Measurement 174 (2021) 109051

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

A probabilistic estimation of traffic congestion using Bayesian network

Tanzina Afrin *, Nita Yodo

Department of Industrial and Manufacturing Engineering, North Dakota State University, 1410 14th Avenue North, Fargo, ND 58102, USA

A R T I C L E I N F O

Keywords:
Traffic congestion
Bayesian network
Probabilistic estimation
Recurring
Nonrecurring

A B S T R A C T

For ensuring a robust traffic management system, monitoring traffic conditions promptly by estimating the congestion level is crucial. The current measures can only represent the variations of specific standard parameters and do not consider the probabilistic property. In this paper, a Bayesian Network (BN) based probabilistic congestion estimation approach is proposed. The proposed BN-based approach considers both speed and volume-related measures and provides a probabilistic estimation of the probable congestion state. For recurring and nonrecurring congestion, two different BN models were developed and implemented in realtime datasets. The case study results showed that the proposed BN models could quantify the probable congestion level in terms of a probability for each state in a variable, at the presence of different combinations of prior variables' state. Further, the proposed BN based approach can be employed in the decision-making process that involves the probabilistic estimation of traffic congestion with a vision of the realtime circumstances.

1. Introduction

In this era of continuous urbanization and population growth, traffic congestion has been increasing across the world day by day. Traffic congestion is worsening nowadays due to high population density, an increasing number of infrastructures, emerging technological advancements and growth in motor vehicles, and the proliferation of rideshare and delivery services [1,2]. In general, congestion can be defined as the traffic condition when the travel demand exceeds road capacity [3]. During congestion, the normal flow of traffic is interrupted due to high vehicle density resulted in excess travel time [4]. Traffic congestion in urban metropolitans could be recurring or nonrecurring, depending on the causes that induce congestion [5]. Recurring congestion could occur due to excess demand, insufficient infrastructures, variation in traffic flow, and signal [5,6]. On the other hand, nonrecurring congestion could be a result of incidents, work zones, weather-related, and special events [5,7]. In recent years, the social, economic, and environmental impacts of traffic congestion have been increased significantly with the growth of the population. Due to the more significant amount of travel delay and cost caused by congestion, the urban transportation system is affected considerably. In 2014, people in the United States (US) traveled 6.9 billion extra hours and purchased 3.1 billion additional gallons of fuel, extra costing a total of $160 billion [8,9]. As the existing infrastructures are unable to accommodate the increasing number of automobiles, congestion also increases. In 2017, the INRIX Roadway Analytics estimated the cost to drivers was about $480 billion in the most congested 25 cities of the US due to lost time, wasted fuel, and carbon emitted during congestion [10]. In 2018, the total cost of lost productivity in the US due to congestion was $87 billion [11]. In 2019, the average American drivers were estimated to lose 99 h due to traffic congestion, which equals to about $1377 in monetary value [12].

Moreover, congestion is partially responsible for the physical degradation of transportation infrastructures, and consequently, for the reduction of network performance [13]. Several network recovery strategies have been developed to recover the damaged network performance, such as periphery recovery or preferential recovery [14,15]. However, most of these strategies may not be sufficient enough to improve the network resilience or applicable to traffic failure induced by congested traffic conditions [16]. For many years, many attempts have been taken to monitor and minimize losses due to congestion with different approaches [17–20]. A variety of congestion detection and re-routing strategies based on fuzzy logic and neural network classifier was developed to minimize congestion. Bhandari et al. [17] performed a survey on these strategies. A reinforcement learning-based variable speed limit (VSL) control strategy was developed to reduce travel time at freeway bottlenecks [18]. Another commonly applied method of minimizing traffic congestion is to assign tolls to streets and roads so that drivers are induced to take alternative routes [19]. This method helps in enhancing the distribution of traffic across the road network. One of the significant reasons for the enhancement of congestion is inefficient

* Corresponding author.
E-mail address: tanzina.afrin@ndsu.edu (T. Afrin).

https://doi.org/10.1016/j.measurement.2021.109051
Received 10 August 2020; Received in revised form 9 January 2021; Accepted 13 January 2021
Available online 21 January 2021

vehicles. Assessing vehicles ensures a sustainable transportation system. Mitropoulos et al. [21] proposed a method that combines fuzzy logic and Monte Carlo Simulation (MCS) to assess urban transportation vehicles to ensure the sustainability performance of vehicles.

For the last few years, predictive analysis has been extensively explored in the research field of intelligent transportation systems. A variety of machine learning and neural network algorithms have been applied for this purpose. Chen et al. [22] proposed a long short-term memory model for predicting traffic model using open-source online data. The proposed model showed better prediction compared to the multilayer perceptron model, decision tree model, and support vector machine model. A decentralized deep learning-based method was proposed by Fouldgar [23]. Based on their approach, the congestion state of a node in realtime can be predicted based on the congestion state of the neighboring nodes or stations. Chen and Yu [24] developed a novel deep convolutional neural network-based method by modeling periodic traffic data for short-term traffic congestion prediction. The traffic state of a transportation network varies with time. Thus, it is necessary to consider the temporal correlations of a transportation network while predicting traffic congestion. Zhang et al. [25] proposed a deep autoencoder-based neural network model to learn the temporal correlations of a transportation network and predicting traffic congestion. Predictive analysis using traffic data has not only been used to predict traffic congestion but also has been used for other applications. Such as accident predictions [26,27], traffic demand analysis [28,29], and predicting traffic speed and flow [30–32].

To build a resilient traffic control management system, proper monitoring of the traffic conditions is necessary. By doing this, the congestion levels can be quantified promptly, and preventive actions can be initiated before the peak of the congestion hours. In recent years, researchers and transportation experts have developed different approaches to estimate traffic congestion. The current measures are established depending on various standard parameters, such as speed, travel time, delay, level of services, or other indices. A survey on the current traffic measurements was performed as the authors' previous work [33]. The applicability of these measures varies in countries, as well as institutions [34,35]. For example, the Roadway Congestion Index was used by the Texas Transportation Institute in the 1994 urban mobility report on the US [36]. In the congestion report of 2006, the Washington State Transportation Department used the average peak travel time [37]. The LoS (level of services) developed in the Highway Capacity Manual (HCM) [34] defines six levels in the US. In Japan, the LoS is set to be three levels. For the annual congestion trend reports of 2016–2018, the US Department of Transportation (DoT) used congestion hours, travel time index, and planning time index [38–40]. Although various measurement approaches have been used in different authorized departments, these approaches are limited to their category. For example, the speed performance index (SPI) can measure speed variation with time, and V/C can represent vehicle count or volume variation. However, it is necessary to observe all the possible standard parameters, such as speed, travel time, delay, and level of services, which cannot be achieved by the available measures. That is why a measurement approach that is more comprehensive in the sense of representing different important standard parameters needs to be developed. Additionally, most available measures provide a deterministic estimation of the congestion level with discrete values which could lack in presenting the real uncertain scenario in many cases.

In this paper, a Bayesian network-based probabilistic traffic congestion estimation approach is developed. A BN is a probabilistic approach that can estimate the probability of occurring an event considering the effects of different variables. Traffic congestion can also be modeled using BN, which could account for various performances. The BN has been used for predicting traffic congestion and many other applications for the past few years [41–44]. Kim and Wang [45] used a Bayesian network to predict traffic congestion at the presence and absence of sudden incidents. A dynamic Bayesian network-based congestion prediction model was developed by Fan et al. [46] to predict the diffusion of congestion in transportation networks. The dynamics of traffic using probe data, which is the byproduct of sensors data and other application data, were investigated by Hofleitner et al. [42]. BN has also been used for predicting the duration of traffic accidents [27]. Moreover, a Bayesian modeling framework was developed to analyze the crash severity effects on the traffic management system [47]. Zhu et al. [48] used Bayesian networks for predicting short-term traffic flow. Another significant use of BN was detecting nonrecurring congestion. Li et al. [49] proposed a coupled scalable Bayesian robust tensor factorization model to detect nonrecurring congestion. In addition, Bayesian inference was used for model selection and applications to urban mobility [50]. Although BN has been implemented in various traffic-related applications, combining different performances to make a probabilistic estimation of the congestion level is a new area to be explored.

This paper aims to address the current challenges and utilize the potentiality of the Bayesian network approach to make a probabilistic estimation of congestion. The objectives of this research are to (1) develop the BN model for estimating both recurring and nonrecurring traffic congestion, (2) combine different performance measures while assessing probable congestion level, and (3) implement the developed BN network with a realtime dataset. The rest of the paper is organized as follows. Section 2 illustrates the concept of the Bayesian network. The proposed BN-based approach for estimating recurring and nonrecurring congestion in a probabilistic manner is elaborated in Section 3. There are ten scenarios presented for each congestion scenario in Section 3. Section 4 discusses the benefits and limitations of the BN approach. Finally, Section 5 summarizes the key findings with the conclusion.

2. Bayesian network

A Bayesian network (BN) is a directed acyclic graph (DAG) with a collection of nodes and arcs that represents probabilistic relationships between different variables [51–53]. These variables and the dependency relationships between them are represented by the nodes and the arcs, respectively. The directions of arcs connecting pairs of nodes represent the type of dependencies between the variables. Consider a graph $G = (V, E)$ as a BN with a set of nodes (variables) $V = \{X_1, X_2, \ldots\ldots X_n\}$ and a set of arcs (links) E as shown in Fig. 1. A link is directed from node X_i to X_j, indicates the states of X_j are dependent on the states of X_i. In this case, X_i is the parent of X_j, and X_j is the child of X_i. The set of all parents of X_i could be defined as par(X_i). Nodes that do not have any child are called the leaf node and do not have any parent are called root vertex. The BN is also known as a belief network.

The BN uses Bayesian inference for probability computations. As BN aims to model conditional dependence between variables, the dependency relationships among variables are quantified by conditional

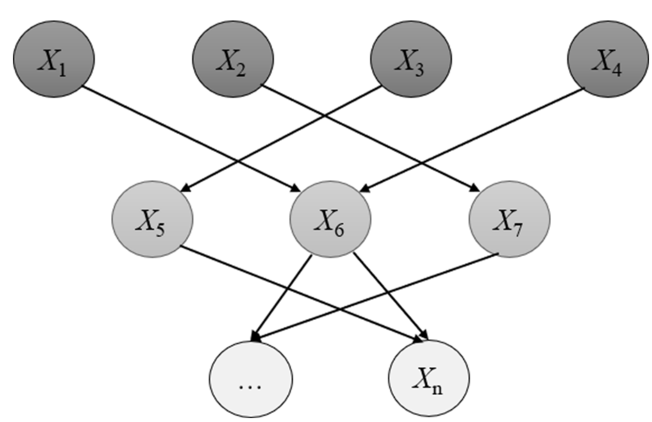

Fig. 1. A general structure of a Bayesian Network with n nodes.

probability distributions, which can be obtained from conditional probability tables (CPT). The conditional probability distributions for X_i can be represented as $P(X_i|par(X_i))$. These probabilities can be found in different ways, such as direct measurement, learning from data, expert knowledge, and from the combination of prior knowledge and data [54–56]. According to the chain rule of probability, the conditional probability distributions associated with all the variables can be used to calculate the joint probability distribution of all the variables specified as,

$$P(X_1, X_2 \cdots\cdots\cdots X_n) = \prod_{i=1}^{n} P(X_i|par(X_i)) \tag{1}$$

$$P(X_1, X_2 \cdots\cdots\cdots X_n) = P(X_1|par(X_1))P(X_2|par(X_2))\cdots\cdots P(X_n|par(X_n))$$

There are three basic steps to construct a BN model: Step 1 defining variables (nodes) with their corresponding states, Step 2 specifying variable dependencies (arcs), and Step 3 generate conditional probability distribution table (CPT). The structure and standard parameters of the BN can be specified manually based on domain expert knowledge and in an automated manner by using a machine learning technique. The conditional probabilities of each state of the variables are calculated using Bayesian inference. For example, the probability of variable X_5 which is dependent on variable X_3 in Fig. 1 can be obtained by the conditional probability between X_5 with X_3 as:

$$P(X_3|X_5) = \frac{P(X_5|X_3)P(X_3)}{P(X_5)} \tag{2}$$

Bayesian inference derives the posterior probability, $P(X_3| X_5)$ as a consequence from a prior probability $P(X_5)$ and a likelihood function $P(X_5| X_3)$, where $P(X_5| X_3)$ is the probability of X_5 given X_3, $P(X_3| X_5)$ is the probability of observing X_3 given X_5. $P(X_3)$ and $P(X_5)$ are the probability of occurring for variable X_3 and variable X_5, respectively. Assuming variable X_3 is an observable variable with two states *true* or *false*, the sum probability of $P(X_{3=true})$ and $P(X_{3=false})$ is equal to 1. If variable X_5 is observed to be *true*, then the Eq (2) becomes $P(X_3| X_{5=true})$ and is updated with $P(X_{5=true})$.

As BN is capable of modeling relationships between variables in complex systems efficiently, it is being used in many real-world applications, including forecasting [30,57], predictive analysis [26,58], risk management [59,60], and many other applications. This paper is mainly focused on employing BN models in probabilistic analysis in transportation and logistics research areas.

3. Bayesian network-based congestion estimation approach

Measuring traffic congestion is the first step towards a reliable traffic management system as it helps taking necessary steps in mitigating congestion in the least possible period. From the analysis of various currently available congestion measurement approaches, it was observed that these measures represent individual performance (speed, volume, and time) [33]. However, all these performances contribute to indicating congestion. That is why different measures are being used simultaneously to portray the real traffic condition. In this section, a Bayesian network-based approach is developed for estimating both recurring and nonrecurring congestion. One of the most significant advantages of using a Bayesian network is that it can determine the probability of occurring different states of a target variable while considering the prior parameter values. Also, it can incorporate previous information to update the present standard parameters. Considering the potentialities of the Bayesian network, a probabilistic analysis of the congestion level estimation is proposed. The flow chart of the proposed framework for the BN-based congestion estimation approach is illustrated in Fig. 2.

This framework can be implemented for both recurring and nonrecurring congestion. For either of the type of congestion, the first step is to select the suitable congestion measures to find the combined estimation. The next step is to define the combined congestion state. The

Table 1
Speed performance index with traffic state.

Speed Performance Index	Traffic State Level	Description of Traffic State
[0,25]	Heavy Congestion	Low average speed, poor road traffic state
(25,50]	Mild Congestion	Lower average speed, road traffic state is weak
(50,75]	Smooth	Higher the average speed, road traffic state is better
(75,100]	Very Smooth	High average speed, road traffic state is good

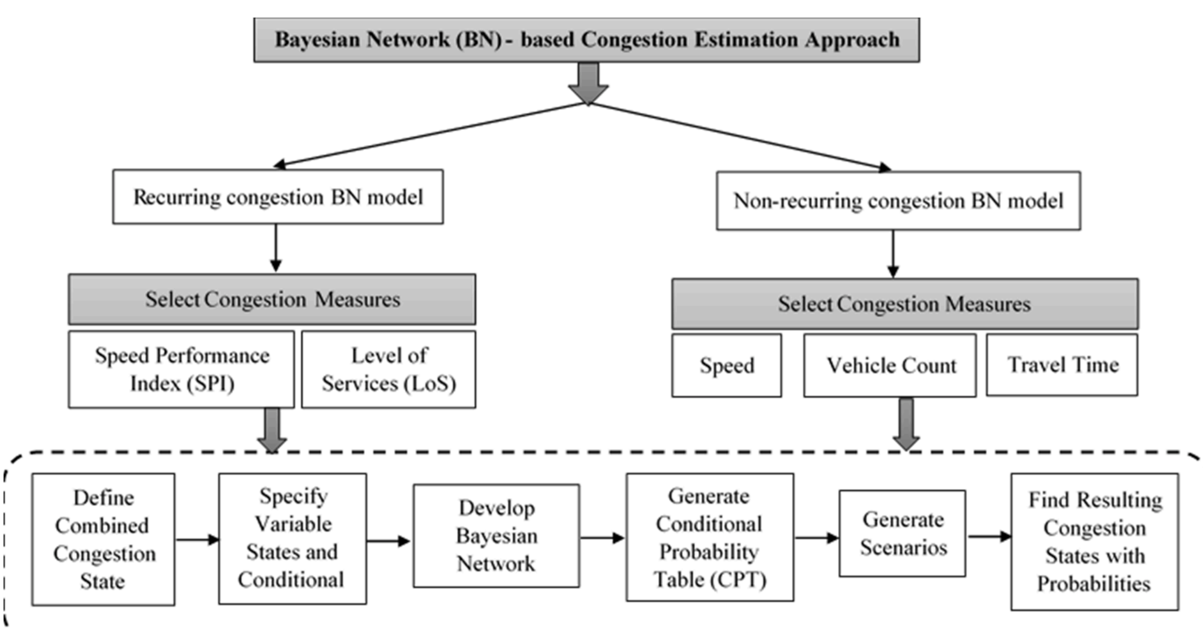

Fig. 2. Bayesian Network (BN) – based congestion estimation framework.

Table 2
LoS classes based on the corresponding V/C ratio and operating conditions.

LoS Class	Traffic state and condition	V/C ratio
A	Free flow	0–0.60
B	Stable flow with unaffected speed	0.61–0.70
C	Stable flow but speed is affected	0.71–0.80
D	High-density but the stable flow	0.81–0.90
E	Traffic volume near or at capacity level with low speed	0.91–1.00
F	Breakdown flow	>1.00

most significant phase of the framework is to develop the Bayesian network (BN). Prior to a suitable BN can be developed, the relevant variables and the conditional relationships between the variables need to be specified. Depending on the structure of the BN, resulting in a CPT table can be obtained for one or more than one variables. After a BN model is finalized a variety of scenarios can be generated to perform qualitative and quantitative analysis. In this paper, assessing the developed BN for different traffic scenarios, the resulting congestion state with the probability of each state occuring can be quantified.

3.1. Bayesian network for recurring congestion

From the variety of different measurement approaches for recurring congestion, speed performance index (SPI), and level of services (Los) were selected to be included in the BN model. These two measures consider the most basic standard parameters, which are speed and volume, that can represent the traffic condition in the most significant way. In general, for road traffic to be considered in the congestion state, the average traffic speed is significantly lower, and volume is higher compared to the normal traffic condition. The justification behind employing both SPI and LoS rather than just either one of them is that the combination of these measures is a better representation of traffic state since the root cause of congestion involves both speed and vehicle volume. Both speed and volume are crucial measures to determine the congestion state as using only one would not be sufficient for this purpose. Moreover, these measures have been used in different works of literature as well as in annual traffic reports. On the basis of adhering with the measurements employed in formal traffic reports, these two measures were selected and deemed to be sufficient to define various traffic states. But there is still no standard for determining the scope of parameter values for different traffic levels.

***a. Speed Performance Index (SPI)*:** SPI is the ratio between average vehicle speed and the maximum permissible speed, as shown in Eq. (3) [34]. The value of SPI ranges from 0 to 100. The traffic state level can be classified with three threshold values (25, 50, and 75) and four levels, as shown in Table 1.

$$SPI = \left(v_{avg}/v_{max}\right) \times 100 \tag{3}$$

where SPI denotes the speed performance index, v_{avg} indicates the average travel speed, and v_{max} denotes the maximum permissible road speed.

***b. Level of Services (LoS)*:** The LoS approach was introduced in the Highway Capacity Manual (HCM) [61]. LoS is a popular method in determining traffic states due to its simplicity. The volume-to-capacity ratio (V/C) is one of the methods used to estimate the LoS of a roadway. The scale intervals of the volume-to-capacity ratio (V/C) are shown in Table 2. The V/C ratio can be calculated by,

$$V/C = N_v/\ N_{max}, \tag{4}$$

where N_v is the spatial mean volume, and N_{max} denotes the maximum number of vehicles that a segment can contain as the capacity [62,63]. It can be further quantified as,

$$N_{max} = (L_s/\ L_v) \times N_l, \tag{5}$$

where L_s is the spatial segment length, L_v is the average vehicle length occupancy, and N_l is the number of lanes. L_v includes vehicle length and safety distance. In general, it is assumed that vehicle length is about 14 ft. (approx. 4.27 m), and safety distance is about 15 ft. (approx. 4.57 m) [63].

a. ***Define Combined Congestion State*:** To determine the recurring congestion state, V/C and SPI levels are combined, as shown in Table 3. For each level of a measure (V/C and SPI), a value is assigned in ascending order. The values assigned for six degrees of services (LoS): A, B, C, D, E, F based on the V/C ratio are 1, 2, 3, 4, 5, and 6, respectively. A value 1 indicates free flow, and 6 shows breakdown flow. Similarly, the values assigned to the SPI levels: very smooth, smooth, mild, and heavy are 1, 2, 3, and 4, respectively. To find the final congestion state measure, these values are added. The

Table 3
The new range for recurring congestion states.

V/C		SPI		Congestion state	
LoS	**Value**	**Level**	**Value**	**Level**	**Value**
A (Free flow)	1	Very smooth	1	Smooth	1–4
B	2	Smooth	2	Mild	5–7
C	3	Mild	3	Heavy	8–10
D	4	Heavy	4	–	–
E	5	–	–	–	–
F (Breakdown)	6	–	–	–	–

Table 4
Variables and state definitions for the BN model for recurring congestion.

Level 1: Attributes		Level 2: Parameters		Level 3: Congestion level	
Variables	**States**	**Variables**	**States**	**Variables**	**States**
Segment (Sg)	G1 = Indiana to I-94 Expressway G2 = I-94 Expressway to Indiana G3 = Lawrence to Kennedy Expressway G4 = Kennedy Expressway to Lawrence	Speed (Sp)	Low Medium High	SPI	Heavy [0,25] Mild (25,50] Smooth (50,75] Very smooth (75,100]
Direction (Dr)	EB - Eastbound WB - Westbound NB - Northbound SB - Southbound	Vehicle count (V)	Low Medium High	V/C	A (0–0.60) B (0.61–0.70) C (0.71–0.80) D (0.81–0.90) E (0.91–1.00) F (>1.00)
Day (D)	Weekday (Mon – Fri) Weekend (Sat, Sun)	–	–	Congestion state **(C)- target variable**	Smooth Mild Heavy
Time (T)	AM Peak (6 – 9 am weekdays) PM Peak (4 – 7 pm weekdays) Off-Peak	–	–	–	–

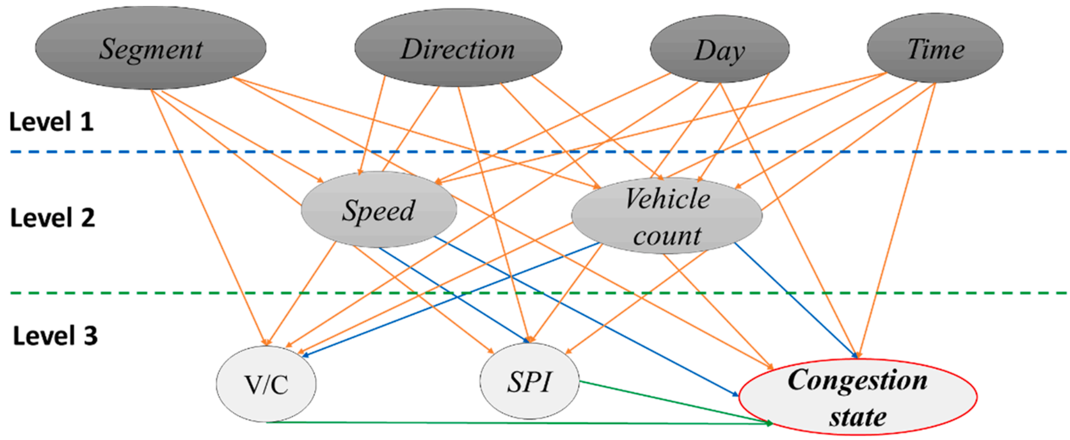

Fig. 3. Graph representation of the proposed BN model for recurring congestion.

new range for congestion states is defined as smooth (1–4), mild (5–7), and heavy (8–10). Table 3 shows a new range of congestion measures after combining both SPI and V/C for primary data processing purposes to obtain congestion values for the conditional probability table to build the initial BN model.

3.1.1. Recurring congestion BN model

The development of BN model for measuring recurring congestion involved nine variables selected as the nodes. These variables are segment, direction, day, time, speed, vehicle count, SPI, V/C, and congestion state. The states of these variables are described in Table 4. The "segment" node specifies the road segments that are considered for the analysis. In this paper, four road segments were taken into account, as given in Table 4. The "direction" node indicates the direction of a road segment, which includes eastbound, westbound, northbound, and southbound. "day" node considers two discrete states weekdays (Mon – Fri) and weekends (Sat, Sun). The "time" node divides the time of the day into three states AM peak, PM peak, and off-peak. Both speed and volume of the vehicles are discretized into low, medium, and high according to the maximum allowable speed level and the road segment capacity. The "speed performance index (SPI)" node indicates four different traffic state levels (heavy, mild, smooth, very smooth). Similarly, the "V/C" node has six levels of service (A, B, C, D, E, F). Finally, the combined congestion state is discretized into three levels (low, medium, high) in the "congestion state" node. All these variables are divided into three levels considering the variable relationships. Level 1: segment, direction, day, time, Level 2: speed, vehicle count, and Level 3: SPI, V/C, congestion state. The congestion state is the target variable.

Once the states of the variable are specified, the next step is to determine the conditional relationships between variables. In this paper, the structure of the BN model was manually defined based on experts' domain knowledge. The proposed BN model for recurring congestion is shown in Fig. 3. The variables in Level 2: speed, vehicle count, and Level 3: SPI, V/C, congestion state is conditionally dependent on the variables in Level 1: segment, direction, day, time. There is also a conditional dependency between Level 2 and Level 3 variables. Besides, the target variable congestion state depends on the V/C and SPI states. The connections (links/lines) between the nodes represent the conditional dependency between the variables. According to the proposed BN model in Fig. 3, the joint probability distribution for all the variables in the network could be written as,

$$P(\mathrm{Sg},\mathrm{Dr},\mathrm{D},\mathrm{T},\mathrm{Sp},\mathrm{V},\mathrm{SPI},\mathrm{V/C},\mathrm{C}) = P(\mathrm{Sg}).P(\mathrm{Dr}).P(\mathrm{D}).P(\mathrm{T}).P(Sp|\mathrm{Sg},\mathrm{Dr},\mathrm{D},\mathrm{T}).P(V|\mathrm{Sg},\mathrm{Dr},\mathrm{D},\mathrm{T})\cdots.P(\mathrm{SPI}|\mathrm{Sg},\mathrm{Dr},\mathrm{D},\mathrm{T},\mathrm{Sp}).P(\mathrm{V/C}|\mathrm{Sg},\mathrm{Dr},\mathrm{D},\mathrm{T},\mathrm{V}).\; P(C|\mathrm{Sg},\mathrm{Dr},\mathrm{D},\mathrm{T},\mathrm{Sp},\mathrm{V},\mathrm{SPI},\mathrm{V/C}) \tag{6}$$

The probability of the congestion state being smooth could be calculated by,

$$\begin{aligned} &P[\mathrm{Sg},\mathrm{Dr},\mathrm{D},\mathrm{T},\mathrm{Sp},\mathrm{V},\mathrm{SPI},\mathrm{V/C},C{=}smooth] \\ &= \sum_{Sg,Dr,T,D,Sp,V,SPI,\frac{V}{C}} P[Sg,Dr,T,D,Sp,V,SPI,\frac{V}{C},C{=}smooth] \\ &= \sum_{Sg,Dr,T,D,Sp,V,SPI,\frac{V}{C}} \left(\begin{array}{c} P\left[C{=}smooth|Sg,Dr,T,D,Sp,V,SPI,\frac{V}{C}\right] P[Sg].P[Dr].P[T]\cdots \\ .P[D].P[Sp].P[V].P[SPI].P[V/C] \end{array} \right) \end{aligned} \tag{7}$$

Eq. (7) could be similarly updated for both C = mild, and C = heavy congestion state.

Segment	Direction	Day	Time	Speed	Vehicle_count	SPI_level	VC	smooth	mild	heavy
G1	EB	Weekday	am peak	low	low	very smooth	B	60	20	20
G1	EB	Weekday	am peak	low	low	very smooth	C	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	very smooth	D	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	very smooth	E	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	very smooth	F	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	smooth	A	60	20	20
G1	EB	Weekday	am peak	low	low	smooth	B	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	smooth	C	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	smooth	D	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	smooth	E	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	smooth	F	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	mild	A	71.429	14.286	14.286
G1	EB	Weekday	am peak	low	low	mild	B	25	50	25
G1	EB	Weekday	am peak	low	low	mild	C	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	mild	D	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	mild	E	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	mild	F	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	heavy	A	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	heavy	B	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	heavy	C	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	heavy	D	33.333	33.333	33.333
G1	EB	Weekday	am peak	low	low	heavy	E	33.333	33.333	33.333

Fig. 4. Partial CPT for the recurring congestion BN model providing different combinations of variables' states.

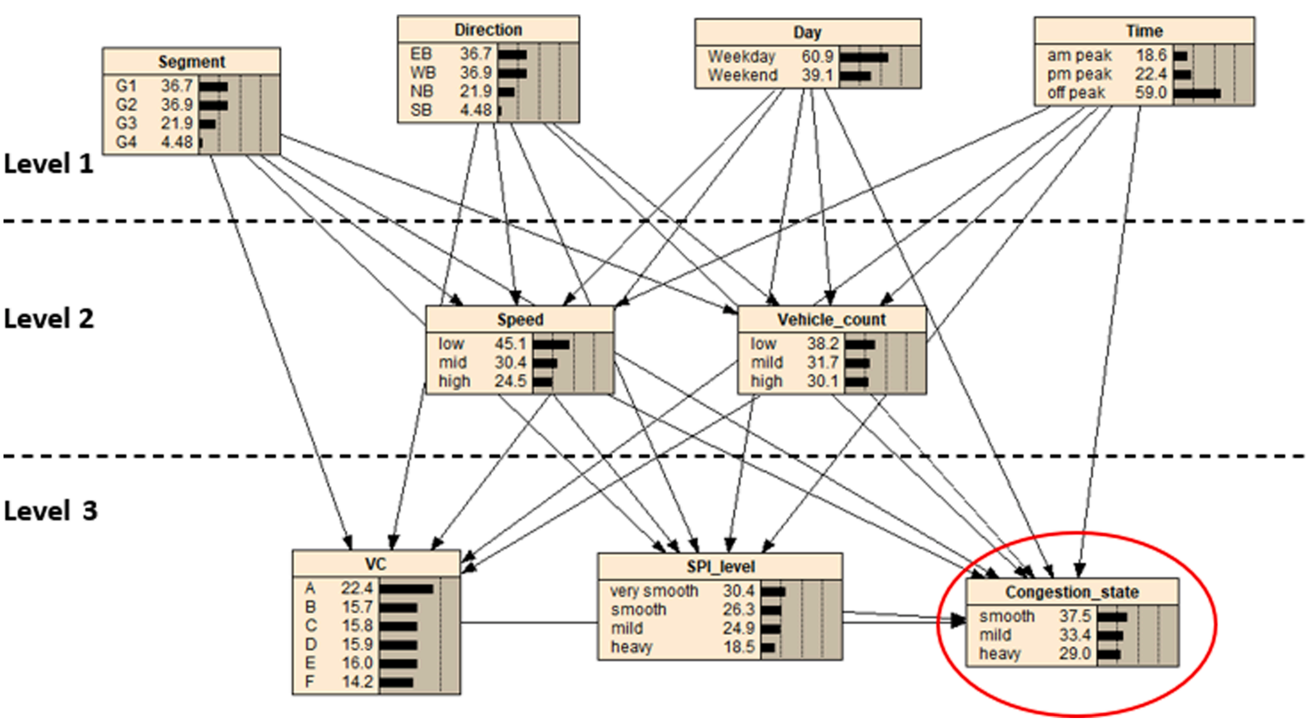

Fig. 5. Graph representation of the proposed recurring congestion BN model when there was no observation made.

3.1.2. Data analysis

The proposed BN model was implemented with an open-source dataset from the Chicago Traffic Tracker [64]. This dataset contains the historical estimation of congestion for over 1000 traffic segments, starting from March 2018. Traffic congestion on Chicago's arterial streets (non-freeway streets) in realtime was estimated by the continuous monitoring and analysis of the received GPS (Global Positioning System) traces from Chicago Transit Authority (CTA) buses. The variables selected from this dataset for analysis are date, time, segment, direction, speed, bus count, message count, segment length. The vehicle count was assumed from the bus count, message count, and the calculated road capacity. By using these variables, SPI and V/C were calculated using Eqs. (3), (4), and (5). In the end, the modified dataset contains the date, time, segment, direction, segment length, speed, vehicle count, SPI, and V/C was used for further analysis. Each of the variable values was taken for every 10 min for seven consecutive days from March 14, 2018, to March 20, 2018. In this paper, four segments: G1 (Indiana to I-94 Expressway), G2 (I-94 Expressway to Indiana), G3 (Lawrence to Kennedy Expressway), G4 (Kennedy Expressway to Lawrence), and four directions: EB (Eastbound), WB (Westbound), NB (Northbound), and SB (Southbound) were considered. Speed and V/C were discretized for each segment according to the maximum allowable speed and the segment capacity. For parameter learning and probabilistic inference, Netica software was used. A part of the conditional probability table (CPT) is shown in Fig. 4. The full CPT table will be made available upon request.

The proposed Bayesian network was implemented, and the probability distribution was estimated based on the provided traffic data. The estimated probability results are shown in Fig. 5. The probability distribution of segment, direction, day, and time of the current dataset provides the actual state of the given dataset. In Level 1, the probability distribution of a given variables x was calculated as, P(A $= x$) = (No. of events A $= x$)/No. of all possible events from the dataset employed. Without any observation made and based on the data employed, in Level 2, the probabilities of the average speed are low, medium, and high are 45.1%, 30.4%, and 24.5%, respectively. The states of vehicle count being low, medium, and high are 38.2%, 31.7%, and 30.1%, respectively.

In Level 3, the V/C levels, A (free flow), B, C, D, E, F (breakdown) indicates traffic states from the free flow to the breakdown state occurred with the probability of 22.4%, 15.7%, 15.8%, 15.9%, 16.0%, and 14.2%. The expectations of very smooth, smooth, mild, and heavy SPI levels are 30.4%, 26.3%, 24.9%, and 18.5%. The found probabilities of final congestion states are 37.5% smooth, 33.4% mild, and 29.0% heavy. This indicates that in the current situation, the congestion state is most likely to be smooth, with a probability of 37.5%. This probability is the highest probability among the three congestion states. The highest likelihood for a smooth congestion state resulted from its child nodes of a very smooth SPI level and A level V/C.

For evaluating the proposed BN model, ten scenarios were generated by changing the probabilities of different variables' states to 100%. The scenario descriptions, along with the changes in expectations of states of V/C, SPI, and the final congestion states, could be found in Table 5, Table 6, and Table 7.

Scenarios 1 – 4 were generated for comparing the congestion states during weekdays and weekends. Scenario 1 was observed for setting both weekends and off-peak to be true (100%), which shows level A (free flow) of V/C with 23.7% and a very smooth level of SPI with 29.4%, which comprises to 38.7% probability for a smooth congestion state. Although the states of V/C, SPI, and the congestion states remain the same in Scenarios 2, 3, and 4, the probabilities vary. From the probability distribution, it was observed that the smooth congestion state is more likely to happen during the weekend and off-peak hours of weekdays.

Scenarios 5, 6, and 7 were generated for observing low speed and varying vehicle count. Scenario 5 (low speed, high volume) indicates a 21.9% probability of V/C level E (Traffic volume near or at capacity level with low speed), 32.1% very smooth SPI, and 39.6% mild congestion state. However, a probability of 35.3% of the heavy congestion state was also observed for this combination. The BN model for Scenario 5 could be found in Fig. 6. The states in Scenario 5 were changed in Scenario 6 (low speed, medium-volume) as of 23.7% V/C level C (Stable flow but speed is affected) and 29.3% smooth SPI comprising 41.6% mild congestion state. In Scenario 7 (low speed, low volume), V/C of level A (free flow) with 35.5% and SPI with 35.5% smooth result 33.4% mild congestion state. These three scenarios demonstrate that low speed with high and medium volume is associated with a mild congestion state.

The rest three Scenarios 8, 9, and 10 were generated for observing high speed and varying vehicle count. For low vehicle count, V/C level A (free flow) with 42.2% and SPI with 43.8% very smooth, and 57.6% smooth congestion state was observed. For medium vehicle count, V/C level C (Stable flow but speed is affected) with 25.6% and SPI with 37.4% very smooth, and 38.1% smooth congestion state was observed. Finally, for high vehicle count, level E (Traffic volume near or at capacity level with low speed) with 26.0% and SPI with 33.7% very smooth, and 33.4% smooth congestion state was observed. These three scenarios demonstrate that high speed with low and medium volume is most likely representing a smooth congestion state.

Table 5
Probability distribution in Scenario 1–4 of recurring congestion BN model.

Scenario	Description	V/C	SPI	Congestion state
1	Weekend, off peak	A (23.7%)	Very smooth (29.4%)	Smooth (38.7%)
2	Weekday, off-peak	A (23.1%)	Very smooth (31.2%)	Smooth (38.3%)
3	Weekday, am peak	A (20.6%)	Very smooth (34.6%)	Smooth (36.1%)
4	Weekday, pm peak	A (22.7%)	Very smooth (29.2%)	Smooth (37.4%)

Table 6
Probability distribution in Scenario 5–7 of recurring congestion BN model.

Scenario	Description	V/C	SPI	Congestion state
5	Low speed, high volume	E (21.9%)	Mild (32.1%)	Mild (39.6%)
6	Low speed, medium volume	C (23.7%)	Smooth (29.3%)	Mild (41.6%)
7	Low speed, low volume	A (35.5%)	Smooth (35.5%)	Mild (33.4%)

Table 7
Probability distribution in Scenario 5–7 of recurring congestion BN model.

Scenario	Description	V/C	SPI	Congestion state
8	High speed, low volume	A (42.2%)	Very smooth (43.8%)	Smooth (57.6%)
9	High speed, medium volume	C (25.6%)	Very smooth (37.4%)	Smooth (38.1%)
10	High speed, high volume	E (26.0%)	Very smooth (33.7%)	Mild (33.4%)

3.2. Bayesian network for nonrecurring congestion

Nonrecurring congestions are an unusual kind of congestion that generally occurs due to unpredictable events. The most common activities that could cause nonrecurring congestion are traffic incidents or accidents, work zones, extreme weather, and special events [5,7]. Due to these events, new congestion during the off-peak periods could be initiated, and the duration of recurring congestion could be increased. Measuring nonrecurring congestion has always been challenging as this kind of congestion does not align with the usual circumstances of recurring congestion. The current congestion measures cannot consider the uncertainties of nonrecurring congestion while measuring the congestion level, so it fails to present the real condition.

In this sub-section, a BN-based approach for the probabilistic estimation of nonrecurring congestion is proposed. The BN-based method is capable of capturing the conditionality between different states of the variables, whether the data points follow a trend or not, and provides a probabilistic estimation of the congestion state. Unlike the recurring

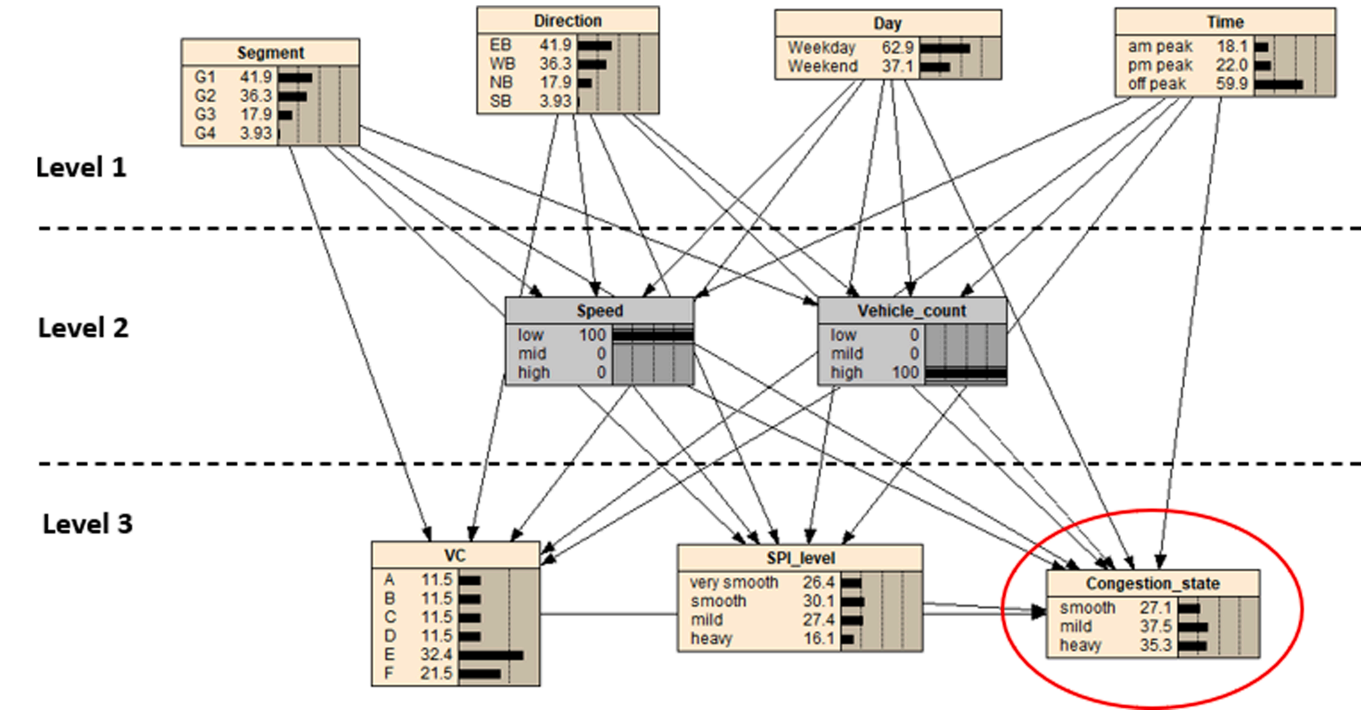

Fig. 6. Variation of probability distributions of variables given that low speed and high vehicle count was observed (Scenario 5).

Table 8
The combined nonrecurring congestion state.

Speed		Vehicle Count		Travel Time		Congestion	
Level	Value	Level	Value	Level	Value	Level	Value
Low (high congestion)	3	Low	1	Low	1	Low	1–3
Medium	2	Medium	2	Medium	2	Medium	4–6
High	1	High (high congestion)	3	High (high congestion)	3	High	7–9

congestion state as defined in Section 3.1, the nonrecurring congestion states can be identified with the standard parameters speed, vehicle count, and travel time, as shown in Table 8. This is because the commonly used measures: SPI and V/C do not apply or intend to be used to measure nonrecurring congestion because of the unpredictable trend and root causes of nonrecurring congestion. SPI and V/C measures used in recurring congestion, capture the variation of speed and volume in normal conditions only. Nonrecurring congestion has unusual condition properties that do not follow the usual trend of speed and volume. SPI and V/C fail to consider this uncertainty in their measures. That is why, instead of SPI and V/C in the recurring congestion, the standard parameters speed, vehicle count, and travel time were adopted in estimating the nonrecurring congestion state.

Table 9
Variables and state definitions for the BN model for nonrecurring congestion.

Level 1: Events		Level 2: Circumstances		Level 3: Parameters and target variables	
Variables	**States**	**Variables**	**States**	**Variables**	**States**
Extreme weather (W)	Yes No	Weather forecast (Wf)	Rain Snow Fog Sunny Cloud	Speed (Sp)	Low Medium High
Disaster (Ds)	Yes No	Population density (Pd)	Low Medium High	Vehicle count (V)	Low Medium High
Incidents (I)	Yes No	Event intensity (Ei)	Low Medium High	Travel time (Tt)	Low Medium High
Special events (S)	Yes No	Action plan (AP)	Yes No	Congestion state (target variable) (C)	Low Medium High
Work zones (Wz)	Yes No	Panic (P)	Yes No	–	–

To define the ranges of final congestion state, speed, vehicle counts, and travel time are divided into three levels as low, medium, and high. The three levels of speed are assigned value as low (3), medium (2), and high (1). A low level of speed indicates high congestion. The assigned values for speed were flipped for vehicle count and travel time because a higher level of vehicle count and travel time are most likely to indicate high congestion. The three levels of vehicle count and travel time are low (1), medium (2), and high (3). The assigned values for speed, vehicle count, and travel time were added to define the range for nonrecurring congestion states. The final nonrecurring congestion states are defined as low (1–3), mild (4–6), heavy (7–9), as shown in Table 8.

3.2.1. *Nonrecurring congestion BN model*

The BN model for measuring nonrecurring congestion was developed by considering 14 variables selected as nodes. These variables were divided into three levels considering the variable relationships; Level 1: weather, disaster, incidents, special events, work zones, Level 2: weather forecasts, population density, event intensity, action plan, panic, and Level 3: speed, vehicle count, travel time, and congestion state. The congestion state is the target variable. The states of these variables are described in Table 9.

The first five nodes in Level 1, “Extreme weather,” “Disaster,”

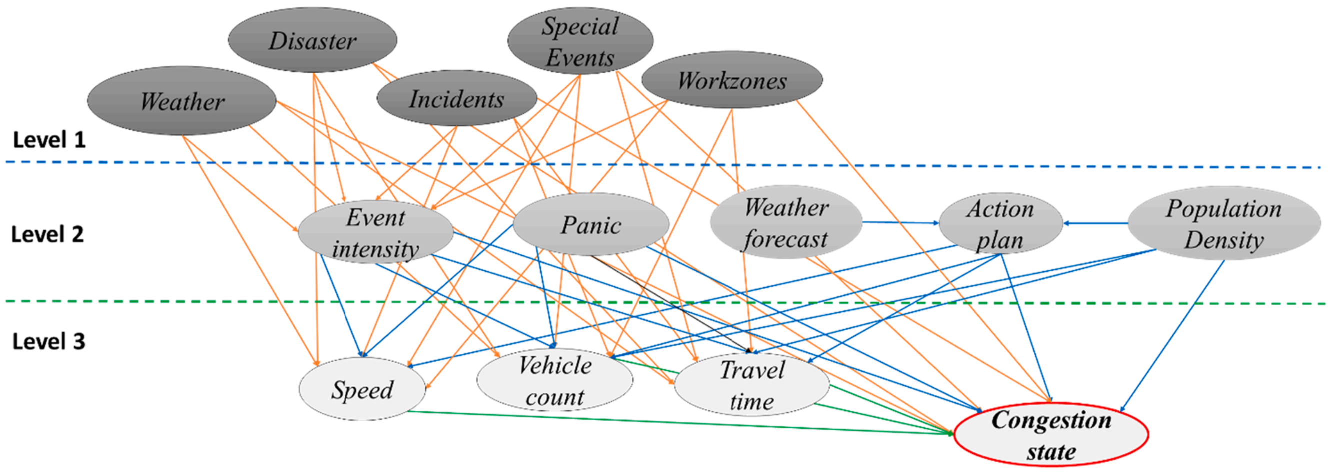

Fig. 7. Graph representation of the proposed BN model for nonrecurring congestion.

"Incidents," "Special events," "Work zones" indicate the possible events that could cause nonrecurring congestion. For the general BN model and to make it flexible to be modified to an event-specific model, these five nodes' states are defined as binary observations, yes or no. In Level 2, several variables that may directly affect the traffic condition, as well as the decision-makers' action to control traffic congestion, such as weather forecast, population density, and event intensity, were taken into consideration. The "Weather forecast" node was set to have five states: rain, snow, fog, sunny, and cloud. Both the "Population" and "Event intensity" nodes were considered to have low, medium, or high state. Another variable that could also affect the traffic condition is whether the passengers are in a panic state or not as it could jeopardize the traffic movement, for example, causing a road accident due to driving recklessly. The "Panic" node has a binary state, yes or no. The next three nodes in Level 3 are the standard congestion parameters: "Speed," "Vehicle count," and "Travel time." The states of these nodes can be defined as a low, medium, and high, which respectively results in a low, medium, or high nonrecurring "Congestion state."

As the states of the nodes are defined, the conditional relationships between nodes were specified. The structure of the BN model for nonrecurring congestion was also manually defined similarly to the recurring congestion BN model depending on experts' domain knowledge. The proposed BN model is shown in Fig. 7. The Level 1 variables (extreme weather, disaster, incidents, special events, work zones) affect the event intensity node in Level 2 and all the variables in Level 3. The variables in Level 2 weather forecast, population density, event intensity, action plan, panic affect the states of the variables in Level 3. Weather forecast and population density in Level 2 affects the action plan within the level as well. The target variable congestion state depends on the other Level 3 variables speed, vehicle count, and travel time. The joint probability distribution of all the variables following the proposed BN model in Fig. 7 can be written as,

$$P(\mathrm{W,D,I,S,Wz,Wf,Pd,Ei,AP,P,Sp,V,Tt,C}) = P(\mathrm{W}).P(\mathrm{D}).P(\mathrm{I}).P(\mathrm{S}).P(\mathrm{Wz}).P(\mathrm{Wf}).P(\mathrm{P}).P(\mathrm{AP}).P(\mathrm{Pd}).$$

$$P(\mathrm{Ei}|\mathrm{W,D,I,S,Wz}).P(AP|\mathrm{Wf,Pd}).P(\mathrm{SP}|\mathrm{W,D,I,S,Wz,Pd,Ei,AP,P}).P(\mathrm{V}|\mathrm{W,D,I,S,Wz,Pd,Ei,AP,P}).$$

$$(\mathrm{Tt}|\mathrm{W,D,I,S,Wz,Pd,Ei,AP,P}).P(C|\mathrm{W,D,I,S,Wz,Wf,Pd,Ei,AP,P,Sp,V,Tt}) \tag{8}$$

The probability of the congestion state being low could be calculated by,

$$P[\mathrm{W,D,I,S,Wz,Wf,Pd,Ei,AP,P,Sp,V,Tt},C = low] = \sum_{W,D,I,S,Wz,Wf,Pd,Ei,AP,P,Sp,V,Tt} P\begin{bmatrix} W,D,I,S,Wz,Wf,Pd,Ei,AP,\cdots \\ P,Sp,V,Tt,C = low \end{bmatrix} \tag{9}$$

$$= \sum_{W,D,I,S,Wz,Wf,Pd,Ei,AP,P,Sp,V,Tt} \left(P[C = low|\mathrm{W,D,I,S,Wz,Wf,Pd,Ei,AP,P,Sp,V,Tt}].P[W].P[D]\cdots.P[I].P[S].P[Wz].P[Pd].P[Ei].P[Ap].P[P].P[Sp].P[V].P[Tt] \right)$$

Eq. (7) could be similarly updated for both C = medium and C = high congestion state.

3.2.1.1. Data analysis. A realtime traffic dataset from the Florida Department of Transportation (FDOT) [65] was employed as the historical data to the proposed BN model for nonrecurring congestion. This dataset contains speed and vehicle count hourly data from September

Event_intensity	Population_d...	Panic	Travel_time	vehicle_count	Speed	Action_plan	Extreme_We...	Disaster	Incidents	Special_events	Workzones	low	mid	high
low	high	No	high	high	high	No	No	No	Yes	No	No	33.333	33.333	33.333
low	high	No	high	high	high	No	No	No	No	Yes	Yes	33.333	33.333	33.333
low	high	No	high	high	high	No	No	No	No	Yes	No	33.333	33.333	33.333
low	high	No	high	high	high	No	No	No	No	No	Yes	33.333	33.333	33.333
low	high	No	high	high	high	No	No	No	No	No	No	33.333	33.333	33.333
mid	low	Yes	low	low	low	Yes	Yes	Yes	Yes	Yes	Yes	33.333	33.333	33.333
mid	low	Yes	low	low	low	Yes	Yes	Yes	Yes	Yes	No	33.333	33.333	33.333
mid	low	Yes	low	low	low	Yes	Yes	Yes	Yes	No	Yes	33.333	33.333	33.333
mid	low	Yes	low	low	low	Yes	Yes	Yes	Yes	No	No	33.333	33.333	33.333
mid	low	Yes	low	low	low	Yes	Yes	Yes	No	Yes	Yes	33.333	33.333	33.333
mid	low	Yes	low	low	low	Yes	Yes	Yes	No	Yes	No	33.333	33.333	33.333
mid	low	Yes	low	low	low	Yes	Yes	Yes	No	No	Yes	33.333	33.333	33.333
mid	low	Yes	low	low	low	Yes	Yes	Yes	No	No	No	33.333	33.333	33.333
mid	low	Yes	low	low	low	Yes	Yes	No	Yes	Yes	Yes	33.333	33.333	33.333
mid	low	Yes	low	low	low	Yes	Yes	No	Yes	Yes	No	33.333	33.333	33.333

Fig. 8. Partial conditional probability table (CPT) for nonrecurring congestion BN model providing different combinations of variables' states.

15, 2017, to September 17, 2017, around US I-75 [65]. This dataset was selected because, during the time Hurricane Irma occurred, nonrecurring traffic congestion occurred. As it was a natural disaster, the state of disaster was set to yes, and the other Level 1 variable states are set to no. The states of the variables of Level 2 event intensity, panic, weather forecast, action plan, and population density were assumed, and the dataset was preprocessed accordingly. The travel time was generated and discretized along with speed and vehicle count using the speed data. The nonrecurring congestion states were found by using the defined range in Table 8, and the Netica software was used for parameter learning and probabilistic inference. A part of the conditional probability table (CPT) is shown in Fig. 8.

The probability distribution was estimated using the dataset, as shown in Fig. 9. At Level 1, the probability of occurring disaster is 99.8%. The likelihood of being the event intensity as low, medium, or high is 37.9%, 32.2%, and 29.9%, respectively.

At Level 2, the probability that the passengers panicked is 71.1%. This higher probability indicates that passengers seemed to panic during any disaster easily. The chance of the forecasted weather to rain, snow, fog, sunny, and cloud happened with a probability 56.2%, 0.20%, 0.20%, 15.2%, and 28.1%, respectively. Before the disaster occurs, experts might take congestion preventive actions considering the weather forecast. The likelihood of taking the action plan was assumed to be 52.2%. Another critical factor that affects the decision-makers' decision as well as the traffic condition is the population density. It could have either of the levels of a low, medium, or high with a probability of 24.9%, 33.2%, and 41.9%.

At Level 3, the probability of the speed level being low is 40.7%, medium is 29.8%, and high with the possibility of 29.5%. The likelihood that the vehicle count is low, medium, and high is 27.8%, 32.3%, and 39.9%, respectively. Similarly, the probability of the travel time being low is 22.9%, medium is 39.7%, and high is 37.4%. Combining the states of speed, vehicle count, and travel time, the probability of the congestion state being low, medium, and high is 28.9%, 34.2%, and 36.9%. Among all three states, a high level of congestion stands out.

From the implemented nonrecurring congestion BN model, ten scenarios were generated by changing the states of the variables, as shown in Table 10 (Scenario 1–2), Table 11(Scenario 3–4), and Table 12 (Scenario 5–6). Scenario 1 (high intensity, high density) and Scenario 2 (low intensity, low density) were generated for varying event intensity and population density. Scenario 1 observed a medium congestion state with a 38.5% probability, and Scenario 2 estimated a low congestion state with 56.4% probability. These two scenarios indicate that both event intensity and population density may induce medium nonrecurring traffic congestion.

Scenario 3 and 4 were generated for varying event intensity and panic states. In Scenario 3 (high intensity, panicked), a high congestion state with a probability of 45.0% is observed. And in Scenario 4 (low intensity, not panicked), a medium congestion state with an expectation

Table 10
Probability distribution in Scenario 1 and Scenario 2 of nonrecurring congestion BN model.

Scenario	Description	Speed	Vehicle Count	Travel Time	Congestion state
1	High intensity, high density	Low (41.0%)	High (49.5%)	Medium (39.4%)	Medium (38.5%)
2	Low intensity, low density	High (71.1%)	Low (67.8%)	Low (45.1%)	Low (56.4%)

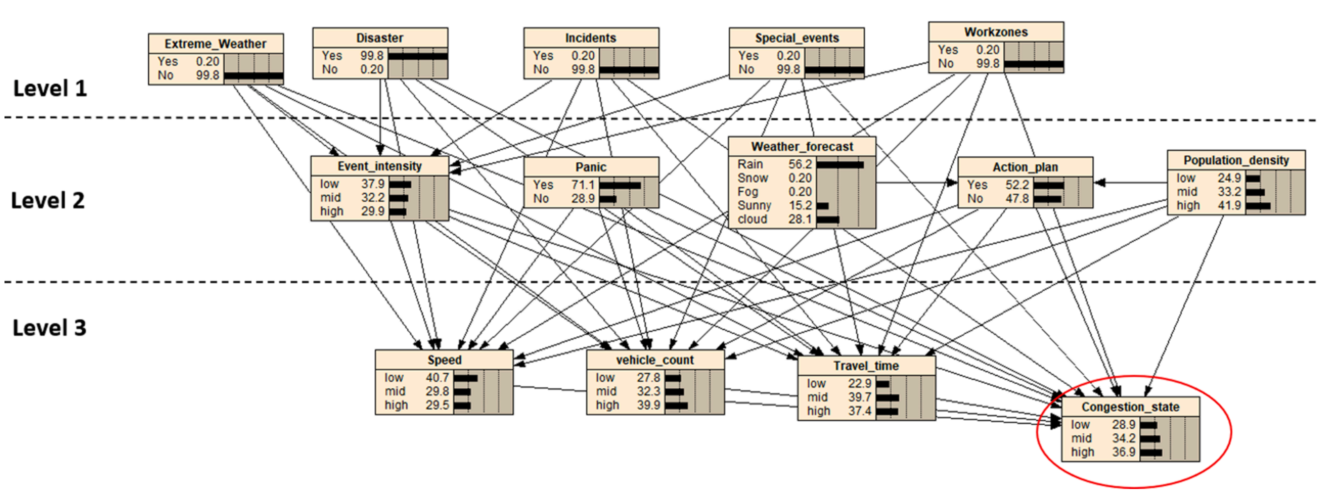

Fig. 9. Graph representation of the proposed nonrecurring congestion BN model with some observation made in Level 1.

Table 11
Probability distribution in Scenario 3 and Scenario 4 of nonrecurring congestion BN model.

Scenario	Description	Speed	Vehicle Count	Travel Time	Congestion state
3	High intensity, panicked	Low (56.6%)	High (56.9%)	High (49.7%)	High (45.0%)
4	Low intensity, not panicked	Medium (54.9%)	Medium (43.8%)	Medium (57.7%)	Medium (49.2%)

Table 12
Probability distribution in Scenario 5 and Scenario 6 of nonrecurring congestion BN model.

Scenario	Description	Speed	Vehicle Count	Travel Time	Congestion state
5	Rain forecast, action taken, high density	Medium (44.1%)	Medium (41.6%)	Medium (45.9%)	Medium (40.1%)
6	Rain forecast, action not taken, high density	Low (42.8%)	High (44.9%)	Medium (62.1%)	High (40.8%)

Table 13
Probability distribution in Scenario 7–10 of nonrecurring congestion BN model.

Scenario	Description	Speed	Vehicle Count	Travel Time	Congestion state
7	Low speed, high vehicle count, high travel time	Low (100%)	High (100%)	High (100%)	High (84.5%)
8	High speed, low vehicle count, low travel time	High (100%)	Low (100%)	Low (100%)	Low (74.6%)
9	Medium speed, medium vehicle count, medium-travel time	Medium (100%)	Medium (100%)	Medium (100%)	Medium (64.1%)
10	No disaster	Equal (33.3%)	Equal (33.3%)	Equal (33.3%)	Equal (33.3%)

of 49.2% is observed. The effects of passenger behavior during a disaster on traffic congestion can be seen from these scenarios. If the passengers panic easily when the disaster intensity is high, this could result in a high congestion state.

Scenario 5 and 6 were generated for varying action plan states with rain forecast and high population density to show the effect of taking the action plan. In Scenario 5 (rain forecast, the action is taken, high density) generated a medium congestion state with 40.1% probability, and Scenario 6 (rain forecast, action not taken, high density) predicted a high congestion state with 40.8% probability. These scenarios demonstrate if the weather forecast is accurate, and the proper action plan is implemented, the congestion could be reduced during a high event intensity.

Scenario 7 (low speed, high vehicle count, high travel time), Scenario 8 (high speed, low vehicle count, low travel time), and Scenario 9 (medium-speed, medium vehicle count, medium-travel time) were generated for varying state combinations of speed, vehicle count, and travel time as shown in Table 13. These three scenarios predicted a congestion state of high with 84.5%, low with 74.6%, and medium with 64.1% probability occurrence, respectively. The probability distribution of Scenario 7 is shown in Fig. 10. These three scenarios demonstrate the effects of different combinations of the states of speed, vehicle count, and travel time to nonrecurring congestion state. A high congestion state is most likely to occur for low speed, high vehicle count, and high travel time, as expected.

Finally, Scenario 10 is observed for no disaster, and no other events occurred. In this scenario, the equal probability distribution (33.3%) of all the states of the Level 3 variables is observed. This indicates that if no such unusual event occurs, the scenario coincides with a general circumstance when traffic congestion could be in any state.

4. Discussion

Traffic congestion is such a universal urban issue that is under constant observation and control of the transportation departments. Transportation experts develop annual reports each year that include the variation of congestion measures over the year. Not only the DOTs, but researchers have also been using a variety of measures for estimating congestion as the first step of monitoring traffic congestion. Because of the low computational complexity and data availability, these measures are applied broadly. However, these measures are parameter specific and do not consider the impacts of various parameter states on the congestion states. A BN-based approach was proposed in this paper to overcome these limitations of the current procedures. The proposed method provides a probabilistic estimation of traffic congestion, which

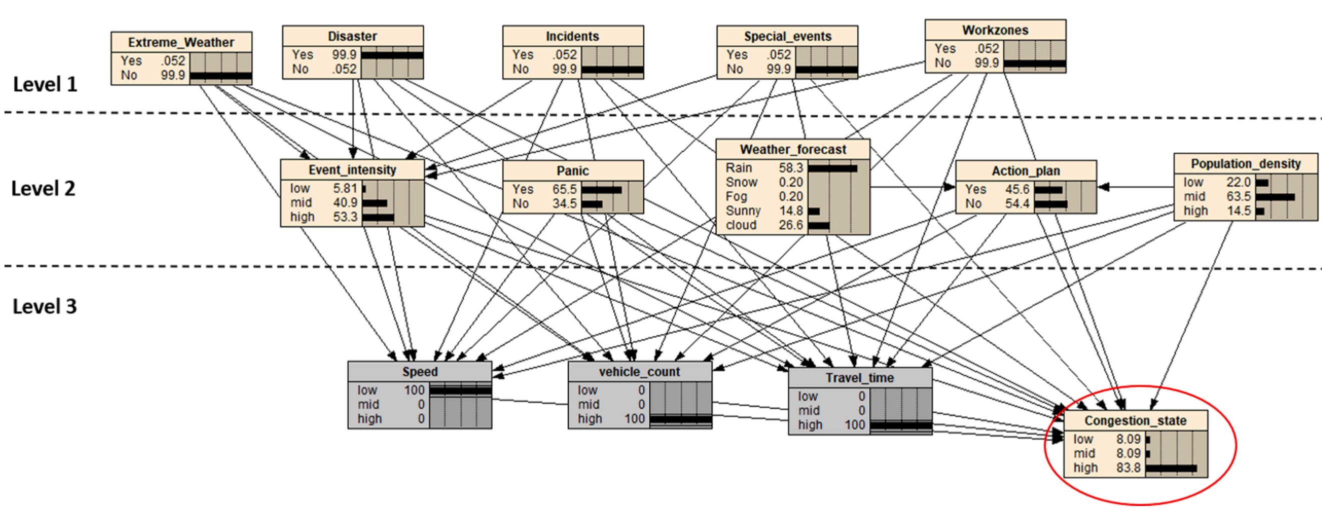

Fig. 10. Variation of probability distributions of variables given that low speed, high vehicle count, and high travel time was observed (Scenario 7).

is a new attempt in this field of study.

One of the most significant advantages of BN is that it can graphically represent the relationships among the variables involved in a circumstance. By using the nodes and arcs, the conditional dependencies between the variables could be easily presented and interpreted. Measuring these dependencies can be done by using conditional probabilities. In this paper, two BN models for recurring and nonrecurring congestion were developed and demonstrated in Sections 3.1.1 and 3.2.1. These BN models were developed by considering the variables that may trigger a specific type of traffic obstruction. Although the number of variables differs in both models, the BN approach is able to estimate the target variable states without any obstacle. Because BN considers all the variables present in the network and there is no limit to the number of variables. From the BN models, the conditional dependencies among these variables are easily comprehensible. Moreover, the BN approach is capable of generating a variety of scenarios by changing the variable states. This attribute helps to view possible changes in an assessment that may aid in the decision-making process.

In this paper, ten scenarios were generated for each of the two proposed BN models. These scenarios indicate how the states of the prior variables can affect the overall congestion state. Evaluating different scenarios through BN is a straightforward and non-time-consuming approach, which makes the BN approach an efficient decision-making tool for users. Although only ten scenarios were generated from each of the models in this paper, more scenarios can be generated depending on the BN models, decision-makers' requirements, and the purpose of the assessment. There is no requirement of a specific amount of data points to develop a BN model. The dataset that was used for recurring congestion contains data points every 10 min, while the dataset that was used for nonrecurring congestion contains hourly data points. However, both models can be used to estimate congestion with varying data points.

One of the aims of the proposed approach was to consider different standard parameters to evaluate the probable congestion level. In the recurring model, V/C and SPI were combined. On the other hand, standard parameters: speed, vehicle count, and travel time were combined to measure the overall nonrecurring congestion state. The most important advantage of using BN for estimating the congestion level is that it considers the probabilities of occurring different variable states and determines the congestion states in a probabilistic manner. This property makes the proposed approach applicable to measure both recurring and nonrecurring congestion. Further, this attribute of the proposed method makes it robust and unique compared to the existing approaches in providing a combined view of speed and vehicle count.

Although the proposed approach offers considerable benefits, there are certain limitations as well. While implementing the BN model, significant data processing may require an extensive amount of time and effort because the raw data needs to be discretized and modified in the form of states for each variable. In addition, traffic congestion is a dynamic phenomenon, as it could change with time as well as space. However, the proposed static BN does not really consider the spatial and temporal aspects of the analysis but purely based on historical data. Thus, in the future, the spatial–temporal traffic propagation throughout a transportation network will be investigated by implementing a dynamic BN for estimating traffic congestion. The spatial–temporal traffic propagation indicates the change of traffic state with time and spreading of congestion hotspots in the traffic network. Dynamic BN may be a good candidate as it can relate variables in the BN to each other at different time steps with the temporal arcs [56]. Moreover, the predictive analysis for a resilient smart traffic management system [62] is another potential future research direction. Although predictive analysis for recurring congestion has been commonly addressed over the years, predicting and detecting nonrecurring congestion requires more attention. For future research, more effort will be allocated towards exploring the root causes and possible solutions for nonrecurring congestions.

5. Conclusion

Despite constant monitoring and investments in transportation systems, traffic congestion remains in society and still increasing with the growing population and infrastructures. Measuring the congestion level provides a view of the traffic state as well as helps to ensure a robust traffic management system. Although a variety of congestion measurement approaches exist, they are still lacking in considering all the traffic standard parameters and applicability. Thus, a BN-based approach is proposed in this paper to overcome the limitations of the current methods. Two BN models for recurring and nonrecurring congestion were developed and implemented using two different datasets due to the nature of traffic congestion behavior. A realtime traffic tracker dataset for recurring congestion and a traffic monitoring dataset for nonrecurring congestion was implemented in the proposed BN models. Ten sensitivity analysis scenarios were generated and analyzed for each recurring and nonrecurring congestion to ensure the effectiveness of the proposed method. The proposed BN approach is a versatile approach that is able to perform qualitative and quantitative analysis. For qualitative analysis, the proposed BN model is able to show the cause and effect relationship between two or more dependent variables. In addition, for quantitative analysis, BN model is able to quantify the likelihood probability of an event happening based on its prior contributing factors. Overall, the proposed BN-based method is beneficial in providing a comprehensive vision of the states of speed and vehicle count with a probabilistic estimation of the congestion state.

CRediT authorship contribution statement

Tanzina Afrin: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Visualization. **Nita Yodo:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was partially supported by NDUS EPSCoR under the award #FAR0021960.

References

[1] T.K. Reed, Joshua, 2018 Global Traffic Scorecard, 2019.
[2] T. Reed, 2019 INRIX Global Traffic Scorecard, 2020.
[3] M. Aftabuzzaman, Measuring traffic congestion-a critical review, Australasian Transport Research Forum (ATRF) 30TH (2007).
[4] C. Systematics, Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation, United States. Federal Highway Administration, 2005.
[5] J.C. Falcocchio, H.S. Levinson, Road traffic congestion: a concise guide. Vol. 7. 2015, Springer.
[6] Y. Wang, et al., Reasons and countermeasures of traffic congestion under urban land redevelopment, Procedia-Soc. Behav. Sci. 96 (2013) 2164–2172.
[7] J.C. Falcocchio, H.S. Levinson, Managing nonrecurring congestion, in: Road Traffic Congestion: A Concise Guide, Springer, 2015, pp. 197–211.
[8] D. Schrank, et al., 2015 urban mobility scorecard, 2015.
[9] A. Dubey, J. White, Dxnat-deep neural networks for explaining non-recurring traffic congestion. arXiv preprint arXiv:1802.00002, 2018.
[10] B. Pishue, US Traffic Hot Spots: Measuring the Impact of Congestion in the United States, 2017.
[11] Traffic congestion cost the US economy nearly $87 billion in 2018. 2019; Available from: https://www.weforum.org/agenda/2019/03/traffic-congestion-cost-the-us-economy-nearly-87-billion-in-2018/.
[12] T. Reed, J. Kidd, Global Traffic Scorecard, INRIX Research, Altrincham, 2019.
[13] G. Karakostas, et al., On the degradation of performance for traffic networks with oblivious users, Transport. Res. B: Methodol. 45 (2) (2011) 364–371.
[14] T. Afrin, N. Yodo, G. Caldarelli, A concise survey of advancements in recovery strategies for resilient complex networks, J. Complex Networks, 2018.

[15] T. Afrin, N. Yodo, Resilience-based recovery assessments of networked infrastructure systems under localized attacks, Infrastructures 4 (1) (2019) 11.
[16] N. Yodo, T. Arfin, A resilience assessment of an interdependent multi-energy system with microgrids, Sustain. Resilient Infrastruct. (2020) 1–14.
[17] A. Bhandari, V. Patel, M. Patel, A survey on traffic congestion detection and rerouting strategies. In: 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI). 2018. IEEE.
[18] Z. Li, et al., Reinforcement learning-based variable speed limit control strategy to reduce traffic congestion at freeway recurrent bottlenecks, IEEE Trans. Intell. Transport. Syst. 18 (11) (2017) 3204–3217.
[19] F. Stefanello, et al., On the minimization of traffic congestion in road networks with tolls, Ann. Operat. Res. 249 (1–2) (2017) 119–139.
[20] F. Harrou, A. Zeroual, Y. Sun, Traffic congestion monitoring using an improved kNN strategy, Measurement 156 (2020) 107534.
[21] L.K. Mitropoulos, et al., A Fuzzy and a Monte Carlo simulation approach to assess sustainability and rank vehicles in urban environment, Transport. Res. Procedia 24 (2017) 296–303.
[22] Y.-y. Chen, et al., Long short-term memory model for traffic congestion prediction with online open data, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). 2016. IEEE.
[23] M. Fouladgar, et al. Scalable deep traffic flow neural networks for urban traffic congestion prediction. In: 2017 International Joint Conference on Neural Networks (IJCNN). 2017. IEEE.
[24] M. Chen, X. Yu, Y. Liu, PCNN: Deep convolutional networks for short-term traffic congestion prediction, IEEE Trans. Intell. Transport. Syst. 19 (11) (2018) 3550–3559.
[25] S. Zhang, et al., Deep autoencoder neural networks for short-term traffic congestion prediction of transportation networks, Sensors 19 (10) (2019) 2229.
[26] B. Ghosh, Predicting the duration and impact of the non-recurring road incidents on the transportation network, 2019.
[27] L. Kuang, et al., Predicting duration of traffic accidents based on cost-sensitive Bayesian network and weighted K-nearest neighbor, J. Intell. Transport. Syst. 23 (2) (2019) 161–174.
[28] T. Watanatada, M. Ben-Akiva, Forecasting urban travel demand for quick policy analysis with disaggregate choice models: a Monte Carlo simulation approach, Transport. Res. A: General 13 (4) (1979) 241–248.
[29] D. Wang, Y. Yang, S. Ning. Deepstcl: A deep spatio-temporal convlstm for travel demand prediction. In: 2018 international joint conference on neural networks (IJCNN). 2018. IEEE.
[30] S. Sun, C. Zhang, Y. Zhang, Traffic flow forecasting using a spatio-temporal bayesian network predictor. arXiv preprint arXiv:1712.08883, 2017.
[31] J. Tang, et al., An improved fuzzy neural network for traffic speed prediction considering periodic characteristic, IEEE Trans. Intell. Transport. Syst. 18 (9) (2017) 2340–2350.
[32] J. Zhao, et al., Traffic speed prediction under non-recurrent congestion: based on LSTM method and BeiDou navigation satellite system data, IEEE Intell. Transport. Syst. Magazine 11 (2) (2019) 70–81.
[33] T. Afrin, N. Yodo, A survey of road traffic congestion measures towards a sustainable and resilient transportation system, Sustainability 12 (11) (2020) 4660.
[34] F. He, et al., A traffic congestion assessment method for urban road networks based on speed performance index, Procedia Eng. 137 (2016) 425–433.
[35] A.M. Rao, K.R. Rao, Measuring urban traffic congestion-a review. Int. J. Traffic Transp. Eng., 2012. 2(4).
[36] T.J. Lomax, D.L. Schrank, The 2005 urban mobility report. 2005.
[37] H. Bar-Gera, Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel, Transport. Res. C: Emerg. Technol. 15 (6) (2007) 380–391.
[38] US Department of Transportation - Federal Highway Administration. 2016 Urban Congestion Trends, 2016.
[39] US Department of Transportation - Federal Highway Administration, 2017 Urban Congestion Trends, 2017.
[40] US Department of Transportation – Federal Highway Administration, 2018 Urban Congestion Trends, 2018.
[41] S. Sun, et al., Bayesian network methods for traffic flow forecasting with incomplete data. European Conference on Machine Learning, Springer, 2004.
[42] A. Hofleitner, et al., Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network, IEEE Trans. Intell. Transport. Syst. 13 (4) (2012) 1679–1693.
[43] P.K. Hoong, et al., Road traffic prediction using, Bayesian Networks (2012).
[44] Y. Ma, et al., Bridge remaining strength prediction integrated with Bayesian network and in situ load testing, J. Bridge Eng. 19 (10) (2014) 04014037.
[45] J. Kim, G. Wang, Diagnosis and prediction of traffic congestion on urban road networks using Bayesian networks, Transport. Res. Rec. 2595 (1) (2016) 108–118.
[46] X. Fan, J. Zhang, Q. Shen, Prediction of road congestion diffusion based on dynamic Bayesian Networks. Journal of Physics: Conference Series, IOP Publishing, 2019.
[47] F. Zong, et al., Analyzing traffic crash severity with combination of information entropy and Bayesian network, IEEE Access 7 (2019) 63288–63302.
[48] Z. Zhu, et al., Short-term traffic flow prediction with linear conditional Gaussian Bayesian network, J. Adv. Transport. 50 (6) (2016) 1111–1123.
[49] Q. Li, et al., Non-recurrent Traffic Congestion Detection with a Coupled Scalable Bayesian Robust Tensor Factorization Model. arXiv preprint arXiv:2005.04567, 2020.
[50] L. Martino, et al., Cooperative parallel particle filters for online model selection and applications to urban mobility, Digit. Signal Process. 60 (2017) 172–185.
[51] D. Heckerman, A tutorial on learning with Bayesian networks, in: Innovations in Bayesian networks, Springer, 2008, pp. 33–82.
[52] F.V. Jensen, T. Nielsen, Bayesian networks and decision graphs Springer-Verlag, New york, 2007.
[53] N. Yodo, P. Wang, Resilience modeling and quantification for engineered systems using Bayesian networks, J. Mech. Des. 138 (3) (2016), 031404 031404.
[54] L. Uusitalo, Advantages and challenges of Bayesian networks in environmental modelling, Ecol. Model. 203 (3–4) (2007) 312–318.
[55] D. Heckerman, J.S. Breese, Causal independence for probability assessment and inference using Bayesian networks, IEEE Trans. Syst. Man Cybernet.-Part A: Syst. Humans 26 (6) (1996) 826–831.
[56] N. Yodo, P. Wang, Z. Zhou, Predictive resilience analysis of complex systems using dynamic Bayesian networks, IEEE Trans. Reliability 66 (3) (2017) 761–770.
[57] R. Zhang, et al., Data-driven photovoltaic generation forecasting based on a Bayesian network with spatial-temporal correlation analysis, IEEE Trans. Indus. Informat. 16 (3) (2019) 1635–1644.
[58] J. Lessan, L. Fu, C. Wen, A hybrid Bayesian network model for predicting delays in train operations, Comput. Industr. Eng. 127 (2019) 1214–1222.
[59] Y. Pan, et al., Improved fuzzy Bayesian network-based risk analysis with interval-valued fuzzy sets and DS evidence theory, IEEE Trans. Fuzzy Syst. (2019).
[60] B.G. Marcot, T.D. Penman, Advances in Bayesian network modelling: integration of modelling technologies, Environ. Model. Software 111 (2019) 386–393.
[61] H.C. Manual, Transportation Research Board Special Report 209. Washington, DC, 1994.
[62] C. Wan, et al., Resilience in transportation systems: a systematic review and future directions. 2018;38(4):479–498.
[63] J. Tang, H.R. Heinimann, A resilience-oriented approach for quantitatively assessing recurrent spatial-temporal congestion on urban roads, PloS one 13 (1) (2018), e0190616.
[64] Chicago Traffic Tracker - Historical Congestion Estimates by Segment - 2018-Current - Data.gov. s; Available from: https://catalog.data.gov/dataset/chicago-traffic-tracker-historical-congestion-estimates-by-segment-2018-current.
[65] Transportation, F.D.o. Traffic Data. 05/12/2020; Available from: https://www.fdot.gov/statistics/trafficdata/default.shtm.