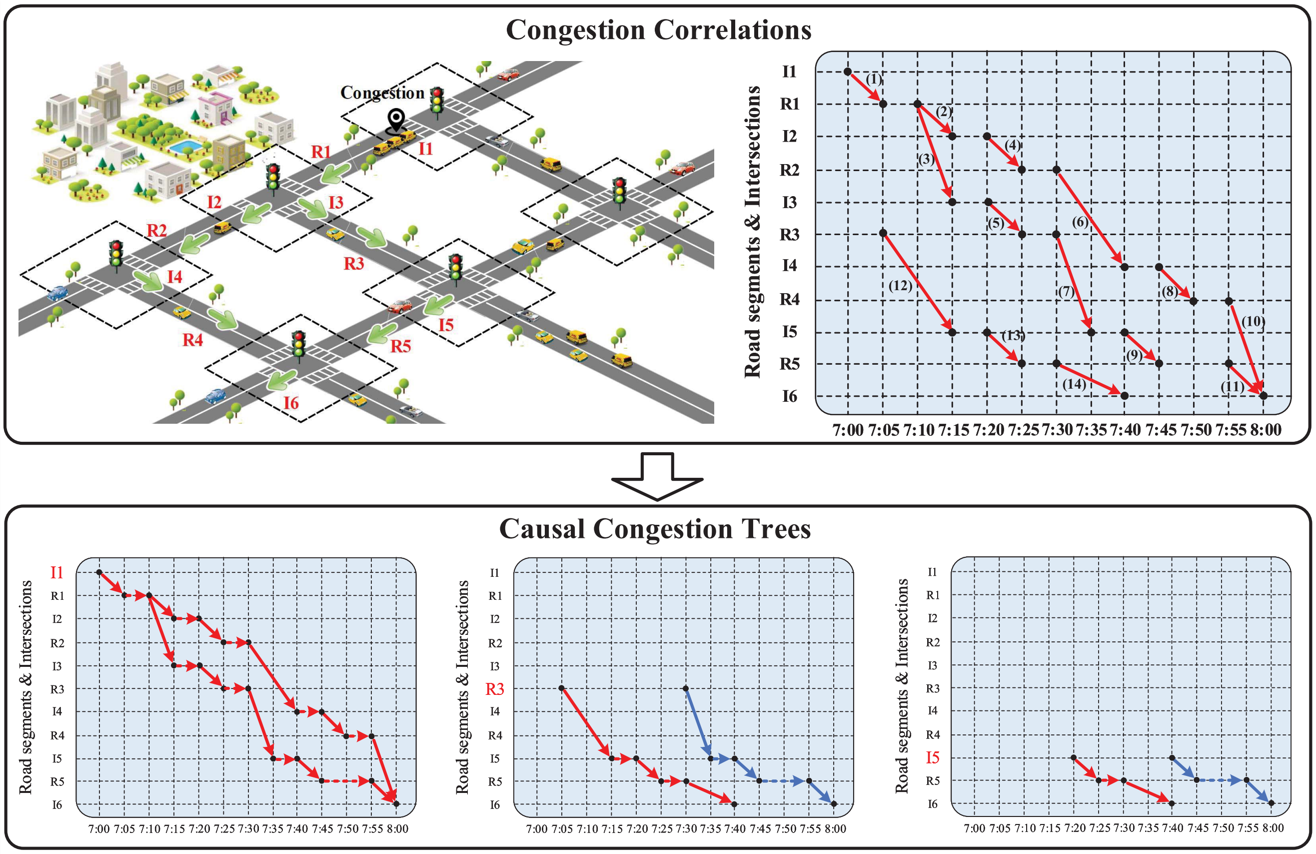

Fig. 4. Congestion correlations and causal congestion trees.

Definition 2 (CoI): An intersection is regarded to be congested if the queue length is longer than the w% of its average waiting queue length.

In this paper, w varies between 110 and 190. An illustration of congestion on an intersection with $w = 170$ is given in Fig. 2, where the black line indicates the average queue length at an intersection and the blue dotted line refers to 170% of the average waiting queue length. Thus, according to Definition 2, when the real-time queue length at an intersection is longer than the predetermined threshold, the intersection can be considered as congested.

Furthermore, when congestion occurs on a road, it tends to propagate to its neighboring roads and leads to more widespread congestion [37]. Specifically, considering the difference between congestion on intersections and road segments, as illustrated in Fig. 2, when congestion occurs at an intersection, it may lead to congestion on the linked road segment and then it is likely to affect the traffic flows of upstream intersections resulting in congestion propagation in road networks. In this case, to analyze the relationship between congestion on road segments and at intersections in urban traffic networks, congestion propagation between road segments and intersections is defined as follows.

Definition 3 (Congestion Propagation): Congestion propagation between intersections and road segments occurs if one of the following requirements is satisfied:

- Congestion occurs at an intersection at time t_0 and another congestion occurs on a road segment of the same road within time $t_0 + T$, where T is the temporal threshold indicating the time duration between congestion on adjacent intersections and road segments;
- Congestion occurs on a road segment at time t_0 and another congestion occurs on its neighboring upstream intersections within time $t_0 + T$.

B. Causal Congestion Trees

To depict congestion propagation in the road network, we propose a method to construct spatio-temporal causal congestion trees by using the obtained congestion correlations above, which can uncover the congestion effects of a certain root cause to the other road segments and intersections in a road network. Specifically, the definition of causal congestion tree (CCT) is presented in Definition 4.

Definition 4 (Causal Congestion Tree): A causal congestion tree describes the propagation pattern from congestion on a root cause to the neighboring road segments and intersections in urban road networks.

Fig. 4 presents an example to demonstrate the construction of spatio-temporal causal congestion trees. Congestion propagates from intersection $I1$ to other road segments and intersections in the road network gradually. We can obtain a set of congestion propagation pairs (CPPs) according to Definition 3. As shown in the upper figure of Fig. 4, each black dot indicates congestion on the road segment or intersection at the corresponding congestion occurrence time,

Algorithm 1 Constructing Spatio-Temporal Causal Congestion Trees

```
Input: A set of CPPs among intersections and road segments with congestion occurrence time. There is a road segment and an intersection in a CPP indicating that congestion from the first road segment or intersection (parent node) to another road segment or intersection (child node).
Output: A set of spatio-temporal causal congestion trees for each intersection or road segment.
for Each CPP_i (i ∈ (1, ..., H)) do
  Trees ← an empty set;
  Trees ← Trees ⋃ FindChildren (CPP_i);
  return Trees;
end for

function FindChildren (CPP_i)
if the congestion occurrence time of the child node in CPP_i is the last minute in sampling data then
  return CPP_i.children;
end if
CPP_i.children ← an empty set;
for each CPP_u (u ∈ (i + 1, ..., H)) do
  if the congestion occurrence time of CPP_i.child is later than that of CPP_u.parent then
    continue;
  end if
  if CPP_u.child ∈ CPP_i.children then
    continue;
  end if
  if (CPP_i.child == CPP_u.parent) then
    CPP_i.children ← CPP_i.children ⋃ FindChildren (CPP_u);
  end if
end for
end function
```

whereas each directed edge indicates congestion propagation among road segments and intersections. Then, considering the spatio-temporal relationships among different congestion propagation pairs, we utilize each road segment or intersection as a root node respectively to construct the causal congestion tree for each root node by using Algorithm 1. In this case, when we connect the spatio-temporal causal congestion trees with the same root node together, we can obtain the congestion propagation pattern of a road segment or an intersection in the whole road network. This strategy can not only illustrate the congestion propagation paths and influence areas of a certain road segment or intersection, but help to quantify congestion costs for all road segments and intersections to identify the most significant road segments or intersections in terms of congestion propagation. Especially, the causal congestion trees for intersection $I1$, road segment $R3$ and intersection $I5$ are presented respectively in the lower figure of Fig. 4.

C. Root Cause Identification

According to the constructed causal congestion trees, congestion propagation effects of each intersection or road segment can be modeled. In this subsection, we first propose a definition of congestion costs considering both congestion weights on intersections or road segments themselves and congestion propagation costs that congestion may propagate to other intersections and road segments, as shown in Definition 5. Then, based on fuzzy logic, we develop a method to quantify congestion costs for all intersections and road segments in a road network.

Definition 5 (Congestion Cost): Congestion cost of an intersection or a road segment can be calculated according to the sum of its congestion weight and congestion propagation cost. The congestion weight indicates the normalized congestion level on each intersection or road segment itself, which is determined according to congestion degree [2], road importance and average congestion time in a day. On the other hand, the congestion propagation cost demonstrates congestion on an intersection or a road segment leads to congestion on its neighboring intersections or road segments, which can be calculated according to congestion propagation probabilities among congestion on intersections and road segments and congestion weights of the involved intersections and road segments.

Specifically, a number of metrics are utilized to quantify road congestion weights. For example, travel speed, traffic flow and road occupancy are often adopted to describe the real-time road conditions; road importance is included to indicate congestion effect of a particular road, because the same level of congestion on different intersections or road segments can lead to different impacts on a road network; meanwhile, to illustrate long-term effects of congestion for vehicle travel reliability, the average congestion time in a day is also considered for congestion quantification [1], [38]. In summary, in this paper, we measure the congestion weight of each intersection or road segment with consideration of average congestion degree $\mathcal{CD}$, road importance $\mathcal{RI}$ and average congestion time in a day $\mathcal{CT}$.

However, it is an NP-complete problem to analyze the incorporation among the three influential factors above and evaluate their effects on road congestion weight [39]. Since the efficiency of solving the NP-complete problem, the fuzzy logic is applied in this paper for the quantification of congestion weights.

A fuzzy logic system is composed of fuzzification, fuzzy control rule base, fuzzy interference and defuzzification. The process of converting numerical inputs into linguistic values by using fuzzy membership functions is called "fuzzification." Then, the fuzzy interference calculates the linguistic outputs according to the matching rules in fuzzy control rule base. Finally, defuzzification coverts the fuzzy outputs to numerical values.

1) Fuzzification: By using a triangular membership function in (1), the average congestion degree, road importance and average congestion time in a day can be converted into a number of fuzzy inputs. In this case, the average congestion degree $\mathcal{CD}$ can be mapped into

[2]Congestion degree for intersections are quantified according to w in Definition 2 and are calculated based on r in Definition 1 for road segments.

TABLE I
IF-THEN RULE BASE

Rule No.	IF (AND)			THEN
	Average congestion degree ($\mathcal{CD}$)	*Road importance* ($\mathcal{RI}$)	*Average congestion time* ($\mathcal{CT}$)	*Congestion weights* ($\mathcal{W}$)
1	H	L	L	VH
2	H	L	M	H
3	H	L	S	H
4	H	M	L	VH
5	H	M	M	H
6	H	M	S	M
7	H	S	L	H
8	H	S	M	M
9	H	S	S	M
10	M	L	L	H
11	M	L	M	M
12	M	L	S	L
13	M	M	L	H
14	M	M	M	M
15	M	M	S	L
16	M	S	L	M
17	M	S	M	L
18	M	S	S	L
19	L	L	L	H
20	L	L	M	L
21	L	L	S	VL
22	L	M	L	M
23	L	M	M	L
24	L	M	S	VL
25	L	S	L	M
26	L	S	M	L
27	L	S	S	VL

a fuzzy set $\{High\ (H), Medium\ (M), Low\ (L)\}$, the road importance $\mathcal{RI}$ can be mapped into a fuzzy set $\{Large\ (L), Medium\ (M), Small\ (S)\}$ and the average congestion time in a day $\mathcal{CT}$ can be also converted to a fuzzy set $\{Long\ (L), Medium\ (M), Short\ (S)\}$. Specifically, in this paper, (α, β, γ) for $\mathcal{CD}^{\mathcal{H}}$, $\mathcal{CD}^{\mathcal{M}}$ and $\mathcal{CD}^{\mathcal{S}}$ are set as $\{(-0.5, 0, 0.5), (0, 0.5, 1), (0.5, 1, 1.5)\}$, (α, β, γ) for $\mathcal{RI}^{\mathcal{L}}$, $\mathcal{RI}^{\mathcal{M}}$ and $\mathcal{RI}^{\mathcal{S}}$ are set as $\{(-0.4, 0, 0.4), (0, 0.4, 0.8), (0.4, 0.8, 1.2)\}$ and (α, β, γ) for $\mathcal{CT}^{\mathcal{L}}$, $\mathcal{CT}^{\mathcal{M}}$ and $\mathcal{CT}^{\mathcal{S}}$ are set as $\{(-0.5, 0, 0.5), (0, 0.5, 1), (0.5, 1, 1.5)\}$, respectively.

$$\mu(x) = \begin{cases} \dfrac{x-\alpha}{\beta-\alpha}, & \alpha < x < \beta \\ \dfrac{\gamma-x}{\gamma-\beta}, & \beta < x < \gamma \\ 0, & otherwise. \end{cases} \tag{1}$$

2) Fuzzy Interference: According to IF-THEN rules, the fuzzy interference can compute the linguistic outputs for congestion weights for intersections and road segments. The design of IF-THEN rules is often based on the experience of experts [40], [41] or generated from numerical and nominal data [42], [43]. In this paper, the knowledge-based rules are obtained according to the understanding of characteristics for congestion at intersections and on road segments. As shown in Table I, the linguistic outputs of congestion weights for intersections and road segments are defined as a fuzzy set $\{Very\ High\ (VH),\ High\ (H),\ Medium\ (M),\ Low\ (L),\ Very\ Low\ (VL)\}$. For example, if fuzzy inputs of the average congestion degree ($\mathcal{CD}$) is High, road importance ($\mathcal{RI}$) is Large and the average congestion time in a day ($\mathcal{CT}$) is Long, then the fuzzy congestion weight ($\mathcal{W}$) of the intersection or road segment is Very High.

3) Defuzzication: The defuzzification can be utilized to convert the fuzzy outputs into normalized congestion weights of intersections or road segments according to a centroid of gravity method [44], as represented in (2). In this case, congestion weights of all k intersections and road segments in a traffic network can be calculated as $\overrightarrow{W} = [W_1, W_2, \cdots, W_k]$.

$$W = \frac{\int \mu(x) \times x dx}{\int \mu(x) dx}, \tag{2}$$

where $\mu(x)$ is represented in (1) and (α, β, γ) for $\mathcal{W}^{\mathcal{VH}}$, $\mathcal{W}^{\mathcal{H}}$, $\mathcal{W}^{\mathcal{M}}$, $\mathcal{W}^{\mathcal{L}}$ and $\mathcal{W}^{\mathcal{VL}}$ are set as $\{(-0.25, 0, 0.25), (0, 0.25, 0.5), (0.25, 0.5, 0.75), (0.5, 0.75, 1), (0.75, 1, 1.25)\}$, respectively.

On the other hand, in this paper, we quantify the congestion propagation costs for each intersection or road segment based on the normalized congestion weights and congestion propagation probabilities among intersections and road segments. Specifically, let $R_t = 1$ be the event that congestion occurs on road segment R at time t and $I_{t+\tau} = 1$ be the event that intersection I is congested at time $t+\tau$. Then, the congestion propagation probability between road segment R and intersection I can be calculated by using a conditional probability $P_{RI} = P(I_{t+\tau} = 1 | R_t = 1)$ and τ should be fulfilled the condition for the temporal threshold in Definition 3. In this case, the congestion propagation cost $C_{RI}^{\mathcal{P}}$ from road segment R to intersection I can be written as

$$C_{RI}^{\mathcal{P}} = P_{RI} * W_I, \tag{3}$$

where W_I is the congestion weight of intersection I. When there are N intersections located upstream of road segment R, the total congestion propagation cost $C_R^{\mathcal{P}}$ of road segment R can be presented as

$$C_R^{\mathcal{P}} = \sum_{i=1}^{N} P_{RI_i} * W_{I_i}, \tag{4}$$

where P_{RI_i} is the congestion propagation probability from road segment R to its i^{th} upstream intersection and W_{I_i} is the congestion weight of intersection I_i. Therefore, according to Definition 5, the total congestion cost C_R of road segment R is

$$C_R = C_R^{\mathcal{P}} + W_R. \tag{5}$$

To illustrate the calculation of congestion cost, we take the causal congestion tree in Fig. 5 as an example, where there are six intersections or road segments with congestion weights $[W_A, W_B, W_C, W_D, W_E, W_F]$ and six directed edges with congestion propagation probabilities $[P_{AB}, P_{AC}, P_{BD}, P_{CE}, P_{DF}, P_{EF}]$. As shown in Fig. 5(a), vertex A is the root of the causal congestion tree and congestion from vertex A propagates to intersections and road segments B, C, D, E and F gradually along the directed tree. In order to calculate the total congestion costs of all intersections and road segments in the directed tree, we first find the vertexes in a tree whose outdegrees are 0, such as vertex F in Fig. 5(a). The total congestion cost of vertex F, C_F can be

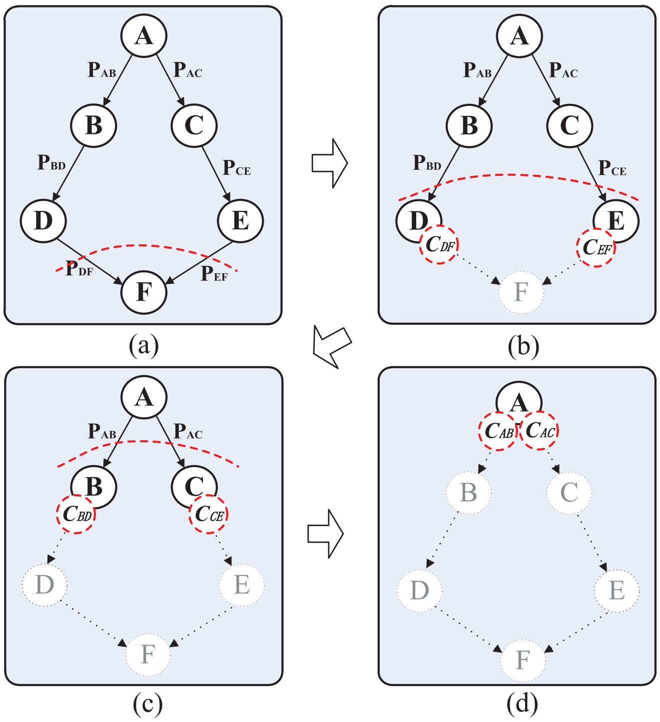

Fig. 5. An illustration of congestion cost calculation for each intersection or road segment in a CCT.

obtained directly and is equal to its congestion weight W_F. Then, we can calculate the congestion propagation costs $C^{\mathcal{P}}_{DF}$ and $C^{\mathcal{P}}_{EF}$ from vertex D to vertex F and from vertex E to vertex F, respectively by using the congestion weight W_F and congestion propagation probabilities P_{DF} and P_{EF}. According to (3), the congestion propagation costs $C^{\mathcal{P}}_{DF}$ and $C^{\mathcal{P}}_{EF}$ are $C^{\mathcal{P}}_{DF} = P_{DF} * W_F$ and $C^{\mathcal{P}}_{EF} = P_{EF} * W_F$, respectively. Thus the total congestion costs of vertexes D and E can be represented as $C_D = C^{\mathcal{P}}_{DF} + W_D$ and $C_E = C^{\mathcal{P}}_{EF} + W_E$. After that, we delete vertex F and the edges connected to vertex F in Fig. 5(a) and obtain the Fig. 5(b). In this case, we can recursively calculate the congestion costs for vertexes B and C, $C_B = P_{BD} * C_D + W_B$ and $C_C = P_{CE} * C_E + W_C$ in this causal congestion tree and eventually obtain the total congestion cost for vertex A, $C_A = P_{AB} * C_B + P_{AC} * C_C + W_A$.

Similar operations can be made to calculate congestion costs for each intersection and road segment in a traffic network. For the intersections or road segments with higher congestion costs, they are more likely to be the root causes for congestion in urban areas. Traffic control strategies, such as widening road infrastructures or optimizing signal control, can be further applied to mitigate congestion in the traffic network.

D. Markov Model

According to the proposed method above, we can get a number of intersections or road segments with higher congestion costs. However, due to the congestion correlation among neighboring intersections and road segments, it is difficult to identify the most critical congestion cause in a road network if congestion costs of these neighboring road infrastructures and successive intersections are comparatively higher. Therefore, in this subsection, we divide these neighboring road segments and intersections into several groups according to their locations in urban areas and propose a critical congestion cause identification method by using the Markov model to identify the most critical root causes for congestion in each group. Specifically, the method consists of three steps: 1) calculate the transition probability and get the Markov chain; 2) fuse the states whose congestion state of a certain intersection or road segment is 1; 3) calculate the transition probabilities from the other states to the fused state.

Specifically, we set the congestion state as 1 if there is congestion at an intersection or on a road segment. Otherwise, the state is 0. Then, according to the inductive loop detector data and congestion propagation probability in subsection (III-C), we can calculate the transition probabilities between two states and the corresponding Markov chain can also be obtained.

Let the state space as set S and let $X = \{X_n, n \in \mathbb{N}\}$ on a state space S as a discrete-time Markov chain. For each state $j \in S$, we denote by N_j the total number of visits to state j except the initial state, that is:

$$N_j = \sum_{n=t_0}^{t_0+T} 1_{\{X_n=j\}}, \tag{6}$$

where t_0 is the first congestion time for state j and T should be fulfilled the condition for the temporal threshold in Definition 3.

Let the random variable $\tau(j)$ that counts the number of transitions necessary to reach state j be:

$$\tau(j) = inf\{n \geq 1 \mid X_n = j\}. \tag{7}$$

For all $i, j \in S$, we calculate the congestion propagation probability from state i to state j during time t_0 and $t_0 + T$ and that is:

$$\begin{aligned}\mathbb{P}\{N_j > 0 \mid X_0 = i\} &= \sum_{n=t_0}^{t_0+T} \mathbb{P}\{N_j > 0, \tau(j) = n | X_0 = i\}\\ &= \sum_{n=t_0}^{t_0+T} \mathbb{P}\{\tau(j) = n | X_0 = i\}\\ &= \sum_{n=t_0}^{t_0+T} f_{ij}^{(n)},\end{aligned} \tag{8}$$

where $f_{ij}^{(n)}$ is the probability, starting from i, that the first visit to state j occurs at time n.

$$f_{ij}^{(n)} = \begin{cases} P_{ij}, & if\ n = 1 \\ \sum_{l \in S \setminus \{j\}} P_{il} f_{lj}^{(n-1)}. & if\ n \geq 2 \end{cases} \tag{9}$$

P_{ij} is the transition probability from state i to state j in Markov chain, which is

$$P_{ij} = \mathbb{P}\{X_1 = j \mid X_0 = i\}. \tag{10}$$

Then, if we need to get which road infrastructures are more likely to lead to congestion on a certain road infrastructure,

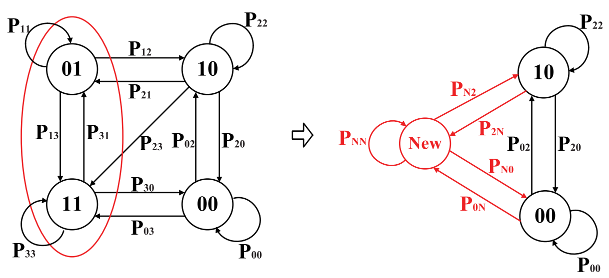

Fig. 6. Markov chain fusion.

we can fuse the states where the congestion state of the certain road infrastructure is 1 and construct a new Markov chain according to their transition probabilities and stationary probabilities. Using the new constructed Markov chain, we can analyze congestion state transition probabilities and obtain the causal relationships among congestion on different road infrastructures in a group. As shown in Fig. 6, there are two road infrastructures (road infrastructures 1 and 2) and four congestion states (00, 01, 10, 11) in the Markov chain with the stationary probabilities (P_0, P_1, P_2, P_3). If we need to get which road infrastructures are more likely to lead to congestion on road infrastructure 2, we fuse the states where the congestion state of road infrastructure 2 is 1 (state 01, 11). Then, using Algorithm 2, we can calculate the state transition probabilities of the new Markov chain. The transition probabilities from other states to the new state are $p_{2N} = p_{21} + p_{23}$ and $p_{0N} = p_{03}$ (line 16). The transition probabilities from the new state to other states are $p_{N0} = \frac{p_{30} * P_3}{P_1 + P_3}$ and $p_{N2} = \frac{p_{12} * P_1}{P_1 + P_3}$ (line 17). The transition probability from the new state to itself is $p_{NN} = \frac{p_{13} * P_1}{P_1 + P_3} + \frac{p_{31} * P_3}{P_1 + P_3} + \frac{p_{11} * P_1}{P_1 + P_3} + \frac{p_{33} * P_3}{P_1 + P_3}$ (line 23). Thus we can obtain a new Markov chain and analyze the most correlated road infrastructure that causes congestion on road infrastructure 2. Similarly, we can also analyze the most correlated road infrastructure that causes congestion on road infrastructure 1. After analyzing the congestion propagation relationships between all road infrastructures on each group, we can identify the most critical road infrastructures for urban congestion and further determine the root causes for congestion in a traffic network.

IV. Congestion Root Cause Prediction

In this section, as shown in Fig. 1, we utilize gradient boosting decision trees to predict the average travel speed of road networks according to a number of influential factors (*e.g.*, traffic flows, road topology, number of lanes and signal control strategies). Then, we further prioritize the relative importance of these influential factors on congestion root cause prediction in road networks.

A. Gradient Boosting Decision Trees

Let y be the average travel speed of a road network which can be utilized to indicate traffic conditions in the whole road network and $f(x)$ be an approximation function of y according to a set of predictor variables x. The squared error

Algorithm 2 Markov State Fusion

1: **Input:** X is a state space set; $P_X(i, j)$ is the transition probability from state i to state j in X; Z is a stationary probability set; $P_Z(i)$ is the stationary probability of the $i-th$ state; $\| \bullet \|$ is the length of a set.
2: **Output:** A Markov chain after fusing some states.
3: **for** $i = 1$ to Number of road infrastructures in a group **do**
4:   **for** $j = 1$ to $\| X \|$ **do**
5:     $Y \leftarrow$ an empty set;
6:     **if** the $i-th$ road infrastructure is 1 in state j **then**
7:       Put state j in set Y.
8:     **end if**
9:   **end for**
10:   **for** $\lambda = 1$ to $\| Y \|$ **do**
11:     $sumprobability = sumprobability + P_Z[Y(\lambda)]$.
12:   **end for**
13:   **for** $\mu = 1$ to $\| X \|$ **do**
14:     **if** $X(\mu) \notin Y$ **then**
15:       **for** $\eta = 1$ to $\| Y \|$ **do**
16:         $P_X[X(\mu), newstate] = P_X[X(\mu), newstate] + P_X[X(\mu), Y(\eta)]$.
17:         $P_X[newstate, X(\mu)] = P_X[newstate, X(\mu)] + P_X[Y(\eta), X(\mu)] * P_Z[Y(\eta)]/sumprobability$.
18:       **end for**
19:     **end if**
20:   **end for**
21:   **for** $\gamma = 1$ to $\| Y \|$ **do**
22:     **for** $\rho = 1$ to $\| Y \|$ **do**
23:       $P_X[newstate, newstate] = P_X[newstate, newstate] + P_X[Y(\gamma), Y(\rho)] * P_Z[Y(\gamma)]/sumprobability$.
24:     **end for**
25:   **end for**
26: **end for**

function (11) is utilized as the loss function to measure the approximation error [45].

$$L(y, f(x)) = [y - f(x)]^2 . \tag{11}$$

Let U be the total number of training data and the GBDT model can be established as follows:

Step 1: The initial approximation function can be calculated as

$$f_0(x) = \arg\min_c \sum_{i=1}^{U} L(y_i, c) . \tag{12}$$

Step 2: Constructing M $(m = 1, 2, \ldots, M)$ different regression trees iteratively.

1) Let r_{im} be the residual error of the i^{th} training data of the m^{th} regression tree and calculate r_{im} for all the training data according to (13).

$$r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)}, \quad i = 1, 2, \ldots, N. \tag{13}$$

2) Let (x_i, r_{im}) $(i = 1, 2, \ldots, U)$ be a set of new training data and construct the m^{th} regression tree. Assuming the

number of splits for each regression tree is J and each regression tree can partition the input space into J disjoint regions $R_{1m}, R_{2m}, \ldots, R_{Jm}$.

3) Calculating the optimal fitting value Υ_{jm} for each region R_{jm} by using (14).

$$\Upsilon_{jm} = \arg\min_{\Upsilon} \sum_{x_i \in R_{jm}}^{J} L\left(y_i, f_{m-1}(x_i) + \Upsilon\right). \tag{14}$$

4) Updating the GBDT model.

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J} \Upsilon_{jm} I\left(x \in R_{jm}\right), \tag{15}$$

where $I\left(x \in R_{jm}\right) = 1,\ if\ x \in R_{jm}$ and $I\left(x \in R_{jm}\right) = 0,\ if\ x \notin R_{jm}$.

Step 3: Outputting the final GBDT model, as shown in (16).

$$f(x) = f_0(x) + \sum_{m=1}^{M} \sum_{j=1}^{J} \Upsilon_{jm} I\left(x \in R_{jm}\right). \tag{16}$$

To prevent over-fitting and improve the prediction accuracy, a shrinkage strategy is utilized to shrink the impact of each regression tree by introducing a learning rate lr $(0 < lr < 1)$. In this case, the GBDT model can be written as

$$f(x) = f_0(x) + lr * \sum_{m=1}^{M} \sum_{j=1}^{J} \Upsilon_{jm} I\left(x \in R_{jm}\right). \tag{17}$$

B. Importance of Influential Factors

Owing to the advantages of GBDT in prioritizing the influence of each factor on the root causes of road congestion [46], we utilize the GBDT model to predict and rank the importance of influential factors, which can provide an intuitive identification of root causes for road congestion.

For the $m-th$ decision tree T_m in GBDT, we can calculate the relative importance $I_k^2(T_m)$ of the influential factor x_k in T_m by using (18).

$$I_k^2(T_m) = \sum_{t=1}^{J-1} \tau_t^2 I\left(v(t) = x_k\right), \tag{18}$$

where $I\left(v(t) = x_k\right)$ is the indicator function that x_k is the splitting influential factor associated with non-terminal node t and τ_t^2 is the squared error improvement that using x_k as a splitting factor of the non-terminal node t. For a set of decision trees $T_1, T_2, \ldots, T_M$, the relative importance of influential factor x_k can be represented as

$$I_k^2 = \frac{1}{M} \sum_{m=1}^{M} I_k^2(T_m). \tag{19}$$

V. Performance Evaluation

In this section, we conduct simulations based on a road network of the City of Sioux Falls, South Dakota, USA by using a traffic simulator SUMO to validate the effectiveness of our proposed root cause identification method.

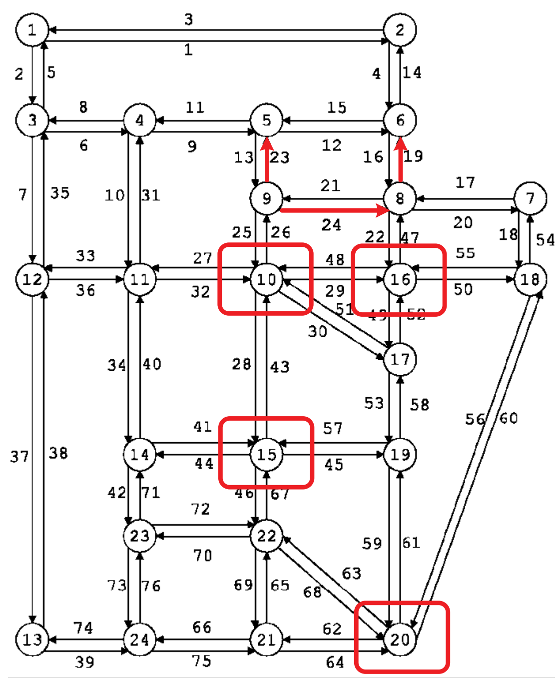

Fig. 7. Sioux falls network.

A. Sioux Falls Network

As a popular benchmark for transport researches [1], [47], Sioux Falls network is utilized as a test scenario for our proposed root cause identification method. As illustrated in Fig. 7, there are 24 intersections and 76 directed edges in the network. The lengths of all road segments are set equal to the Euclidian distances between the respective two intersections in the real road network and each road segment is set with 2 lanes. A fixed-time signal control strategy is used at each intersection.

B. Root Cause Identification Based on Our Proposed Method

In this subsection, we first evaluate the performance of the proposed method under different thresholds for congestion on intersections to identify the optimal threshold for root cause identification. Then, using the threshold, we identify the root causes of congestion in the road network by using the proposed approach under different traffic flows. Next, we increase the number of each road segment or utilize traffic responsive signal control[3] at each intersection and compare the improvement of average traffic speed in a road network before and after each change to validate the proposed root cause identification approach. Moreover, we also utilize the Markov model to locate the most critical congestion root causes among a number of adjacent intersections and road segments. Finally, we compare our proposed root cause identification method with existing methods under different traffic flows.

1) Thresholds for Congestion on Intersections: To evaluate the performance of root cause identification under different

[3]The responsive signal control strategy used in this paper can automatically respond to the prevailing traffic conditions and adjusts the green-time proportion according to the queue lengths at each intersection [7], [21].

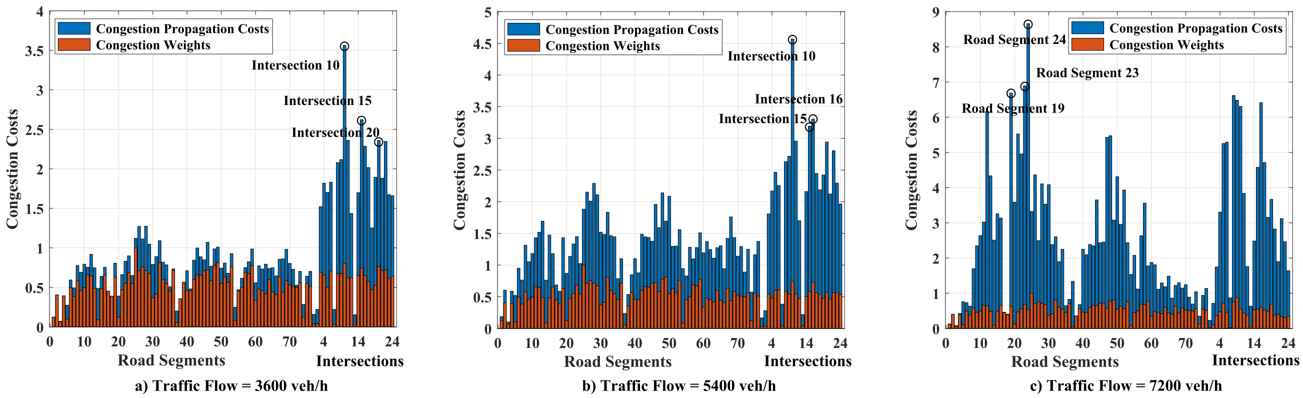

Fig. 8. Congestion costs of road segments and intersections in Sioux Falls network under traffic flow 3600 veh/h, 5400 veh/h and 7200 veh/h.

TABLE II

TRAVEL SPEED IMPROVEMENT UNDER DIFFERENT CONGESTION THRESHOLDS ON INTERSECTIONS

Percentage (%)	110	120	130	140	150	160	170	180	190
Speed Improvement (%)	15.2	15.2	15.2	15.2	15.5	15.5	**17.1**	17.1	17.1

thresholds for congestion on intersections, we identify root causes of congestion under different thresholds by using our proposed approach, which varies from 110% to 190% of their average waiting queue lengths. Then, we investigate travel speed improvement of the road network under different congestion thresholds on intersections by comparing the average travel speed before and after increasing the number of lanes on identified road segments or improving signal control strategy (from fixed-time signal control to responsive signal control) at identified intersections to determine the optimal congestion threshold on intersections for root cause identification. As illustrated in Table II, when the threshold is set over 170%, our proposed root cause identification method can achieve the highest improvement of the average travel speed in Sioux Falls network. Thus, in this paper, we determine 170% of the average waiting queue length as the metric to identify congestion on intersections.

2) Congestion Root Cause Verification: Using the determined threshold, we identify root causes of congestion in Sioux Falls network using the proposed method under traffic flow 3600 veh/h, 5400 veh/h and 7200 veh/h, respectively and verify the simulation results by comparing the average travel speed improvement before and after increasing the number of lanes on road segments or improving signal control strategy at intersections.

As shown in Fig. 8, the horizontal axis indicates each road segment and intersection in Sioux Falls network and the vertical axis demonstrates the calculated congestion costs of all road segments and intersections by using our proposed approach under different traffic flow in the road network. The red bars indicate congestion weights of road segments or intersections themselves, the blue bars indicate congestion propagation costs of road segments or intersections that lead to congestion on their neighboring intersections or road segments. Following Definition 5, the sum of congestion weights and congestion propagation costs indicate that congestion costs of road segments or intersections in the road network, which can demonstrate the importance of all road segments and intersections to congestion in the whole road network. Specifically, when traffic flow in the road network is 3600 veh/h, intersections 10, 15 and 20 are identified as root causes of congestion, when traffic flow in the road network is 5400 veh/h, intersections 10, 15 and 16 are identified as root causes of congestion and when traffic flow in the road network is 7200 veh/h, road segments 19, 23 and 24 are more likely to be root causes of congestion in Sioux Falls network, as shown in Fig. 7. We can see that congestion costs of road segments and intersections become larger with the increases of traffic flows and congestion costs of road segments grow faster than that of intersections. These results demonstrate that when the traffic flow is relatively small and less than the capacity of the road network, congestion in the road network mainly occurs at intersections and a suitable signal control strategy will improve traffic conditions of the road network. When the traffic flow increases, root causes of congestion in the road network transfers from intersections to road segments gradually and it is more effective for congestion mitigation to improve road infrastructures.

An intuitive verification for congestion root causes is that traffic improvement at identified root causes can bring the most significant network-wide traffic condition improvement. Therefore, in this paper, to validate the effectiveness of our proposed root cause identification method, we increase the number of lanes from 2 to 3 on each road segment using SUMO, replace the original fixed-time signal control strategy with a responsive signal control strategy [7], [21] at each intersection respectively, and further compare the percentage of travel speed improvement in the road network to illustrate the effects of each road segment or intersection on congestion in the whole road network. As illustrated in Fig. 9, when traffic flow is 3600 veh/h, the average travel speed of the road network improves 4%, 3.8% and 3.3% respectively after improving the signal control strategy at intersections 10, 15 and 20 respectively. When traffic flow is 5400 veh/h,

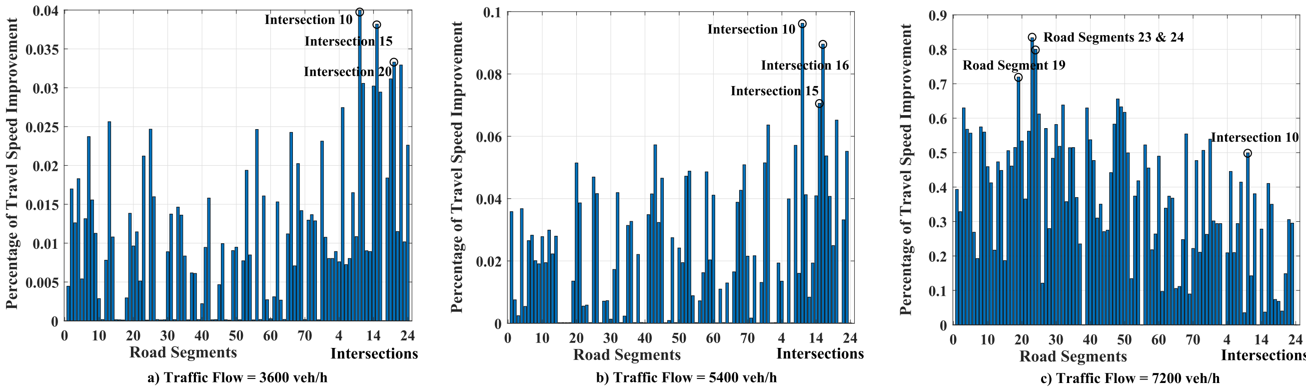

Fig. 9. Percentage of travel speed improvement in Sioux Falls network after increasing the number of lanes on each road segment or improving signal control strategy at each intersection under traffic flow 3600 veh/h, 5400 veh/h and 7200 veh/h.

traffic enhancement at intersections 10, 15 and 16 indicate the most travel speed improvement, which are 9.5%, 7.1% and 8.8% respectively. When traffic flow of the road network reaches 7200 veh/h, there are significant improvements of the average travel speed in the road network after increasing the number of lanes on road segments 19, 23 and 24, which are 71%, 83% and 80% respectively. Comparing with the calculated congestion cost results in Fig. 8, these simulation results based on SUMO demonstrate the effectiveness of our proposed method in identifying and locating the root causes of congestion in a road network. Furthermore, when traffic flow is 7200 veh/h, we also compare simulation results (in the road network with signal control at intersections) with the results in [1] (in the road network without signal control at intersections). We can see that road segments 28 and 51 are considered as root causes of congestion in the road network without signal control [1] and after differentiating congestion effects on road segments and intersections, road segments 19, 23 and 24 are more likely to be the root causes in the road network with signal control at intersections. This result suggests that if there is no signal control at each intersection, congestion tends to occur at upstream and downstream of an intersection with higher degree centrality, such as intersection 10 in Fig. 7. When signal control strategy is introduced at each intersection and traffic congestion can be better relieved at intersections, congestion is more likely to occur on road segments with lower road capacities, such as road segments 19, 23 and 24.

3) Markov Transition Probability: To quantify the relationship among congestion on multiple identified root causes, we utilize the Markov model to obtain the most critical congestion root cause for two groups of identified root causes {intersections 10, 15, 16, 20} under traffic flow 3600 veh/h and {road segments 19, 23 and 24, intersections 5 and 6} under traffic flow 7200 veh/h, respectively. Thus, according to our proposed method in subsection III-D, there should be 16 states in the first group and 32 states in the second group, respectively. In order to decrease the uncertainty and occasionality of state transitions, we remove a number of states with few occurrence times. Then, according to Algorithm 2, we can calculate the transition probabilities from the other states to the fused state.

TABLE III

FUSED STATE TRANSITION PROBABILITIES OF INTERSECTIONS 10, 15, 16 AND 20

Intersection 10		Intersection 15		Intersection 16		Intersection 20	
State	P	State	P	State	P	State	P
0000	0.064	0000	0.104	0000	0.299	0000	0.385
0001	0.066	0001	0.109	0001	0.208	0010	0.353
0010	0.065	0010	0.107	0100	0.477	0100	0.484
0011	0.066	0011	0.108	-	-	1010	0.496
0100	0.066	1010	0.108	-	-	-	-

Specifically, after removing the states with few occurrence times, there are 6 states of the first root cause group {intersections 10, 15, 16, 20}, which are 0000, 0001, 0010, 0011. 0100 and 1010. If we need to obtain the most relevant state to intersection 20, we should fuse states 0001 and 0011, and calculate transition probabilities from the other 4 states to the fused state. In this case, we can obtain the transition probabilities from the other states to the fused state, which are shown in Table III. We can see that the most critical state that causes the congestion on intersection 16 is state 0100 and the most critical state that causes the congestion on intersection 20 are states 0100 and 1010. Thus intersection 10 and 15 are more likely to lead to congestion on intersection 16 and 20. However, the transition probabilities from other states to intersection 10 is less than that from other states to road 15, which illustrates that congestion is more likely to propagate to other roads from intersection 10 than that from intersection 15. In this case, intersection 10 are more likely to be the most critical root cause of congestion in the road network under traffic flow 3600 veh/h.

Similarly, as shown in Table IV, there are 9 states for the root cause group {road segments 19, 23 and 24, intersections 5 and 6}, which are 00000, 00001, 00100, 00101, 01000, 01001, 10000, 10001 and 10100. We can see that there is not any states indicating the congestion at intersection 5 after removing the states with few occurrence times, thus congestion at intersection 5 is not likely to influence congestion on road segments 19, 23, 24 and intersection 6 and we can only consider the relationships among congestion on the other 3 road segments and 1 intersection. As we can see that, the most critical state that causes congestion on road segment 19 is

TABLE IV
FUSED STATE TRANSITION PROBABILITIES OF ROAD SEGMENTS 19, 23 AND 24, INTERSECTIONS 5 AND 6

Road Segment 19		Road Segment 23		Road Segment 24		Intersection 6	
State	P	State	P	State	P	State	P
00000	0.323	00000	0.121	00000	0.403	00000	0.019
00001	0.249	00001	0.116	00001	0.396	00010	0.009
00010	0.357	00010	0.121	01000	0.496	01000	0.123
00011	0.238	00011	0.118	01001	0.509	10000	0.004
01000	0.477	10000	0.108	10000	0.348	10010	0.002
01001	0.184	10001	0.118	10001	0.441	-	-
-	-	10010	0.119	-	-	-	-

state 01000, the most critical state that causes congestion on road segment 23 are states 01000 and 01001 and the most critical state that causes congestion on intersection 6 is state 01000. Therefore, road segment 23 is more likely to cause congestion on road segments 19, 24 and intersection 6, and it is considered as the most critical root cause of congestion in the road network under traffic flow 7200 veh/h.

4) Comparison With Existing Methods: In this subsection, we compare our proposed method with congestion propagation based method [1], congestion level based method and spatial cross area based method [17] under different traffic flows in Sioux Falls network. Especially, take the road network under traffic flow 7200 veh/h as an example, as illustrated in Fig. 8, road segments 23 and 24 are considered as root causes of congestion by using our proposed method. In terms of the congestion level based method, which regards the road segments with higher congestion weights as root causes of congestion in the whole road network, as shown in Fig. 8, road segments 25 and 32 are considered as root causes. For the spatial cross area based method, the root causes of congestion are more likely to locate at the spatial cross area of two congestion propagation patterns, thus road segments 51 and 63 are seen as the root causes in the road network. Finally, for the congestion propagation based method in [1], which does not differentiate congestion effects of road segments and intersections, however, due to the major causes for congestion tend to lie in road infrastructure under higher traffic flows, road segments 23 and 24 can also be identified as root causes in the road network.

Then, in order to compare the effectiveness of these root cause identification methods above, we increase the number of lanes on each identified road segment or improving the signal control strategy at each identified intersection and compare the average travel speed in the entire road network under multiple traffic flows. As shown in Fig. 10, the horizontal axis demonstrates traffic flows of the road network and the vertical axis indicates the average travel speed of the network. We can see that when traffic flow of the road network is small relatively, improving traffic conditions at the identified congestion root causes by using the proposed method can achieve superior performance in travel speed improvement than the other methods. This is expected because as shown in Fig. 8(a), when the traffic flow is small relatively, the main root causes of congestion are more likely to lie in intersections. The existing methods can only locate the root causes at road

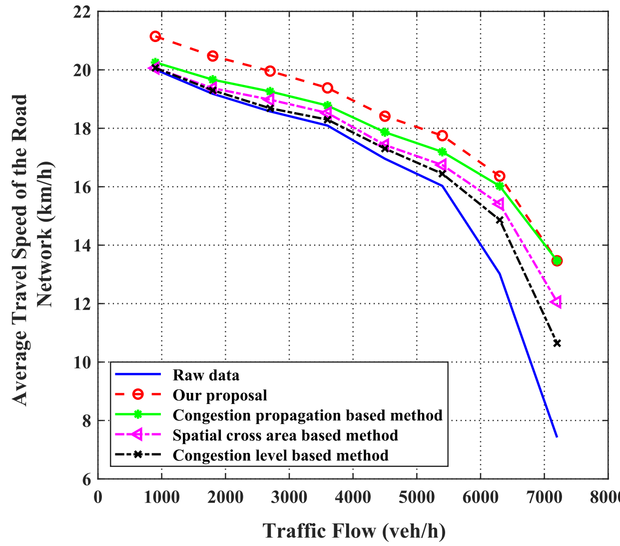

Fig. 10. Average travel speed of the road network after increasing the number of lanes or improving signal control strategy on identified root causes of congestion based on existing methods and our proposed method.

segments, while the proposed method differentiates congestion effects between road segments and intersections and further identify the major contributed intersections to congestion in the road network. This result can illustrate the superiority of our proposed method in classifying congestion effects between road segments and intersections, especially under low traffic flow circumstances. Moreover, with the increase of network traffic flow, improving the traffic conditions on root causes identified by our proposed method can lead to more travel speed improvement than improving the traffic conditions on root causes identified by the spatial cross area based method and congestion level based method. Especially, when traffic flow in the road network is 7200 veh/h, the average network travel speed can be improved by 81.5% when increasing the number of lanes on road segments identified by the proposed method, which are 62.5% and 43.5% by increasing the number of lanes on road segments identified according to the spatial cross area based method and congestion level based method, respectively. While compared with the congestion propagation based method, our proposed method can achieve better performance when there is a light traffic flow in the road network and the major causes of congestion in a road network lie in intersections. When traffic flow in the road network increases, the root causes of congestion transfer from intersections to road segments gradually and our proposed method obtains the similar identification results compared with the congestion propagation based method, especially under traffic flow 7200 veh/h (where road infrastructures are the dominant causes of congestion in a road network). In summary, these results suggest that our proposed method by differentiating congestion effects of road segments and intersections can provide a more effective way to identify and locate root causes of congestion in road networks.

TABLE V
PERFORMANCE OF GBDT MODEL WITH DIFFERENT COMBINATIONS OF LEARNING RATE AND MAXIMUM TREE DEPTH

Learning rate	R^2 score on test dataset			
	Maximum depth = 2	Maximum depth = 3	Maximum depth = 4	Maximum depth = 5
0.05	0.819	0.825	0.845	0.807
0.1	0.833	0.843	0.846	0.815
0.4	0.766	**0.864**	0.808	0.769
0.7	0.659	0.738	0.679	0.661
1	0.537	0.426	0.587	0.457

C. Influential Factors Prioritizing Based on GBDT Method

In this subsection, we present a GBDT based method to predict the average travel speed of a whole road network according to signal control strategies, traffic flows and road topology, and further prioritize these influential factors by calculating their relative importance to identify the root causes of congestion. Specifically, we initiate the number of lanes (2 or 3 lanes) for each road segment and signal control strategies (without signal control, fixed-time signal control and responsive signal control) at each intersection randomly with a certain traffic flow in Sioux Falls network based on SUMO, which is considered as a sample for this study. In this paper, we obtain 1000 samples under different traffic flows and 80% of sample data are used for training and the other 20% of samples are utilized as a test data set.

1) Optimization of Parameters: To determine the optimal learning rate lr and maximum tree depth in Section IV, we utilize the coefficient of determination R^2 as a measure [48] to evaluate the accuracy for average travel speed estimation of GBDT, which is defined as (20) and the score of R^2 closer to 1 indicates a higher accuracy of the model.

$$R^2 = \frac{\sum_i \left(\hat{y}_i - \bar{y}\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2}, \tag{20}$$

where y_i is the measured average travel speed of the road network for the $i-th$ sample, $\hat{y}_i$ is the estimation for y_i based on GBDT and $\bar{y}$ is the mean value of travel speeds for all samples.

To test the model performance under different combinations of parameters, a series of GBDT models are constructed with various learning rates (lr varies from 0.01 to 1 with a step of 0.01) and maximum tree depth (values from 1 to 50 with a step of 1) by fitting a maximum of 1000 estimators. In this study, the maximum number of estimator is specified as 1000, which can guarantee the convergence of GBDT models. Then, we test the performance of GBDT models based on different combinations of learning rate and maximum tree depth. As shown in Table V, we can see that when the learning rate is set as 0.4 and maximum tree depth is set to be 3, the GBDT model can achieve a highest R^2 score 0.864.

Furthermore, we also compare the proposed GBDT based method with other algorithms including Multilayer Perceptron (MLP) and random forest (RF). Specifically, the Multilayer Perceptron used in this paper has two hidden layers (the dimension of the first hidden layer is 100 and the second is 11) and we utilize the ReLU activation after each hidden layer. Stochastic gradient descent (SGD) is also utilized to train the network. For RF algorithm, the number of estimators and the maximum depth of subtrees are important for the accuracy of this model. In this paper, they are set to be 200 and 3 respectively after searching in parameter space. As illustrated in Table VI, we compare R^2 score by using MLP, RF and the proposed GBDT model under traffic flows 3600, 5400 and 7200 veh/h and hybrid traffic flows, respectively to validate the effectiveness of the proposed method. We can see that the proposed GBDT based method can achieve a better performance at different traffic flows compared with other algorithms, especially when there is a heavy traffic flow in a road network. Moreover, we also integrate all samples under different traffic flows together to verify the effectiveness of our proposed method considering the impact of different traffic flows on road conditions. As shown in Table VI, the proposed method can achieve a highest R^2 score 0.864 compared with other algorithms, which suggests that the proposed GBDT based method is more adaptive to different traffic demands in road networks and has the potential to be applied for various road networks to evaluate their traffic conditions and further locate root causes of congestion according to a number of easily-measured influential factors.

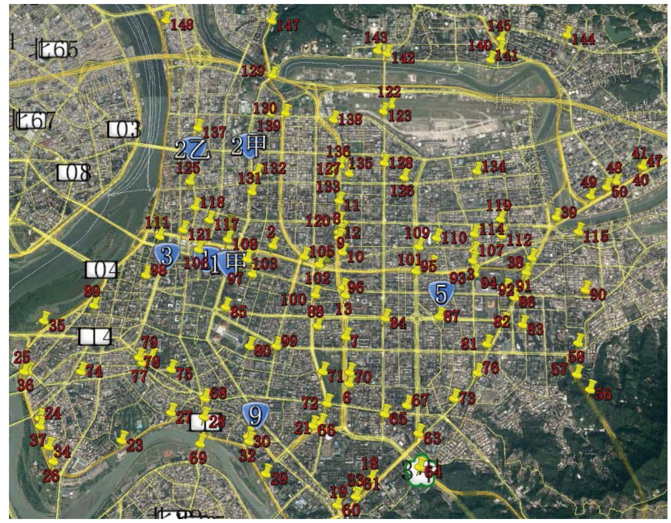

Fig. 11. Inductive loop detectors in the road network of Taipei.

TABLE VI
COMPARISON WITH OTHER ALGORITHMS

Traffic flow	Prediction performance (R^2 score)		
	MLP	RF	GBDT
3600 veh/h	0.786	0.733	0.942
5400 veh/h	0.639	0.696	0.855
7200 veh/h	0.455	0.358	0.807
Entire traffic flows	0.622	0.677	0.864

2) Relative Importance of Influential Factors: In order to explore the different influences of each estimation variables on road condition under different traffic flows and further predict root causes of congestion in traffic networks, the relative contributions of influential factors are calculated using the optimal model and a higher value of relative importance

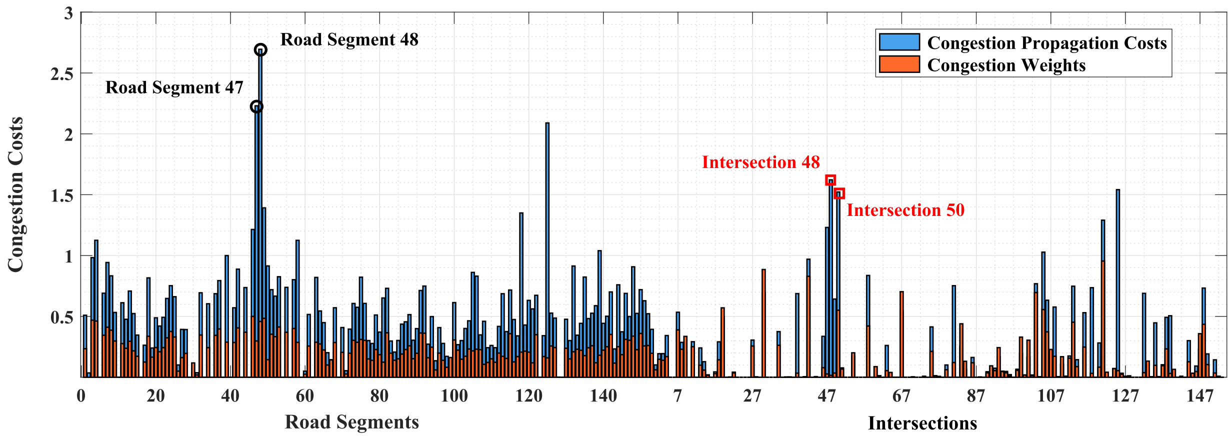

Fig. 12. Congestion costs of road segments and intersections in Taipei.

TABLE VII
RELATIVE IMPORTANCE OF DIFFERENT INFLUENTIAL FACTORS

Categories	Variables	Rank	Relative importance
Traffic flow	Hourly traffic volume	1	80.0%
Intersection	Signal control strategy	2	15.3%
Road segment	Number of lanes	3	4.7%

indicates a stronger impact of these influential factors on traffic condition in the entire road network. Specifically, as shown in Table VII, traffic flow contributes most in traffic conditions with relative importance of 80.0%, which is consistent with simulation results in Subsection V-B. Moreover, signal control strategies at intersections contribute a total relative importance of 15.3% and among that, intersection 10 is the most contributed intersection for traffic conditions in the road network, which is also verified according to the simulation results in Subsection V-B. On the other hand, we can that the number of lanes on road segments only contributes 4.7% to traffic conditions, which indicates that the impact of number of lanes on road segments is less sensitive for traffic conditions in road networks. This may be due to the fact that there are a variety of factors on road segments that contribute traffic conditions in a road network, such as number of lanes, road segment lengths, centrality and betweenness, thus considering the contribution of number of lanes on road segments only has a relatively limited impact on traffic condition in the whole road network.

D. Experiment Results

Experiments are further carried out using the inductive loop detector data in Taipei, Taiwan to identify root causes of the road network. As shown in Fig. 11, there are 153 loop detectors in the urban road network of Taipei. The average travel speed and road occupancy data are sampled over 1-minute intervals from 1 April, 2013 to 30 April, 2013. Because the spatio-temporal correlations among traffic flows vary with time of the day and traffic conditions in peak hours can significantly reflect the congestion level of a road network [49]. In this paper, we choose the data from 7:00 to 10:00 and from 18:00 to 21:00 to analyze the root cause of congestion in the road network of Taipei.

Specifically, we utilize the average travel speed as an indication of the real-time traffic condition on each road segment. Moreover, because inductive loop detectors are located at the interaction of each link and the real-time road occupancy is proportional to the queue length at each intersection [50], in this paper, the change rate of road occupancy is used to approximate the change rate of average waiting queue length. In this case, using the causal congestion tree constructing algorithm and congestion cost calculation, we can obtain the total congestion costs of all road segments and intersections. As illustrated in Fig. 12, the horizontal axis describes each road segment and intersection in the road network and the vertical axis indicates congestion propagation costs and congestion weights of road segments and intersections, respectively. It can be seen that congestion costs of road segments are generally higher than the costs of congestion at intersections. According to the simulation results in Fig. 8 and Fig. 9, this result suggests that the traffic load is relatively high in urban road networks of Taipei. Moreover, congestion costs of road segments 47 and 48, intersections 48 and 50 are higher than most of the other road segments and intersections, which are more likely to be the root causes of congestion in the road networks. In addition, we also map the road segments and intersections using different color labels onto the road network of Taipei according to their congestion costs. As shown in Fig. 13, road segments and intersections with congestion costs from 1 to 1.5 are marked by red labels; road segments and intersections with congestion costs from 0.5 to 1 are marked by green labels; road segments and intersections are marked by white labels when their congestion costs are less than 0.5. Specifically, road segments whose congestion costs are larger than 1.5 are marked by "vehicle" labels, such as road segments 47, 48 and 125. Intersections whose congestion costs are larger than 1.5 are marked by "warning" labels, such as intersections 48, 50 and 125.

Furthermore, as illustrated in Fig. 13, road segments 47 and 48, intersections 48 and 50 are located in the northeast of Taipei road networks. To identify the root cause

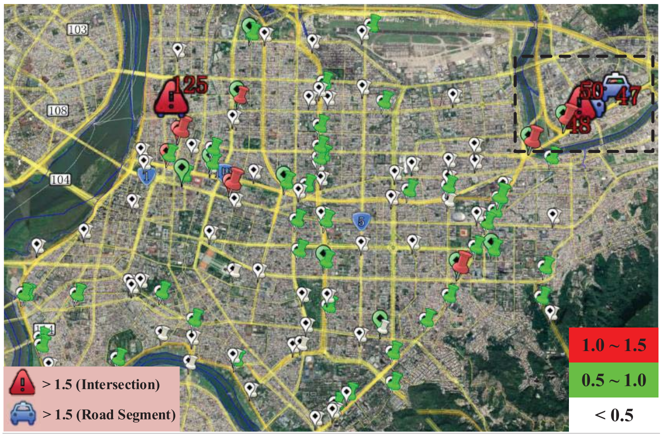

Fig. 13. An illustration of congestion costs for road segments and intersections in the road networks of Taipei.

TABLE VIII

FUSED STATE TRANSITION PROBABILITIES OF ROAD SEGMENTS 47 AND 48, INTERSECTIONS 48 AND 50

Road Segment 47		Road Segment 48		Intersection 48		Intersection 50	
State	P	State	P	State	P	State	P
0000	0.262	0000	0.109	0000	0.269	0000	0.289
0001	0.281	0001	0.105	0001	0.283	0010	0.249
0010	0.216	0010	0.141	0100	0.284	0100	0.336
0011	0.219	0011	0.146	1000	0.232	1000	0.307
0100	0.292	1000	0.139	-	-	-	-

of congestion among these road segments and intersections, we further utilize the Markov model to analyze the causal relationships of congestion between these road segments and intersections. As shown in Table VIII, the most critical state causing congestion on road segment 47 is state 0100, the most critical state that causes congestion on intersection 48 is state 0100 and the most critical state that causes congestion on intersection 50 is also state 0100. Therefore, road segment 48 is more likely to be the most critical root cause of congestion in the road network and increasing road capacity of the road segment is most effective to improve traffic efficiency of Taipei road networks effectively.

VI. CONCLUSION

In this paper, in order to identify root causes of congestion in urban traffic networks, we first differentiated the congestion effects between road segments and intersections and presented definitions about congestion at road segments and intersections and congestion correlation between them, respectively. Then, we constructed causal congestion trees to model congestion propagation and calculated congestion costs for all road segments and intersections in the whole road network to locate root causes of congestion. After that, a Markov model was developed to analyze congestion relationships among multiple road segments and intersections simultaneously and locate the most critical root cause in a road network. Simulations were conducted based on a road network of the City of Sioux Falls using SUMO. Simulation results suggested that when the traffic flow is relatively small in a road network, congestion in the road network mainly occurs at intersections and a suitable signal control strategy will improve traffic condition of the road network; when there is a heavy traffic flow, root causes of congestion transfers from intersections to road segments gradually and it is more effective for congestion mitigation to improve road infrastructures; and the proposed method can provide a more effective way to identify root causes of congestion in road networks compared to existing methods. Furthermore, a GBDT based method was presented to predict the root causes of congestion which provided an intuitive approach to predict the major causes of congestion in a whole road network according to a number of easily-measured congestion influential factors, such as signal control strategies, traffic flows and road topology and the simulation results indicated that the GBDT based method can achieve superior performance in root cause identification compared with existing approaches.

The proposed method in this paper can be utilized to identify major contributors of congestion and further predict the root causes of congestion for different road networks according to traffic flows, signal control strategies and road topology, which provides a specific guidance for road administrators to mitigate traffic congestion in urban road networks. This paper compared the average travel speed in a road network with the utilization of fixed-time and responsive signal control strategies respectively to demonstrate the importance of signal control at each intersection for traffic congestion mitigation. In the future, we would like to investigate suitable signal control strategy at each intersection to evaluate its network-wide traffic improvement.

REFERENCES

[1] C. Li, W. Yue, G. Mao, and Z. Xu, "Congestion propagation based bottleneck identification in urban road networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 4827–4841, May 2020.

[2] A. Ahmad, R. Arshad, S. A. Mahmud, G. M. Khan, and H. S. 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.

[3] D. Schrank, B. Eisele, T. Lomax, and J. Bak, "2015 urban mobility scorecard," Texas A&M Transp. Inst., College Station, TX, USA. Accessed: Sep. 2015.

[4] R. Claes, T. Holvoet, and D. Weyns, "A decentralized approach for anticipatory vehicle routing using delegate multiagent systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 364–373, Jun. 2011.

[5] J. Pan, I. S. Popa, K. Zeitouni, and C. Borcea, "Proactive vehicular traffic rerouting for lower travel time," *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 3551–3568, Oct. 2013.

[6] P. Hunt, D. Robertson, R. Bretherton, and R. Winton, "SCOOT: A traffic responsive method of coordinating signals," Transp. Road Res. Lab., Crowthorne, U.K., Tech. Rep. TRRL-LR-1014, 1981.

[7] P. Lowrie, "The Sydney coordinated adaptive traffic system," in *Proc. Int. Conf. Road Traffic Signal.*, 1982, pp. 67–70.

[8] *A6 Clapham Bypass*. Accessed: Jan. 2000. [Online]. Available: http://www.highways.gov.uk/roads/projects/6006.aspx

[9] K. Hymel, "Does traffic congestion reduce employment growth?" *J. Urban Econ.*, vol. 65, no. 2, pp. 127–135, Mar. 2009.

[10] *Public Expenditure Statistical Analyses*. Accessed: Jul. 2009. [Online]. Available: https://www.gov.uk/government/collections/public-expenditure-statistical-analyses-pesa

[11] D. Schrank, B. Eisele, and T. Lomax, "2012 urban mobility report," Texas A&M Transp. Inst., College Station, TX, USA, Tech. Rep., 2012.

[12] R. Sanchis-Guarner, "Driving up wages: The effects of road construction in Great Britain," SERC Discuss. Paper 120, 2012.

[13] *Road Infrastructure*. Accessed: Sep. 2008. [Online]. Available: http://www.lgam.info/road-infrastructure

[14] N. Mineta, "National strategy to reduce congestion on America's transportation network," U.S. Dept. Transp., Washington, DC, USA, Tech. Rep., 2006.

[15] D. Hale *et al.*, "Traffic bottlenecks: Identification and solutions," Federal Highway Admin. (FHWA), Washington, DC, USA, 2016.
[16] J. Gong and W. Yang, "The traffic bottleneck analysis on urban expressway under information condition," in *Proc. 2nd Int. Conf. Power Electron. Intell. Transp. Syst. (PEITS)*, Dec. 2009, pp. 400–403.
[17] W.-H. Lee, S.-S. Tseng, J.-L. Shieh, and H.-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.
[18] J. Ma, C. Li, Z. Liu, Y. Duan, Y. Lei, and L. Xiong, "On traffic bottleneck in green ITS navigation: An identification method," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–5.
[19] B. S. Kerner, *Introduction to Modern Traffic Flow Theory and Control: The Long Road to Three-Phase Traffic Theory*. Berlin, Germany: Springer, 2009, pp. 9–39.
[20] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," *Proc. IEEE*, vol. 91, no. 12, pp. 2043–2067, Dec. 2003.
[21] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "Store-and-forward based methods for the signal control problem in large-scale congested urban road networks," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 2, pp. 163–174, Apr. 2009.
[22] A. Kouvelas, K. Aboudolas, M. Papageorgiou, and E. B. 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.
[23] *Traffic Bottlenecks: Identification and Solutions*, US Dept. Transp., Washington, DC, USA, 2016.
[24] B. Anbaroglu, B. Heydecker, and T. Cheng, "Spatio-temporal clustering for non-recurrent traffic congestion detection on urban road networks," *Transp. Res. C, Emerg. Technol.*, vol. 48, pp. 47–65, Nov. 2014.
[25] J. Kwon, M. Mauch, and P. Varaiya, "Components of congestion: Delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1959, no. 1, pp. 84–91, Jan. 2006.
[26] A. H. F. 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, Dec. 2014.
[27] 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. CICTP*, Jun. 2014, pp. 1890–1901.
[28] 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.
[29] Y.-K. Ki and D.-K. Baik, "Model for accurate speed measurement using double-loop detectors," *IEEE Trans. Veh. Technol.*, vol. 55, no. 4, pp. 1094–1101, Jul. 2006.
[30] H. Sun, J. Wu, D. Ma, and J. Long, "Spatial distribution complexities of traffic congestion and bottlenecks in different network topologies," *Appl. Math. Model.*, vol. 38, no. 2, pp. 496–505, Jan. 2014.
[31] T. Chu, J. Wang, L. Codeca, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1086–1095, Mar. 2020.
[32] W. Liu, G. Qin, Y. He, and F. Jiang, "Distributed cooperative reinforcement learning-based traffic signal control that integrates V2X networks' dynamic clustering," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8667–8681, Oct. 2017.
[33] P. La and S. Bhatnagar, "Reinforcement learning with function approximation for traffic signal control," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 412–421, Jun. 2011.
[34] L. Wei and D. Hong-Ying, "Real-time road congestion detection based on image texture analysis," *Procedia Eng.*, vol. 137, pp. 196–201, Jan. 2016.
[35] 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.
[36] X. Luo, D. Ma, S. Jin, Y. Gong, and D. Wang, "Queue length estimation for signalized intersections using license plate recognition data," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 3, pp. 209–220, Jun. 2019.
[37] T. Cheng, J. Wang, J. Haworth, B. Heydecker, and A. Chow, "A dynamic spatial weight matrix and localized space-time autoregressive integrated moving average for network modeling," *Geographical Anal.*, vol. 46, no. 1, pp. 75–97, Jan. 2014.
[38] S. Wang, S. Djahel, Z. Zhang, and J. Mcmanis, "A multiagent system for mitigating unexpected urban traffic congestion," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2888–2899, Oct. 2016.
[39] V. V. Vazirani, *Approximation Algorithms*. Atlanta, GA, USA: Springer, 2013.
[40] Q. Luo, C. Li, T. H. Luan, and W. Shi, "EdgeVCD: Intelligent algorithm-inspired content distribution in vehicular edge computing network," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5562–5579, Jun. 2020.
[41] L. Altoaimy and I. Mahgoub, "Fuzzy logic based localization for vehicular ad hoc networks," in *Proc. IEEE Symp. Comput. Intell. Vehicles Transp. Syst. (CIVTS)*, Dec. 2014, pp. 121–128.
[42] L.-C. Dutu, G. Mauris, and P. Bolon, "A fast and accurate rule-base generation method for Mamdani fuzzy systems," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 715–733, Apr. 2018.
[43] X.-Z. Wang and C.-R. Dong, "Improving generalization of fuzzy IF-THEN rules by maximizing fuzzy entropy," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 3, pp. 556–567, Jun. 2009.
[44] G. E. M. Zhioua, N. Tabbane, H. Labiod, and S. Tabbane, "A fuzzy multi-metric QoS-balancing gateway selection algorithm in a clustered VANET to LTE advanced hybrid cellular network," *IEEE Trans. Veh. Technol.*, vol. 64, no. 2, pp. 804–817, Feb. 2015.
[45] J. H. Friedman, "Stochastic gradient boosting," *Comput. Statist. Data Anal.*, vol. 38, no. 4, pp. 367–378, Feb. 2002.
[46] X. Ma, C. Ding, S. Luan, Y. Wang, and Y. Wang, "Prioritizing influential factors for freeway incident clearance time prediction using the gradient boosting decision trees method," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2303–2310, Sep. 2017.
[47] H. Bar-Gera, F. Hellman, and M. Patriksson, "Computational precision of traffic equilibria sensitivities in automatic network design and road pricing," *Transp. Res. B, Methodol.*, vol. 57, pp. 485–500, Nov. 2013.
[48] N. J. D. Nagelkerke, "A note on a general definition of the coefficient of determination," *Biometrika*, vol. 78, no. 3, pp. 691–692, 1991.
[49] P. Duan, G. Mao, W. Liang, and D. Zhang, "A unified spatio-temporal model for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3212–3223, Sep. 2019.
[50] 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.

Wenwei Yue (Member, IEEE) received the B.E. and Ph.D. degrees in communication and information systems from Xidian University, Xi'an, China, in 2015 and 2020, respectively. From 2017 to 2019, he was a Visiting Ph.D. Student with the School of Computing and Communications, University of Technology Sydney. He is currently an Associate Professor with the State Key Laboratory of Integrated Services Networks, Xidian University. His research interests include intelligent transportation systems, vehicular networks, wireless sensor networks, and big data.

Changle Li (Senior Member, IEEE) received the Ph.D. degree in communication and information system from Xidian University, China, in 2005. He conducted his post-doctoral research in Canada and the National Institute of Information and Communications Technology, Japan, respectively. He had been a Visiting Scholar with the University of Technology Sydney. He 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 *ad hoc* networks, and wireless sensor networks.

Yue Chen received the B.E. degree in electronic and information engineering from Xidian University, Xi'an, China, in 2019, where he is currently pursuing the master's degree in communication and information systems. His current research interests include multi-agent reinforcement learning and its application in intelligence transportation systems.

Peibo Duan (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Northeastern University, Shenyang, China, in 2011, 2013, and 2017, respectively, and the Ph.D. degree from the University of Technology, Sydney (UTS), Australia, in 2020. He is currently working as an Associate Professor with the School of Software, Northeastern University. His current research interests include machine learning and intelligent transportation systems.

Guoqiang Mao (Fellow, IEEE) is currently a Distinguished Professor and the Dean of the Research Institute of Smart Transportation, Xidian University. Before that, he was with the University of Technology Sydney and the University of Sydney. He has published over 200 papers in international conferences and journals that have been cited more than 10,000 times. His research interests include intelligent transport systems, the Internet of Things, wireless localization techniques, wireless sensor networks, and applied graph theory and its applications in telecommunications. He is a fellow of IET. He received the "Top Editor" Award for outstanding contributions to the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY in 2011, 2014, and 2015. He was the Co-Chair of IEEE Intelligent Transport Systems Society Technical Committee on Communication Networks. He has served as a TPC member, the chair, and the co-chair for a number of international conferences. From 2010 to 2020, he was an Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. From 2014 to 2019, he was an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. Since 2018, he has been an Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.