IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.0322000

Modeling traffic congestion spreading using a topology-based SIR epidemic model

ASSEMGUL KOZHABEK[1], WEI KOONG CHAI[1] (Senior Member, IEEE), and GE ZHENG[2]
[1]Department of Computing and Informatics, Bournemouth University, Fern Barrow, Poole, BH12 5BB, UK
[2]Department of Engineering, University of Cambridge, Trumpington St, Cambridge, CB2 1PZ, UK

Corresponding author: Assemgul Kozhabek (e-mail: akozhabek@bournemouth.ac.uk).

This work was supported by the Bournemouth Christchurch Poole (BCP) Council.

ABSTRACT The continuous urbanisation and increase in vehicle ownership have increasingly exacerbated traffic congestion problems. In this paper, we advocate the use of epidemic theory to model the spreading of traffic congestion in urban cities. Specifically, we use the Susceptible-Infected-Recovered (SIR) model but propose to explicitly consider the road network structure in the model to understand the contagion process of road congestion. This departs from the classical SIR model where homogeneous mixing based on the law of mass action is assumed. For this purpose, we adopt the N-intertwined modeling framework for the SIR model based on continuous-time Markov chain analysis. In our evaluation, we used two real-world traffic datasets collected in California and Los Angeles. We compare our results against both classical and average-degree-based SIR models. Our results show better agreement between the model and actual congestion conditions and shed light on how congestion propagates across a road network. We see the potential application of insights gained from this work on the development of traffic congestion mitigation strategies.

INDEX TERMS Congestion spread modeling, epidemics, SIR model, topology,traffic congestion, urban road networks.

I. INTRODUCTION

Traffic congestion has been a persistent societal challenge despite continuous efforts and technological advancements in the last decades. The problem is exacerbated by rapid urbanisation and the growth of motor vehicle ownership. It not only increases journey times but also escalates excess fuel consumption and environmental pollution as well as decreases workforce productivity with increasing time loss [1]. The UK Department for Transport [2] highlighted that the rise in urban congestion is becoming a substantial problem for cities. City transportation authorities worldwide introduced various policies and campaigns for congestion mitigation (e.g., motivating people to walk and cycle by introducing congestion charge zones [3], improving public transport efficiency [4], connecting infrastructure with smart traffic signal optimisations [5] and dynamic traffic forecasting [6]).

One important aspect of these is understanding how congestion spreads over time and space in urban road networks [7], [8]. It offers city municipalities valuable insights on which road segments will be impacted and when. Fundamentally, the question is on how an inherently challenging process like traffic congestion propagates and dissipates over time in urban transport systems. It is known that traffic oscillation itself has two components: formation and propagation [9], [10]. The former can be caused by lane-changing activity or any kind of moving bottleneck [9], [11]. The latter is traffic propagation, the subject of this paper, which remains an interesting research topic especially when traffic congestion is a non-linear, non-equilibrium dynamic process that exhibits emergent behaviours [12], [13]. Different approaches to model the spread of traffic congestion have been investigated in the literature (see Setion II).

In this paper, we follow the approach based on epidemiology models where traffic congestion evolution is modeled as a disease-spreading process. Basic yet versatile compartmental epidemic models such as the SI (Susceptible-Infected), SIS (Susceptible-Infected-Susceptible), and SIR (Susceptible-Infected-Recovered) models have been applied for simulating the spreading processes in the networks. Wu et al. [14] proposed to describe traffic congestion spreading with the SIR model whereby a congested road junction or segment is said to be "infected" and conversely, "susceptible" if not congested and finally, the road junction/segment that becomes uncongested after an episode of congestion can be modeled as a "recovered" state. Furthermore, Saberi et al. [7] provided empirical validation of traffic congestion prop-

agation and its dissipation as a network epidemic spreading phenomenon and discussed the different possible interpretations of basic epidemic models to represent traffic congestion spreading process. We further extend this line of research. While previous works have largely taken the basic simplifying assumption of homogeneous mixing of nodes [15], in this paper, we eschew this assumption and instead explicitly consider the road network topology to model the congestion spreading. Specifically, we adopt an individual-based mean-field approach to model the congestion spreading and apply the model on two real-world traffic datasets.

The rest of this paper is organised as follows. We review the related work in Section II. In Section III, we develop the topology-aware model advocated in this paper. Section IV presents the evaluation of the model and compares it against classical and average-degree-based SIR models. We show better performance of our model using real-world traffic datasets from two different cities. This is followed by the conclusion of our work in Section V.

II. RELATED WORK

The literature on road congestion formation and spreading has a long history and the topic has been investigated from different perspectives. We broadly segregate them into three levels, namely vehicle-level, link-level, and network-level. Work adopting the vehicle-level perspective incorporates individual vehicle's movement and its interactions with other vehicles into consideration. Ahmed et al. [16] identified three approaches to model these interactions: (1) car following models (e.g., [17]; [18]; [19]) where the models take into account how drivers behave when they are driving behind another vehicle in proximity, (2) lane changing models (e.g., [20]) where the models endeavour to capture the idea of replicating the driver's decision to change lanes influenced by diverse factors and (3) gap acceptance models (e.g., [21]) where the models attempt to mimic human behaviour and take into account the reasons for driving decisions. These models focus more on the question of congestion formation but are not able to provide further insights into how congestion spreads in a road network.

The work adopting the link-level view focuses on individual road segments instead and studies traffic as a continuous and aggregated stream. One direction adopting this view employs queueing theory in which the road segment is modeled as a queue with the vehicles traveling across it as the customers [22], [23], [24]. Alternatively, Lighthill and Whitham [25], Richards [26], Newell [27] advocated kinematic wave theories to describe the vehicles' behaviour across a road segment in which traffic is modeled as shockwaves in a two-dimensional time-space diagram either as forward or backward moving. Such models are often applied to understand traffic on a single freeway and can be extended to small road networks. While they offer insights on the traffic evolution within a road segment, they are unable to capture the complex congestion evolution or behaviour of large-scale urban road networks, such as those found in major cities, where the road network topology offers rich possibilities of traveling routes as well as opportunities of disruptions.

Finally, to understand how traffic congestion evolves over time in large-scale road networks, network-level perspective is adopted. The structure of road networks is known to affect the traffic flow at city scale [28], [29], and [30]. Levinson [31] underlined the importance of understanding and monitoring urban street network structures for planning and evaluation of the effectiveness of investments and land use changes within cities. Considering the network-level traffic congestion spread is more challenging as it evolves in multiple directions over space and time. Some researchers attempted to model this by abstracting it as a physical system [32], [33]. For instance, Payne [34] and Whitham [35] developed a traffic flow model using fluid dynamics law where the flow of vehicles is represented as the flow of liquid in a pipe. Helbing [36], on the other hand, developed a gas-kinetic equation-based macroscopic traffic model while Vandaele et al. [37] proposed the use of speed-flow and speed-density diagrams for this purpose. Mahmassani et al. [38] investigate the application of the network fundamental diagram (NFD) as a means to observe, comprehend, and model the loading and unloading processes of urban traffic. Furthermore, the authors of [39], [40], and [41] explore congestion propagation through the lens of network theory, assessing the severity of congestion through clustering methods. However, these works do not address the question on understanding the dynamics of congestion propagation which we look into in this paper.

Meanwhile, recent advancements in artificial intelligence have given rise to new machine learning (ML) and deep learning (DL) models developed for traffic prediction. We refer readers to Zheng et al. [42] and the references therein for an overview of the evolution of the research along this direction from simple ML approaches (e.g., K-nearest neighbour (KNN) algorithm [43], Artificial Neural Network (ANN) [44], [45] to DL models (e.g., Recurrent Neural Networks (RNN) and its variants (Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) [46], Convolutional Neural Network (CNN) [47], and Graph Convolutional Neural Network (GCN) [48]. Here, rather than understanding the congestion spreading, the focus is on predicting the state of the road traffic in the next prediction horizon (e.g., could be set to 5-min or 10-min in the future) based on historical data. The models, especially the recent ones, are usually complex, entail high computational costs, and are dependent on the size and quality of the dataset.

Finally, traffic congestion spreading has also been investigated as a contagion process (e.g., [49], [50], [51], [52], [10]). In this approach, the road network is abstracted and represented as a graph consisting of a collection of nodes (road junctions/intersections in our case) and links (road segments). Traffic congestion evolution can then be replicated as a spreading process on the graph via tools such as epidemic theory which originated from biological studies on contagious disease spreading and has found applications in

various domains (e.g., in communication networks [53], in (online) social network [54], in hub protein and brain structure [55]). The analogy in our case is quite intuitive where a congested road segment can "infect" its adjacent road due to traffic queues extending and overflowing to the next junctions. Collizza et al. [56] associated the epidemic spreading with the traffic flow based on the metapopulation model. Wu et al. [57], on the other hand, described traffic congestion spread with the SIR epidemic model and conducted simulation studies. They highlighted the importance of considering the topological properties of the road network in affecting the behaviour of the traffic system. Furthermore, there is already evidence in the transportation domain that congestion propagation patterns show similarity with the virus contagion process in the real world [58], [59], and [60]. Specifically, Saberi et al. [7] used empirical data from Google for different cities and showed that congestion spreading behaves akin to an epidemic process. Following the above, the literature has highlighted the potential of the epidemic model in replicating congestion spreading and at the same time, the model should take into account the road network while not incurring expensive computational costs or requiring large volume of data.

In this paper, we adopt this contagion view and model the congestion spreading as a contagion process and we explicitly consider the impact of the road connectivity into consideration as we take as input the network topology, allowing us to consider each road segment individually. The model is also scalable via a mean-field approximation.

III. MODELING TRAFFIC CONGESTION SPREADING WITH SIR TOPOLOGY-AWARENESS

As prior mentioned, there are three fundamental contagious models of infectious diseases: SI, SIS, and SIR. From the literature, the SIR model has been adopted as the suitable model for replicating congestion propagation and provides a realistic representation [57], [59], [7]. As such, we also adopt the SIR epidemic model in this work.

A. PRELIMINARY

The SIR model [61] is a widely used epidemic model in epidemiology. By compartmentalisation, a disease is broken down into three distinct stages, namely Susceptible (state S) to represent healthy individuals, Infected (state I) to represent infected individuals who are also infectious and Removed (state R) to represent an individual who has recovered from an infection. As indicated in Section II, this model has gained significant attention in the field of traffic congestion spreading due to its simplicity and performance in capturing congestion dynamics. In the context of this paper, the SIR model is applied to represent the following states:

- Susceptible (S) – nodes which are not congested (free flow traffic) but susceptible to be congested,
- Infected (I) – nodes which are congested, and the congestion may over-spill to the neighbouring nodes,
- Recovered (R) – nodes that suffered from congestion, but the congestion has since dissipated.

Using the law of mass action, the SIR model could be modeled with the following system of ordinary differential equations (ODEs) [61] and [15]:

$$\frac{ds(t)}{dt} = -\beta s(t) i(t) \tag{1}$$

$$\frac{di(t)}{dt} = \beta s(t) i(t) - \gamma i(t) \tag{2}$$

$$\frac{dr(t)}{dt} = \gamma i(t) \tag{3}$$

where $s(t)$, $i(t)$ and $r(t)$ denote the number of nodes in susceptible, infected and recovered state at time t respectively while β and γ are the propagation and dissipation rates. Hereafter, in this paper, we refer to the above model as the *classical SIR* model.

Using the homogeneous mixing hypothesis, the classical SIR model can further consider including the average contacts of each node in a network into the system of equations as follows [33]:

$$\frac{ds(t)}{dt} = - < k > \beta s(t) i(t) \tag{4}$$

$$\frac{di(t)}{dt} = \beta s(t) i(t) - < k > \gamma i(t) \tag{5}$$

$$\frac{dr(t)}{dt} = \gamma i(t) \tag{6}$$

where $< k >$ denotes the average degree of the network. Hereafter, we refer to this model as *average-degree-based SIR* model. Both the above models neglect the heterogeneity of nodes and thus, have not consider the network structure.

B. TOPOLOGY-BASED SIR MODEL

To account for the heterogeneity of the individual road segment in a road network, we adopt the modeling framework proposed in [62], named the N-intertwined epidemic model. The framework employs a continuous-time Markov chain analysis to model spreading behavior. It has been studied and extended in various directions (e.g., [63]; [64]; [60]; [65]; [66]). The framework is general and applicable to different problem domains. It allows us to study how the underlying network structure affects congestion spreading patterns.

In the context of our problem, consider a road network represented by an undirected graph, $G(V, E)$, where V and E are the set of N nodes and L links. In our case, the nodes represent the sensors (e.g., inductive loops, roadside traffic cameras, etc.) deployed at the roads to monitor traffic while links represent road segments connecting neighbouring sensors. Graph, G, can be represented by A, the $N \times N$ adjacency matrix, with $a_{n,m} = 1$ if there exists a link between nodes n and m, and 0 otherwise.

We say a node gets infected when a node previously having free-flowing traffic becomes congested (i.e., a transition from

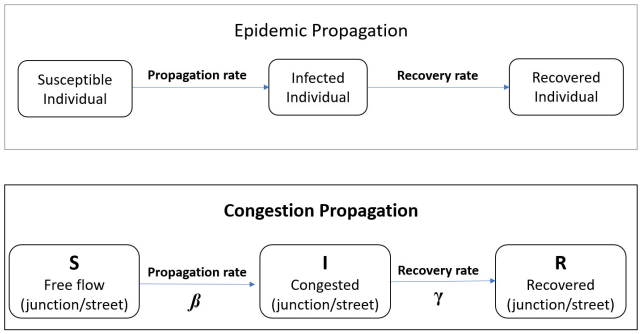


FIGURE 1: Physical mapping of congestion spreading on the SIR model

state S to state I). Consistent with the previous models introduced in Section III-A, this infection process takes place at an average rate of β per unit of time. Similarly, a node is said to have recovered from congestion (i.e., a transition from state I to state R) when the traffic returns to a normal free-flowing state (see Figure 1). Again, we use γ as the average recovery rate per unit of time.

Next, we formally define the notion of "congestion". Given the average vehicle speed recorded at node n between time t and $t-\Delta$, $v_n(t)$ where Δ is the time interval dependent on the dataset (e.g., 5-min) and the speed limit at node n, v_n^{max}, we define the following:

$$\lambda_n(t) = \frac{v_n(t)}{v_n^{max}} \tag{7}$$

as the measure of traffic flow at the node n. A low value of $\lambda_n(t)$ means the vehicle is moving slowly or non-moving if $\lambda_n(t) = 0$. Conversely, the traffic is free-flowing when $\lambda_n(t)$ is high. Node n is considered as congested (i.e., "infected" in the epidemiology terminology) if $\lambda_i(t) < \rho$ where ρ represents different congestion levels [40]. Otherwise, a node is considered not congested (i.e., $\lambda_i(t) \geq \rho$). The number of congested nodes grows as ρ increases. ρ is a tuneable threshold value (cf. Section IV on the impact of different values of ρ).

Let $s_n(t)$, $i_n(t)$, and $r_n(t)$ denote the probability of node n being in the susceptible, infected, and recovered state at time, t, respectively. Since each node can be in one and only one of the three possible states at any one time, then we have:

$$s_n(t) + i_n(t) + r_n(t) = 1 \tag{8}$$

and

$$\frac{ds_n(t)}{dt} + \frac{di_n(t)}{dt} + \frac{dr_n(t)}{dt} = 0. \tag{9}$$

The complexity of the solution by applying the Markov theory directly to the entire network is exponential (i.e., $O(3^N)$) since all possible combinations of states for each and every node have to be considered. This results in the infinitesimal generator of the system, $Q(t)$ to have the dimension of $3^N \times 3^N$. To address this issue, we advocate the use of the N-intertwined epidemic framework ([63], [64], [60], [65], [66]) for the SIR model which approaches the problem by considering each node individually. Now, applying the Markov theory, we will get N infinitesimal generators, $Q_n(t)$ of the three-state continuous Markov chain; one for each node as follows:

$$Q_n(t) = \begin{bmatrix} -q_{1,2;n} & q_{1,2;n} & 0 \\ 0 & -q_{2,3;n} & q_{2,3;n} \\ 0 & 0 & 0 \end{bmatrix} \tag{10}$$

where $q_{1,2;n}$ is dependent on the states of other nodes within the network. One way to account for this dependency is to condition $q_{1,2;n}$ with all possible combinations of states for all nodes. However, this reverts back to the exact Markov chain solution of exponential complexity. Hence, we apply a mean field approximation to the random variable $q_{1,2;n}$ and compute its expected rate. By this, we remove the random nature and it allows us to reduce the complexity of the solution to polynomial ($O(N)$). The effective infinitesimal generator, $\overline{Q_n(t)}$, is then as follows:

$$\overline{Q_n(t)} = \begin{bmatrix} -E[q_{1,2;n}] & E[q_{1,2;n}] & 0 \\ 0 & -\gamma & \gamma \\ 0 & 0 & 0 \end{bmatrix} \tag{11}$$

where $E[q_{1,2;n}] = \sum_{m=1}^{N} a_{m,n} i_n(t)$.

We can now obtain the following system of non-linear governing differential equations for our case,

$$\frac{ds_n(t)}{dt} = -s_n(t)\beta \sum_{m=1}^{N} a_{m,n} i_n(t) \tag{12}$$

$$\frac{di_n(t)}{dt} = s_n(t)\beta \sum_{m=1}^{N} a_{m,n} i_n(t) - \gamma i_n(t) \tag{13}$$

$$\frac{dr_n(t)}{dt} = \gamma i_n(t) \tag{14}$$

Note the explicit consideration on the road network topology via the inclusion of the adjacency matrix elements in the system. This allows us to validate the model against real road networks and congestion data (cf. Section IV). Another advantage of adopting this framework is that we can now consider the state transitions of each node individually, as opposed to the classical methods where the mean aggregate behaviour of all nodes is considered.

We can now solve the above system of differential equations to obtain the instantaneous evolution of the nodes for the three distinct states to discover the congestion spreading patterns of road networks. By additionally using Eqs. 8 and 9, we can reduce the problem from solving $3 \times N$ simultaneous equations to $2 \times N$ equations.

Following the above, we can then obtain the epidemic prevalence (i.e., the fraction of nodes in congested state (infected)) at time t as follows:

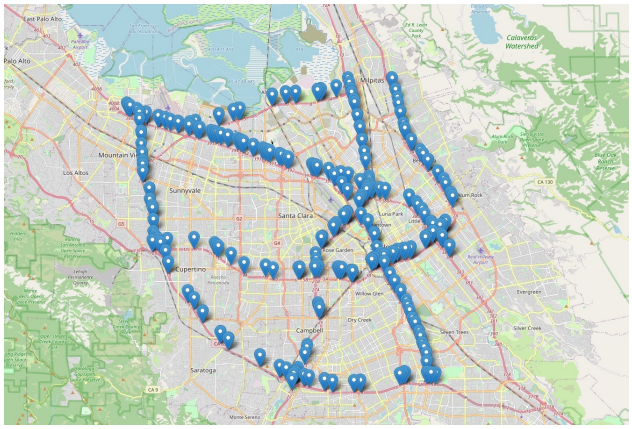


FIGURE 2: Sensor distribution of the PEMS-BAY dataset

$$P(t) = \frac{1}{N}\sum_{n=1}^{N} i_n(t). \tag{15}$$

At steady-state ($\frac{di_n(t)}{dt}|_{t\to\infty} = 0$), $i_{n\infty} \equiv \lim_{t\to\infty} i_n(t) = 0$ and consequently, $\sum_n^N s_n + \sum_n^N r_n = N$.

IV. EVALUATION WITH REAL-WORLD DATA

A. DATA DESCRIPTION

For our study, two real-world datasets, namely PEMS-BAY and METR-LA [67] are used. The PEMS-BAY dataset is collected from the California Transportation Agencies (CalTrans) Performance Measurement System (PeMS) [13]. It recorded traffic speed data from 325 sensor stations in Bay Area. The locations of the sensor stations are shown in Figure 2. The traffic measurements cover a period of six months between the 1st of January and the 30th of June in 2017. The time interval for the data is 5 minutes and the total number of observed traffic data points is 16,937,700 ($= 52,116 \times 325$).

The second real-world dataset, METR-LA, shown in Figure 3, was collected from loop detectors in the highways of Los Angeles County [68]. It includes 207 sensors and covers traffic data for a period of four months from the 1st of March to the 30th of June in 2012. The time interval between data points is also 5 minutes, and the total number of observed traffic data points is 7,094,304 ($= 34,272 \times 207$) with 8.11% missed data points due to sensor incidents.

TABLE 1: Traffic speed characteristics of datasets

Datasets	Max	Min	Mean	Std	Var	Size
PEMS-BAY	85.1	0.0	62.62	8.56	85.41	136 MB
METR-LA	70.0	0.0	53.72	19.2	374.9	57 MB

Table 1 provides a summary of the basic statistics of both datasets including maximum (Max), minimum (Min), mean value (Mean), standard deviation (Std), and variance (Var) of traffic speed data as well as the size of the dataset (in MByte). From the table, it can be noted that METR-LA has higher traffic fluctuations with larger standard deviation and variance than PEMS-BAY.

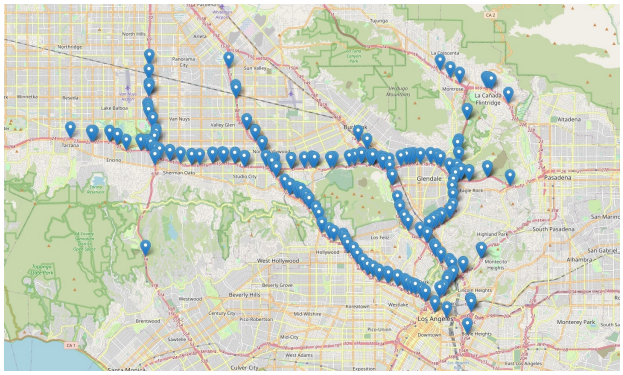


FIGURE 3: Sensor distribution of the METR-LA dataset

Since we advocate the consideration of the actual road topology, we also constructed the adjacency matrix for both datasets based on road connectivity. The basic characteristics of the road network topology derived from their respective adjacency matrices are presented in Table 2.

TABLE 2: Characteristics of adjacency matrices

Datasets	Nodes	Edges	Average node degree
PEMS-BAY	325	398	2.098
METR-LA	207	264	2.058

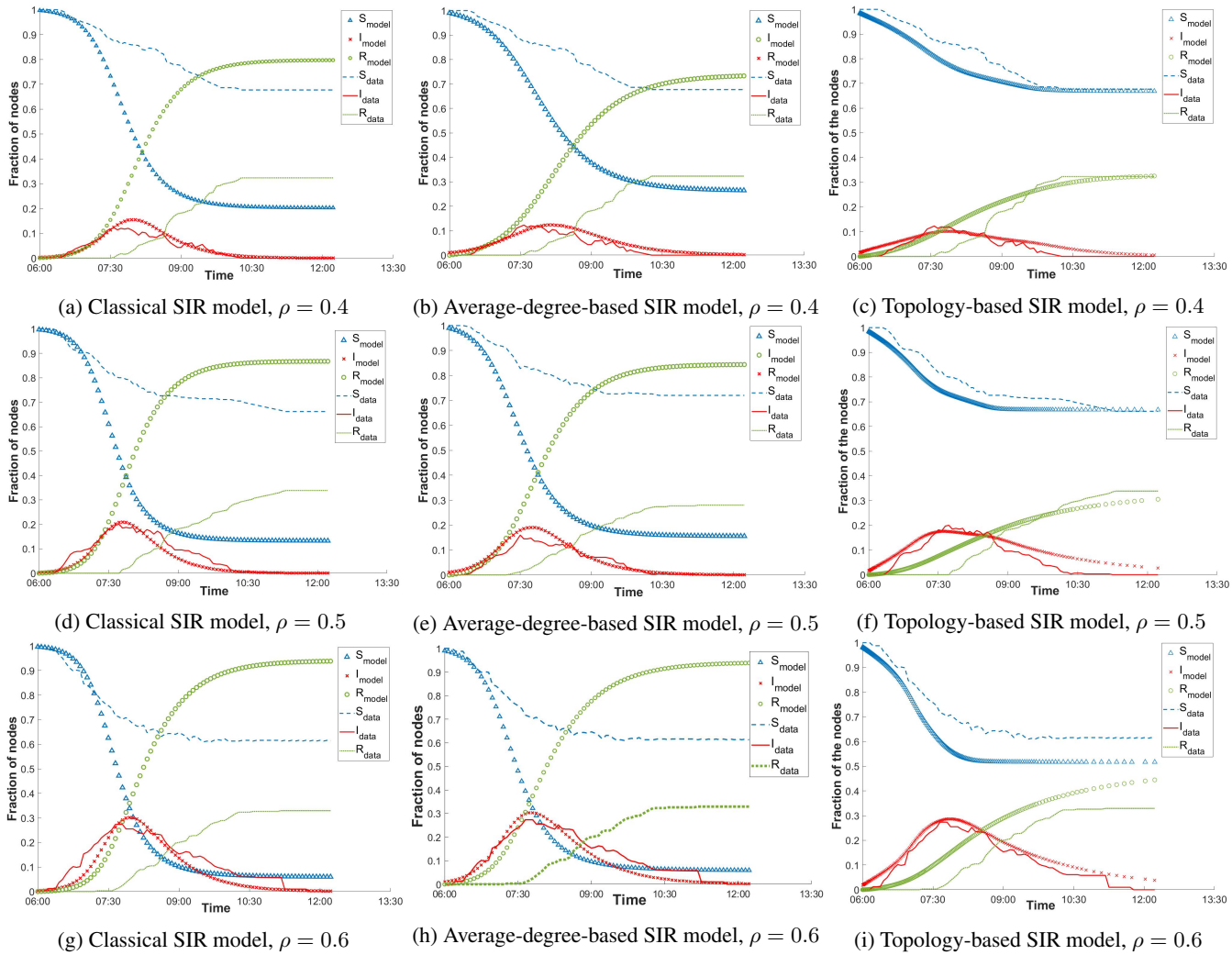
B. PARAMETER ESTIMATIONS

To determine the propagation (β) and dissipation (γ) parameters, we follow [69] and use an ordinary least squares (OLS) approach along with a pattern search algorithm. The goal of the estimation process is to minimise the root-mean-squared error (RMSE) between the observed and modeled $i(t)$ values over the specified study period.

C. TEMPORAL EVOLUTION OF TRAFFIC CONGESTION

We compare the predictive capacity of the three models, i.e., (1) classical SIR, (2) average-degree-based SIR, and (3) topology-based SIR models, over time. For this, we first show how the models track the network-wide congestion state focusing on the early peak hour period (i.e., between 06:00am to 12:00pm) of a workday.

We show in Figure 4 for PEMS-BAY and Figure 5 for METR-LA the fraction of nodes in different states over the specified time period over three different congestion thresholds (i.e., $\rho = 0.4, 0.5, 0.6$). The S_{data}, I_{data}, and R_{data} curves are the actual traffic recorded in the dataset. Specifically, we present the average traffic speed data taken on all Wednesdays across the measurement period as representative observations. The curves are qualitatively similar for other days of the week. On the other hand, the S_{model}, I_{model}, and R_{model} curves are computed by the respective models described in Section III. For both figures, the first column includes the performance of the classical SIR model, the second column incorporates the average-degree-based SIR model's outcomes and the third column presents the results of our topology-based SIR model.



(a) Classical SIR model, $\rho = 0.4$ (b) Average-degree-based SIR model, $\rho = 0.4$ (c) Topology-based SIR model, $\rho = 0.4$

(d) Classical SIR model, $\rho = 0.5$ (e) Average-degree-based SIR model, $\rho = 0.5$ (f) Topology-based SIR model, $\rho = 0.5$

(g) Classical SIR model, $\rho = 0.6$ (h) Average-degree-based SIR model, $\rho = 0.6$ (i) Topology-based SIR model, $\rho = 0.6$

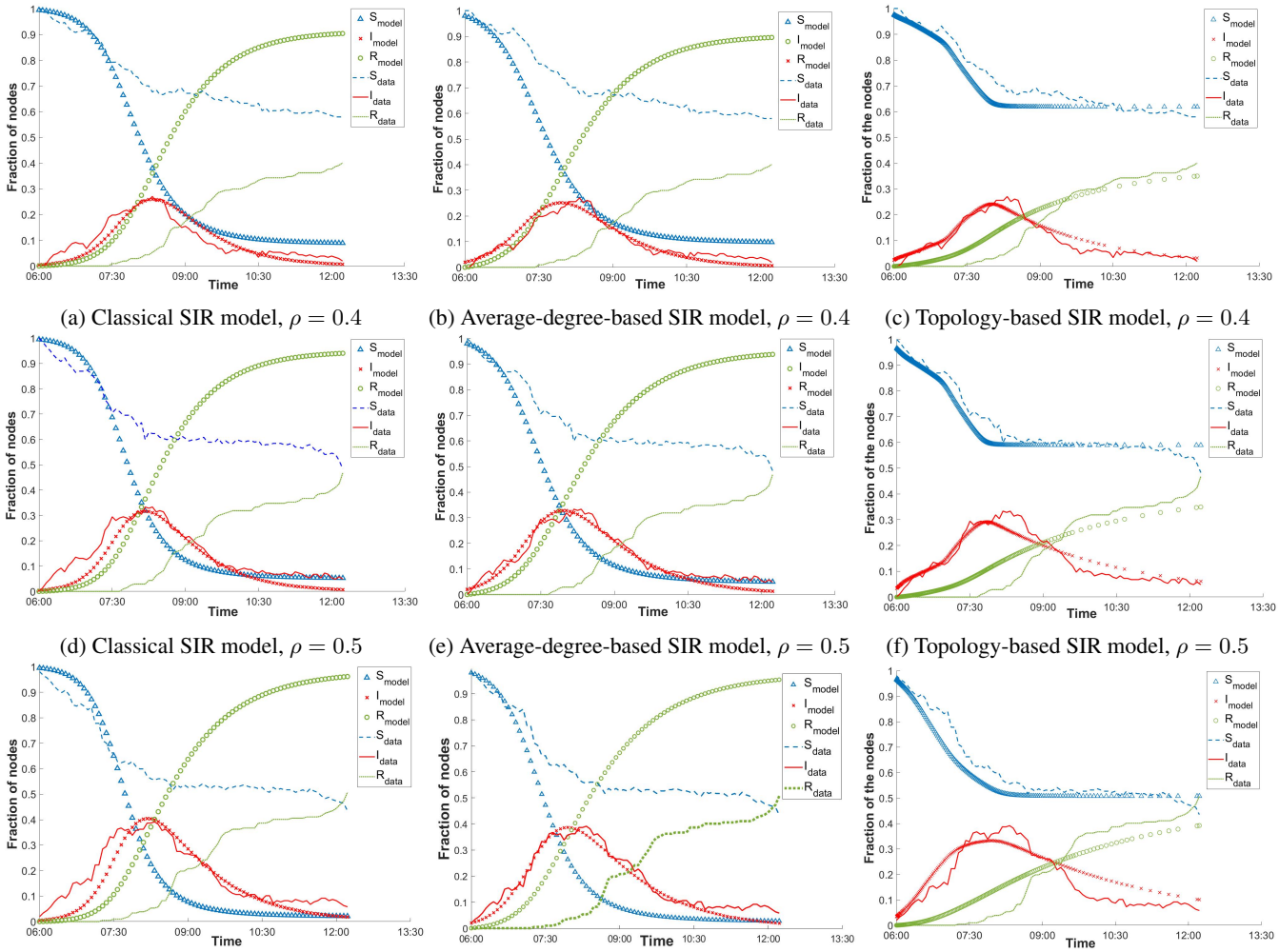
FIGURE 4: Models vs Data – Time evolution of node states for $\rho = 0.4, 0.5, 0.6$ on the PEMS-BAY dataset.

From the figures, we see that, in comparison, the proposed topology-based SIR model captures most accurately all three traffic states over the entire period and can generally track the evolution of the congestion closely for the different ρ considered as well as across both road networks. Taking PEMS-BAY network at $\rho = 0.4$ as example, the real traffic data indicates that there are approximately 70% of nodes in the network that never suffered traffic congestion (conversely, the remaining 30% have recovered from congestion at the end of the peak period). The classical SIR model, however, predicted only 20% of nodes avoided congestion during this period (i.e., a 50% discrepancy from the ground truth). The average-degree-based SIR model performed slightly better, computed about 30% of nodes not congested. The topology-based SIR model advocated here managed to correctly predict the outcome at 70% of nodes. Further, despite having higher traffic volatility in the METR-LA dataset, our topology-based SIR model can still track the congestion state closely.

By definition, when ρ is small, fewer nodes are considered to be congested (i.e., vehicles must be traveling at a lower speed or being stationary at the junction to be classified as congested). We see this across both road networks. All three models correctly capture this phenomena (i.e., higher peak for I_{model} when ρ is higher). However, by not explicitly considering the network topology, both classical and average-degree-based SIR models showed higher node recovery, resulting in disproportionately inflating the number of congested nodes over the considered time period.

Comparing the two cities, despite having different road topology and traffic demand patterns, they have a similar number of congested nodes at peak for both ρ = 0.4 and 0.5 (see Figures 4 and 5). For a large ρ (ρ = 0.6), the observed difference between the data and the model grows. This is mainly due to the looser definition of congestion when ρ is large, more nodes are considered congested even though they are relatively still flowing at a decent speed. Therefore, we expect to see a divergence between cities as ρ increases.

We show in Figure 6 (a)-(b) the fraction of nodes recovered from congestion at the end of the observation period across different ρ for PEMS-BAY and METR-LA. In both networks,



(a) Classical SIR model, $\rho = 0.4$ (b) Average-degree-based SIR model, $\rho = 0.4$ (c) Topology-based SIR model, $\rho = 0.4$

(d) Classical SIR model, $\rho = 0.5$ (e) Average-degree-based SIR model, $\rho = 0.5$ (f) Topology-based SIR model, $\rho = 0.5$

(g) Classical SIR model, $\rho = 0.6$ (h) Average-degree-based SIR model, $\rho = 0.6$ (i) Topology-based SIR model, $\rho = 0.6$

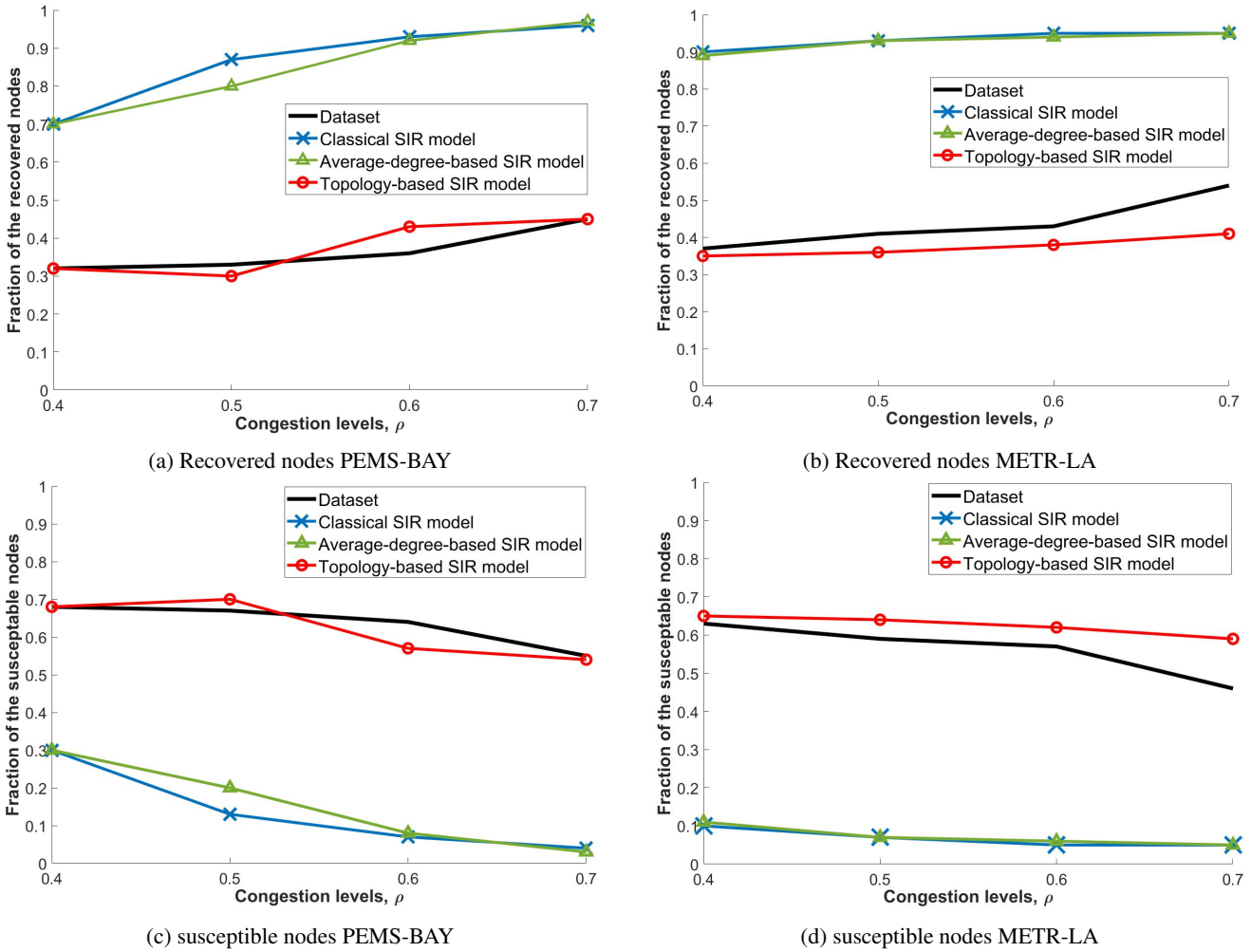
FIGURE 5: Model vs Data – Time evolution of node states for $\rho = 0.4, 0.5, 0.6$ on the METR-LA dataset.

we see the topology-based SIR model closely follows the ground truth. The other two models, which have disregarded the topology, have failed to correctly predict the congestion state. Both the classical and average-degree-based SIR models have made similar predictions. The discrepancies with the ground truth are significant. For instance, for PEMS-BAY, the average difference between the actual traffic conditions and that computed by the topology-based SIR model is 0.025. In contrast, the differences for classical and average-degree-based SIR models are 0.5 and 0.48 respectively. Interestingly, for PEMS-BAY, we see that the model predicted a higher number of recovered nodes despite predicting a lower peak (cf. Figure 4). This is because the model computed that the congestion spreads to more nodes but is distributed more evenly over the period; thus, resulting in the peak being not as high but having more congested nodes. Our work further corroborates the literature (e.g., [58]; [14]; [59]; [7]) that traffic congestion spreading mimics the contagion process of an epidemic.

We correspondingly show in Figure 6 (c)-(d) the fraction of nodes that never get congested across different ρ for both networks. Similar insights could be found here whereby both classical and average-degree-based SIR models unable to track the ground truth while our topology-based SIR model made successful predictions (the average discrepancies are 0.0275 and 0.07 for PEMS-BAY and METR-LA respectively). In addition, we also note that the discrepancies for our topology-based SIR model are higher in METR-LA than in PEMS-BAY. This is due to the fact that traffic in METR-LA has higher volatility (as indicated in Table 1).

D. SPATIO-TEMPORAL EVOLUTION OF CONGESTION

One feature of the topology-based SIR model is the ability to compute the state of each node individually. As such, we can create congestion map over time to show spatially the spread of traffic congestion. This is not possible for both the classical and average-degree-based SIR models since they treat the network as a whole and compute the aggregated fraction of



(a) Recovered nodes PEMS-BAY

(b) Recovered nodes METR-LA

(c) susceptible nodes PEMS-BAY

(d) susceptible nodes METR-LA

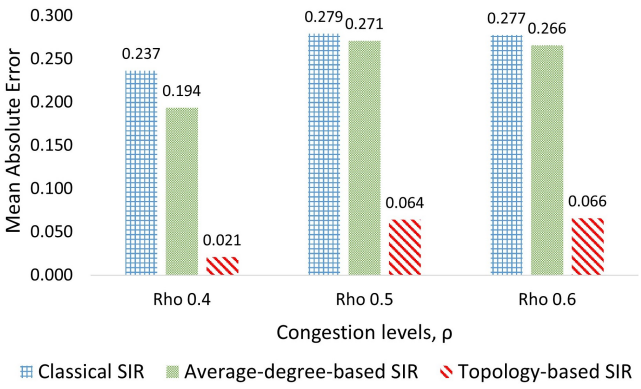
FIGURE 6: (a)-(b) fraction of the recovered nodes from datasets and models at $\rho = 0.4 - 0.7$ for PEMS-BAY and METR-LA, (c)-(d) fraction of susceptible nodes from data and models at $\rho = 0.4 - 0.7$ for PEMS-BAY and METR-LA respectively.

nodes in each state.

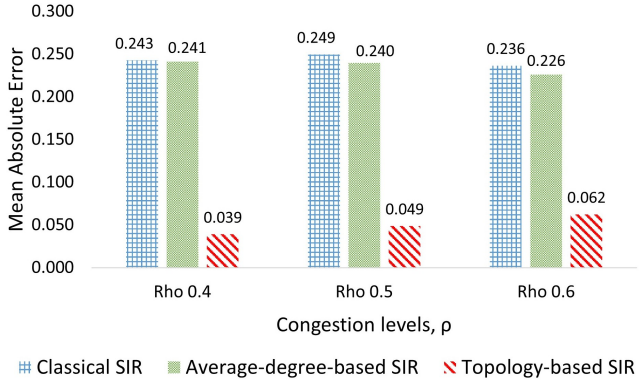
In the following, we present the mean absolute error (MAE) and root mean square error (RMSE) comparing the models against the ground truth in Figure 7 (a)-(b) and Figure 7 (c)-(d) respectively. From the figures, we can see that the average-degree-based SIR model marginally performs better than the classical SIR model but our topology-based SIR model achieves significantly lower errors against the other two models.

Finally, we illustrate the spatio-temporal evolution of traffic congestion by showing how the state of the nodes is distributed geographically on the map. We use a series of hourly congestion maps comparing the actual congestion (derived from the dataset) and that computed by the topology-based SIR model in Figure 8 and Figure 9 for PEMS-BAY and METR-LA respectively. For both cases, the initial congested nodes are based on the dataset (seed nodes depicted in black in the maps). We set the same seed nodes for the model. In the figures, nodes in green indicate the free flow traffic state, nodes in red indicate congested locations and nodes in yellow indicate locations that have recovered from the congestion state during the considered period.

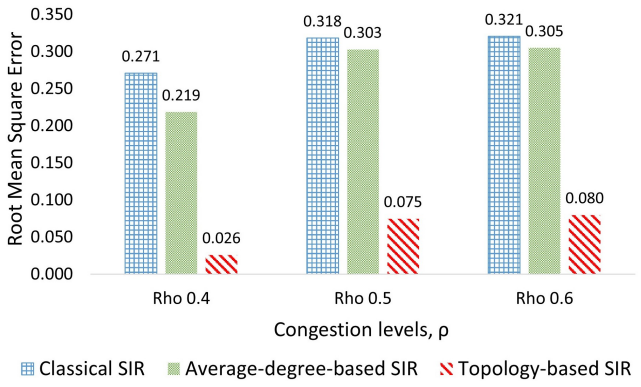
In general, we can see that the model offers a good approximation of the conditions of the traffic at different areas or stretches of roads. For PEMS-BAY, we can see that at 6am, the network is largely free flowing but by 7am, several roads are already congested (e.g., top right corner). Beginning from 8am, the congestion gradually dissipates (more nodes in yellow appearing). These are all captured by the model (see the corresponding hourly maps Figure 8 (b), (d), (f), (h) and (j)). For METR-LA, the data showed a slower build-up of congestion in which the network stayed relatively free-flowing between 6am-7am and congestion only gradually appearing in the later hours. We can see from the counterpart figures computed by the model equally up to the task in capturing this spreading. Therefore, the model can provide a good overall estimation of the traffic conditions in terms of geographical areas but we also note that there are some inaccuracies when specific nodes are of interest.



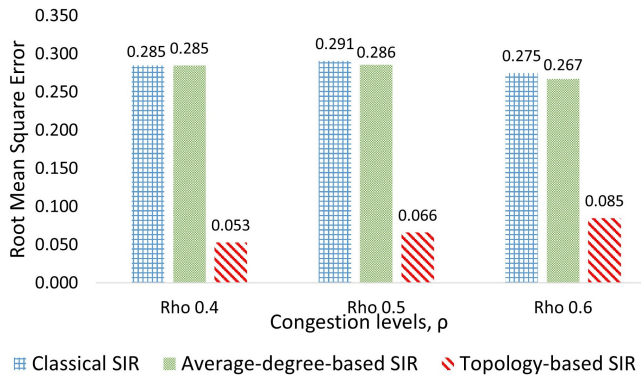
(a) PEMS-BAY



(b) METR-LA

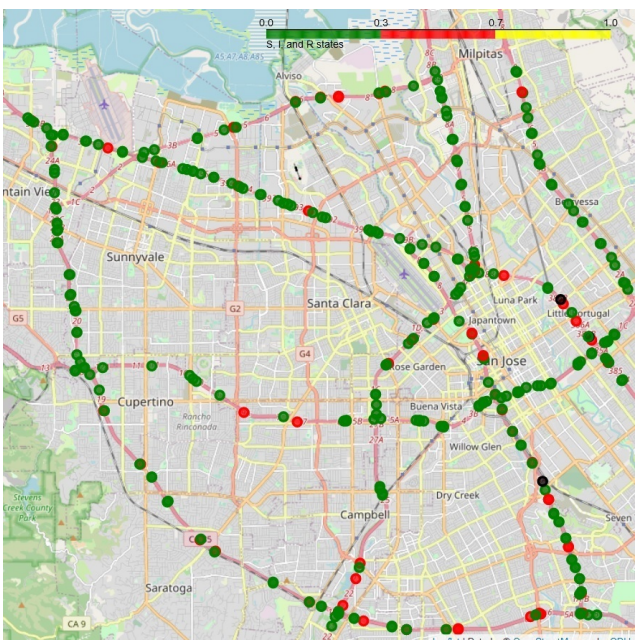


(c) PEMS-BAY

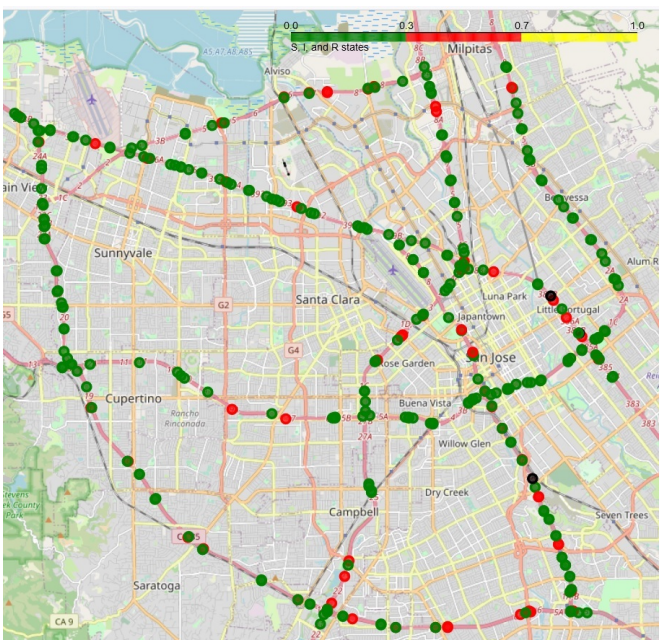


(d) METR-LA

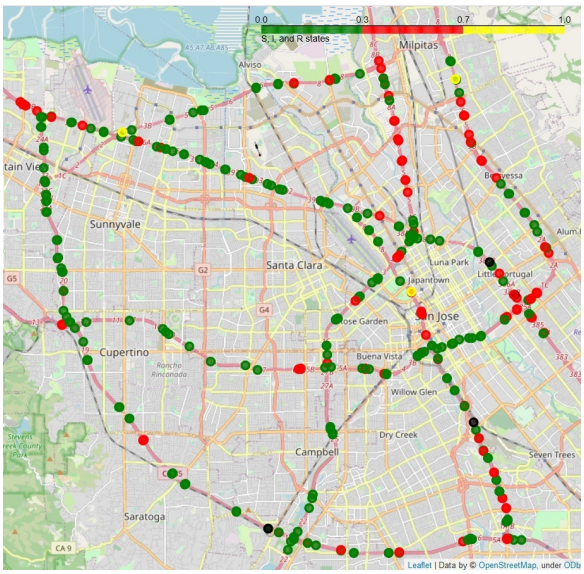
FIGURE 7: Mean Absolute Error (MAE) of three models: (a) PEMS-BAY, (b) METR-LA and Root Mean Square Error (RMSE) of three models: (a) PEMS-BAY, (b) METR-LA.



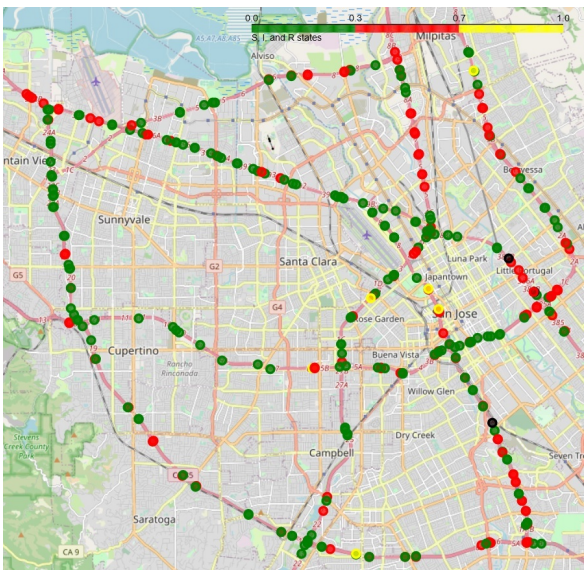
(a) Congestion Map from the dataset 06:00 am



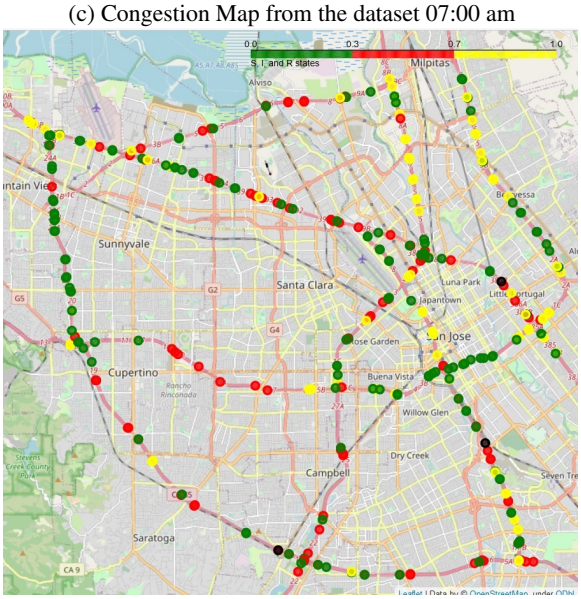
(b) Congestion Map based on the model 06:00 am



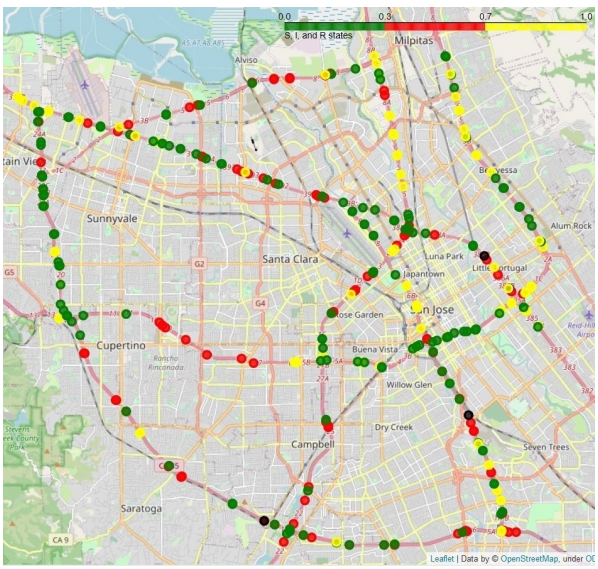
(c) Congestion Map from the dataset 07:00 am



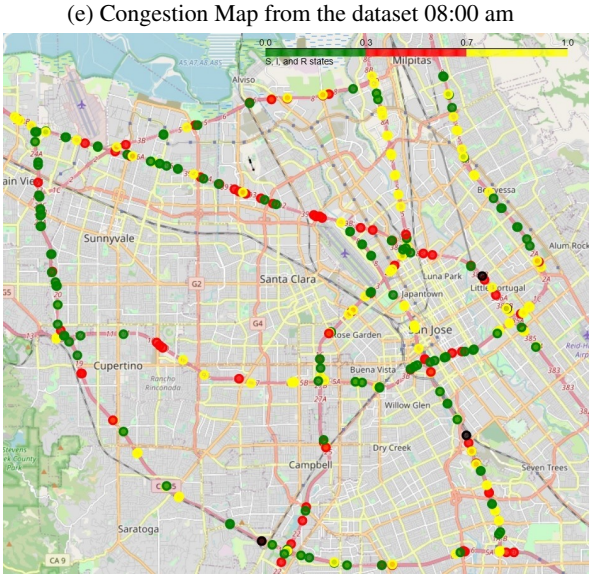
(d) Congestion Map based on the model 07:00 am



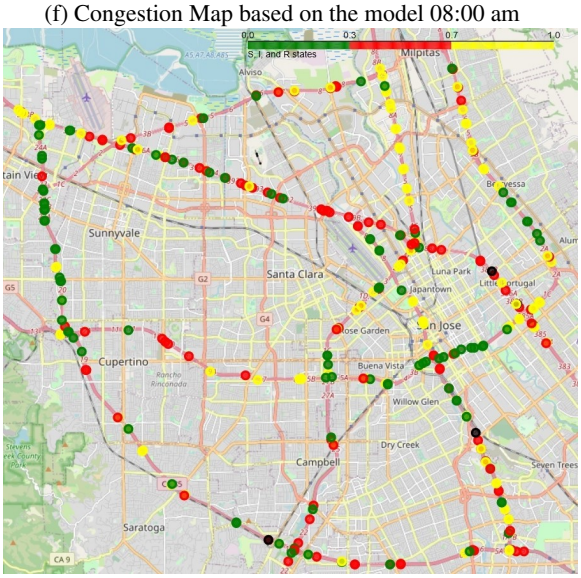
(e) Congestion Map from the dataset 08:00 am



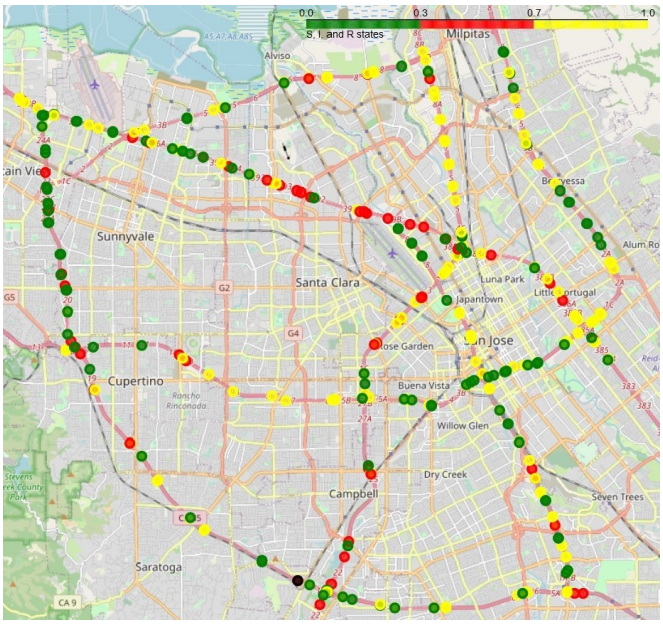
(f) Congestion Map based on the model 08:00 am



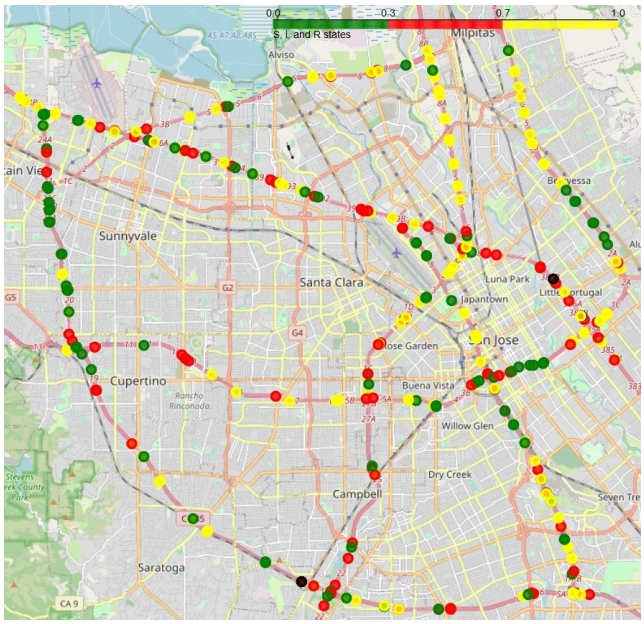
(g) Congestion Map from the dataset 09:00 am



(h) Congestion Map based on the model 09:00 am

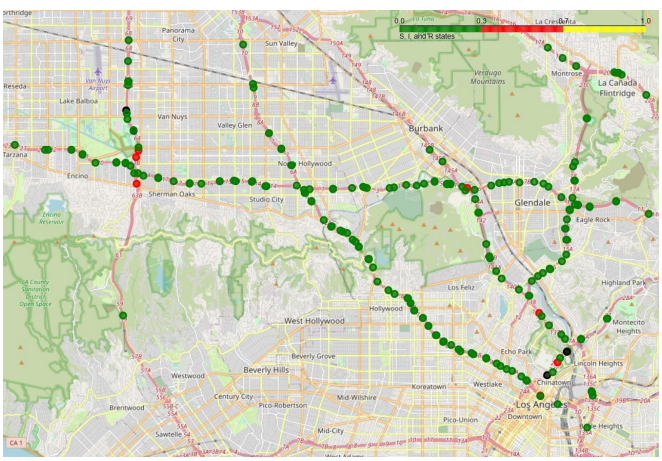


(i) Congestion Map from the dataset 10:00 am

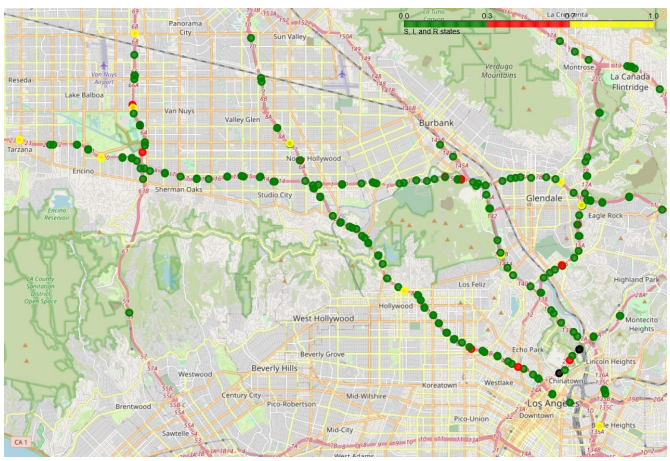


(j) Congestion Map based on the model 10:00 am

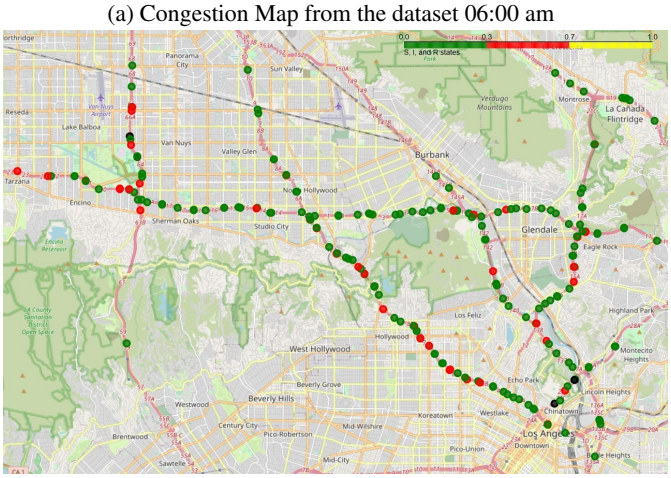
FIGURE 8: Hourly congestion map based on dataset and the model for PEMS-BAY. The first column represents dataset and the second column illustrates output of the model.



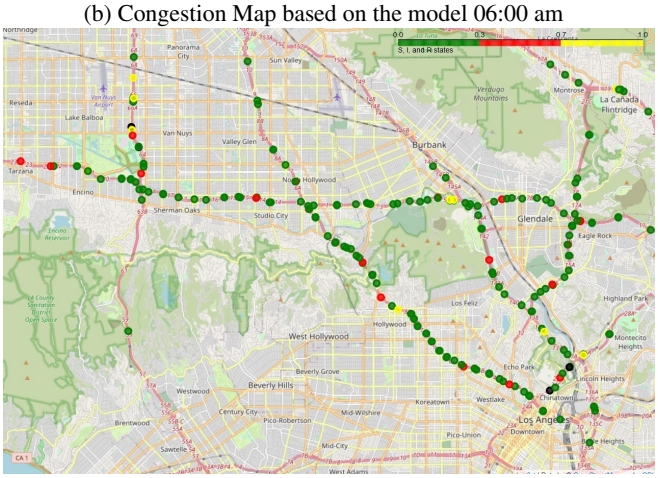
(a) Congestion Map from the dataset 06:00 am



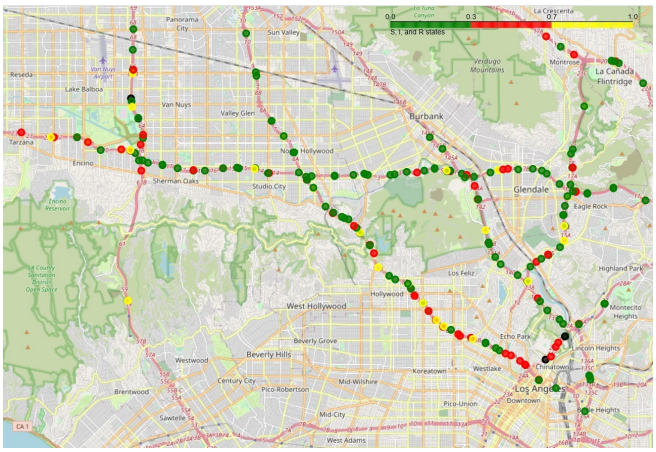
(b) Congestion Map based on the model 06:00 am



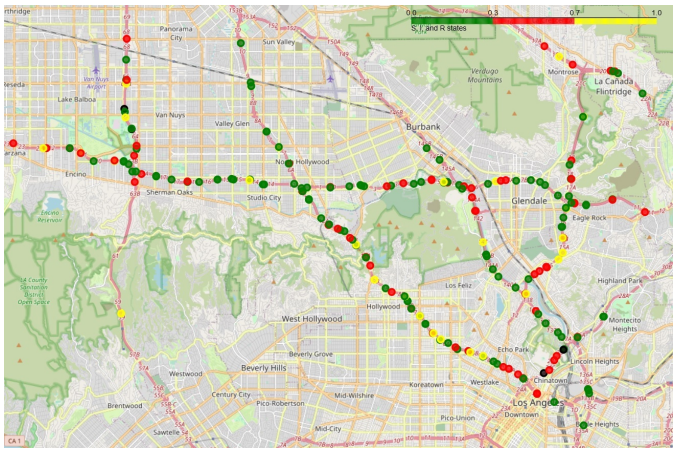
(c) Congestion Map from the dataset 07:00 am



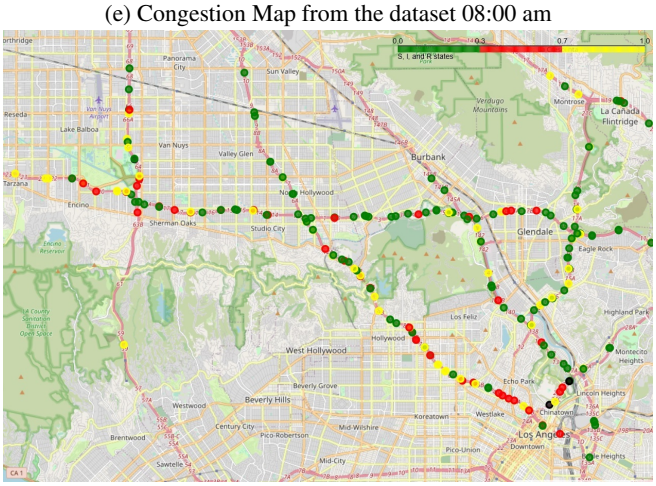
(d) Congestion Map based on the model 07:00 am



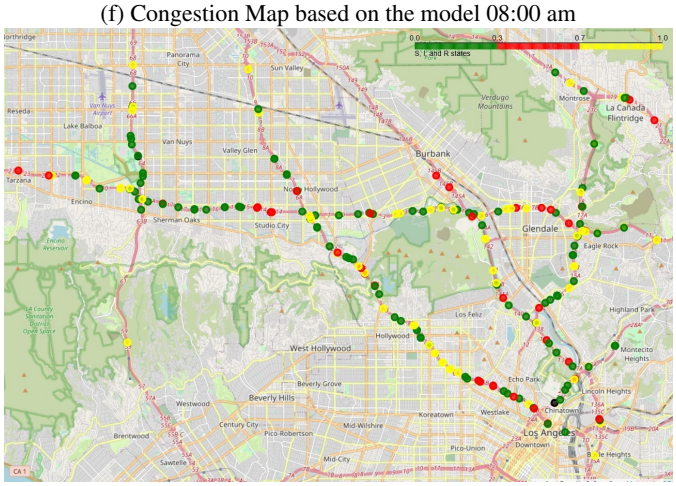
(e) Congestion Map from the dataset 08:00 am



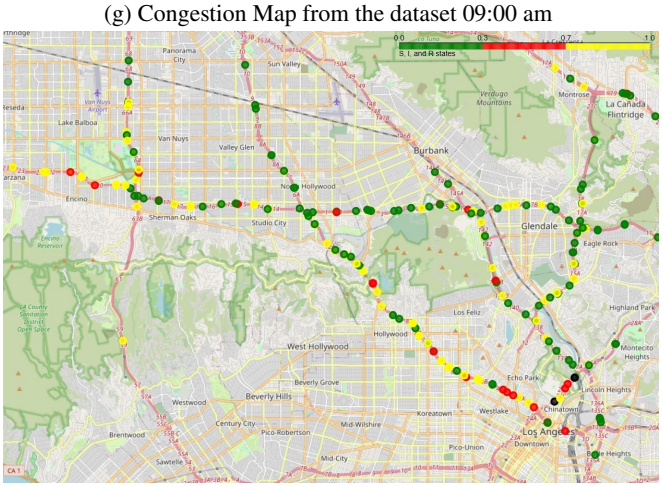
(f) Congestion Map based on the model 08:00 am



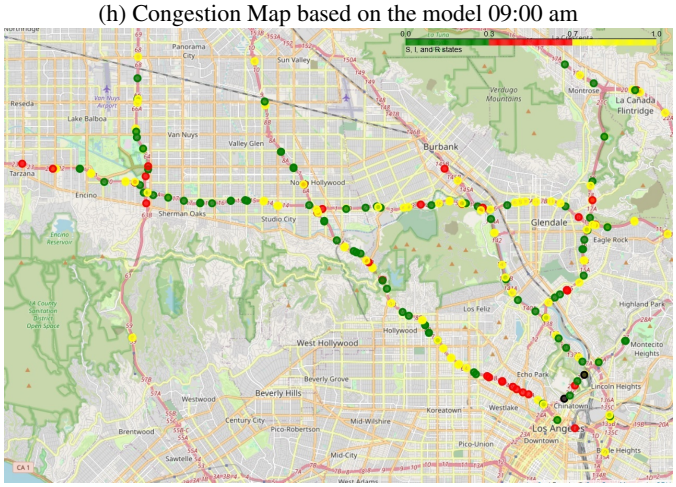
(g) Congestion Map from the dataset 09:00 am



(h) Congestion Map based on the model 09:00 am



(i) Congestion Map from the dataset 10:00 am



(j) Congestion Map based on the model 10:00 am

FIGURE 9: Hourly congestion map based on the dataset and the model for METR-LA. The first column represents dataset and the second column illustrates output of the model.

V. CONCLUSIONS

In this paper, we first advocate the use of the epidemic model for modeling the spreading of traffic congestion. Despite the complex nature of urban traffic, we show in this work that the contagion-like process of traffic congestion can be modeled using the SIR epidemic model that includes propagation and dissipation characteristics of traffic dependent on time-varying travel demand. Departing from the classical and average-degree-based SIR models, which mostly consider homogeneous mixing based on the law of mass action ([70]; [61]), we propose the explicit consideration of the road network structure via the use of the corresponding adjacency matrix that represents the network topology.

In this work, we tested the model against traffic data from two cities, namely California and Los Angeles, and compared the performance of the model against the classical SIR and average-degree-based SIR models. We show that the proposed topology-based SIR model outperforms these two SIR models by a significant margin. Our results also demonstrate better agreement, both temporally and spatially, between the model and data despite the two cities having different traffic profiles. We also illustrate the congestion spreading spatially over time; a feature of our model that the other models are incapable of computing. Our model can be applied to develop control strategies with different objectives, such as minimizing the total duration of congestion, minimizing the total number of congested roads and junctions, and minimizing the recovery time.

ACKNOWLEDGMENT

The authors are grateful to the Bournemouth Christchurch Poole (BCP) Council.

REFERENCES

[1] J. C. Falcocchio and H. S. Levinson. *Road traffic congestion: a concise guide*, volume 7. Springer, 2015.
[2] UK Department for Transport. National road traffic projections, 2022.
[3] Transport for London. Congestion charge marks 20 years of keeping london moving sustainably, 2023.
[4] Chinese Ministry of Transport. Guidelines for the construction of urban public transportation smart application demonstration projects, 2014.
[5] Taipei City Traffic Engineering Office. Traffic lights get smarter and driving gets more efficient, 2019.
[6] Boston Consulting Group. Barcelona urban lab: using the city as a testing ground for innovation.
[7] M. Saberi, H. Hamedmoghadam, M. Ashfaq, S.A. Hosseini, Z. Gu, S. Shafiei, D.J. Nair, V. Dixit, L. Gardner, S.T. Waller, et al. A simple contagion process describes spreading of traffic jams in urban networks. *Nature Communications*, 11(1):1616, 2020.
[8] A.M. Nagy and V. Simon. Traffic congestion propagation identification method in smart cities. *Infocommunications Journal*, 13(1):45–57, 2021.
[9] A. Abdi, M. Saffarzadeh, and A. Salehikalam. Identifying and analyzing stop and go traffic based on asymmetric theory of driving behavior in acceleration and deceleration. *International Journal of Transportation Engineering*, 3(4):237–251, 2016.
[10] S.T Zheng, R. Jiang, J. Tian, X. Li, B. Jia, Z. Gao, and S. Yu. A comparison study on the growth pattern of traffic oscillations in car-following experiments. *Transportmetrica B: Transport Dynamics*, 11(1):706–724, 2023.
[11] C. Li, W. Yue, G. Mao, and Z. Xu. Congestion propagation based bottleneck identification in urban road networks. *IEEE Transactions on Vehicular Technology*, 69(5):4827–4841, 2020.
[12] J.A. Laval and L. Leclercq. A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928):4519–4541, 2010.
[13] C. Chen. *Freeway performance measurement system (PeMS)*. University of California, Berkeley, 2002.
[14] J.J. Wu, Z.Y. Gao, H.J. Sun, and H.J. Huang. Congestion in different topologies of traffic networks. *Europhysics letters*, 74(3):560, 2006.
[15] R.M. Anderson and R.M. May. *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1991.
[16] H.U. Ahmed, Y. Huang, and P. Lu. A review of car-following models and modeling tools for human and autonomous-ready driving behaviors in micro-simulation. *Smart Cities*, 4(1):314–335, 2021.
[17] L.A. Pipes. An operational analysis of traffic dynamics. *Journal of Applied Physics*, 24(3):274–281, 1953.
[18] M. Brackstone and M. McDonald. Car-following: a historical review. *Transportation Research Part F: Traffic Psychology and Behaviour*, 2(4):181–196, 1999.
[19] M. Saifuzzaman and Z. Zheng. Incorporating human-factors in car-following models: a review of recent developments and research needs. *Transportation Research Part C: Rmerging Technologies*, 48:379–403, 2014.
[20] M. Rahman, M. Chowdhury, Y. Xie, and Y. He. Review of microscopic lane-changing models and future research opportunities. *IEEE Transactions on intelligent transportation systems*, 14(4):1942–1956, 2013.
[21] R. Akçelik. A review of gap-acceptance capacity models. In *Conference of Australian Institutes of Transport Research (CAITR), 29th, 2007, Adelaida, South Australia, Australia*, 2007.
[22] W.S. Vickrey. Congestion theory and transport investment. *The American Economic Review*, pages 251–260, 1969.
[23] R. Arnott. A bathtub model of downtown traffic congestion. *Journal of Urban Economics*, 76:110–121, 2013.
[24] J. Zhu, L.Hu, Z.Chen, and H. Xie. A queuing model for mixed traffic flows on highways considering fluctuations in traffic demand. *Journal of Advanced Transportation*, 2022, 2022.
[25] M.J. Lighthill and G.B. Whitham. On kinematic waves ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. series a. mathematical and physical sciences*, 229(1178):317–345, 1955.
[26] P.I. Richards. Shock waves on the highway. *Operations research*, 4(1):42–51, 1956.
[27] G.F. Newell. A simplified theory of kinematic waves in highway traffic, part ii: Queueing at freeway bottlenecks. *Transportation Research Part B: Methodological*, 27(4):289–303, 1993.
[28] M. Kurant and P. Thiran. Layered complex networks. *Physical Review Letters*, 96(13):138701, 2006.
[29] L.E. Olmos and J.D. Muñoz. Traffic gridlock on a honeycomb city. *Physical Review E*, 95(3):032320, 2017.
[30] V. Verbavatz and M. Barthelemy. The growth equation of cities. *Nature*, 587(7834):397–401, 2020.
[31] D. Levinson. Network structure and city size. *PloS one*, 7(1):e29721, 2012.
[32] B. Piccoli and M. Garavello. Traffic flow on networks. *American Institute of Mathematical Sciences*, 2006.
[33] A.L. Barabási. *Network Science Spreading Phenomena*. Cambridge University Press, 2016.
[34] H.J. Payne. (1971) models of freeway traffic and control. *Mathematical Models of Public Systems*, 1:51, 1971.
[35] G.B. Whitham. Linear and nonlinear waves(book). *New York, Wiley-Interscience, 1974. 651 p*, 1974.
[36] D. Helbing. Gas-kinetic derivation of navier-stokes-like traffic equations. *Physical Review E*, 53(3):2366, 1996.
[37] N. Vandaele, T. Van Woensel, and A. Verbruggen. A queueing-based traffic flow model. *Transportation Research Part D: Transport and Environment*, 5(2):121–135, 2000.
[38] H.S. Mahmassani, M. Saberi, and A. Zockaie. Urban network gridlock: Theory, characteristics, and dynamics. *Procedia-Social and Behavioral Sciences*, 80:79–98, 2013.
[39] G. Zeng, D. Li, S. Guo, L. Gao, Z. Gao, H.E. Stanley, and S. Havlin. Switch between critical percolation modes in city traffic dynamics. *Proceedings of the National Academy of Sciences*, 116(1):23–28, 2019.
[40] D. Li, B. Fu, Y. Wang, G. Lu, Y. Berezin, H.E. Stanley, and S. Havlin. Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proceedings of the National Academy of Sciences*, 112(3):669–672, 2015.

[41] B. Anbaroglu, B. Heydecker, and T. Cheng. Spatio-temporal clustering for non-recurrent traffic congestion detection on urban road networks. *Transportation Research Part C: Emerging Technologies*, 48:47–65, 2014.

[42] G. Zheng, W.K. Chai, J.K. Duanmu, and V. Katos. Hybrid deep learning models for traffic prediction in large-scale road networks. *Information Fusion*, 92:93–114, 2023.

[43] E. Ko, J. Ahn, and E.Y. Kim. 3d markov process for traffic flow prediction in real-time. *Sensors*, 16(2):147, 2016.

[44] G.B. Huang. An insight into extreme learning machines: random neurons, random features and kernels. *Cognitive Computation*, 6:376–390, 2014.

[45] D. Zhu, H. Du, Y. Sun, and N. Cao. Research on path planning model based on short-term traffic flow prediction in intelligent transportation system. *Sensors*, 18(12):4275, 2018.

[46] R. Fu, Z. Zhang, and L. Li. Using lstm and gru neural network methods for traffic flow prediction. In *31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pages 324–328. IEEE, 2016.

[47] R. Toncharoen and M. Piantanakulchai. Traffic state prediction using convolutional neural network. In *2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pages 1–6. IEEE, 2018.

[48] F. Diehl, T. Brunner, M.T. Le, and A. Knoll. Graph neural networks for modelling traffic participant interaction. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 695–701. IEEE, 2019.

[49] J.F. Zheng, Z.Y. Gao, and X. Zhao. Modeling cascading failures in congested complex networks. *Physica A: Statistical Mechanics and its Applications*, 385(2):700–706, 2007.

[50] Y. Chen, J. Mao, Z. Zhang, H. Huang, W. Lu, Q. Yan, and L. Liu. A quasi-contagion process modeling and characteristic analysis for real-world urban traffic network congestion patterns. *Physica A: Statistical Mechanics and its Applications*, 603:127729, 2022.

[51] Y. Li, L. Zhao, Z. Yu, and S. Wang. Traffic flow prediction with big data: A learning approach based on sis-complex networks. In *2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pages 550–554. IEEE, 2017.

[52] Y. Zhang, S. Li, J. Zhang, J. Ma, and H. An. Traffic-driven si epidemic spreading on scale-free networks. *International Journal of Modern Physics C*, 33(08):2250111, 2022.

[53] S. Meloni, A. Arenas, and Y. Moreno. Traffic-driven epidemic spreading in finite-size scale-free networks. *Proceedings of the National Academy of Sciences*, 106(40):16897–16902, 2009.

[54] Li H., Cheng X., and Liu J. Understanding video sharing propagation in social networks: Measurement and analysis. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 10(4):1–20, 2014.

[55] Y. Iturria-Medina, R.C. Sotero, P.J. Toussaint, A.C. Evans, and Alzheimer's Disease Neuroimaging Initiative. Epidemic spreading model to characterize misfolded proteins propagation in aging and associated neurodegenerative disorders. *PLoS computational biology*, 10(11):e1003956, 2014.

[56] V. Colizza, A. Barrat, M. Barthelemy, A.J. Valleron, and A. Vespignani. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS medicine*, 4(1):e13, 2007.

[57] Wu J., Gao Z., and Sun H. Simulation of traffic congestion with sir model. *Modern Physics Letters B*, 18(30):1537–1542, 2004.

[58] B. Jiang and C. Claramunt. Topological analysis of urban street networks. *Environment and Planning B: Planning and design*, 31(1):151–162, 2004.

[59] D. Brockmann and D. Helbing. The hidden geometry of complex, network-driven contagion phenomena. *science*, 342(6164):1337–1342, 2013.

[60] F.D. Sahneh, C. Scoglio, and P. Van Mieghem. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Networking*, 21(5):1609–1620, 2013.

[61] W.O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[62] P. Van Mieghem, J. Omic, and R. Kooij. Virus spread in networks. *IEEE/ACM Transactions On Networking*, 17(1):1–14, 2008.

[63] P. Van Mieghem. The n-intertwined sis epidemic network model. *Computing*, 93(2-4):147–169, 2011.

[64] M. Youssef and C. Scoglio. An individual-based approach to sir epidemics in contact networks. *Journal of Theoretical Biology*, 283(1):136–144, 2011.

[65] W.K. Chai and G. Pavlou. Path-based epidemic spreading in networks. *IEEE/ACM Transactions on Networking*, 25(1):565–578, 2017.

[66] W.K. Chai. Modelling spreading process induced by agent mobility in complex networks. *IEEE Transactions on Network Science and Engineering*, 5(4):336–349, 2017.

[67] Li Y., Yu R., C. Shahabi, and Liu Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017.

[68] H.V. Jagadish, J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J.M. Patel, R. Ramakrishnan, and C. Shahabi. Big data and its technical challenges. *Communications of the ACM*, 57(7):86–94, 2014.

[69] T.T.Marinov, R.S. Marinova, J. Omojola, and M. Jackson. Inverse problem for coefficient identification in sir epidemic models. *Computers and Mathematics with Applications*, 67(12):2218–2227, 2014.

[70] D.J. Daley and J. Gani. *Epidemic modelling: an introduction*, volume 15. Cambridge University Press, 2001.

ASSEMGUL KOZHABEK received a master's degree in engineering business management from the University of Warwick, UK, in 2016. She is currently a Ph.D. candidate at Bournemouth University (BU). Assemgul received a match-funded scholarship from the Bournemouth Christchurch Poole (BCP) Council and Bournemouth University, UK. She received her MSc degree in Engineering Business Management from the University of Warwick, UK. Assemgul holds BSc degree in Mathematical and Computer Modelling from AL-Farabi KazNU, KZ. Prior to joining BU, she was working as teaching and research assistant at Nazarbayev University, Kz. Assemgul is an OpenBright grant awardee for women in computer science. Her research interests are network science, transportation planning and management.

DR. WEI KOONG CHAI holds the position of Deputy Head of Department in the Department of Computing and Informatics at Bournemouth University (BU), UK. He is responsible for leading the Future and Complex Networks Research Group (FlexNet) within the department. Additionally, he serves as a visiting academic in the Department of Electronic and Electrical Engineering at University College London (UCL), UK. Previously, he worked as a Senior Research Associate at UCL, where he led research projects, collaborated with other researchers, and contributed to teaching and supervising undergraduate and postgraduate students. He has successfully secured research funding from both EU and UK funding bodies. His research interests include information-centric networking (ICN), communications in cyber-physical systems, and network science.

DR. GE ZHENG. currently is a Research Associate working for the Manufacturing Analytics Group (MAG) of the Distributed Information and Automation Laboratory (DIAL) at the Institute for Manufacturing (IfM), Department of Engineering, University of Cambridge, UK. Before moving to Cambridge, Ge did her PhD with Dr. Wei Koong Chai and Professor Vasilis Katos in the Department of Computing and Informatics at Bournemouth University. Her research interests include predictive systems for supply chain operations, pattern recognition and classification, intelligent transportation systems, and healthcare applications.