Transportation Research Part B 131 (2020) 160–190

Contents lists available at ScienceDirect

Transportation Research Part B

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

An epidemiological diffusion framework for vehicular messaging in general transportation networks

Jungyeol Kim[a,*], Saswati Sarkar[a], Santosh S. Venkatesh[a], Megan Smirti Ryerson[a,b], David Starobinski[c]

[a] Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, United States
[b] Department of City and Regional Planning, University of Pennsylvania, Philadelphia, PA 19104, United States
[c] Division of Systems Engineering, Boston University, Boston, MA, 02215, United States

ARTICLE INFO

Article history:
Received 16 December 2018
Revised 20 November 2019
Accepted 22 November 2019

Keywords:
V2V communication
Mobility
Information propagation
Transportation network
Shared transportation
Trajectory data

ABSTRACT

Emerging Vehicle-to-Vehicle (V2V) technologies are expected to significantly contribute to the safety and growth of shared transportation provided challenges towards their deployment can be overcome. This paper focuses on one such challenge: characterizing the fraction of vehicles which have received a message, as a function of space and time, and operating under different traffic and communication conditions. V2V technologies bridge two infrastructures: communication and transportation. These infrastructures are interconnected and interdependent. To capture this inter-dependence, which may vary in time and space, we propose a new methodology for modeling information propagation between V2V-enabled vehicles. The model is based on a continuous-time Markov chain which is shown to converge, under appropriate conditions, to a set of clustered epidemiological differential equations. The fraction of vehicles which have received a message, as a function of space and time may be obtained as a solution of these differential equations, which can be solved efficiently, independently of the number of vehicles. Such characterizations can form the basis of assessing several attributes of V2V systems, some of which we demonstrate. The characterizations lend themselves to a variety of generalizations and capture various interdependencies between communication and mobility. As tests of the model we provide applications both in real-world settings using microscopic traffic traces and in postulated scenarios of outages and system perturbations: we find good model agreement with microscopic trajectory from two actual trajectory datasets, as well as a synthetic trajectory dataset generated from the origin/destination matrix.

1. Introduction

1.1. Motivation and applications

Vehicle-to-Vehicle (V2V) technology is poised to significantly impact the functioning and management of transportation networks (Harding et al., 2014; NHTSA, 2016). With V2V, vehicles can communicate directly with each other, or with bikes,

* Corresponding author.
E-mail addresses: jungyeol@seas.upenn.edu (J. Kim), swati@seas.upenn.edu (S. Sarkar), venkates@seas.upenn.edu (S.S. Venkatesh), mryerson@design.upenn.edu (M.S. Ryerson), staro@bu.edu (D. Starobinski).

https://doi.org/10.1016/j.trb.2019.11.004
0191-2615/

wheelchairs, and devices held by pedestrians, to share information about road conditions ahead. We consider all these communications within the purview of V2V. In 2016, the National Highway Traffic Safety Administration (NHTSA) of the U.S. Department of Transportation proposed to mandate the integration of dedicated short-range communications (DSRC) for V2V on all lightweight vehicles (NHTSA, 2016); a final decision on the proposal is awaited. In 2019, the European Commission implemented new rules stepping up the deployment of Cooperative Intelligent Transport Systems (C-ITS) which enables V2V communication (European Commission, 2019). Several automakers are already deploying V2V; for example, General Motors has already incorporated V2V technology into the 2017 Cadillac CTS sedan (Bonelli, 2017), and by March 2018 there were over 100,000 Toyota and Lexus vehicles equipped with V2V in Japan (Toyota, 2018). Volkswagen has also announced plans to begin deploying V2V technology in Europe starting in 2019 (Cain, 2018). In view of these market and regulatory forces, we may safely anticipate a rise in market penetration of V2V in the near future. Thus, it is wise to both plan for its immediate deployment and currently unforeseen future uses in traffic congestion, safety, and monitoring.

V2V can substantially enhance public safety. It can provide advance warnings to drivers about developing dangerous situations, through for example forward collision warning, intersection movement assist, left turn assist, and blind-spot warning (Harding et al., 2014); the latter two applications alone could prevent an estimated 600,000 crashes and save over 1000 lives each year in the U.S. (Bertini et al., 2016). V2V can inform drivers about the incipient arrival of emergency and service vehicles such as police cars, ambulances, and fire brigades. For example, police vehicles in an active pursuit could use V2V communication to pull over vehicles or deliver warnings to drivers, and ambulances with critical patients and fire services could send messages to clear roadways in advance of their arrival — it is difficult for drivers to determine the direction of approach of an emergency vehicle and react appropriately based on auditory cues of traditional sirens alone, since auditory localization of warning alarms is imprecise within the confines of a closed vehicle (Caelli and Porter, 1980). V2V can also propagate warnings of disruptive conditions—flash floods, flooded roads, and damaged bridges, for instance—to vehicles approaching from a distance.

V2V may be used to alleviate traffic congestion as several works in the existing state of the art suggests. Beyond the usual suspects such as poorly designed narrow streets, traffic accidents, and inefficient traffic signals, traffic congestion arises from small perturbative effects when the traffic density is high. For example, in *phantom jam* (documented through aerial photography (Helbing, 2001) and confirmed experimentally (Sugiyama et al., 2008)) a small perturbation initiated through braking by an individual vehicle slows down the speed of the vehicles behind and, when the traffic density is high, this can cause a chain reaction with the phenomenon amplifying as the waves spread farther back (Qian et al., 2017; Patire and Cassidy, 2011). This can quickly result in severe traffic congestion throughout a relatively large area, especially in peak traffic. V2V communications can mitigate such phenomena by timely and efficient distributed communication. In recent work, Won et al. (2017), for instance, have proposed an efficient phantom jam control protocol leveraging V2V communication. Autonomous vehicles may be particularly susceptible to such perturbations because of higher traffic densities enabled by automation. V2V-equipped autonomous vehicles will however be able to anticipate and mitigate phantom traffic jams by receiving relevant information to automatically adjust speed, space headway, and routes.

For effective remediation, congestion must be detected early and be communicated to relevant parties efficiently. Congestion can be detected by floating vehicles equipped with GPS (Wang et al., 2019; Hellinga et al., 2008; He et al., 2017b) or cellular devices (Wang et al., 2012; Demissie et al., 2013). However, traffic status estimation based on floating vehicles is not accurate enough (Mandal et al., 2011; Yong-chuan et al., 2011) and the instantaneity and stability of urban traffic congestion estimation remain challenging (Kong et al., 2016). V2V can facilitate smoother traffic flow overall by quickly disseminating relevant traffic information, e.g., vehicle speed, acceleration, and location of neighboring vehicles, to detect traffic congestion and estimate its severity as it builds (Chen et al., 2006; Lakas and Chaqfeh, 2010; Chou et al., 2011; Knorr et al., 2012; Forster et al., 2014; Jiang et al., 2014; de Souza et al., 2015; Wang et al., 2016). V2V can also facilitate autonomous exchanges of routing plans between vehicles and suggest alternative routes to optimize overall traffic flow. For example, Gupte and Younis (2012) proposes to factor in microscopic data incorporating the destination and routes of nearby vehicles in deciding whether rerouting is advisable, in contrast to centralized schemes that only provide macroscopic information on congestion.

Traffic conditions are known to vary widely and sometimes abruptly in a variety of unexpected events. For example, at the end of a major sporting event, unexpected victories can cause sudden increases in traffic in areas of celebration. Unscheduled personal visits by high-ranking officials (e.g., presidents) and celebrities can also significantly increase local traffic volumes. The congestion from such hot spots can spread all over the city, including the peripheries. This congestion can be ameliorated through the use of V2V communications originating from vehicles at or near the epicenter of the disturbance allowing the recipients the option to bypass the congested region.

All in all, V2V is likely to have widespread applications in ameliorating congestion, improving routing, enhancing safety, and improving profitability. Many of these applications save time and money of participants of ride-sharing services, who are likely to use transportation networks more than the rest of the populace. Through participation in these applications, shared-ride services also contribute to society by improving the travel experience of the overall populace. Early adopters of V2V or connected and automated vehicles (CAV) may be commercial vehicles including ride-sharing, shared taxis and shuttles and transit services etc., rather than private vehicles of individual consumers (Sweatman, 2017): individual consumers tend to hold on to personal vehicles for many years; consider that the average age of cars and light trucks reached 11.6 years in 2016 (U.S. Department of Transportation, 2017). As to the ride-sharing companies, Uber has invested heavily in CAV research and development (McGuckin et al., 2017). Since May 2018 any one with the Lyft app in Las Vegas can hail a V2V-enabled CAV

(Blanco, 2019). As of January 2019, they currently have 30 CAVs and have provided over 30,000 rides. These developments suggest that CAVs enabled with V2V may be expected to facilitate growth and change in commercial shipping and in the shared mobility industry. Once a critical mass is achieved, the exchange of information between such commercial vehicles will also benefit unaffiliated or otherwise affiliated vehicles, especially in the context of safety and traffic conditions.

V2V technologies need not however constitute an unmitigated blessing - they can also bring new threats, e.g., by serving as vectors to infect vehicles with malware (Checkoway et al., 2011; Koscher et al., 2010; Miller and Valasek, 2015). Infected vehicles can proceed to infect other vehicles with the malware. Malware can disrupt functionalities of vehicles, the automated ones certainly, but also manually driven ones, and thereby constitute a public-safety hazard. Malware may contaminate the information that vehicles receive or transmit. Routing information could be manipulated to direct the vehicles to locations causing maximum harm, e.g., 1) to roads with poor visibility condition, thereby causing accidents, 2) to areas of heavy traffic, for example, to block the roads surrounding a busy sports coliseum at the conclusion of a game, 3) to congregate to maximize malware propagation, for example, to launch a denial of service (DoS) on the communications infrastructure associated with critical facilities (e.g., transportation hubs). Alternatively, the information communicated to other vehicles informing of precarious road conditions ahead may be suppressed, leading to a chain of collisions. Malware can cause serious malfunctions in assisted driving, e.g., adaptive cruise control, forward collision warning, lane departure warning (Miller and Valasek (2015), p. 11), leading to accidents. Malware may also be able to record and leak private information of individuals in the vehicles, e.g., their driving patterns, the address books in their phones, their pictures (Miller and Valasek (2015), p. 15).

V2V can also constitute part of the solution for this security threat. It can be used for sharing security patches that would render vehicles immune to malware (of types that are anticipated or detected), and certificates for privacy and authenticity, that would help blacklist infected vehicles and enable other vehicles to ignore messages from vehicles which do not have these certificates. Certification Authorities (CAs) issue certificates to trusted vehicles, and certificates of untrusted vehicles must be immediately revoked. The Certificate Revocation Lists (CRLs) must be continuously updated, thoroughly secured, and quickly delivered over large areas like entire cities. A recent report from the U.S. Federal Highway Administration (FHWA) (Green et al., 2018) noted that epidemic routing over V2V may be utilized for the distribution and revocation of cryptographic materials. Various other studies have also advocated the secure distribution of CRLs using V2V (Haas et al., 2011; Chen et al., 2011; Laberteaux et al., 2008).

Characterizing the spread of malware, security patches, and CRLs over V2V constitute the first step in countering the spread of the malware and enhancing the efficacy of the defense mechanisms, both of which would help shared transportation companies to consider their vulnerability to cyber attacks and devise plans to protect against intrusions.

We seek to characterize the spatio-temporal propagation of information (whether it be hazards, traffic conditions, safety, emergency protocols, malware, CRLs, or something else) through local messaging between vehicles, regardless of the specific content of the messages. Such characterizations would reveal the fraction of vehicles who have the information of interest, i.e., the fraction of informed vehicles, as a function of time and location. Obtaining such a characterization is of central importance for ensuring the efficacy and security of all the applications in question. Given the diverse set of applications of V2V, the models for such characterizations need to apply regardless of specific content, be flexible, easy to compute and scale efficiently with size.

1.2. Challenges

Different applications that utilize V2V would need to populate vehicles over different areas with the pertinent messages - we refer to such an area for a given application as its *Region of Interest* or *RoI*. Geocast can be used to provide information only to vehicles in the RoI (Navas and Imielinski, 1997). The RoI will be of different shapes and sizes for different applications. The characterization of the spatio-temporal propagation of information would be of value provided the RoI spans at least a few blocks. In many applications of V2V, the RoI varies from an area spanning a few blocks to a large area spanning the entire city, or areas of intermediate sizes. From domain knowledge, we now suggest the sizes of the RoIs for different applications mentioned in Section 1.1. In urban areas, such as Manhattan or Hong Kong, which are always suffering from traffic congestion, information pertaining to 1) incipient arrival of emergency vehicles such as police cars, ambulances, and fire brigades, 2) sudden development of disruptive conditions—flash floods, flooded roads, ice formation, damaged bridges etc., or 3) possibility of a phantom jam, must be transmitted in advance to a region of at least 10 blocks radius around the epicenter of the event or the current location of the emergency vehicle in question. Thus, the RoI is of the order of 100 blocks for such applications. The relevant information may also have to be shared in certain key arteries connecting a larger area to warn vehicles directed towards the epicenter of the events in question or to clear arterial roads in advance of the arrival of the emergency vehicles. In these cases, the size of the RoI would be larger, especially in the city center and at peak traffic, because it takes longer to clear arterial roads. In urban areas with heavy traffic congestion, even informed vehicles may require more time to change trajectories rather than following the road ahead. Incidentally, information has to be spread to the vehicles not only traversing the same road as the emergency vehicles, but also those traversing the intersecting roads; otherwise, vehicles in these intersecting roads may congest by arriving unaware into the road taken by the emergency vehicle. Thus, information must be spread over two dimensions. The same observation applies for messages pertaining to the development of emergency conditions or phantom jams. The RoI for these local events may not however exceed radii of 20 blocks or so.

The RoI for security applications may constitute entire cities, large counties or nationwide highways, etc. since malware can quickly spread throughout large areas causing large scale instability and chaos. On the other hand, the RoI for the spread of information concerning routes and congestions due to unanticipated events, such as unexpected victories in major sporting events, unscheduled personal visits by dignitaries and celebrities, may span anywhere between radii of 10−20 blocks to entire cities, counties, long highways, etc.

Any model developed must lend itself to simple computation regardless of scale. Consider that the number of vehicles registered in Los Angeles in 2017 was around 8 million (California Department of Motor Vehicles, 2017) and the number of registered vehicles in Shanghai in 2015 was about 2.5 million (Ningning, 2016). Although not all the registered vehicles may simultaneously ply a RoI, the number of vehicles plying the RoIs, particularly in urban areas and highways is large. For example, considering multiple lanes, at peak travel times at city centers, more than one thousand vehicles may simultaneously ply even a modest-sized two-dimensional RoI of radius of five blocks (the area encompasses 25 blocks or so). This inhibits a simulation-based study because memory usage and execution time exponentially increases with the number of simulation vehicles in most of the V2V simulators (e.g., in VEINS, composed of OMNeT++ and SUMO, which is designed to capture microscopic aspects such as details of communication network protocols etc.); as such, these can simulate only transportation networks with far fewer number of vehicles. Thus, given the sheer number of vehicles and adding a multitude of non-motorized travelers to the mix, designing easy-to-compute analytical characterizations for the spatio-temporal spread, in which computation time gracefully scales with the number of entities, assumes paramount importance.[1]

Frustratingly, not only the size of the RoI but also the heterogeneity of vehicular networks complicates the task of obtaining a computationally simple mathematically tractable model for characterizing the fraction of informed vehicles as a function of time and space. Both vehicle mobility and wireless communication influence information propagation in the vehicular network, and both conditions vary temporally and spatially. Heterogeneity in vehicular mobility arises, for example, due to (1) different forms of transportation networks including grid networks (e.g., Manhattan), radial and circular road topologies (e.g., Paris, Moscow), and irregular narrow streets (e.g., medieval town centers); (2) regional characteristics, such as urban or rural; and (3) time of day, such as morning and evening rush hour. Heterogeneity in communication conditions can primarily be attributed to: (1) Obstacles like buildings and trees; (2) adverse weather conditions like heavy storms; (3) frequency of communication; and (4) user density and available bandwidth. Different admixtures of these transportation and communication conditions have different consequences for the dynamics of information propagation. Also, both mobility patterns and communication conditions continuously evolve over space and time because of (1) evolving road topologies (e.g., due to road constructions and temporary roadblocks), (2) varying traffic conditions (e.g., due to congestion and accidents), and (3) fluctuating link quality. Any model should therefore be readily adaptable to this temporal and spatial evolution to capture its impact on information propagation dynamics. Also, it is not sufficient to study how mobility (e.g., vehicle speed, traffic density, routing) and communications independently influence information propagation, as these factors mutually interact with each other and consequentially affect information propagation in a more complicated way.

Finally, variations of mobility and communication in a vehicular network are inherently stochastic. A characterization of V2V information flow must hence begin by considering the stochastic components of the system. In such settings, the classical theory of even well-behaved stochastic systems governed by Markov processes provides computation approaches only for the steady-state distributions (i.e., distribution as time approaches infinity) of the number of informed vehicles. These computation approaches are also computation-intensive as these rely on the inversion of a $Q \times Q$ transition probability matrix, where Q is the number of states, i.e., Q is at least J^K, where J is the maximum number of vehicles in a city block and K is the number of blocks in the RoI. Such computations become intractable even for modest size RoIs. Obtaining

[1] A question that arises is if for large RoIs like entire cities, etc., the communication will necessarily have to be over cellular networks, rather than over V2V. Towards that end, note that DSRC technology for V2V can utilize the specified authenticated bandwidth to enable reliable communication, and it can also provide high-speed data transmission (Ban and Li, 2018). For security applications, as noted before, several research papers and U.S. Federal Highway Administration (FHWA) make the case for transmitting CRLs over V2V; the medium for the propagation of malware will be determined by the designer and as such V2V can not be ruled out. According to a recent FHWA report (see Green et al. (2018)), multi-hop dedicated short-range communication (DSRC) V2V is an example of a Mobile Ad-hoc Network (MANET), and that "A MANET can serve many needs that traditional physical infrastructure-dependent networks cannot. For example, MANETs can provide cellular network offloading. Network offloading can be crucial in situations where the number of users on a network approaches or exceeds the network capacity, causing delays and other interruptions in service. MANETs are capable of temporarily diverting traffic from traditional network infrastructure to reestablish service for, or increase the number of, users within a typical area of coverage. In a similar light, MANETs are able to expand the coverage of a given network to regions beyond what an infrastructure-dependent network usually covers. MANETs can also provide communication and information dissemination capabilities in areas that temporarily or permanently lack an efficient communication infrastructure." Next, note that, V2V is only intended to provide autonomous communication in transportation environments and offers control over priority, channel, transmit power. Thus, special purpose V2V messaging networks may be optimized for inexpensive and efficient delivery of the specific messages they need to cater to, as opposed to the general-purpose cellular data networks which can not be optimized for the requirements of the transmission of any one specific category of messages. Thus, it is quite possible that transmission over V2V may cost travelers less. In fact, V2V may well be offered as public utility in due course since it provides public service by enhancing road safety and reducing congestion; if so, transmission over V2V may become free, while Cellular communication will cost users in the foreseeable future. Thus, a cellular network-based system may be replaced or at least supplemented, for at least data communication pertaining to transportation requirements, as market penetration of V2V increases, as is expected given the interests shown by automakers and regulators. In such an eventuality, equivalents of the popular tools like google maps and crowdsourcing-based apps like Waze, which are currently utilized over Cellular networks, for applications like congestion control and route selection, may be developed for V2V; their current design for utilization only over Cellular networks may be the artifact of the current reality that Cellular network is the only available network as V2V is not widely deployed. All the above constitute reasonable presumptions now, and the applications would become clearer as the V2V technology is widely deployed.

distributions at a given finite time is even more hopeless. On the other hand, one can envision more computationally tractable deterministic characterizations, but the challenge then becomes to show these as limits or some other mathematical derivations or statistic such as expectations of the governing stochastic process (e.g., Kim et al. (2016) notes the need to consider "stochastic modeling to introduce greater realism", as future work).

1.3. Positioning our contributions with respect to the state of the art

Prior research on V2V has studied how to leverage the technology to smooth out congestion speeds in a transportation problem with communication exogenous (Scholliers et al., 2016). Other work in a dual setting of a communications problem with transportation exogenous has focused on the relationship between the latency of different communication mechanisms and their impact on traffic flow and on vulnerable road users (Whyte et al., 2013; Zhang et al., 2014a; Greenberg, 2015). The state of the art that considers both communication and mobility as endogenous has largely focused on characterizing and evaluating message propagation speeds. Here, certain destinations points are identified, and the speed is characterized as the ratio of the distance between the source and destination and the time message takes to reach the destination; the latter is referred to as the delay. Some existing works compute average message propagation speed (or delay) for various system attributes such as traffic densities, vehicle speed, etc.; some others obtain, using renewal processes, long-term average propagation speed, i.e., $\lim_{t \to \infty}$(distance/time).

Most of the analytical works in the above genre have considered movement along a one-dimensional road, which we briefly review next. Wu et al. (2009) studied message dissemination on one-dimensional road, considering one- and two-way traffic, based on the inter-vehicle gap and vehicle speed distributions. For a one-dimensional road with bidirectional traffic, Agarwal et al. (2012) derived bounds on the average speed of information propagation, while Liu et al. (2013) and Saleet et al. (2011) investigated the influence on delay of traffic density, vehicle speed, and delivery distance. Baccelli et al. (2012) provided analysis of the information propagation speed in bidirectional vehicular delay tolerant networks, and showed that a phase transition occurs in the propagation speed with respect to vehicle density. Kesting et al. (2010) considered bidirectional traffic and presented an analytical model to study probability distributions for message transmission times, assuming an exponential distribution of inter-vehicle distances. Yin et al. (2013) also proposed an analytical model for the expected distance of information propagation on two parallel roads, considering the distance between the two roads and the general distributions of vehicle headways. Zhang et al. (2014b) considered multiple lanes of the one-dimensional road with bidirectional traffic and assessed the effect of traffic density and distribution of vehicle speed in different lanes on the speed of message dispersion.

The literature on two-dimensional traffic models is much sparser and more recent. A recent paper studies the delay in forwarding messages along a selected path in a two-dimensional road topology assuming that the speed of all vehicles in the same direction on a road segment is identical (He et al., 2017a). In this paper, the authors considered a unicast scenario and, assuming communication delays may be ignored, introduced an algorithm to choose the path with the minimum expected delay. In another recent work Kim et al. (2016) study the speed of the information propagation "wave", approximating it at each time and location as a function of only the deterministic traffic flow at that time and in the vicinity of the location, and evaluate the approximation through comparison with a synthetic stochastic process.

To summarize, the state of the art has focused mostly on analyzing expected propagation speed (or expected delivery delays) for given vehicle speed, traffic density, and distance between vehicles, with investigations limited primarily to one-dimensional roads and, rarer, two-dimensional networks. To the best of our knowledge, characterizations of fractions of informed vehicles at a given time and location, that have been verified through either mathematical proofs, or synthetic simulations, or empirical studies involving actual microscopic trajectory data, have remained elusive for both one and two-dimensional vehicular networks, regardless of size. Such fractions are more informative of the message propagation process than expected propagation speed (or expected delivery delays), in the same manner as cumulative probability distributions or densities are more informative of a stochastic process than an expectation. We obtain such characterizations following rigorous mathematical proofs, and evaluate the same through extensive synthetic simulations, and utilization of several actual microscopic trajectory data. Such characterizations can form the basis of assessing several attributes of V2V systems, some of which we demonstrate. The characterizations are computationally tractable and lend themselves to a variety of generalizations and capture various different interdependencies between communication and mobility. We elaborate our contributions in the next Section.

1.4. Our contribution

We start with by modeling V2V information flow in a transportation network, i.e., the number of informed and uninformed vehicles as a function of time and space, as a continuous-time Markov chain (CTMC) following the extensive precedents of utilization of CTMCs in transportation networks, e.g., in estimation of freeway travel time in both routine and perturbed states (Ramezani and Geroliminis, 2012; Dong and Mahmassani, 2009; Geroliminis and Skabardonis, 2005; Alfa and Neuts, 1995). We are immediately confronted with the perils of the stochastic essence of such systems. As mentioned before, even well-behaved stochastic processes like CTMCs are known to be computationally challenging in providing probability distributions which are only accentuated in the context of information flow in vehicular networks (last paragraph of

Section 1.2). Our key insight to bypass the computational bottleneck in this setting is to finesse the computational limitations by appeal to the ergodic theorem to argue that the propagation process converges, in the mean field limit, to a solution of a set of *clustered epidemiological differential equations* which provide the fraction of informed vehicles at any given time and space, and also lend themselves to fast computation. One of the main virtues of this modeling approach is that, in the mean field limit, the computation time needed to solve the equations does not depend on the number of entities including vehicles, pedestrians, bikes, and wheelchairs, while computation time increases only modestly with topographical complexity (we demonstrate the computational tractability in Section 5). In fact, as the number of vehicles in the system grows, the better is the fit of the model to the underlying traffic patterns, which renders our tool suitable for most of the applications that arise in practice. In Section 2, we also show how our framework caters to various interdependence between mobility and communication that arise in practice-namely, temporal variation of the traffic density and routing to capture the vehicular movement pattern during rush hour, a location-dependent mobility model that reflects speed limits applied differently depending on the region, and finally a traffic density dependent mobility model that reflects reduced vehicle speed due to high traffic density.

Continuum limit and mean field models have been used in the ride-sharing literature to control shared transportation systems for ride-sharing services. The goal of these platforms is to maximize fulfilled demands, revenue, or other objectives by employing tools from optimal control (Banerjee et al., 2018). Specifically, Banerjee et al. (2018) focuses on the scheduling policy, where the platform can determine which vehicle to allocate in response to an incoming ride request. They proposed policies to minimize the proportion of dropped requests. In Braverman et al. (2019) and Braverman et al. (2017), empty-car routing in a ride-sharing network was considered under a condition where supply and demand for vehicles tend to infinity. They provided a comprehensive analysis of the design of an optimal empty-car routing policy based on an asymptotic fluid analysis. And Yang et al. (2018) studied the equilibrium behavior of nomadic agents for different resources depending on time and location, arising in ride-sharing economies. The utilization of these models in the context of information flow in transportation networks is new to our knowledge.

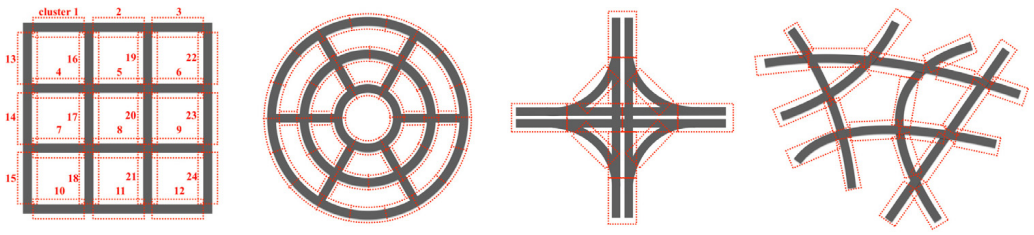
The analytically sanitized theoretical model lives in the mean field limit of an idealized propagation process which converges to the solution of a coupled system of ordinary differential equations when vehicles follow exponential sojourn times and the number of vehicles goes to infinity. In contrast, in practice, we deal with a finite number of vehicles with sojourn times governed by an unknown underlying process. It is legitimate to wonder then how closely our models hew to reality. We therefore conduct detailed empirical verifications of our model, intentionally considering settings which stress the model assumptions (Section 3). For this purpose, we use synthetic stochastic models and microscopic traffic traces - two actual trajectory datasets on highways, two-way roads with intersections, and one synthetic trajectory dataset gleaned from origin/destination information on a roundabout. Our empirical validations confirm that, even for a moderate number of vehicles and a variety of road topologies, the output of the differential equations matches the results of the trace propagation process quite well even in the settings of the microscopic trajectory data where there are a finite number of vehicles and there is no compelling reason to believe that vehicles have exponential sojourn times, and even when vehicles follow shortest-path destination-specific routes (as opposed to our modeling abstraction of using routing probabilities agnostic of the destination). We have made publicly available a software that can model the spread of V2V messages in arbitrary transportation networks using the clustered epidemiological differential equations (Kim, 2019).

We next present case-studies demonstrating the usefulness of our model to characterize the behavior of V2V systems in various practical situations (Section 4). Here is a sampling of the results. We show how traffic congestion can be alleviated through an intelligent application of V2V technology. We also assess how quickly information about the location of disruptive changes (i.e., temporary roadblocks) can be disseminated. Lastly, we examine how the initial location of informed vehicles determines the spread of information throughout the transportation network. Our study reveals a counter-intuitive phenomenon: message propagation is not necessarily accelerated if the initially informed vehicles are centrally located.

Finally, we show how our mathematical framework can be readily generalized to accommodate information propagation under the condition of destination-dependent vehicle routing mechanisms (e.g., the shortest path routing) and also handle the simultaneous spread of multiple pieces of information (Section 6). We discuss the various computation tradeoffs associated with the generalizations. We conclude by summarizing our research findings in Section 7.

2. Model formulation

Section 2 introduces models pertaining to various forms of vehicle mobility and communication. We first introduce a general mathematical framework for the propagation of messages in V2V systems (Section 2.1). We subsequently show how the general framework caters to various specialized cases that arise in practice (Section 2.2): temporal variation of the traffic density and routing to capture the vehicular movement pattern during rush hour (Section 2.2.1), a location dependent mobility model that reflects speed limits applied differently depending on the region (Section 2.2.2), and finally a traffic density dependent mobility model that reflects reduced vehicle speed due to high traffic density (Section 2.2.3). These various scenarios can occur simultaneously, which can easily be represented by combining these models. All models are based on continuous-time Markov chains. We show that information propagation based on the Markov chain can be very well approximated by differential equations for the various vehicular mobility and communication patterns mentioned above.



(a) Grid roads (b) Radial and circular roads (c) Diamond interchange (d) Irregular roads

Fig. 1. Clustered road topologies. The figures represent various road topologies such as grid roads, radial with circular roads, and irregular roads. In arbitrary types of topologies, roads can be divided into multiple smaller segments, so clusters can be defined as shown by the red dotted rectangle in the illustration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.1. Clustered epidemiological differential equation model

We develop tools to model information propagation under general types of transportation network and various communication conditions. Transportation networks exist in various forms such as highways connecting cities; coastal roads; and urban roads. Unlike highways or coastal roads, which are relatively simple one-dimensional forms, urban roads exist in complex networks of different types depending on the characteristics of the area. In most areas, however, these road topologies are superposed in a complex manner. We introduce mathematical tools that can be used to model and analyze information flow across arbitrary complex road networks as shown in Fig. 1, which can potentially be used for prediction of information propagation through vehicle-to-vehicle communication.

The extent to which information is spread between moving vehicles is determined through vehicle movement and wireless communication. The mobility of vehicles on the transportation network depends on topology which will continuously evolve (addition of new roads, blockage of existing roads due to maintenance), traffic conditions (traffic congestion, the presence of an accident), time of day, and the characteristics of the individual travelers (urban, rural, land use interactions). Communication on these transportation networks is influenced by traffic conditions (packet collisions due to high traffic density) and communication conditions (frequency of communication between vehicles, fading due to obstacles such as buildings, trees, etc., vehicle occlusion, multipath transmissions, rate and power control, the hidden terminal problem, etc). We present a mathematical model that can capture information flow in arbitrary vehicular network that contains all of these complex elements.

We divide the entire RoI into a collection of J clusters, with each cluster corresponding to a specific region of the road; thus each vehicle is located in one of the J clusters. One possible way to form a cluster is to set the cluster size to the communication range as it is natural to assume that the vehicles located within the same cluster are within the V2V communication range. However, vehicles located at the boundary of one cluster can communicate with vehicles in other adjacent clusters if they are within the communication range. Our model can cater to this communication between vehicles located near the boundary of two adjacent clusters by considering that the vehicles located in the different clusters can communicate at a reduced rate. When considering the communication of vehicles in adjacent clusters, it is reasonable to apply the reduced rate because not all cars in the two adjacent clusters can communicate with each other and only vehicles located near the boundary can communicate.

Vehicles can both communicate and move across clusters or within clusters. Clearly, a vehicle cannot communicate between every pair of clusters because of the technical limitations of the wireless communication range. Similarly, a vehicle cannot move between each pair of clusters due to the nature of the road and traffic rules. Both of the above are influenced, but not solely determined, by geography. For example, although two clusters are close enough to permit communication, the vehicle may not be able to move between them. Fig. 2 shows cluster i and j located on the same road segment of the two-way roads; these vehicles are close enough to permit communication but traffic rules do not allow them to travel across the median. Similarly, even between two adjacent clusters where vehicles can move, the success rate of communication between two clusters may become low or 0 due to obstacles (buildings, trees, etc.).

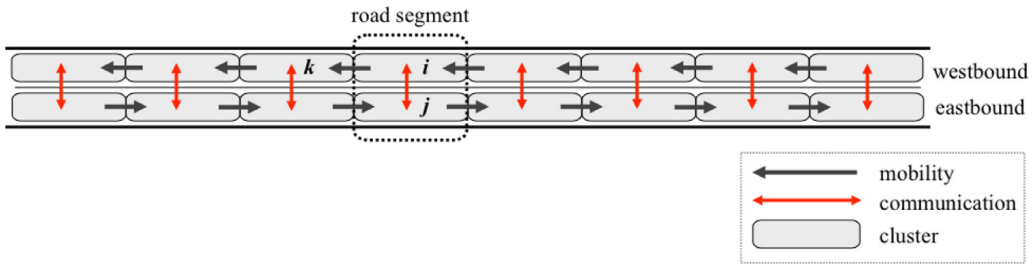


Fig. 2. Mobility and communication networks. The mobility network is a directed network and the communication network is an undirected network. The edges of these two networks may overlap but need not be exactly the same.

Table 1
Mathematical notation.

$G(V, E)$	Directed network of mobility on road topology
$G'(V, E')$	Undirected network of communication on road topology
V	Set of clusters, $\|V\| = J$
E	Set of directed mobility edges
E'	Set of undirected communication edges
N	Total number of vehicles
$n_j^I(t)$	Number of informed vehicles in cluster j at time t
$n_j^S(t)$	Number of non-informed vehicles in cluster j at time t
$X(t)$	Continuous-time Markov process
$X_N(t)$	Scaled Markov process, $X(t)/N$
$\lambda_{ij}^I(\cdot)$	Mobility rate from cluster i to j for informed vehicles
$\lambda_{ij}^S(\cdot)$	Mobility rate from cluster i to j for non-informed vehicles
λ	Upper bound for mobility rate
β_{ij}/N	Communication rate between a vehicle located in cluster i and a vechicle in j
$N_c(j)$	Neighborhood of cluster j; set of clusters connected from cluster j
p_{jk}	Probability that a vehicle in cluster j move to k
$\rho(t)$	Proportion of informed vehicles at time t

We define two networks and corresponding adjacency metrics: communication network and mobility network. While both depend on the geographical characteristics of the roads, they do not necessarily have to be the same. We first describe the mobility network. Let $G = (V, E)$ be the directed mobility network on the road topology, and the directed network G consists of a set of nodes $V = \{1, 2, \ldots, J\}$ corresponding to clusters and a set of mobility edge set E. If clusters $j, k \in V$ are adjacent roads and vehicle movement is possible from cluster j to k, the directed edge is specified as the edge $e \in E$ from j to k (equivalently, e $:= j \to k$). The corresponding adjacency matrix of G is the $J \times J$ matrix $A = (a_{jk})$ where $a_{jk} = 1$ if $j \to k \in E$ and $a_{jk} = 0$ otherwise. We now introduce the communication network. Let $G' = (V, E')$ be the undirected network of communication with the set of same nodes V and set of wireless communication edge set E'. If two clusters $j, k \in V$ can directly communicate amongst each other then there exist an edge $e \in E'$ between j and k (equivalently, e $:= j \leftrightarrow k$). The corresponding adjacency matrix G' is the $J \times J$ symmetric matrix $a' = (a'_{jk})$ where $a'_{jk} = a'_{kj} = 1$ if $j \leftrightarrow k \in E'$ and $a'_{jk} = a'_{kj} = 0$ otherwise. Through the adjacency matrices G and G' we have discussed, the characteristics of any road topology can be extracted.

We model the information propagation in transportation networks based on the Susceptible-Infective(SI) epidemiological model that has been adopted in numerous infectious disease and information propagation research. This epidemiological model assumes that susceptible individuals have not yet incurred the disease but are vulnerable to it, and susceptible individuals can become infected after receiving the disease through contact with infected individuals. These newly infected individuals can spread the disease to susceptible individuals. In this study, we use the mathematical formulation where the vehicles which carry information are referred to as infective, vehicles which do not yet have the information are called susceptible. From hence, we denote the vehicle that carries the information as informed vehicles, and vehicles that have not received the information as non-informed vehicles. Table 1 summarizes the mathematical notations used in this paper.

Suppose that N vehicles are located in the network with each vehicle in one of the J clusters corresponding to different road segments. We will suppose for now that the network is closed and that there are no exogenous arrivals into, or departures from, the system. Let $n_j^I(t)$ and $n_j^S(t)$ respectively represent the number of informed and non-informed vehicles in cluster $j \in V$ at time t. The $2J$-dimensional lattice vector

$$\left(\mathbf{n}^I(t), \mathbf{n}^S(t)\right) = \left(n_1^I(t), n_2^I(t), \ldots, n_J^I(t);\ n_1^S(t), n_2^S(t), \ldots, n_J^S(t)\right)$$

then represents the instantaneous state of the system, semicolon and extra spacing have been added merely for visual separation of informed and non-informed vehicular counts in the various clusters. The state space on which we model the dynamics of information propagation accordingly is the set of lattice points in $\mathbb{Z}^J \times \mathbb{Z}^J$ satisfying

$$S^N := \left\{ (\mathbf{n}^I, \mathbf{n}^S) \mid n_j^I \geq 0,\ n_j^S \geq 0,\ j = 1, \ldots, J;\ \sum_{j=1}^{J} \left(n_j^I + n_j^S\right) = N \right\}.$$

The basic transitions in this state space capture one of three types of phenomena: the movement of an informed vehicle to a neighbouring cluster; the movement of a non-informed vehicle to a neighbouring cluster; and the conversion of a non-informed vehicle to an informed vehicle by the successful transmission and receipt of information. A little notation helps grease the wheels: write $\mathbf{k} = (\mathbf{n}^I, \mathbf{n}^S)$ for the current state and let $\mathbf{1}_j$ represent the $2J$-dimensional unit vector whose jth element is 1 with all other elements being 0. For $j, k \in \{1, \ldots, J\}$ with $j \neq k$, the state transition $\mathbf{k} \to \mathbf{k} - \mathbf{1}_j + \mathbf{1}_k$ captures the movement of an informed vehicle from cluster j to cluster k; the state transition $\mathbf{k} \to \mathbf{k} - \mathbf{1}_{J+j} + \mathbf{1}_{J+k}$ represents the movement of a non-informed vehicle from cluster j to cluster k; and the state transition $\mathbf{k} \to \mathbf{k} + \mathbf{1}_k - \mathbf{1}_{J+k}$ represents a successful communication of information to an uninformed vehicle in cluster k which now joins the informed ranks in that cluster with a concomitant reduction of the non-informed ranks in that cluster.

We now develop the stochastic underpinnings of the time evolution of the state process $X(t) = (\mathbf{n}^I(t), \mathbf{n}^S(t))$. We model mobility delays by assuming that the time taken by a vehicle to move to a neighboring cluster is exponentially distributed with possibly state-dependent parameters. Likewise, we model communication delays, within and across clusters, by supposing that the time taken for a successful transmission of information from an informed vehicle to a non-informed vehicle is exponentially distributed, again with possibly state-dependent parameters. Under these assumptions, the state evolution process $X(t)$ forms a continuous-time Markov chain (CTMC). We flesh out the structure of the CTMC in what follows.

We recall that the CTMC exhibits the following three types of state transitions: (1) an informed vehicle moves from cluster j to cluster k, $k \neq j$; (2) a non-informed vehicle moves from cluster j to cluster k, $k \neq j$; and (3) a non-informed vehicle in a cluster k receives a successful transmission from an informed vehicle located in the same cluster or in a different cluster.

The first two types of transition capture vehicle mobility. Write $\lambda^I_{jk}(\cdot)$ for the rate at which informed vehicles from cluster j migrate to cluster k, and $\lambda^S_{jk}(\cdot)$ for the rate at which non-informed vehicles migrate from cluster j to cluster k. These rates may be the same but there is no cost in the model to assuming potentially different mobility rates for informed and non-informed vehicles and we may as well do so. We assume that both $\lambda^I_{jk}(\cdot)$ and $\lambda^S_{jk}(\cdot)$ are bounded functions of $\frac{1}{N}(\mathbf{n}^I, \mathbf{n}^S)$ if $a_{jk} = 1$ and are 0 otherwise. In other words, the model permits mobility-based transitions only between neighboring clusters, the transition rates between neighboring clusters are permitted to vary boundedly across clusters as a function of both the (geographic location of) the clusters as well as the density of vehicles in the clusters, and these rates may depend on whether the vehicle is informed or non-informed.

The third type of state transition that we encounter deals with a successful communication of information from an informed vehicle to a non-informed vehicle resulting in the non-informed vehicle attaining informed status. We posit fixed, non-negative constants β_{jj} and β_{jk}, for each j and k, such that intra-cluster communications between vehicles in a cluster j occur at rate β_{jj}/N while inter-cluster transmissions of information from cluster j to a distinct cluster k occur at rate β_{jk}/N. (To keep away from unnecessarily burdening notation, we suppose that $\beta_{jk} = 0$ if $a'_{jk} = 0$, that is to say, there is no direct communication across clusters not connected by a wireless communication link.) The model explicitly captures the phenomenon that communication rates diminish due to reductions in shared bandwidth as the number of vehicles in the clusters increase. There is an implicit ergodic model assumption here: effectively, we assume that the number of vehicles in each cluster is proportional to the total number N of vehicles in the network where the proportion of the population that is captured within each cluster may be cluster-dependent—these cluster-dependent constants of proportionality may be folded into the specification of the parameters β_{jj} and β_{jk}. This type of phenomenon is familiar in ergodic chains where, with a large population N, the occupancy in each cluster will be close to its expected value. The implicit mobility network modelling assumption here, of course, is that the model clusters represent settings in which road segments may be reasonably considered to have an ergodic character where, with a sufficiently large population of vehicles, each cluster sees a non-trivial vehicular occupancy.

The communication parameter will be drawn from the underlying, well-studied communication pathways, communication conditions, capacities, and protocols. The capabilities of the physical devices such as antennas can provide the maximum rate that can be transmitted, and also the maximum capacity of the medium restricts the amount of information that can be transmitted over a certain period of time. In this respect, the capabilities and capacity which are closely related to the network performance can also be reflected in the communication parameter. Besides, one can analyze the communication protocol and draw the corresponding parameter based on what is feasible in the given protocol. Hartenstein and Laberteaux (2008) present a tutorial survey on the network protocols, communication technologies, potential applications, and challenges of vehicular ad hoc network (VANET), and many protocols for VANET have also been discussed (Almalag et al., 2013). Any given combination of protocols would translate to a certain attainable communication parameter. Furthermore, as protocols and physical devices are redesigned for better network performance, we can get a corresponding higher communication rate, which can be used to inform our communication parameters.

To summarize, state transitions in the Markov chain are governed by exponential processes, the transitions from a given state $\mathbf{k} = (\mathbf{n}^I, \mathbf{n}^S)$ to a state $\mathbf{k}' = \mathbf{k} + \mathbf{h}$ occurring at a rate

$$q(\mathbf{k}, \mathbf{k}+\mathbf{h}) = \begin{cases} \lambda^I_{jk}\left(\dfrac{\mathbf{k}}{N}\right) \cdot n^I_j & \text{if } \mathbf{h} = -\mathbf{1}_j + \mathbf{1}_k \text{ and } j \neq k, \\ \lambda^S_{jk}\left(\dfrac{\mathbf{k}}{N}\right) \cdot n^S_j & \text{if } \mathbf{h} = -\mathbf{1}_{J+j} + \mathbf{1}_{J+k} \text{ and } j \neq k, \\ \dfrac{\beta_{jk}}{N} \cdot n^I_j \cdot n^S_k & \text{if } \mathbf{h} = \mathbf{1}_k - \mathbf{1}_{J+k}, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

This follows a well-worn pathway in the theory of continuous-time Markov chains. The key to a dramatic asymptotic simplification in our setting is that the transition rates given by (1) have a certain density-dependent property which reduces considerations via the ergodic theorem to a system of ordinary differential equations in the continuum.

Proceed to the continuum limit and introduce the set $E := \{(\mathbf{I}, \mathbf{S}) \mid I_i \geq 0, S_i \geq 0, i = 1, 2, \ldots, J;\ \sum_{i=1}^J (I_i + S_i) = 1\}$ where, in the natural vector notation, we write $(\mathbf{I}, \mathbf{S}) = (I_1, I_2, \ldots, I_J;\ S_1, S_2, \ldots, S_J)$. The continuous analog of (1) is a continuous

function $f(\mathbf{x}, \mathbf{h})$ on $E \times \mathbb{Z}^{2J}$ given, for each $\mathbf{x} = (\mathbf{I}, \mathbf{S}) \in E$ and $\mathbf{h} \in \mathbb{Z}^{2J}$, by

$$f(\mathbf{x}, \mathbf{h}) = \begin{cases} \lambda^I_{jk}(\mathbf{x}) \cdot I_j & \text{if } \mathbf{h} = -\mathbf{1}_j + \mathbf{1}_k \text{ and } j \neq k, \\ \lambda^S_{jk}(\mathbf{x}) \cdot S_j & \text{if } \mathbf{h} = -\mathbf{1}_{J+j} + \mathbf{1}_{J+k} \text{ and } j \neq k, \\ \beta_{jk} \cdot I_j \cdot S_k & \text{if } \mathbf{h} = \mathbf{1}_k - \mathbf{1}_{J+k}, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The discrete formulation (1) can be imbedded in the continuous formulation (2) by the simple observation that $q(\mathbf{k}, \mathbf{k} + \mathbf{h}) = Nf\left(\frac{\mathbf{k}}{N}, \mathbf{h}\right)$, hence the connection to a continuum density—in the nomenclature introduced by Kurtz, we say that the Markov chain is *density-dependent*. In such cases a very general theorem of Kurtz (Kurtz, 1970) asserts that, in the asymptotic limit as $N \to \infty$, state evolution in the CTMC may be represented by a system of ordinary differential equations.

Introduce the formal notation

$$\lim_{N\to\infty} \frac{\mathbf{n}^I(t)}{N} = \mathbf{I}(t), \quad \lim_{N\to\infty} \frac{\mathbf{n}^S(t)}{N} = \mathbf{S}(t).$$

The formal quantities $\mathbf{I}(t)$ and $\mathbf{S}(t)$ represent the asymptotic fraction of informed and non-informed vehicles, respectively, in each cluster. The following is the key consequence of Kurtz's theorem adapted to our model assumptions.

Under the conditions of our model, for a given choice of initial conditions $(\mathbf{I}(0), \mathbf{S}(0))$, *the time-evolution*, $(\mathbf{I}(t), \mathbf{S}(t))$, *of the distribution of the asymptotic fraction of informed and non-informed vehicles across clusters is governed by the following system of ordinary differential equations:*

$$\begin{aligned} \dot{I}_j(t) &= -\sum_{k\neq j}^{J} \lambda^I_{jk}(\mathbf{I}, \mathbf{S}) \cdot I_j + \sum_{k=1}^{J} \beta_{kj} \cdot I_k \cdot S_j + \sum_{k\neq j}^{J} \lambda^I_{kj}(\mathbf{I}, \mathbf{S}) \cdot I_k \qquad (j = 1, 2, \ldots, J), \\ \dot{S}_j(t) &= -\sum_{k\neq j}^{J} \lambda^S_{jk}(\mathbf{I}, \mathbf{S}) \cdot S_j - \sum_{k=1}^{J} \beta_{kj} \cdot I_k \cdot S_j + \sum_{k\neq j}^{J} \lambda^S_{kj}(\mathbf{I}, \mathbf{S}) \cdot S_k \qquad (j = 1, 2, \ldots, J). \end{aligned} \tag{3}$$

This reduction to a system of differential equations is the jumping off point for our model analysis and a reader who is primarily interested in seeing applications of the model in diverse settings can begin with (3) and read on. The theoretically inclined reader who would like to see details of how Kurtz's theorem, adapted to our setting, results in (3) will find the analysis and proofs in the Appendix.

2.2. Specialization

The general framework, that is the set of differential Eq. (3), referred to as the clustered epidemiological differential equations (CEDE), cater to several special cases that arise in practice. For this, we consider the grid road topology in Fig. 3 with six avenues and streets. In the center of the network is a representation of a Central Business District (CBD): the CBD acts as an attractor of trips from the surrounding locations in the city (periphery). In this network, we assume that all roads are two-way and allow vehicles to move in both directions, and a road segment consists of two clusters corresponding to the opposite directional roads.

2.2.1. Temporal variation of traffic density and routing

We now show how the differential Eq. (3) cater to the temporal variation of the traffic density and routing pertaining to vehicle movement during rush hour. We capture the morning rush from the periphery to the CBD in the morning and then the reverse at the conclusion of the work day. In the directed mobile network $G(V, E)$, the neighborhood of cluster j is defined as the set of clusters connected from $j \in V$ through a directed edge, denoted $N_G(j)$. Let D be the set of clusters in the CBD, and $O = D^c$ be the set of clusters in the periphery. As shown in Fig. 3, the shaded area of the city center is generally located in the center of the city. Let p_{jk} be the probability that vehicles in cluster j move to cluster $k \in N_G(j)$. In that case, the mobility rate set in this model is

$$\lambda^I_{jk}(\cdot) = \lambda^S_{jk}(\cdot) = p_{jk}\lambda \tag{4}$$

where λ is constant. Since the mobility rates for $j = 1, 2, \ldots, J$ and $k \in N_G(j)$ are constant, it is clearly Lipschitz continuous on E. Therefore, by results in the Appendix, the behavior of propagation process based on this temporal variation of traffic density and routing can be approximated by ordinary differential Eq. (3).

In the rest of this subsection, we describe how p_{jk} can be computed. Suppose that, for the starting point $j \in \{1, 2, \ldots, J\}$, this probability is classified into two types: the probability p^j_d of moving towards the CBD, and the probability p^j_o of moving towards the periphery. Concretely, if the direction from cluster j to k is in the direction moving toward the periphery, p_{jk} corresponds to p^j_o, and in the opposite case, p_{jk} corresponds to p^j_d. We introduce the γ parameter defined by the ratio of probability p^j_d to p^j_o to control the temporal variation of the routing for these two directions, resulting in $p^j_d = \gamma p^j_o$. For example, if $\gamma = 5$, the probability of moving towards the CBD is five times higher than the probability of moving towards

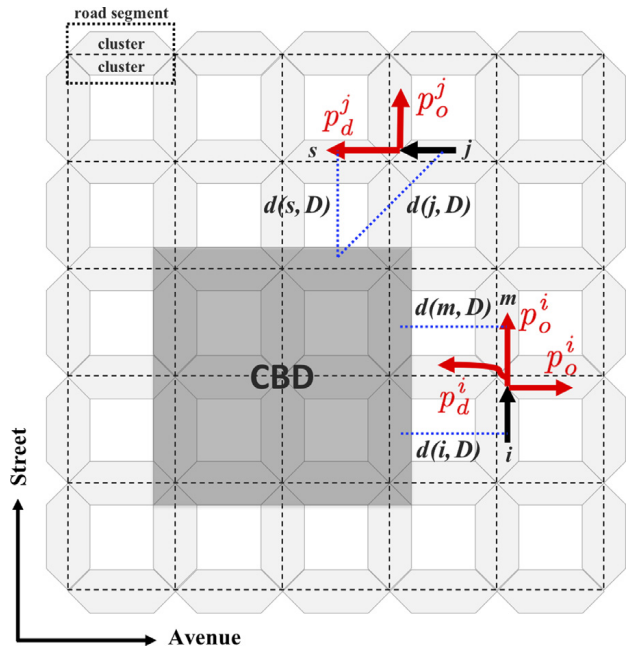


Fig. 3. Clustered grid road topology. We capture probabilities of a vehicle moving between the CBD and the periphery. As for the movement from j to s, the direction is classified as to the direction towards the CBD since $d(s, D) < d(j, D)$. Movement from i to m is classified as movement towards the periphery because the vehicle stays in the periphery, and does not move in the direction toward the CBD.

the periphery. Thus, $\gamma > 1$ indicates a movement pattern in which peripheral dwellers move into the city center, e.g., during morning commute time; $0 < \gamma < 1$ indicates a movement pattern in which workers leave the CBD and go back to the periphery, e.g., during evening commute time.

We define a decision rule for determining whether movement from one cluster to another cluster is toward or away from the CBD. Let $d(x, y)$ denote the Euclidean distance between two geometric centers of $x \in V$ and $y \in V$, and let $d(x, D)$ denote the minimum distance between a cluster $x \in V$ and a cluster in D as follows.

$$d(x, D) = \min\{d(x, y) \mid y \in D\}$$

When vehicles move from cluster j to k, if $d(j, D) < d(k, D)$ then it is classified as the movement towards the periphery, resulting in probability p_{jk} being p_o^j. If $d(j, D) > d(k, D)$, it is classified as the movement towards the CBD, and as a result, the probability p_{jk} becomes $p_d^j = \gamma p_o^j$. In the case of $d(j, D) = d(k, D)$, the type of movement direction is classified according to the position of the origin cluster and the destination cluster. Specifically, if $d(j, D) = d(k, D)$ and origin cluster j and destination cluster k are both in the CBD, p_{jk} is considered to be $p_d^j = \gamma p_o^j$, since this direction corresponds to the movement to stay in the CBD rather than leaving the CBD. Similarly, if $d(j, D) = d(k, D)$ and origin cluster j and destination cluster k are both in the periphery, p_{jk} is regarded as p_o^j because it is a type of movement that stays in the periphery, not in the direction toward the CBD. Therefore, for all cluster $j \in V$ and its neighborhood $k \in N_G(j)$, we have

$$p_{jk} = \begin{cases} p_d^j, & \text{if } d(j, D) > d(k, D) \\ p_d^j, & \text{if } d(j, D) = d(k, D) \text{ and } j, k \in D \\ p_o^j, & \text{if } d(j, D) < d(k, D) \\ p_o^j, & \text{if } d(j, D) = d(k, D) \text{ and } j, k \in O \end{cases} \tag{5}$$

From this decision rule, for cluster $j \in V$ and $k \in N_G(j)$, p_{jk} is classified into one of p_d^j and p_o^j, then these two probabilities p_d^j and p_o^j are determined by $\sum_{k \in N_G(j)} p_{jk} = 1$. More specifically, we now show how p_d^j and p_o^j can be computed for all $j \in V$. First, consider an arbitrary node $j \in D$ and let A be the set of clusters $k \in N_G(j)$ such that $d(j, D) \geq d(k, D)$. By using $\sum_{k \in N_G(j)} p_{jk} = 1$, we have $|A| p_d^j + |N_G(j) \backslash A| p_o^j = |A| \gamma p_o^j + |N_G(j) \backslash A| p_o^j = 1$, resulting in $p_o^j = \frac{1}{|A|\gamma + |N_G(j) \backslash A|}$ and $p_d^j = \frac{\gamma}{|A|\gamma + |N_G(j) \backslash A|}$. Similarly, for arbitrary node $j \in O$, let A' be the set of clusters $k \in N_G(j)$ such that $d(j, D) \leq d(k, D)$. Through the same approach, we have $p_o^j = \frac{1}{|A'| + |N_G(j) \backslash A'| \gamma}$ and $p_d^j = \frac{\gamma}{|A'| + |N_G(j) \backslash A'| \gamma}$. For example, in Fig. 3, when a vehicle in cluster i moves to neighborhood clusters k, n, and m, the p_{ik}, p_{in}, and p_{im} are classified as p_o^i, p_d^i, and p_o^i, respectively. Therefore, $p_o^i = 1/(\gamma + 2)$ and $p_d^i = \gamma/(\gamma + 2)$, resulting in $p_{ik} = 1/(\gamma + 2)$, $p_{in} = \gamma/(\gamma + 2)$, and $p_{im} = 1/(\gamma + 2)$.

2.2.2. Location-dependent mobility model

In this subsection, we show that how the differential Eq. (3) can capture location dependent mobility model that reflects different speed limits for various regions. Speed limits are applied differently depending on the local characteristics of each city and the type of road, and as a result the average speed of the vehicles will depend on these characteristics. In this model, two different mobility rates are applied to reflect different speed limits applied to the CBD and the periphery roads, respectively, and the mobility rate of the CBD is set to a lower value. Let λ_d be the mobility rate from a cluster located in the CBD to the neighboring clusters, so that the average time to stay in the cluster before moving to a neighboring cluster is $1 / \lambda_d$. Similarly, let λ_o be the mobility rate from a cluster located in the periphery to the neighboring clusters, so that the average time to stay in the cluster before moving to a neighboring cluster is $1 / \lambda_o$. The routing probability associated with γ is also applied. The location dependent mobility rate can be described as

$$\lambda^I_{jk}(\cdot) = \lambda^S_{jk}(\cdot) = p_{jk}\lambda_j,$$
$$\lambda_j = \begin{cases} \lambda_d & \text{if cluster } j \text{ is in the CBD} \\ \lambda_o & \text{if cluster } j \text{ is in the periphery,} \end{cases} \tag{6}$$

where λ_d and λ_o are constants, and p_{jk} is computed by the decision rule (5). Since the mobility rates for $j = 1, 2, \ldots, J$ and $k \in N_G(j)$ are constants, the behavior of propagation process can be approximated by ordinary differential equations (3) for the same reason as the previous model (Section 2.2.1).

2.2.3. Traffic density-dependent mobility model

As in the previous two cases, the differential Eq. (3) can cater to traffic density dependent mobility model. Vehicle speed depends on the traffic density on the road with density being inversely related to speed; this is described by models such as the Greenshields model, the Drew model, and the Pipes-Munjal model. The generalized form of these models (Haefner and Li, 1998; Kühne and Rödiger, 1991; Wang et al., 2009) can be expressed as $v = v_f[1 - (k/k_{jam})^a]^b$ where v_f is free flow speed, k_{jam} is jam density, and v and k are speed and density respectively.

Motivated by these studies, we introduce the traffic density-dependent mobility rate $\lambda_{jk}(\cdot)$ from cluster j to $k \in N_G(j)$, which depends on the relative density of both the clusters j and k, where the relative density is defined as the fraction of vehicles located in the clusters j and k. As the fraction of vehicles located in clusters j and k increase, the movement from cluster j to k is slowed down, which implies that the mobility rate from the origin cluster j to the neighboring cluster k decreases. Concretely, when the traffic density of the origin cluster j is high, movement beyond this region is restricted, and if the traffic density of the destination cluster k is high, it is also difficult to enter this region. The mobility rate reflecting this mobility characteristic is set to

$$\lambda^I_{jk}(\cdot) = \lambda^S_{jk}(\cdot) = \lambda \cdot p_{jk} \cdot \left[1 - \left(\sum_{i \in \{j,k\}} (I_i + S_i)\right)^a\right]^b \tag{7}$$

where λ is constant, p_{jk} is computed by the decision rule (5), and a, $b \geq 1$. By controlling parameters a and b, we can reflect a general form of relationship between vehicular speed and traffic density. The mobility rate function λ_{jk} is Lipschitz continuous on E for $j = 1, 2, \ldots, J$ and $k \in N_G(j)$ since the function is continuously differentiable on E. Thus, by results in the Appendix, the dynamics of the information propagation converge to the solution of the differential Eq. (3).

3. Results of empirical validation

We now empirically validate the mathematical model (Section 2) when the underline assumptions of the model are relaxed. The analytical result in the previous section ensures convergence of the propagation process results to solutions of differential equations only when the number of vehicles goes to infinity. In practice, however, the number of vehicles is finite. Therefore, we investigate the effect of a finite number of vehicles on various mobility and communication characteristics. We first consider statistical models which we call synthetic models (Section 3.1), and subsequently consider the two actual and one synthetic vehicle trajectory data collected on different road topologies (Section 3.2). In the latter case, the statistical communication process is superimposed on the trajectory data since there is no data currently available for the communication process. With a finite number of vehicles, we show that there is an excellent match between the result of the propagation process simulation and the model solution even for a moderate number of vehicles in the statistical models. We additionally show there is an acceptable match even for the trajectory data that does not satisfy the statistical assumption of exponential sojourn time under which convergence is guaranteed. Throughout this result section, the ordinary differential equations were solved using `ode` function from the **deSolve** package of R.

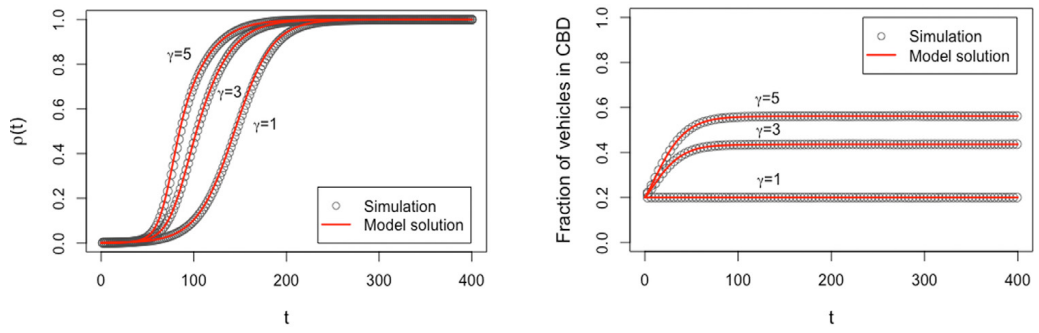
3.1. Synthetic models

We show that the output of the differential Eq. (3) matches simulations for different statistical mobility models with a finite number of vehicles. In Section 3.1.1, we consider the mobility model of Section 2.2.1 which captures the temporal variation of the traffic density and routing. In Section 3.1.2, we apply the location dependent mobility model of Section 2.2.2.

In Section 3.1.3, we apply the traffic density dependent model of Section 2.2.3. In Section 3.1.4, we consider that each vehicle chooses its destination randomly among the clusters, follows the shortest path towards it, and upon reaching chooses another destination, again uniformly. For the validation, we consider the grid road topology introduced in Section 2.2 (We will consider the other road topology in Section 3.2). As shown in Fig. 3, there are six avenues and six streets. The CBD is the shaded area in Fig. 3 and the rest is the periphery. All road segments are assumed to be two-way roads which are set to be composed of two clusters corresponding to the opposite directional roads. Since two clusters on the same road segment are sufficiently close to each other, the vehicles located in these can communicate. Therefore, the adjacency matrix of the communication network $G(V, E')$ is given by $a'_{ij} = a'_{ji} = 1$ if i and j are in the same road segment, otherwise $a'_{ij} = a'_{ji} = 0$. We set the communication parameter to $\beta_{ij} = 3$ if $a'_{ij} = 1$ and $\beta_{ij} = 0$ otherwise. At initial time, N vehicles are uniformly distributed in J clusters, where $J = 120$. Thus, each cluster has $n = N/J$ vehicles at initial time. The information of interest initially begins to propagate from 10% of vehicles located in the lower left cluster (equivalently, $0.1n$ vehicles), which are located in peripheral areas. To study the degree of information propagation, we introduce $\rho(t)$, the fraction of overall vehicles that are informed at time t.

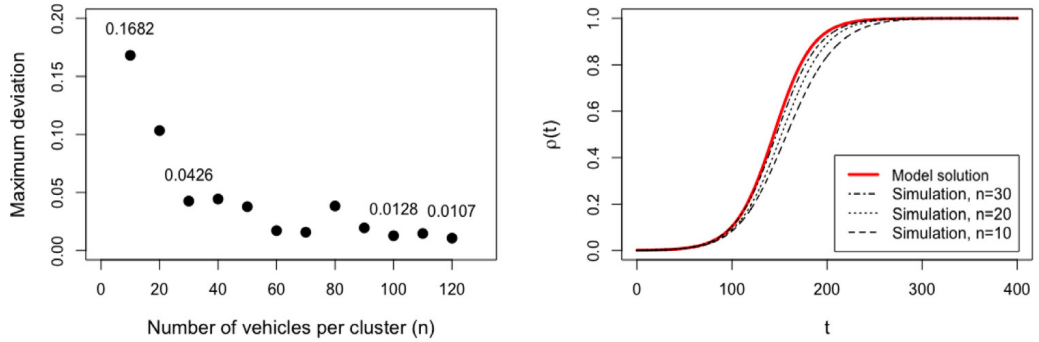
3.1.1. Temporal variation of traffic density and routing

We show that the simulation of the propagation process with the mobility model in Section 2.2.1 closely matches the output of the differential Eq. (3) for a finite number of vehicles. We subsequently use the differential equations to understand the temporal and spatial propagation of the information. The mobility parameter λ in (4) is set to $\lambda = 0.1$. We set $n = 100$, thus $N = n \cdot J = 12000$. Under these settings, we compare the solutions of the corresponding differential equations with an average of 200 runs of the propagation process simulations. Then, the results of the propagation process and the solution of the corresponding differential equations are compared. We consider different γ values reflecting the temporal variation of traffic density and routing. Fig. 4a shows that the simulations of the propagation process closely match the solutions of the differential equations. The largest deviation between them is 0.0256 for $\gamma = 1$, 0.0101 for $\gamma = 3$, and 0.0222 for $\gamma = 5$. Now, we investigate the impact of a relatively small number of vehicles on our model. For non-rush hour ($\gamma = 1$), Fig. 5a shows that the greater the number of cars, the smaller the maximum deviation between the two results. Even for a small n, the solution of the differential equations well approximates the simulation of the propagation process: the maximum deviation between the two is (a) 0.0426 for $n = 30$ and (b) 0.1682 for $n = 10$. This means that asymptotic results are obtained even for a very small number in practice.



(a) Fraction of informed vehicles (b) Fraction of vehicles in CBD

Fig. 4. The gray points represent the average of 200 simulation runs of the information propagation, and the red lines are the solutions of the ordinary differential equations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a) Maximum deviation (b) $\rho(t)$ of changing the number of vehicles per cluster

Fig. 5. Both (a) and (b) show that with the increase in the number of vehicles per cluster, simulation results approach the solutions of the differential equations. We consider the average of 200 simulation runs in both cases. Even for $n = 30$, the maximum deviation is as little as about 0.0426.

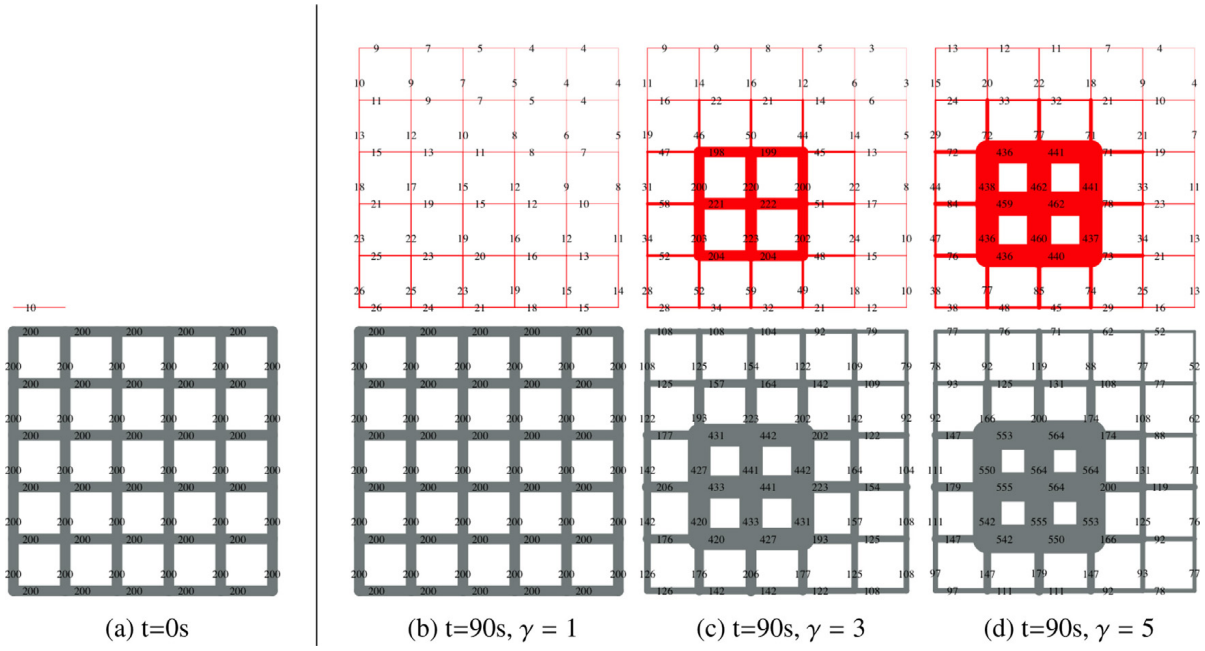
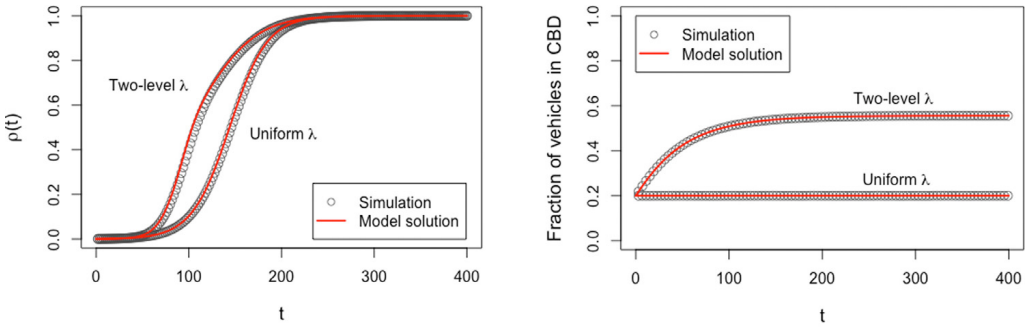


Fig. 6. Geographical representation of traffic density and information propagation for various γs under the same initial conditions with $n = 100$ (i.e., 200 vehicles per road segment). The thickness of the road segment is linearly proportional to the number of vehicles, that is, when the thickness corresponding to one vehicle is x, the thickness of the vehicle v is expressed by $v \cdot x$. The number written on each road segment indicates the number of vehicles. The upper row of red corresponds to the number of vehicles informed in each road segment, and the lower row of gray corresponds to the total number of vehicles in each road segment. The first column shows the initial distribution of vehicles, which is the same for all γ. The remaining columns show the distribution of vehicles at time $t = 90$s for different γ. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Now that we have verified that the differential Eq. (3) can capture information propagation reasonably accurately even with a moderate number of vehicles, we now use them to understand the characteristics of information propagation. In the case of non-rush hour ($\gamma = 1$), as shown in Fig. 4b, the information slowly spreads over the entire area without an upsurge in particular areas. On the other hand, in the case of morning rush hour ($\gamma > 1$), there is a significant increase in the number of vehicles in the CBD, and as a result, information spreads very quickly. The larger the γ, the higher the traffic density in the CBD; thus, the information spreads quicker in the CBD than in the periphery, as shown in the geographical representation (Fig. 6).

3.1.2. Location dependent mobility rate

We now consider the location dependent mobility model introduced in Section 2.2.2. In this subsection, two different mobility rate values are applied to roads in the CBD and the periphery respectively, which we call *two-level mobility rate*. We set the mobility rate of the CBD to a lower value, reflecting a lower speed limit in the CBD. For the two-level mobility rate case, we choose $\lambda_d = 0.05$ for the CBD and $\lambda_o = 0.1$ for the periphery in (6); for the uniform mobility rate case, we choose $\lambda_d = \lambda_o = 0.1$ for both the CBD and the periphery. The number of vehicles per cluster is set to $n = 100$ at $t = 0$. Also, γ is assumed to be 1 (non-rush hour). Fig. 7a shows that the propagation process for this two-level mobility rate is also



(a) Fraction of informed vehicles (b) Fraction of vehicles in CBD

Fig. 7. The curves and points of the two-level λ are the result of applying a different λ value depending on whether the region is CBD or peripheral region, and the uniform λ case is the result of applying the same λ regardless of region. The simulation results are averaged over 200 runs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

well approximated by solutions of the corresponding differential Eq. (3): The maximum deviation between them is only 0.0484. We now compare the propagation of information under the two-level mobility rate with that for uniform mobility rate. Since the mobility rate in the CBD is lower than in the periphery, the average speed of vehicles in the CBD is slower. Therefore, even in the case of non-rush hour with $\gamma = 1$, once vehicles enter the CBD, the vehicles in the CBD take longer to move, resulting in the concentration of vehicles in the CBD (Fig. 7b). Concentrated traffic leads to faster information propagation as shown in Fig. 7a.

3.1.3. Traffic density dependent mobility model

Next we consider the mobility model of Section 2.2.3 where the mobility rate depends on the traffic density of both the origin and destination clusters. We have the mobility rate depend on the traffic densities of both the origin and destination clusters. For (7), we choose $\lambda = 0.1$. Recall that the sensitivity of the mobility rate to traffic density is determined by parameters a and b of (7). As Fig. 8a shows for fixed b, the mobility rate λ_{jk} decreases with traffic density of cluster j and k; this is sharp for relatively small a and more gradual as a increases. To illustrate, we fix the parameter $b = 40$ and compare the results for $a = 1$ and $a = 5$. In both cases, we consider the movement pattern of the morning rush hour ($\gamma = 5$). As Fig. 8b shows, the differential equations closely approximate the propagation process; the maximum deviation between them is 0.0304 at $a = 1$ and 0.0208 at $a = 5$.

Now using the geographical representation of information propagation in Fig. 9, we investigate how information propagates for the traffic density dependent model. Since we are considering morning rush hour, the vehicles congregate in the CBD over time, thus information propagates faster therein. This phenomenon is pronounced for larger values of a as the mobility rate increases with a given fixed b and traffic density as can be seen in Fig. 8a.

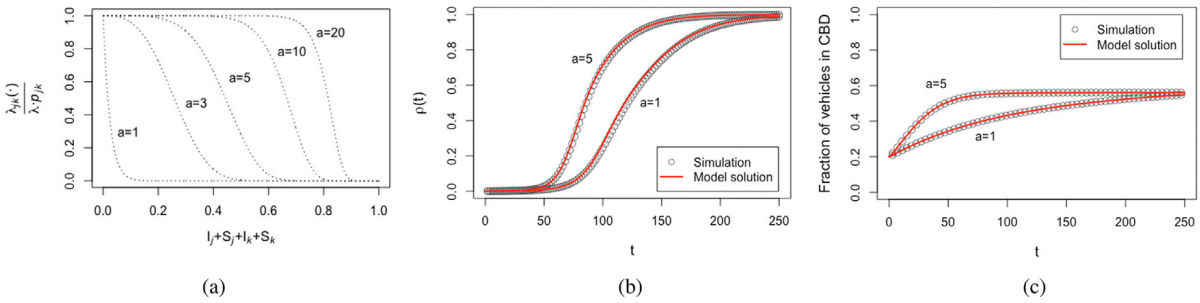


Fig. 8. (a) The relation between mobility rate divided by a constant factor ($\frac{\lambda_{jk}(\cdot)}{\lambda \cdot p_{jk}}$) and traffic density ($I_j + S_j + I_k + S_k$), for a given origin cluster j and destination cluster k. (Recall that the mobility rate in Section 2.2.3 is given by $\lambda_{jk}(\cdot) = \lambda \cdot p_{jk} \cdot [1 - (I_j + S_j + I_k + S_k)^a]^b$) (b) Fraction of informed vehicles over time. (c) Fraction of vehicles located in CBD over time. For both (b) and (c), the gray points are simulation results of averaging 200 simulation runs, and the red lines are the solutions of the approximate ordinary differential equations. For all three figure, $b = 40$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

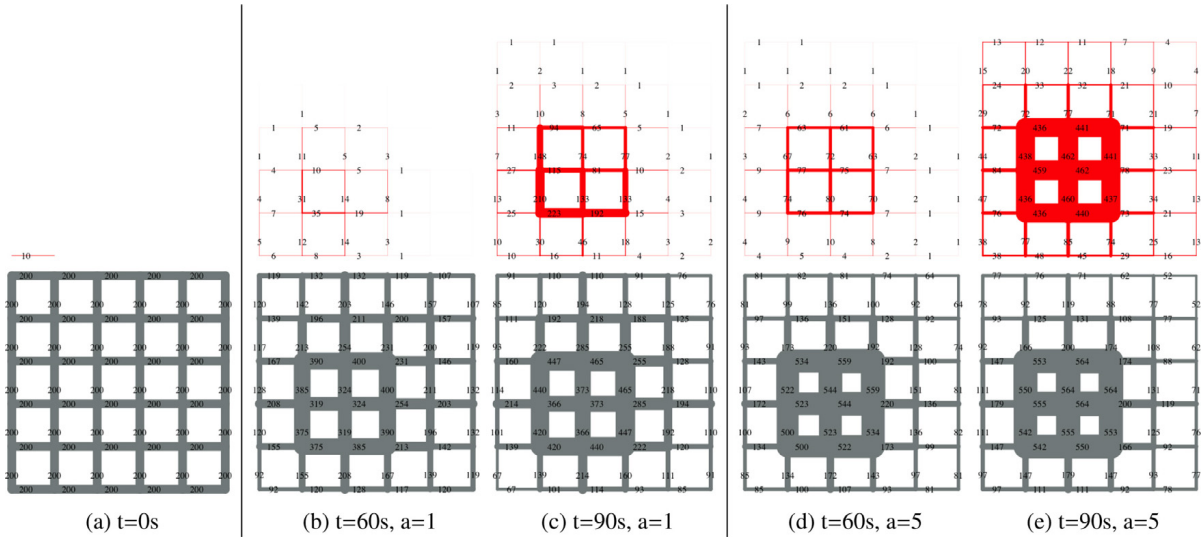


Fig. 9. The upper row of red corresponds to the number of informed vehicles in each cluster, and the lower row of gray corresponds to the number of vehicles in each cluster. The thickness of the road segment is linearly proportional to the number of vehicles, that is, when the thickness corresponding to one vehicle is x, the thickness of the vehicle v is expressed by $v \cdot x$. The first column represents the initial distribution of vehicles, which is the same for all values of a. The remaining columns show the times and the values of a as indicated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1.4. Shortest-path routing

In our model, the routing probabilities of the vehicles at the clusters depend only on the cluster but not on the destinations of the vehicles. Thus, with positive probability, vehicles retrace their paths. Retracing does arise in practice, for example, when people drive around for sight-seeing, trying to locate some eatery rather than a specific eatery, or an available parking spot. A compilation of ten studies in eight cities between 1927 and 2011, reveal that an average of 34% of drivers cruise for parking in congested downtowns (Pierce and Shoup, 2013). Next, in Carpooling services such as Uber Pool and Lyft Line, vehicles frequently move back and forth to pick-up and drop-off various riders who share the ride at their different sources and destinations. Then again vehicles that deliver goods or transport passengers, e.g., shared ride vehicles, taxis, (even without car-pooling) choose different destinations in quick succession, to transport different goods and different passengers. In the process, although they do not retrace their paths while traversing to a destination, with a positive probability they retrace an earlier path while traversing to another destination, depending on the choice of successive destinations. Thus, paths are retraced over time.

We next compare the solution of our model with the characteristics that emerges when vehicles do not retrace their paths. Towards that end, we consider the propagation of messages when every vehicle follows the shortest path from its source to its destination. This is a special case of an extensively utilized probabilistic routing model in the transportation community (Yperman et al., 2005; Gentile, 2015) in which a vehicle randomly chooses its next hop at each node, with the probabilities of the choice depending on the destination of the vehicle in question. In our special case, each vehicle chooses at each node with probability 1 the next hop corresponding to the shortest path to its destination. We consider that the destination is chosen at the start of the travel, with uniform probability among all clusters. We also assume that when a vehicle arrives at its destination, the vehicle chooses another random destination cluster, uniformly, and travels along the shortest path route again. This is for example consistent with traveling patterns of delivery vehicles, shared rides (without car pooling) and taxis. These vehicles proceed along a recommended path to their chosen destination, which is often the shortest path or close to it, and after reaching the destination, choose another destination for transporting their goods or passengers, and follow the shortest path to it, and so on.

We set $\lambda = 0.05, n = 50$, thus $N = nJ = 6000$. We average over 10 runs of the shortest path routing simulations. We then compare the solutions of the corresponding differential equations with the simulation result of the shortest path propagation process, averaged over 10 runs. We extract the routing probabilities from the average of the 10 runs of the shortest path propagation process, and substitute them into the differential equations to get the model solution. As can be seen in Fig. 10a, considering the averages over all clusters, the differential equations closely approximate the propagation process. Next, we focus on two specific example clusters respectively located in the center and upper-right corner. As Fig. 10b shows, despite fluctuations in the number of informed vehicles over time in both clusters, the spatial and temporal propagation of information is relatively well approximated by the theoretical model that operates under the assumption of memoryless probabilistic routing at each cluster. Fig. 10c shows that the match only improves, and the fluctuation of the simulation results decreases, as we increase the number of runs to 100. The match is close because, as mentioned previously, vehicles that choose different destinations in quick succession retrace some of their paths over time (with positive probability). The close match also suggests that in the memoryless probabilistic routing that we have assumed the probability that routes are retraced is not high.

3.2. Empirical validation with traffic trace data

When the mobility process is exponential, even for a finite number of vehicles, the solution of the differential equations closely approximates the dynamics of the propagation process (Section 3.1). In this section, we show that the temporal and spatial flow of information is well approximated by the solution of differential equations even when the mobility process is not exponential, and the number of vehicles N is finite. Towards that end, we use microscopic vehicle trajectory data

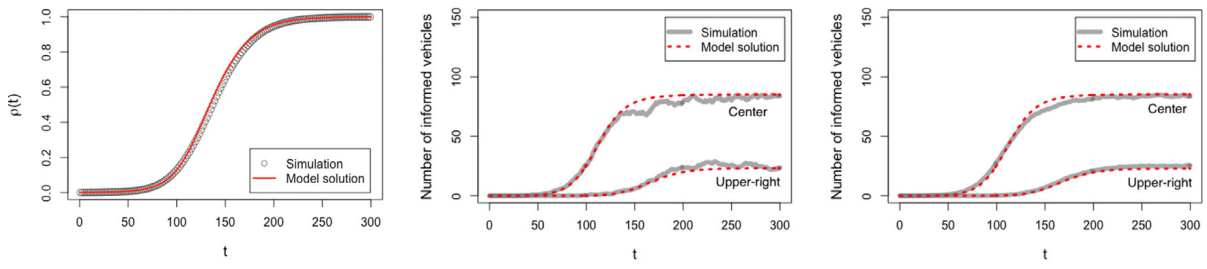


Fig. 10. (a) Fraction of informed vehicles over time from the average of 10 runs of the shortest path routing. (b) The number of informed vehicles over time in the cluster located at the center and the upper right corner. The gray lines represent an average of 10 simulation runs, and the red lines represent the model solution obtained by using the routing probabilities from the average of the 10 runs of the shortest path propagation process. (c) As we increase the number of runs to 100, the match improves, and the fluctuation of the simulation results decreases. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

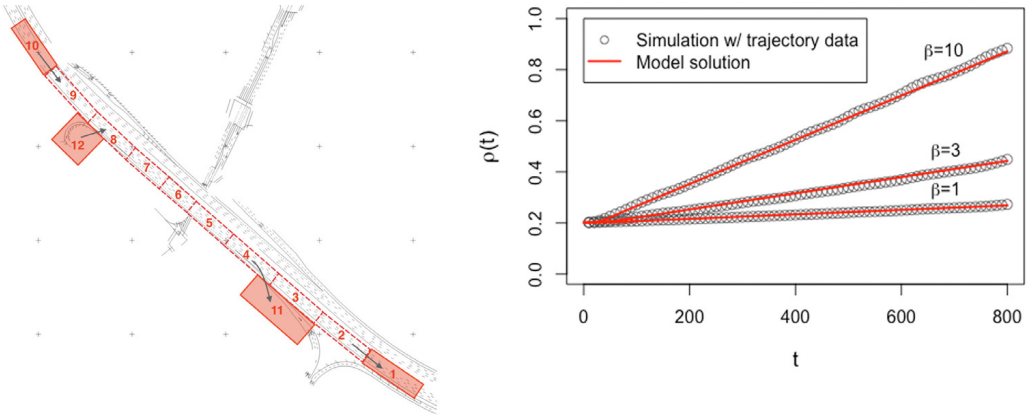
collected to empirically validate our model. Thus far, we have only considered grid topology, but we will also consider other topologies in this section. In this case, we show that the output of the differential equations reasonably matches the dynamics of information propagation for various road topologies.

The three vehicle trajectory data we use are collected from different road topologies of varying complexity, unlike the grid topologies considered earlier; U.S. Highway 101 in Los Angeles, California (Section 3.2.1), Peachtree Street in Atlanta, Georgia (Section 3.2.2), and Europarc Roundabout in Creteil, France (Section 3.2.3). We manually divide each topology into J clusters, and define the mobility network G and the communication network G'. The directed mobility network G, which consists of directed edges between clusters, is determined by the existence of a trajectory in which the vehicle moves from one cluster to another. To indicate that vehicle-to-vehicle communication in the cluster is possible, the diagonal element of the adjacency matrix corresponding to the communication network is set to 1 as a baseline for all three data sets. In addition, if the distance between neighboring clusters is close enough, the corresponding element is set to 1 to enable inter-cluster communication according to the characteristics of the road. The mobility rate between clusters is extracted from the vehicle trajectory data, and is applied to the ordinary differential equation of our model to estimate the information propagation. We superimpose the statistical communication process on the trajectory data, and compare the result with the solution of the differential equations. As a result, not only does this show that our model is applicable to arbitrary road topologies, but it also shows that simulation results using even trajectory data are well approximated by model solutions. To extract the mobility rates from the data, we will first classify the clusters into three categories. Let S be the set of clusters of the study area where information propagation occurs. Note that the mathematical model considers the fixed set of vehicles in the system. In the real transportation network, there would be entrances from outside and also exit to outside. To incorporate the impact of the entrances and exits, we introduce a set of virtual clusters A and B where A and B are the respective sets of clusters corresponding to the entry and exit roads respectively entering and leaving the study area. Let $O = S \cup A \cup B$. Here we consider that the mobility rate of a vehicle does not depend on whether it is informed or non-informed.

We now describe how we obtain mobility rate $\lambda_{i,j}$ of the analytical model from the trajectory data ($\lambda_{i,j}$ moving from cluster $i \in S$ to $j \in O$, such that $i \neq j$). First, from the trajectory data, we compute λ_i, which is the reciprocal of the average time of staying in cluster i. Then we compute the fraction of vehicles that move to cluster j among the vehicles located in cluster i, and denote this as $p_{i,j}$. By multiplying λ_i by probability $p_{i,j}$, the mobility rate from cluster i to j, $\lambda_{i,j}$, can be computed. If the origin cluster of the movement is the entry road, it is assumed that a vehicle enter the study area at a fixed rate, regardless of the number of vehicles in the entry cluster. In this case, the mobility rate moving from cluster $i \in A$ to $j \in S$, such that $i \neq j$, is given by $\widetilde{\lambda}_{i,j}/N$ where N is the total number of vehicles and $\widetilde{\lambda}_{i,j}$ denotes the number of vehicles that enter cluster $j \in S$ from cluster $i \in A$ per unit time.

3.2.1. U.S. Highway 101 in Los Angeles, California

We use the microscopic actual vehicle trajectory data of the southbound U.S. highway 101 in Los Angeles, California, that had been collected under the auspices of the Next Generation Simulation (NGSIM) program (dataset U.S. Department of Transportation, 2017). This data includes geographical location information for each vehicle on the southbound U.S. highway 101 of 600 m in length. We consider the total number of vehicles that exist in the system at any point during the observation period ($N = 1993$). As shown in Fig. 11a, we divide the road into $J = 12$ clusters. The study area in which we conducted information propagation studies is a set of clusters $S = \{2, 3, \ldots, 9\}$, which is depicted as a dotted rectangle in Fig. 11a. Vehicles enter the system through clusters {8, 9}, and leave the system from clusters {2, 4}. We introduce a set of



(a) (b)

Fig. 11. (a) U.S highway 101 broken into clusters. The vehicles enter the study area from the cluster 10 or 12 and leave the study area toward the cluster 1 or 11. (b) Fraction of informed vehicles over time. The information spreading simulation based on the trajectory data is well approximated by the theoretical predictions from the ordinary differential Eq. (3). The gray points are simulation results of averaging 30 simulation runs for each β, and the red lines are the solutions of the differential equations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

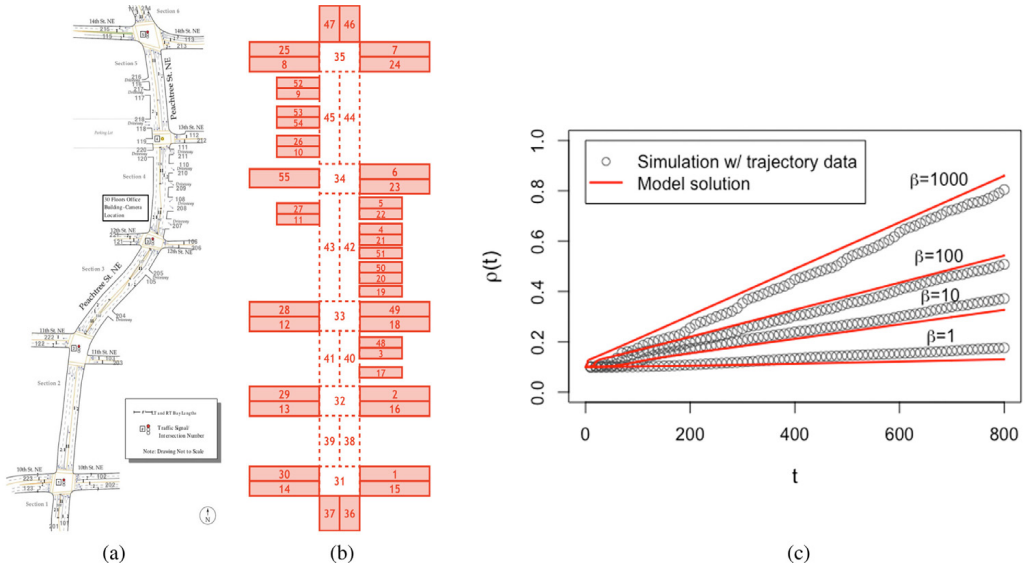


Fig. 12. (a) Peachtree street schematic (dataset U.S. Department of Transportation, 2017). (b) Clustered Peachtree street (c) Fraction of informed vehicles over time. The information spreading simulation based on the actual trajectory data is well approximated by the theoretical predictions from the ordinary differential Eq. (3). The gray points are simulation results of averaging 30 simulation runs for each β, and the red lines are the solutions of the differential equations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

virtual entry clusters and exit clusters as shown in Fig. 11a; thus $A = \{10, 12\}$ and $B = \{1, 11\}$. The set of clusters A and B corresponding to the entry and exit roads respectively are represented by shaded solid line rectangles. The actual trajectory data for the first 831.7 seconds out of the entire data was used for this study, and all vehicles entering the study area were regarded as separate vehicles. There is no vehicle in the study area at an initial time, and we assumed that approximately 20% of all incoming vehicles, that is, 402 out of 1993 incoming vehicles, had already received the information before they entered the study area. The adjacency matrix for the communication network is set to $a'_{ij} = 1$ if $i \in S$ and $i = j$, otherwise $a'_{ij} = 0$. As shown in Fig. 11b, there is an excellent match between the simulation result using the actual trajectory data and the solutions of corresponding differential equations. The maximum deviations between the average of 30 simulation runs and the solution of the differential equation are 0.0039 for $\beta = 1$, 0.0112 for $\beta = 3$, and 0.0151 for $\beta = 10$.

3.2.2. Peachtree Street in Atlanta, Georgia

We use the actual microscopic vehicle trajectory data of the Peachtree street in Atlanta, Georgia, that had been also collected under the auspices of the Next Generation Simulation (NGSIM) program (dataset U.S. Department of Transportation, 2017). This data includes geographical location information for each vehicle on the two-way street of 640 meters in length with 5 intersections, which are more complex than the previous one-way road topology. We consider the total number of vehicles that exist in the system at any point during observation period ($N = 2298$). As shown in Fig. 12b, we divide the road into $J = 55$ clusters. The actual trajectory data for 1044.2 seconds was used for this study, and all vehicles entering the study area were considered separate vehicles.[2] In addition, we clean up the data to exclude obvious instances of data error and vehicles which do not enter the study area; we consequentially use 98.4% of the original number of vehicles. As shown in Fig. 12b, the study area in which we conducted information propagation studies is a set of clusters $S = \{31, 32, \ldots, 35, 38, 39, \ldots, 45\}$. Vehicles enter the study area from the virtual clusters $A = \{1, 2, \ldots, 14, 36, 47, 49, 50, 54, 55\}$, and leaves the study area through the virtual clusters $B = \{15, 16, \ldots, 30, 37, 46, 48, 51, 52, 53\}$. We consider that the vehicles in the study area are not informed at an initial time, but we assumed that 30% of vehicles entering the study area from the cluster $36 \in A$ and $47 \in A$, that is, 231 out of 770 incoming vehicles, had received the information before they enter the study area. The adjacency matrix for the communication network is set to $a'_{ij} = 1$ if $i \in S$ and $i = j$. In addition, given their proximity, $a'_{38,39} = a'_{39,38}$, $a'_{40,41} = a'_{41,40}$, $a'_{42,43} = a'_{43,42}$, and $a'_{44,45} = a'_{45,44}$ are also set to 1, and all other elements are set to zero. Under these conditions, the information propagation simulation using the actual trajectory data on Peachtree street is well approximated by the model solution as shown in Fig. 12c. The maximum deviations between the average of 30 simulation runs and the solution of the differential equations are 0.0456 for $\beta = 1$, 0.0448 for $\beta = 10$, 0.0376 for $\beta = 100$, and 0.0647 for $\beta = 1000$.

[2] In the original dataset (dataset U.S. Department of Transportation, 2017), we used data with `Global_time` attribute values between 1163019100 and 1164063300. The `Total_Frames` attribute represents a total number of frames in which the vehicle appears in the system. The `O_Zone` and `D_Zone` attributes represent the place where the vehicles enter and exit the system respectively. There are records with the same value of `Vehicle_ID` but different values of `Total_Frames`, `O_Zone`, and `D_Zone` attributes. In this study, every time a vehicle enters the system, we have considered it a new vehicle even if the value of `Vehicle_ID` is same.

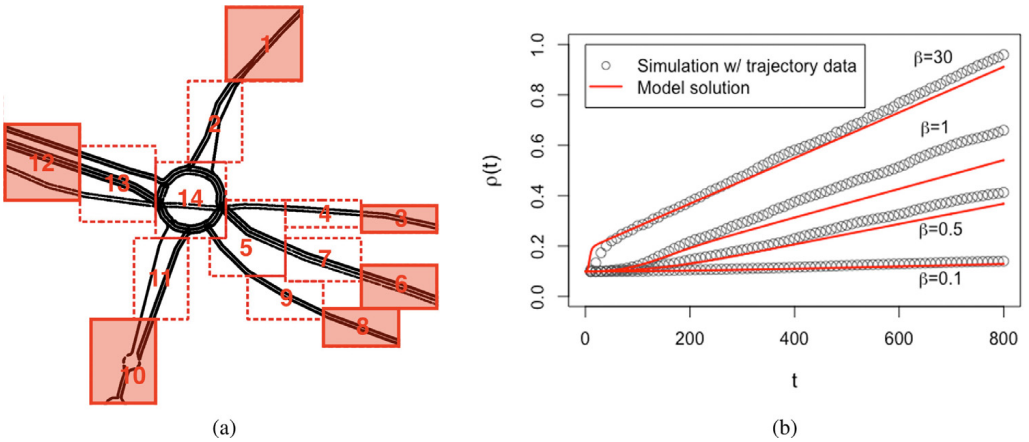


Fig. 13. (a) Clustered Europarc roundabout road. (b) Fraction of informed vehicles over time. The information spreading simulation based on the trajectory data is well approximated by the theoretical predictions from the ordinary differential Eqs. (3). The gray points are simulation results of averaging 30 simulation runs for each β, and the red lines are the solutions of the differential equations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.3. Europarc Roundabout in Creteil, France

We use the Microscopic vehicle trajectory data of the Europarc Roundabout in Creteil, France, that had been generated by Lèbre et al. (2015). According to Lèbre et al. (2015), traffic information for Europarc roundabout was based on actual observations of vehicle flow, and manual counting were performed to generate origin/destination (O/D) matrix. This O/D matrix faithfully mimics the daily movement of the vehicle, from which a realistic synthetic data set of vehicle mobility is presented by Lèbre et al. (2015). We choose 800 seconds from the morning traffic data (7.15 AM to 9.15 AM) which have the peak number of vehicles (the peak occurs between time steps 4200 to 5000). This topology is much more complex than the previous two; specifically, this is not a grid topology, as it includes a roundabout road and 15 traffic lights. As shown in Fig. 13a, we divide the road into $J = 14$ clusters. The study area in which we conducted information propagation studies is a set of clusters $S = \{2, 4, 5, 7, 9, 11, 13, 14\}$. Vehicles enter the road from the set of virtual clusters $A = \{1, 3, 6, 10, 12\}$, and leave the study area toward the set of virtual clusters $B = \{1, 3, 10, 12\}$. Suppose that all vehicles entering the study area were regarded as separate vehicles. We consider that the vehicles in the study area are not informed at an initial time, but we assumed that approximately 11% of vehicles entering the study area from the set of clusters A, that is, 66 out of 591 incoming vehicles, had received the information before they enter the study area. The adjacency matrix for the communication network is set to $a'_{ij} = 1$ if $i \in S$ and $i = j$, otherwise $a'_{ij} = 0$.

As shown in Fig. 13b, there is a reasonable match between the solution of the differential Eq. (3). The maximum deviation between the average of 30 simulation runs and the solution of the differential equation is 0.0124 for $\beta = 0.1$, 0.0507 for $\beta = 0.5$, 0.1185 for $\beta = 1$, and 0.0741 for $\beta = 30$.

The model solution and the simulation results match well for data obtained from U.S. Highway 101, but for the other two trace data sets, there is a higher deviation between the model solution and the simulation. First, the total number of vehicles the for Europarc Roundabout in Creteil, France, is fairly low compared to the previous two data sets. More importantly, unlike the first data obtained on the U.S. Highway 101, the traffic movement becomes pulsed in the other two data sets because of the presence of many traffic lights (5 signals exist on Peachtree Street and 15 signals exist on Europarc Roundabout) in a segment of the topology. Each traffic signal results in traffic synchronization, because, all vehicles stop at a red light and start moving almost simultaneously when the light turns green. This causes a significant divergence between the actual mobility of this trajectory and the exponential mobility process, which our mathematical framework has assumed. The influence of multiple traffic signals is even more complicated due to possible correlations between the durations of red lights and green lights at each.

Modeling information propagation with pulsed traffic, which is due to traffic signals, constitutes an open research challenge. We briefly mention some possible directions for solving this open problem and postpone the details for our future research. We can consider a time-dependent mobility parameter $\lambda_{i,j}(\cdot)$ which equals 0 when the light in the corresponding cluster is red, and is at the normal values when it is green.

We elucidate the impact of traffic synchronization due to traffic lights, utilizing our modeling innovation outlined above, considering a simple example. Consider a one-way road with two traffic signals (Fig. 14a). The road is divided into 16 small segments. The study area where information propagation occurs is the set of clusters $S = \{6, 7, \ldots, 15\}$. Vehicles enter the study area from the set of clusters $A = \{1, 2, \ldots, 5\}$ and leave the study area toward the cluster $B = 16$. Traffic signals are located between clusters 8 and 9 and between clusters 12 and 13. The traffic signals have cycle lengths of 80 s, and 40% of the cycle time is spent in red. Both signals begin their cycles in red at the start of the simulation. 1500 vehicles are uniformly distributed in the entry clusters {1, 2, 3, 4, 5}, and there is no vehicle in the study area and exit road at an initial time. We assume that 10% of all incoming vehicles had already received the information before they entered the study area.

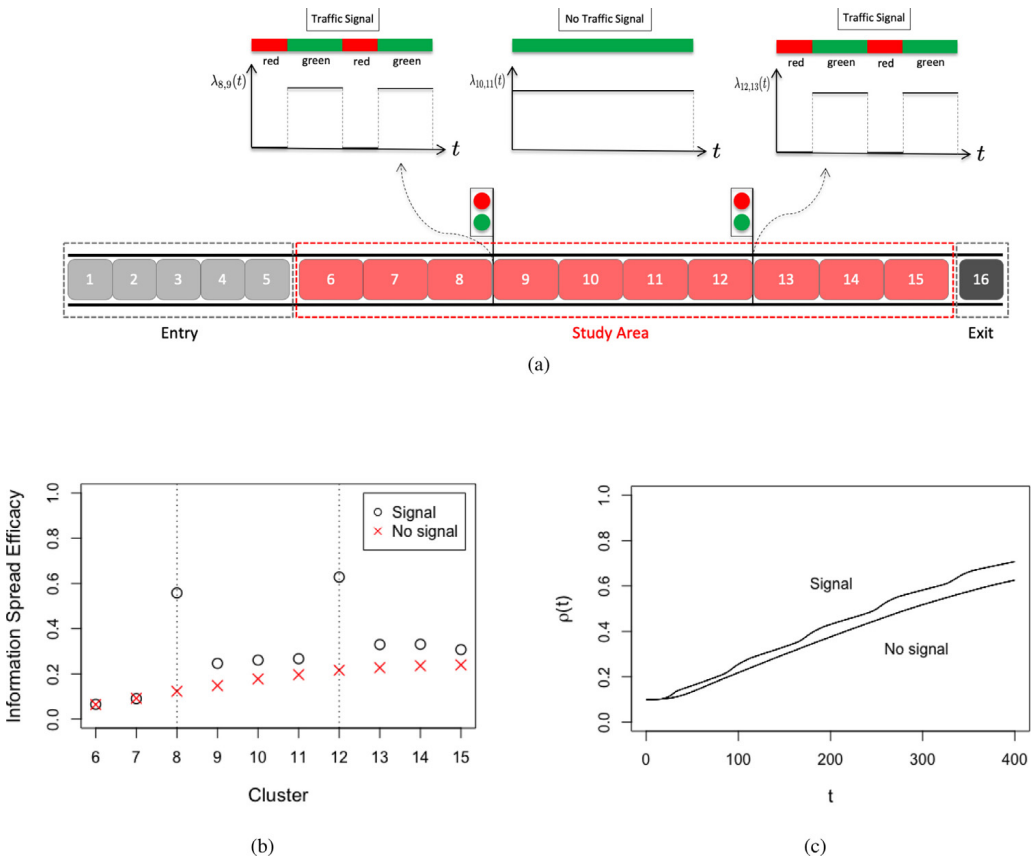


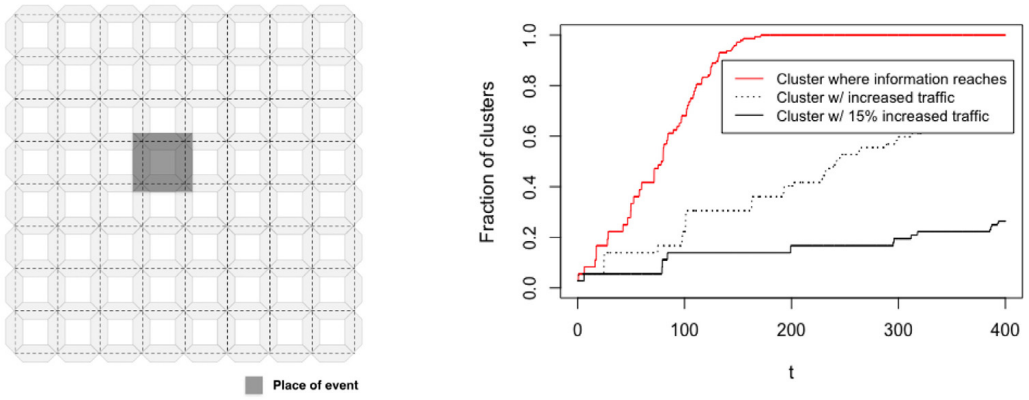
Fig. 14. (a) Clustered road topologies. (b) We plot the ratio of the number of vehicles that are informed at a cluster to the number of non-informed vehicles that pass by it in the entire simulation time horizon, as a function of the cluster number in the study area. The ratio spikes exactly at those clusters 8 and 12 located just before the two traffic signals. (c) We simulate the spread of messages in a straight road segment, with 16 clusters. The lines represent the average of 200 simulation runs of the information propagation.

The mobility parameter $\lambda_{i,j}(\cdot)$, for $i \in \{8, 12\}$ where there are traffic signals, switch along with the lights: they are 0 when the lights are red, and 0.2 when they are green. The mobility parameter $\lambda_{i,j}(\cdot)$, for $i \in S\backslash\{8, 12\}$ where there is no traffic signal in the study area, is set to 0.2, and $\lambda_{i,j}(\cdot)$, for $i \in A$ in the entry clusters, is set to 0.01. The communication parameter β is set to 10.

Fig. 14b shows that V2V messages propagate rapidly among vehicles when they wait at red lights. The ratio of the number of vehicles that are informed at a cluster to the number of non-informed vehicles that pass by it spikes exactly at the clusters just before the two traffic signals. Thus, traffic signals have a significant impact on information propagation. The traffic movement becomes more synchronized, due to the presence of these lights, which results in faster information propagation (Fig. 14c). When there are many traffic signals on more complex roads, multiple traffic signals affect traffic flows in different directions simultaneously, which affects information propagation in a more complicated way. Our future study will investigate this mathematical model and its generalizations, through mathematical analysis, numerical computations and simulations, for a variety of traffic signal designs.

4. Use of the model

We have shown an excellent match between the solution of the approximated differential equation and the propagation process results for a finite number of vehicles in the statistical synthetic models that reflect diverse mobility and communication characteristics (Section 3.1). We also showed that there is a fairly good match between the two results, even for trajectory data that does not satisfy the statistical assumptions under which convergence is guaranteed (Section 3.2). Now, through concrete real world examples, we will explain how these differential equation based models for information propagation can be utilized. As in the result section on the synthetic model, we assume a grid topology consisting of two-way roads, but we consider a larger city consisting of nine avenues and nine streets with the total number of clusters $J = 288$. Likewise, the adjacency matrix of the communication network $G(V, E')$ is given by $a'_{ij} = a'_{ji} = 1$ if i and j are in the same road segment, otherwise $a'_{ij} = a'_{ji} = 0$. We apply the model with temporal variation of the traffic density and routing (Section 2.2.1). The communication parameter is set to $\beta_{ij} = \beta$ if $a'_{ij} = 1$ and $\beta_{ij} = 0$ otherwise, where β is a constant. Since



(a) Location of the event (b) Fraction of clusters

Fig. 15. (a) Clustered grid road topology. Location of the event is represented by a dark shaded square. (b) The red solid line represents the percentage of clusters in which information arrives, the solid black line represents the percentage of congested clusters that have increased by more than 15% over the initial traffic volume. The black dashed line represents the fraction of clusters with increased traffic (i.e. the fraction of clusters increases by more than one from the initial number of vehicles). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

we have verified the validity of the approximation by differential equations, we now use differential equations to understand information propagation characteristics in this section.

4.1. Unexpected events

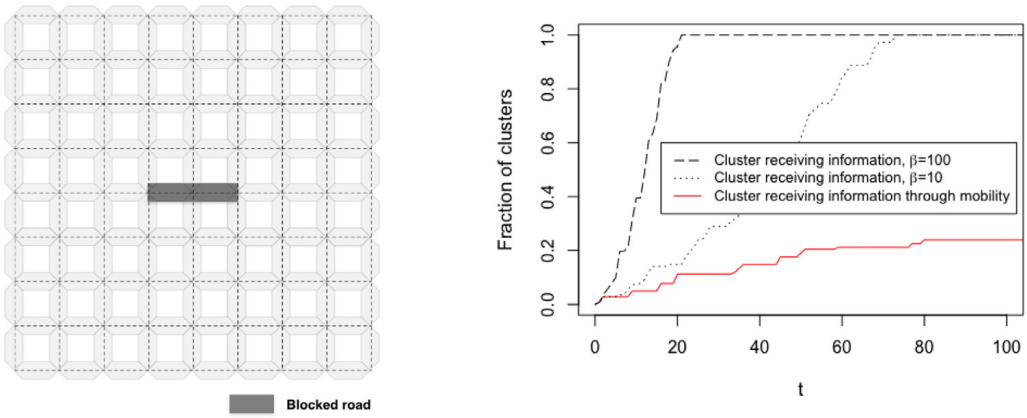
We now assess the efficacy of V2V technology in mitigating traffic congestion due to sudden events e.g., unexpected victories in a major sporting event, unscheduled personal visits by high-ranking officials (e.g., president) and celebrities) by studying how quickly V2V can help spread information about congestion.

We model an unexpected event occurring in a central area of a city with a grid topology (Fig. 15a). It is also assumed that 12.5% of the total number of vehicles $N = 14400$ are located on 4 road segments (equivalently, 8 clusters) surrounding the event site, and 45 vehicles are uniformly distributed in all other clusters. We set the parameters to $\beta = 10$, $\lambda = 0.05$, and $\gamma = 1$ (non-rush hour). Using our model, we can estimate how the traffic changes over time due to the events, and investigate how the information about the traffic propagate during the dispersion of the gathered vehicles. Our numerical computations reveal that information about the expected traffic congestion propagates faster than the spread of traffic congestion itself. As can be seen in Fig. 15b, the number of clusters in which information arrives (i.e., the number of clusters in which more than one vehicle receives information) increases much faster than the number of clusters with more than 15% increase in traffic volume. These results have been demonstrated when the unexpected event occurs in the center of the grid. Because unexpected events can occur anywhere in a city, we investigate the impact of the location of the initial informed vehicles on information propagation in Section 4.3.

4.2. Obstructions

In many cases, roads are obstructed for reasons such as traffic accidents or road maintenance. Vehicles upstream of the obstruction must be detoured while vehicles that are in the queue immediately upstream of the obstruction must be discharged. There are several obvious advantages associated with dispersing the location information of obstructions as early as possible. Drivers who are heading for the obstruction can choose an alternate route even from a distance, reducing their own inconveniences as well as traffic congestion around the area. Towards that end, we assess how quickly the location information of the obstruction can be spread through the utilization of V2V technology.

We assume that two road segments in a central area of the city are blocked (Fig. 16a). Information is propagated from queued vehicles, stopped from the obstruction. Let R be the set of clusters that are blocked. If $j \in R$, $a_{ij} = 0$ in the adjacency matrix of the mobility network, reflecting that the vehicle cannot enter the obstructed clusters. The mobility parameter λ is set to $\lambda = 0.05$, and communication parameter β is set to $\beta = 10$ and $\beta = 100$. The total number of vehicles $N = 14400$ is uniformly distributed in the total number of clusters $J = 288$ at initial time, thus the initial number of vehicles per cluster is set to $n = 50$. In Fig. 16, we plot as a function of time the number of clusters in which (a) vehicles containing information on the obstruction location reach, and (b) the initially informed vehicles located on the obstructed road reach. Fig. 16 reveals that the result of the former is significantly higher than the latter, which indicates that the information propagates much faster through V2V communication, as compared to vehicular mobility alone. Fig. 16 also shows that the difference in the extent to which information has been propagated becomes greater as the communication rate β increases, since the larger the β value, the greater the impact of V2V on information propagation.

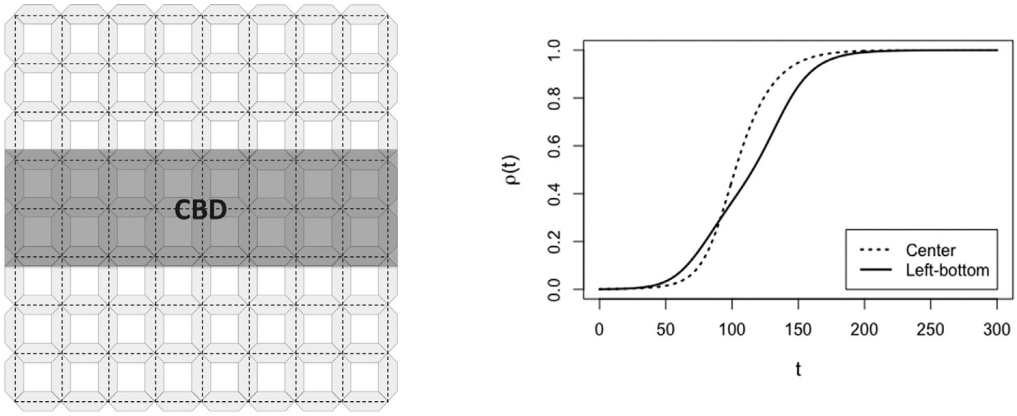


(a) Blocked road (b) Fraction of clusters

Fig. 16. (a) Clustered grid road topology. Blocked road segments are represented by dark shaded squares. (b) The red solid line represents the fraction of clusters in which one or more informed vehicles reach when there is no communication ($\beta = 0$). This represents the rate at which information propagates solely due to mobility. The black dotted and black dashed lines represent the fraction of clusters that one or more informed vehicles reach, in the presence of both communication and mobility, where $\beta = 10$ and $\beta = 100$, respectively. This represents the rate at which information propagates due to the combination of communication and mobility. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Initial locations of informed vehicles

In the previous two examples, we have respectively considered that the unexpected event and the road obstructions occur in the center of the city. However, these can occur in anywhere in the city, so we investigate the impact of the location of initially informed vehicles on information propagation in this subsection. In particular, the impact of the location of initially informed vehicles on information propagation is expected to be more accentuated in habitations that are limited by natural boundaries such as coastal areas (e.g., Hong Kong and Manhattan). Thus, one might think that the message propagation is accelerated if the initially informed vehicles are centrally located because information can propagate simultaneously in all directions towards the boundary; using our framework, we will show that this is not necessarily the case even under the condition that vehicles are uniformly distributed in the entire region at an initial time (equivalently, no initial disparity of traffic density exists). This is because the information propagation speed is also affected by the vehicular movement pattern; we show this counter-intuitive phenomenon through the following example. As an example of a coastal area, we consider the finite grid area like Manhattan and assume that the CBD is located in the middle of the city (Fig. 17a). We set the parameters to $\beta = 10$, $\lambda = 0.1$. The total number of vehicles $N = 14400$ is uniformly distributed at the initial time. We consider the evening rush hour mobility pattern with $\gamma = 0.5$. When information is propagated from the lower left cluster located in the periphery, information propagates rapidly to vehicles located at the bottom of the periphery where traffic density is increasing as can be seen in Fig. 18b. Hence, at an early stage, information propagates faster when it propagates from the lower left corner (Fig. 17b). However, from a certain point onwards, information propagates more quickly when



(a) CBD area (b) Fraction of informed vehicles

Fig. 17. (a) Clustered grid road topology. The CBD area is represented by a dark shade. (b) The fraction of informed vehicles over time for $\gamma = 0.5$ (evening rush hour). The dotted line represents the fraction of informed vehicles when the information propagates from the center, and the solid line represents the fraction of informed vehicles when the information propagates from the bottom-left corner.

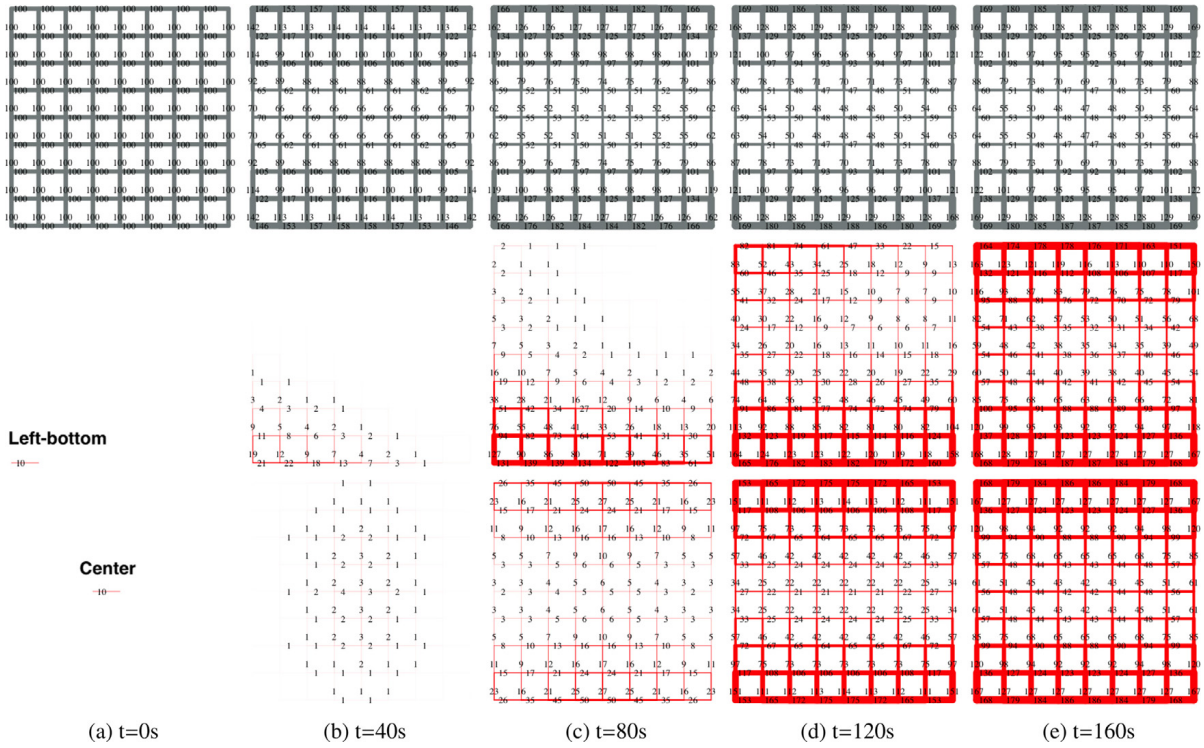


Fig. 18. The first row of gray represents the number of vehicles in each road segment over time, and the second and third rows of red represent the number of vehicles informed in each road segment over time when information is propagated from 10 vehicles located in the lower left cluster and in the center cluster respectively. The thickness of the road segment is linearly proportional to the number of vehicles, that is, when the thickness corresponding to one vehicle is x, the thickness of the vehicle v is expressed by $v \cdot x$. The number indicates the number of vehicles located in each road segment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

information propagates from the center. When the information propagates from the bottom left corner, it takes time until the information is delivered to the top area of the periphery where traffic density is increasing as can be seen in the second row of Fig. 18c. This counter-intuitive phenomenon shows that the effect of the initial location of informed vehicles on the dynamics of the information propagation can vary significantly depending on the temporal variation of the traffic density.

5. Computation time

A significant benefit of our model is that it is computationally tractable regardless of the number of vehicles. The statistics of the information propagation converge to the solution of the differential equations as the number of vehicles goes to infinity. Thus, the result becomes more accurate as the number of vehicles increase, and the computation time to numerically solve the differential equations does not depend on the number of vehicles. The number of variables and the number of differential equations are linear (twice) in the number of clusters. Nonetheless, we show that the computation time is still tractable even for a large number of clusters.

We report the computation time using a computer which is not computationally high-end; a 2.8 GHz Intel Core i7 processor and 16 GB of RAM. We solved the equations using the `ode` function from the **deSolve** package of R with lsoda integration method developed by Petzold (1983). We solve the equations for 200 seconds, producing output every 1 second. We consider the size of a cluster to be the V2V communication range so as to enable all vehicles in the same cluster to communicate with each other. Note that the V2V messages have a range of over 300 meters (NHTSA, 2016). We consider extensive, complex transportation networks with relatively long roads and calculate the number of clusters, and then obtain the computation time. First, a road of approximately 366 km from New York City to Washington, D.C. via the Highway I-95S can consist of approximately 1220 clusters. The computation time required to numerically solve the differential equations for the 1220 clusters is only about 150 seconds. Next, as a more challenging example, consider the entire metropolitan region. The total road network size in Hong Kong is 2107 km (Highway Department of the Hong Kong, 2018), which can be covered with 7024 clusters. The computation time for a one-dimensional road consisting of 7024 clusters is approximately 172 minutes. Lastly, US Highway 101 is a North-South highway that runs through California, Oregon, and Washington State, with a total length of nearly 2,500 km. This entire West coast highway can be covered with 8334 clusters, and the computation time is 240 minutes. These examples show that information propagation between vehicles can be identified within a reasonable time frame even in extreme cases using clustered epidemiological differential equation model.

6. Generalization

We generalize our model to accommodate 1) arbitrary destination dependent probabilistic routing mechanisms (Section 6.1) 2) transmission of multiple messages in overlapping time intervals and RoIs (Section 6.2).

6.1. Destination-based routing

In our model, the routing probabilities of the vehicles at the nodes (clusters) depend only on the node (cluster) but not on the destinations of the vehicles. Some of the extensively utilized probabilistic routing models in the transportation community consider that each vehicle chooses its next arc at each node with a probability that depends on the node in question and destination of the vehicle (Yperman et al., 2005; Gentile, 2015). Considering an important special case of this, namely shortest path routing, we have presented simulations in Section 3.1.4 that reveal that there is a close match with our model, and have explained why the match may be intuited. We do not however have any analytical guarantee on the fit of the model. We therefore generalize our model to accommodate arbitrary destination dependent probabilistic routing mechanisms (as in Yperman et al., 2005; Gentile, 2015). Note, however, that vehicles may still retrace their routes in the arbitrary destination-dependent probabilistic routing mechanisms, as in our model, depending on the choice of the routing probabilities and the network topology.

Recall that we have thus far denoted the fraction of informed and non-informed vehicles in cluster j as $I_j(t)$ and $S_j(t)$ at time t, respectively. We now subdivide each category based on the destinations of the corresponding vehicles. We consider a set of destinations $\{1, 2, \ldots, M\}$ in the transportation network. Each destination corresponds to a class of vehicles headed towards the destination in question, $M+1$ denotes the class of vehicles meandering around rather than seeking a particular destination (e.g., sight-seeing, trying to locate available parking, some eatery rather than a particular one, etc.). We now introduce destination-based mobility rates $\lambda^I_{jk:m}$ and $\lambda^S_{jk:m}$ for informed and non-informed vehicles, respectively, which represents the mobility rate from cluster j to k of vehicles whose destination is $m \in \{1, 2, \ldots, M\}$. These mobility rates incorporate the destination-dependent routing probabilities. As a specific example, if the path from cluster j to k is not in the direction toward the destination m, $m \in \{1, 2, \ldots, M\}$, then the corresponding mobility rates may be chosen as zero, and then these vehicles do not retrace their paths.

Following the approach in Section 2.1, the system may be modeled as a continuous-time Markov chain (if the sojourns of the vehicles in the clusters and the intervals between communications are exponentially distributed). Similar to the proofs in Section Appendix A, it may be shown that Lemma 1 and Kurtz's theorem (Existing result 1) extend to this case, and the asymptotic fraction of informed and non-informed vehicles across clusters converge to the solution of a system of ordinary differential equations. Towards describing these differential equations, we let $I_{j:\,m}(t)$ and $S_{j:\,m}(t)$ denote the fraction of informed and non-informed vehicles of class m in cluster j at time t, for $m \in \{1, 2, \ldots, M+1\}$, respectively. The ordinary differential equations in question are:

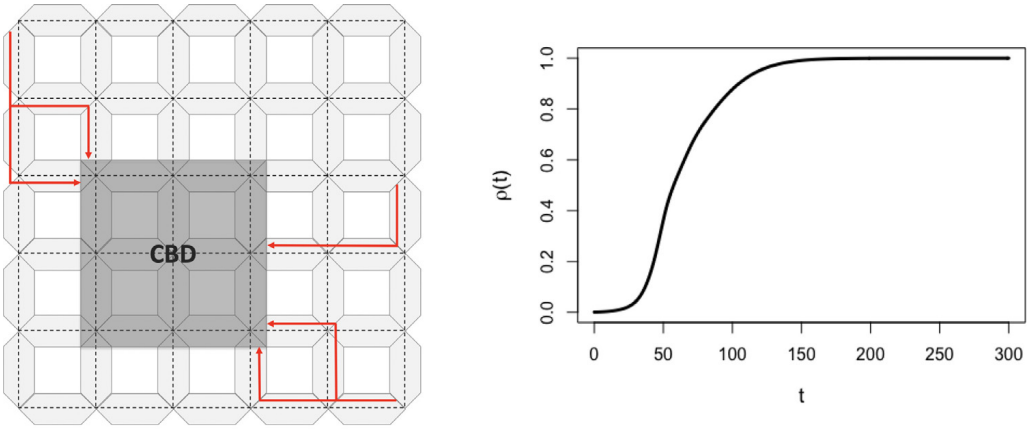
$$\dot{I}_{j:m}(t) = -\sum_{k\neq j}^{J} \lambda^I_{jk:m}(\mathbf{I},\mathbf{S}) \cdot I_{j:m} + \sum_{k=1}^{J} \beta_{kj} \cdot \sum_{m=1}^{M+1} I_{k:m} \cdot S_{j:m} + \sum_{k\neq j}^{J} \lambda^I_{kj:m}(\mathbf{I},\mathbf{S}) \cdot I_{k:m} \qquad (j = 1, 2, \ldots, J),$$

$$\dot{S}_{j:m}(t) = -\sum_{k\neq j}^{J} \lambda^S_{jk:m}(\mathbf{I},\mathbf{S}) \cdot S_{j:m} - \sum_{k=1}^{J} \beta_{kj} \cdot \sum_{m=1}^{M+1} I_{k:m} \cdot S_{j:m} + \sum_{k\neq j}^{J} \lambda^S_{kj:m}(\mathbf{I},\mathbf{S}) \cdot S_{k:m} \qquad (j = 1, 2, \ldots, J).$$

We now consider the computation time for the above system of differential equations. The number of variables and the total number of differential equations are now $2(M+1)J$ each, rather than $2J$ (recall that J is the total number of clusters). Thus, the number of differential equations increase linearly in the number of destinations. Utilizing characteristics of transportation networks that arise in practice, we suggest some values of M, next.

First, in all large cities there exist some important destinations (such as airports, shopping malls, colleges, hospitals, etc.), referred to as trip attractors or special generators in the transportation community (Mamun et al., 2010), that attract most of the traffic. For example, six and thirteen sites were treated as special generators to estimate trip attraction rates in Chittenden County in Vermont, U.S. and Des Moines Area in Iowa, U.S., respectively (Resource Systems Group, 2008; Des Moines Area MPO, 2006). Also, the City of Philadelphia has roughly dozens of such important destinations, such as University of Pennsylvania, Hospital of the University of Pennsylvania, airport, baseball stadium, museum, etc. One may also consider distinct parts of a city as trip attractors, e.g., the Central Business District (CBD), old city, University city, North Philadelphia, etc. in Philadelphia. Then M becomes the number of such major destinations, and the vehicles headed to destinations other than these may be considered to belong to the $M+1$th category. Considering the above realistic values of M, the computation remains tractable.

We now describe how the destination based probabilistic routing may be utilized to characterize the fraction of informed vehicles as a function of time and space, in one simple instance that arises in practice; this instance also helps us elucidate the details of this generalization. We consider a grid road topology with six avenues and streets, as in Section 2.2.1; thus, a total of $J = 120$ clusters. We consider the morning rush hour, and divide vehicles into two classes: Class 1 corresponds to vehicles heading only towards the CBD, and Class 2 corresponds to other vehicles. We obtain the routing probability of Class 1 vehicles, building on the model introduced in Section 2.2.1. Recall that the mobility rate set in Section 2.2.1 is



(a) Morning rush (no back-and-forth) (b) Fraction of informed vehicles

Fig. 19. (a) An Example Destination-based routing. Even if there are multiple routes to the CBD, the vehicle only travels in the CBD direction by selecting one of these routes. (b) The fraction of informed vehicles over time.

$\lambda^I_{jk} = \lambda^S_{jk} = p_{jk}\lambda$. The mobility parameter λ is set to $\lambda = 0.03$ and the communication parameter is set to $\beta_{ij} = 10$. At the initial time, the total number of vehicles $N = 12000$ is uniformly distributed, and 70% of the vehicles in each cluster belong to Class 1. For a vehicle in Class 1, that is currently in cluster j which is outside the CBD, the routing probability p_{jk} is set to 1 if k is the cluster adjoining j that is the closest to the CBD; if multiple clusters adjoining j are closest to the CBD, then the routing probability is equally divided among them; the routing probability is set to 0 towards other adjoining clusters. Thus, vehicles in Class 1 do not retrace their paths, i.e., do not move back and forth, while they are outside the CBD (see Fig. 19a). The existing non-rush hour model (random walk) model of Section 2.2.1 applies to vehicles in 1) Class 1 once they reach the CBD, and 2) Class 2 anywhere. Thus, vehicles in Class 1 (Class 2 respectively) may retrace parts of their paths once they have reached the CBD (anywhere, respectively), to represent cruising to locate a parking spot, a cafetaria, etc.

Fig. 19b shows the fraction of informed vehicles obtained from the generalized differential equations, and Fig. 20 shows the corresponding geographical representation of traffic density and informed vehicles. The second and third rows in

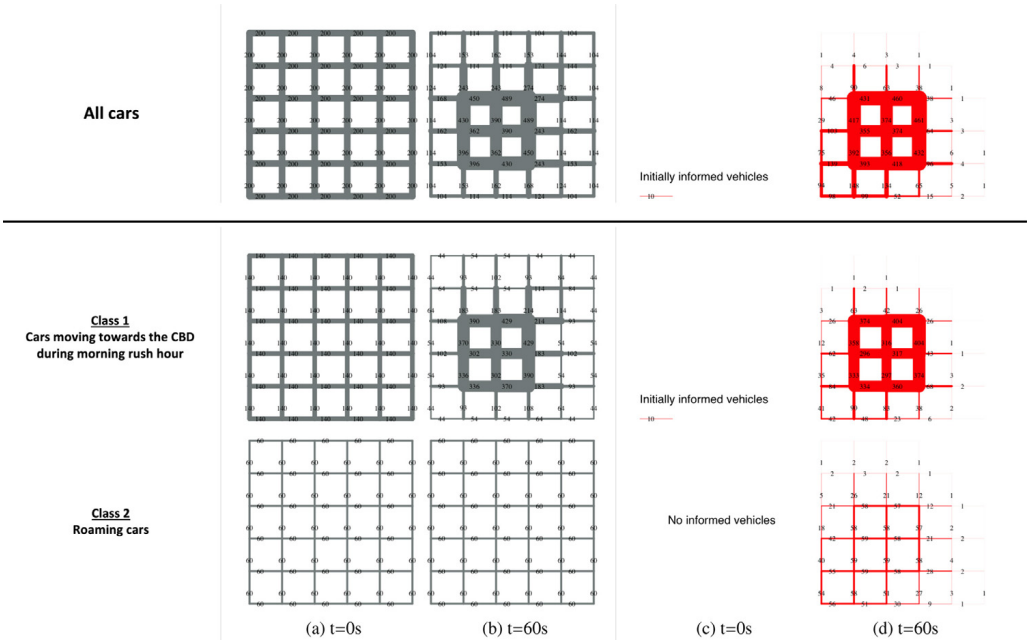


Fig. 20. Geographical representation of traffic density and information propagation, obtained from the generalized differential equations. The first and second columns of gray represent the number of vehicles in each road segment over time, and the third and fourth columns of red represent the number of vehicles informed in each road segment over time when information is propagated from 10 vehicles located in the lower left cluster. Class 1 corresponds to vehicles heading only towards the CBD, and Class 2 corresponds to roaming vehicles. The second row of Class 1 represents the movement of vehicles only towards the CBD and the propagation of information. The third row of Class 2 represents the movement of roaming vehicles and the propagation of information. The thickness of the road segment is linearly proportional to the number of vehicles, that is, when the thickness corresponding to one vehicle is x, the thickness of the vehicle v is expressed by $v \cdot x$. The number indicates the number of vehicles located in each road segment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 20 respectively represent the movement of vehicles in Classes 1 and 2 and the propagation of information. The sum of the second and third rows is the first row that depicts the traffic flow of the vehicles as a whole and the spread of information between all the vehicles.

We conclude with some parting thoughts on the case that the RoI for the propagation of the message is a large city or a long nationwide highway (e.g., for propagation of malware or CRL, as in Section 1.1). In this case, we note that the routes to two blocks in a neighborhood do not usually diverge until a vehicle reaches the neighborhood, e.g., a vehicle traversing to a specific restaurant in Chinatown in a big city, first moves in the direction of the Chinatown regardless of the exact address of its destination, and only after it reaches Chinatown, does it move towards the particular restaurant. Thus, it ought to suffice to consider only the major destinations and the neighborhoods as distinct entities. Thus, M would not be large as discussed before, and the product $2(M+1)J$ would be of the same order as J, therefore retaining computation tractability. But, when the RoI spans relatively smaller regions, e.g., small downtowns, or neighborhoods in large cities (e.g., when messages pertaining to sudden development of disruptive conditions, driving conditions like icy roads, arrival of ambulances etc. are shared), then individual city-blocks ought to be considered as separate destinations as the routes towards these destinations would differ in most of the RoI. But, given that the RoI is small, relatively speaking, in this case, both J and M would be moderate. Thus, computation time would be manageable throughout. For example, neighborhoods like SoHo in Manhattan, New York City, USA and downtown Ottawa, Ontario, Canada, consist of approximately 40 and 50 city blocks, respectively, and there are respectively 7440 and 12700 differential equations. The computation time for the corresponding differential equations is less than 3 hours using even modest computation facilities. Recall that the computation time for a total of 14,048 differential equations (corresponding to $J = 7024$ clusters) is approximately 172 minutes (See Section 5).

6.2. Multiple types of information

Our framework has been primarily designed for characterizing the spread of one kind of information, though in practice, the network has to handle the spread of multiple overlapping pieces of information. This may not however be a significant impediment as the RoIs of the multiple pieces of information that propagate simultaneously do not always overlap, and then the same vehicle would not need to simultaneously act on these pieces of information. Then again, even when the RoIs of different pieces of information overlap, the pieces may be unrelated to the extent that actions that the vehicles need to take based on these may be distinct. For example, traffic accident information and Certificate Revocation Lists (CRLs) are examples of unrelated pieces of information whose RoIs may overlap. In these cases, one can separately formulate and evaluate differential equations concerning these pieces of information, considering one piece of information at one instance.

We now describe how our framework can be generalized to characterize the spatio-temporal characteristics of simultaneous propagation over overlapping RoIs, of K kinds of related information, which together impact choices of a vehicle. Congestions, traffic accidents, driving conditions (e.g., ice formations), road obstructions, arrival of emergency vehicles at relatively close distances would constitute examples of such related information, as route choices would be influenced by simultaneous knowledge of all of these if their RoIs overlap. The generalization would involve increasing the state dimension, which in turn, increases the computation time. In this case, instead of classifying as non-informed and informed vehicles, vehicles can be indexed by a bit map consisting of K bits, in which each bit represents whether the vehicle has received the corresponding piece of information. Accordingly, a vehicle in each cluster is classified as one of 2^K types. In Section 2.1, we have introduced a $2J$-dimensional vector $(\mathbf{I}, \mathbf{S}) = (I_1, I_2, \ldots, I_J; S_1, S_2, \ldots, S_J)$ to model the propagation of a single piece of information on a road topology consisting of a set of J clusters. The $2J$-dimensional vector for the single type of information is expanded to a vector in $2^K J$ dimensions described in the form $(\mathbf{I_0}, \mathbf{I_1}, \cdots, \mathbf{I_{2^K-1}})$, where $\mathbf{I_0}$ represents the fraction of vehicles in different clusters which have not received any information, and plays the role of $\mathbf{S}$ in Section 2.1. Thus, for $i \in \{1, 2, \ldots, J\}$ and $j = \{0, 1, \ldots, 2^K - 1\}$, I_{ij} represents the fraction of informed vehicles located in cluster i and having received the information bits corresponding to j. Considering $K = 2$, the state vector becomes $(\mathbf{I_0}, \mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3}) = (\mathbf{I_{00}}, \mathbf{I_{01}}, \mathbf{I_{10}}, \mathbf{I_{11}})$, e.g., I_{13} (equivalently, $I_{1\{11\}}$) represents the fraction of vehicles in cluster 1 that have received both kinds of information.

Following the approach in Section 2.1, the system may be modeled as a continuous-time Markov chain (if the sojourns of the vehicles in the clusters and the intervals between communications are exponentially distributed). Similar to the proofs in Section Appendix A, it may be shown that Lemma 1 and Kurtz's theorem (Existing result 1) extend to this case, and the fraction of vehicles with different kinds of information across clusters asymptotically converge to the solution of a system of ordinary differential equations involving $(\mathbf{I_0}, \mathbf{I_1}, \cdots, \mathbf{I_{2^K-1}})$. We now introduce the mobility rates $\lambda_{jk}^{(m)}$, which represents the mobility rate from cluster j to k of vehicles having received information bits corresponding to the m-th category for $m \in \{0, 1, \ldots, 2^K - 1\}$. We also introduce the communication parameter $\beta_{jk}^{(r)}$, which corresponds to the communication of each of the bits that is 1 in bit-map r between vehicles in clusters j and k, for $r \in \{1, 2, \ldots, K\}$. We now present the differential equations in the example scenario that $K = 2$:

$$\dot{I}_{j3}(t) = -\sum_{k\neq j}^{J} \lambda_{jk}^{(3)}(\mathbf{I}, \mathbf{S}) \cdot I_{j3} + \sum_{k=1}^{J} \beta_{kj}^{(2)} \cdot I_{k2} \cdot I_{j1} + \sum_{k=1}^{J} \beta_{kj}^{(1)} \cdot I_{k1} \cdot I_{j2} + \sum_{k\neq j}^{J} \lambda_{kj}^{(3)}(\mathbf{I}, \mathbf{S}) \cdot I_{k3} \qquad (j = 1, 2, \ldots, J),$$

$$\dot{I}_{j2}(t) = -\sum_{k\neq j}^{J} \lambda_{jk}^{(2)}(\mathbf{I}, \mathbf{S}) \cdot I_{j2} - \sum_{k=1}^{J} \beta_{kj}^{(1)} \cdot I_{k1} \cdot I_{j2} + \sum_{k=1}^{J} \beta_{kj}^{(2)} \cdot I_{k2} \cdot I_{j0} + \sum_{k\neq j}^{J} \lambda_{kj}^{(2)}(\mathbf{I}, \mathbf{S}) \cdot I_{k2} \qquad (j = 1, 2, \ldots, J),$$

$$\dot{I}_{j1}(t) = -\sum_{k\neq j}^{J} \lambda_{jk}^{(1)}(\mathbf{I},\mathbf{S}) \cdot I_{j1} - \sum_{k=1}^{J} \beta_{kj}^{(2)} \cdot I_{k2} \cdot I_{j1} + \sum_{k=1}^{J} \beta_{kj}^{(1)} \cdot I_{k1} \cdot I_{j0} + \sum_{k\neq j}^{J} \lambda_{kj}^{(1)}(\mathbf{I},\mathbf{S}) \cdot I_{k1} \qquad (j = 1, 2, \ldots, J),$$

$$\dot{I}_{j0}(t) = -\sum_{k\neq j}^{J} \lambda_{jk}^{(0)}(\mathbf{I},\mathbf{S}) \cdot I_{j0} - \sum_{k=1}^{J} \beta_{kj}^{(1)} \cdot I_{k1} \cdot I_{j0} - \sum_{k=1}^{J} \beta_{kj}^{(2)} \cdot I_{k2} \cdot I_{j0} + \sum_{k\neq j}^{J} \lambda_{kj}^{(0)}(\mathbf{I},\mathbf{S}) \cdot I_{k0} \qquad (j = 1, 2, \ldots, J).$$

In many cases, V2V messages pertaining to specific events, e.g., unexpected events, obstructions in roads do not need to propagate beyond a certain RoI or after a particular period. For a single message, i.e., when $K = 1$, our model can directly accommodate this temporal and spatial context, by (1) computing the differential equations only in the time horizon in which the message is relevant, and (2) considering clusters only in the RoI. Thus, Geocast, i.e., propagation of messages in specific designated geographical areas, can be accommodated in our model as above, without increasing the state space and the number of differential equations. But, for multiple related pieces of information, different kinds of information may need to be propagated over different (but overlapping) RoIs and time intervals. Communication parameters associated with each type of information can be set to zero in the cluster outside the corresponding RoI and the time interval in which it needs to propagate. This generalization does not increase the state space and the number of differential equations. Suppose that two different pieces of information about traffic accidents at two distinct locations in Manhattan propagate simultaneously. The RoI of each piece of information may overlap, and each information is only valid for different time horizons until each traffic accident scene is cleared. In this case, the communication parameter $\beta_{jk}^{(2)}$, which corresponds to the communication of information about accident 2, is 1) positive when clusters j and k are located in the RoI of this message, and in the time interval in which the message is relevant, 2) 0 otherwise.

The number of variables and the number of differential equations increase exponentially in the number of messages, by a factor of 2^K to be specific. We however expect that only a few important pieces of related information would need to simultaneously propagate in overlapping RoIs. Thus, K would be small. Moreover, such RoIs would be smaller than entire cities or long highways because the related pieces of information are typically relatively local in scope, e.g., congestions, traffic accidents, driving conditions like ice formations, road obstructions, arrival of emergency vehicles at relatively close distance etc. The kinds of information whose RoIs span entire cities or counties or long highways, like malware, CRL, are not related to the above. Thus, again the product $2^K J$ may not always be prohibitive. We elucidate this using some examples that arise in practice. We consider the situation in which seven different types of information are disseminated in SoHo (40 blocks) in Manhattan and downtown Ottawa (50 blocks); there are respectively 11904 and 16256 differential equations. The computation time would be less than 4 hours using even modest computation facilities. Recall that the computation time for a total of 16,668 differential equations (corresponding to $J = 8334$ clusters) is approximately 240 minutes (See Section 5).

7. Conclusion and future work

V2V technologies bridge two interconnected and interdependent infrastructures: the communications infrastructure and the transportation infrastructure (including the vehicular infrastructure, the sensor infrastructure, and the physical roadway capacity). In this manuscript, proceeding from a continuous-time Markov chain model, utilizing rigorous mathematical proofs, we compute the fraction of informed vehicles as a function of time and space as a solution of a set of clustered epidemiological differential equations which lend itself to fast computation. We then demonstrate the applicability of this model in various scenarios: both real world scenarios, involving several generalizations and interdependence between communication and mobility and hypothesized scenarios of outages and system perturbations. We find that our models match microscopic trajectory data with acceptable error, demonstrating the applicability of our models. Our findings are of critical importance in shared transportation, as many of the current commercial ventures are investing in and considering the deployment of Connected Autonomous Vehicles enabled with V2V technologies. Overall, our work captures the spatio-temporal dynamics of information propagation over connected vehicles, enabling shared mobility services, individuals, and transportation system operators to see the benefits and drawbacks of large-scale V2V-enabled vehicle deployments in different transportation network typologies and for different densities of V2V-enabled vehicles and communication conditions of V2V-enabled vehicles.

We now list some directions for future research. Reduction of the worst case computation times for the generalizations considered in Section 6 through creative formulations that exploit specific temporal and spatial properties of given transportation networks deserves further investigation. Next, the traffic signals result in traffic synchronization, because, all vehicles stop at a red light and start moving almost simultaneously when the light turns green. This synchronization causes a divergence between the actual mobility of the traffic trajectory and the exponential mobility process, which we have assumed. Modeling information propagation with such pulsed traffic remains open. We can for example consider a time-dependent mobility parameter $\lambda_{i,j}(\cdot)$, whereby they are 0 when the lights are red, and at the normal values when they are green. When there are many traffic signals on more complex roads, multiple traffic signals affect traffic flows in different directions simultaneously, which affects information propagation in a complicated manner. We postpone the details, including mathematical analysis, numerical computations and simulations, impact of the locations of traffic signal(s) for future research.

CRediT authorship contribution statement

Jungyeol Kim: Methodology, Writing - original draft, Software, Formal analysis, Visualization. **Saswati Sarkar:** Conceptualization, Methodology, Writing - review & editing, Supervision. **Santosh S. Venkatesh:** Conceptualization, Methodology, Writing - review & editing. **Megan Smirti Ryerson:** Conceptualization, Writing - review & editing, Funding acquisition. **David Starobinski:** Conceptualization, Writing - review & editing, Funding acquisition.

Acknowledgments

This project was supported in part by the U.S. National Science Foundation [grant numbers CNS-1409053, CNS-1908087]; the University of Pennsylvania's Mobility21 National University Transportation Center, in partnership with Carnegie Mellon University, which is sponsored by the US Department of Transportation [grant number 69A3551747111].

Appendix A. Proof of the result in Section 2.1

The stochastic model for the information propagation can be approximated by a solution of ordinary differential equations (3). To this end, we first present the results developed by Kurtz (1970).

Definition 1. One parameter family of Markov chain $X(t)$ with state spaces contained in $\mathbb{Z}^K$ is called density dependent if there exist continuous functions $f(\mathbf{x},\mathbf{h}):\mathbb{R}^K\times\mathbb{Z}^K\to\mathbb{R}$, such that the transition rates $q(\mathbf{k},\mathbf{k}+\mathbf{h})$ from $\mathbf{k}$ to $\mathbf{k}+\mathbf{h}$ is given by

$$q(\mathbf{k},\mathbf{k}+\mathbf{h}) = Nf\left(\frac{\mathbf{k}}{N},\mathbf{h}\right), \quad \mathbf{h}\neq\mathbf{0}.$$

We can obtain a new Markov chain by scaling with N, and the scaled process $X_N(t)$ is defined by

$$X_N(t) := \frac{X(t)}{N} = \frac{1}{N}(\mathbf{n}^I(t),\mathbf{n}^S(t)).$$

If a Markov process $X(t)$ is density dependent, under certain conditions, a scaled process $X_N(t)$ can be approximated by a solution of ordinary differential equations determined by the following function

$$F(\mathbf{x}) := \sum_{\mathbf{h}\neq\mathbf{0}} \mathbf{h} f(\mathbf{x},\mathbf{h}), \tag{A.1}$$

which is the limiting mean increment. The following result (Kurtz, 1981; Ethier and Kurtz, 2009) provides sufficient conditions for convergence of the scaled process $X_N(t)$ to the unique trajectory of an deterministic path when N is large.

Existing Result 1. Suppose for each compact $E\in\mathbb{R}^K$

$$\sum_{\mathbf{h}} |\mathbf{h}| \sup_{\mathbf{x}\in E} f(\mathbf{x},\mathbf{h}) < \infty, \tag{A.2}$$

and there exist $M_E > 0$ such that

$$|F(\mathbf{x}) - F(\mathbf{y})| < M_E|\mathbf{x}-\mathbf{y}| \quad \mathbf{x},\mathbf{y}\in E. \tag{A.3}$$

Suppose $\lim_{N\to\infty} X_N(0) = \mathbf{x}(0)$, and $\mathbf{x}(t)$ satisfies

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t F(\mathbf{x}(s))ds,$$

for all $t\geq 0$ (in particular $\sup_{s\leq t}|\mathbf{x}(s)| < \infty$). Then

$$\lim_{N\to\infty} \sup_{s\leq t} |X_N(s) - \mathbf{x}(s)| = 0 \quad \text{a.s. for all } t > 0.$$

We now use this result to approximate the dynamics of infomation propagation through the solution of ordinary differential equations. Recall a set $E := \{(\mathbf{I},\mathbf{S}) \mid I_i\geq 0,\ S_i\geq 0 : i = 1,2,\ldots,J,\ \sum_{i=1}^J (I_i+S_i) = 1\}$ with $(\mathbf{I},\mathbf{S}) = (I_1,I_2,\ldots,I_J,S_1,S_2,\ldots,S_J)$. Note that S^N/N is a subset of E and the scaled process $X_N(t)$ is contained in E. Also note that $\mathbf{I}$ and $\mathbf{S}$ respectively have physical connotations of fraction of informed and non-informed vehicles in each cluster as discussed. Let the function $f(\mathbf{x},\mathbf{h})$, $\mathbf{x}\in E$, $\mathbf{h}\in\mathbb{Z}^{2J}$, be defined as

$$f(\mathbf{x},\mathbf{h}) = \begin{cases} \lambda^I_{jk}(\mathbf{x})\cdot I_j, & \mathbf{h} = -\mathbf{1}_j + \mathbf{1}_k,\ j\neq k \\ \lambda^S_{jk}(\mathbf{x})\cdot S_j, & \mathbf{h} = -\mathbf{1}_{J+j} + \mathbf{1}_{J+k},\ j\neq k \\ \beta_{jk}\cdot I_j\cdot S_k, & \mathbf{h} = \mathbf{1}_k - \mathbf{1}_{J+k} \\ 0, & \text{otherwise.} \end{cases} \tag{A.4}$$

where $\mathbf{x} = (\mathbf{I}, \mathbf{S})$. Since transition rate (1) can be written as the form of $q(\mathbf{k}, \mathbf{k}+\mathbf{h}) = Nf\left(\frac{\mathbf{k}}{N}, \mathbf{h}\right)$, $\mathbf{h} \neq \mathbf{0}$, the Markov process satisfies density-dependent property (Definition 1). From (A.1) and (A.4), the function $F(\mathbf{x}) = F(\mathbf{I}, \mathbf{S})$, can be expressed as

$$F(\mathbf{I},\mathbf{S}) = \sum_{\mathbf{h}\neq\mathbf{0}} \mathbf{h} f((\mathbf{I},\mathbf{S}),\mathbf{h}) = \begin{cases} -\sum_{k\neq 1}^{J} \lambda^I_{1k}(\mathbf{I},\mathbf{S})\cdot I_1 & +\sum_{k=1}^{J} \beta_{k1} I_k S_1 + \sum_{k\neq 1}^{J} \lambda^I_{k1}(\mathbf{I},\mathbf{S})\cdot I_k \\ & \vdots \\ -\sum_{k\neq J}^{J} \lambda^I_{Jk}(\mathbf{I},\mathbf{S})\cdot I_J & +\sum_{k=1}^{J} \beta_{kJ} I_k S_J + \sum_{k\neq 1}^{J} \lambda^I_{kJ}(\mathbf{I},\mathbf{S})\cdot I_k \\ -\sum_{k\neq 1}^{J} \lambda^S_{1k}(\mathbf{I},\mathbf{S})\cdot S_1 & -\sum_{k=1}^{J} \beta_{k1} I_k S_1 + \sum_{k\neq 1}^{J} \lambda^S_{k1}(\mathbf{I},\mathbf{S})\cdot S_k \\ & \vdots \\ -\sum_{k\neq J}^{J} \lambda^S_{Jk}(\mathbf{I},\mathbf{S})\cdot S_J & -\sum_{k=1}^{J} \beta_{kJ} I_k S_J + \sum_{k\neq J}^{J} \lambda^S_{kJ}(\mathbf{I},\mathbf{S})\cdot S_k. \end{cases}$$

Lemma 1. *Suppose for $i, j = 1, 2, \ldots, J$ and $i \neq j$, mobility rate functions $\lambda^I_{ij} : E \to \mathbb{R}$ and $\lambda^S_{ij} : E \to \mathbb{R}$ are bounded and Lipschitz continuous on E. Then, the function F is Lipschitz continuous on E.*

Proof. Let $\mathbf{x} = (\mathbf{I}, \mathbf{S}) = (I_1, I_2, \ldots, I_J, S_1, S_2, \ldots, S_J)$ and $\mathbf{y} = (\bar{\mathbf{I}}, \bar{\mathbf{S}}) = (\bar{I}_1, \bar{I}_2, \ldots, \bar{I}_J, \bar{S}_1, \bar{S}_2, \ldots, \bar{S}_J)$ be points in E. Starting from $|F_i(\mathbf{x}) - F_i(\mathbf{y})|$ for $i = 1, 2, \ldots, J$, we have

$$\begin{aligned} |F_i(\mathbf{x}) - F_i(\mathbf{y})| &= \left| -\sum_{k\neq i}^{J} \left(\lambda^I_{ik}(\mathbf{x}) I_i - \lambda^I_{ik}(\mathbf{y})\bar{I}_i\right) + \sum_{k=1}^{J} \beta_{ki}(I_k S_i - \bar{I}_k \bar{S}_i) + \sum_{k\neq i}^{J} \left(\lambda^I_{ki}(\mathbf{x}) I_k - \lambda^I_{ki}(\mathbf{y})\bar{I}_k\right) \right| \\ &= \left| -\sum_{k\neq i}^{J} \lambda^I_{ik}(\mathbf{x})(I_i - \bar{I}_i) - \sum_{k\neq i}^{J} \left(\lambda^I_{ik}(\mathbf{x}) - \lambda^I_{ik}(\mathbf{y})\right)\bar{I}_i + \sum_{k=1}^{J} \beta_{ki} I_k (S_i - \bar{S}_i) + \sum_{k=1}^{J} \beta_{ki} \bar{S}_i (I_k - \bar{I}_k) \right. \\ &\quad \left. + \sum_{k\neq i}^{J} \lambda^I_{ki}(\mathbf{x})(I_k - \bar{I}_k) + \sum_{k\neq i}^{J} \left(\lambda^I_{ki}(\mathbf{x}) - \lambda^I_{ki}(\mathbf{y})\right)\bar{I}_k \right|. \end{aligned}$$

for $i = 1, 2, \ldots, J$. Suppose for $i, j = 1, 2, \ldots, J$ and $i \neq j$, the mobility rate function λ^I_{ij} and λ^S_{ij} are bounded above by $\widehat{\lambda^I_{ij}}$ and $\widehat{\lambda^S_{ij}}$ respectively. Suppose further that λ^I_{ij} and λ^S_{ij} are Lipschitz continuous in the sense that $|\lambda^I_{ij}(\mathbf{x}) - \lambda^I_{ij}(\mathbf{y})| \leq L^I_{ij} \cdot |\mathbf{x} - \mathbf{y}|$ with Lipschitz constant L^I_{ij} and $|\lambda^S_{ij}(\mathbf{x}) - \lambda^S_{ij}(\mathbf{y})| \leq L^S_{ij} \cdot |\mathbf{x} - \mathbf{y}|$ with Lipschitz constant L^S_{ij}. Then, we have

$$|F_i(\mathbf{x}) - F_i(\mathbf{y})| \leq \sum_{k\neq i}^{J} \widehat{\lambda^I_{ik}} \left|I_i - \bar{I}_i\right| + \sum_{k\neq i}^{J} L^I_{ik} |\mathbf{x} - \mathbf{y}| + \sum_{k=1}^{J} \beta_{ki} |S_i - \bar{S}_i| + \sum_{k=1}^{J} \beta_{ki} |I_k - \bar{I}_k| + \sum_{k\neq i}^{J} \widehat{\lambda^I_{ki}} |I_k - \bar{I}_k| + \sum_{k\neq i}^{J} L^I_{ki} |\mathbf{x} - \mathbf{y}|.$$

Since $|I_i - \bar{I}_i| \leq |(\mathbf{I}, \mathbf{S}) - (\bar{\mathbf{I}}, \bar{\mathbf{S}})| = |\mathbf{x} - \mathbf{y}|$ and $|S_i - \bar{S}_i| \leq |(\mathbf{I}, \mathbf{S}) - (\bar{\mathbf{I}}, \bar{\mathbf{S}})| = |\mathbf{x} - \mathbf{y}|$ for $i = 1, 2, \ldots, J$, we have

$$|F_i(\mathbf{x}) - F_i(\mathbf{y})| \leq K_i \cdot |\mathbf{x} - \mathbf{y}| \qquad i = 1, 2, \ldots, J$$

where $K_i = \sum_{k\neq i}^{J} \widehat{\lambda^I_{ik}} + \sum_{k\neq i}^{J} L^I_{ik} + 2\sum_{k=1}^{J} \beta_{ki} + \sum_{k\neq i}^{J} \widehat{\lambda^I_{ki}} + \sum_{k\neq i}^{J} L^I_{ki}$. Through the same way, for $i = J+1, J+2, \ldots, 2J$, we have

$$|F_i(\mathbf{x}) - F_i(\mathbf{y})| \leq K_i \cdot |\mathbf{x} - \mathbf{y}| \qquad i = J+1, J+2, \ldots, 2J$$

where $K_i = \sum_{k\neq i}^{J} \widehat{\lambda^S_{ik}} + \sum_{k\neq i}^{J} L^S_{ik} + 2\sum_{k=1}^{J} \beta_{ki} + \sum_{k\neq i}^{J} \widehat{\lambda^S_{ki}} + \sum_{k\neq i}^{J} L^S_{ki}$. Thus, the component functions F_i for $i = 1, 2, \ldots, 2J$ are Lipschitz continuous on E; consequently, the function F is Lipschitz continuous on E. □

By Lemma 1, the mobility rate functions λ^I_{ij} and λ^S_{ij} for $i, j = 1, 2, \ldots, J$ and $i \neq j$, are bounded and Lipschitz continuous on E, so the condition (A.3) is satisfied. The state space S^N is finite and the function $f(\mathbf{x}, \mathbf{h})$ for each $\mathbf{h}$ is bounded, so the condition (A.2) is satisfied. Consequently, if $\lim_{N\to\infty} \frac{1}{N}(\mathbf{n}^I(0), \mathbf{n}^S(0)) = (\mathbf{I}(0), \mathbf{S}(0))$, the scaled process $X_N(t)$ converges to the solution of the ordinary differential equations (3) as the total number of vehicles N increases.

References

Agarwal, A., Starobinski, D., Little, T.D., 2012. Phase transition of message propagation speed in delay-tolerant vehicular networks. IEEE Trans. Intell. Transp.Syst. 13 (1), 249–263.
Alfa, A.S., Neuts, M.F., 1995. Modelling vehicular traffic using the discrete time markovian arrival process. Transp. Sci. 29 (2), 109–117.
Almalag, M.S., Weigle, M.C., Olariu, S., 2013. MAC Protocols for VANET. In: Mobile Ad Hoc Networking. John Wiley & Sons, Inc., pp. 599–618.

Baccelli, E., Jacquet, P., Mans, B., Rodolakis, G., 2012. Highway vehicular delay tolerant networks: information propagation speed properties. IEEE Trans. Inf. Theory 58 (3), 1743–1756.
Ban, X.J., Li, W., 2018. Connected vehicle based traffic signal optimization. Connected Cities for Smart Mobility towards Accessible and Resilient Transportation Center (C2SMART). http://c2smart.engineering.nyu.edu/wp-content/uploads/2018/07/TrafficSignalOptimizationinConnectedCities-FinalReport.pdf.
Banerjee, S., Kanoria, Y., Qian, P., 2018. State dependent control of closed queueing networks with application to ride-hailing. arXiv preprint arXiv: 1803.04959.
Bertini, R.L., Wang, H., Knudson, T., Carstens, K., Rios, E., 2016. Assessing state department of transportation readiness for connected vehicle–cooperative systems deployment: oregon case study. Transp. Res. Rec.: J. Transp. Res. Board 2559 (1), 24–34.
Blanco, S., 2019. Aptiv has data on over 30,000 autonomous Lyft rides in Las Vegas. Forbes. http://www.forbes.com/sites/sebastianblanco/2019/01/29/aptiv-has-data-on-over-30000-autonomous-lyft-rides-in-las-vegas/#79514c791ba7 (Accessed May 7, 2019).
Bonelli, C., 2017. V2V safety technology now standard on Cadillac CTS sedans. Cadillac Pressroom. http://media.cadillac.com/media/us/en/cadillac/news.detail.html/content/Pages/news/us/en/2017/mar/0309-v2v.html (Accessed May 7, 2019).
Braverman, A., Dai, J., Liu, X., Ying, L., 2017. Fluid-model-based car routing for modern ridesharing systems. In: ACM SIGMETRICS Performance Evaluation Review, 45. ACM, pp. 11–12.
Braverman, A., Dai, J.G., Liu, X., Ying, L., 2019. Empty-car routing in ridesharing systems. Oper. Res. 67 (5), 1437–1452.
Caelli, T., Porter, D., 1980. On difficulties in localizing ambulance sirens. Human Factors 22 (6), 719–724.
Cain, H., 2018. DSRC is best suited for collision avoidance and other safety applications. The Eno Center for Transportation. https://www.enotrans.org/article/dsrc-is-best-suited-for-collision-avoidance-and-other-safety-applications/, (Accessed August 26, 2019).
California Department of Motor Vehicles, 2017. Estimated vehicles registered by county. http://www.dmv.ca.gov/portal/wcm/connect/add5eb07-c676-40b4-98b5-8011b059260a/est_fees_pd_by_county.pdf?MOD=AJPERES, (Accessed May 7, 2019).
Checkoway, S., Mccoy, D., Kantor, B., Anderson, D., Shacham, H., Savage, S., Koscher, K., Czeskis, A., Roesner, F., Kohno, T., 2011. Comprehensive experimental analyses of automotive attack surfaces. In: USENIX Security Symposium. San Francisco, pp. 77–92.
Chen, A., Khorashadi, B., Chuah, C.-N., Ghosal, D., Zhang, M., 2006. Smoothing vehicular traffic flow using vehicular-based ad hoc networking & computing grid (VGrid). In: 2006 IEEE Intelligent Transportation Systems Conference. IEEE, pp. 349–354.
Chen, J., Cao, X., Zhang, Y., Xu, W., Sun, Y., 2011. Measuring the performance of movement-assisted certificate revocation list distribution in VANET. Wirel. Commun. Mobile Comput. 11 (7), 888–898.
Chou, L.-D., Li, D.C., Chao, H.-W., 2011. Mitigate traffic congestion with virtual data sink based information dissemination in Intelligent Transportation System. In: 2011 Third International Conference on Ubiquitous and Future Networks (ICUFN). IEEE, pp. 37–42.
dataset U.S. Department of Transportation, 2017. Next Generation Simulation (NGSIM) vehicle trajectories and supporting data. https://data.transportation.gov/Automobiles/Next-Generation-Simulation-NGSIM-Vehicle-Trajector/8ect-6jqj, (Accessed June 10, 2018).
Demissie, M.G., de Almeida Correia, G.H., Bento, C., 2013. Intelligent road traffic status detection system through cellular networks handover information: an exploratory study. Transp. Res. Part C 32, 76–88.
Des Moines Area MPO, 2006. Des Moines Area Travel Demand Model: Documentation & User Guide. The Des Moines Area Metropolitan Planning Organization. https://web.engr.uky.edu/~rsouley/CE%20451/LABS/Lab%2012/DSM_Documentation.pdf.
Dong, J., Mahmassani, H.S., 2009. Flow breakdown and travel time reliability. Transp. Res. Rec. 2124 (1), 203–212.
Ethier, S.N., Kurtz, T.G., 2009. Markov Processes: Characterization and Convergence (Wiley Series in Probability and Mathematical Statistics). John Wiley & Sons.
European Commission, 2019. Road safety: New rules clear way for clean, connected and automated mobility on EU roads. https://ec.europa.eu/commission/presscorner/detail/en/IP_19_1648, (Accessed Nov 16, 2019).
Forster, M., Frank, R., Gerla, M., Engel, T., 2014. A cooperative advanced driver assistance system to mitigate vehicular traffic shock waves. In: IEEE INFOCOM 2014-IEEE Conference on Computer Communications. IEEE, pp. 1968–1976.
Gentile, G., 2015. Using the general link transmission model in a dynamic traffic assignment to simulate congestion on urban networks. Transp. Res. Procedia 5, 66–81.
Geroliminis, N., Skabardonis, A., 2005. Prediction of arrival profiles and queue lengths along signalized arterials by using a markov decision process. Transp. Res. Rec.: J.Transp. Res. Board 1934, 116–124.
Green, M., Thomas, A., Chachich, A., Fehr, W., Towery, N.D., 2018. FHWA White Paper on Mobile Ad Hoc Networks. United States Federal Highway Administration. https://www.fhwa.dot.gov/publications/research/ear/18027/18027.pdf.
Greenberg, A., 2015. After Jeep hack, Chrysler recalls 1.4M vehicles for bug fix. Wired. http://www.wired.com/2015/07/jeep-hack-chrysler-recalls-1-4m-vehicles-bug-fix, (Accessed May 7, 2019).
Gupte, S., Younis, M., 2012. Vehicular networking for intelligent and autonomous traffic management. In: 2012 IEEE International Conference on Communications (ICC). IEEE, pp. 5306–5310.
Haas, J.J., Hu, Y.-C., Laberteaux, K.P., 2011. Efficient certificate revocation list organization and distribution. IEEE J. Sel. Areas Commun. 29 (3), 595–604.
Haefner, L.E., Li, M.-S., 1998. Traffic flow simulation for an urban freeway corridor. In: Proceedings of the 1998 Transportation Conference. Citeseer, pp. 1–6.
Harding, J., Powell, G., Yoon, R., Fikentscher, J., Doyle, C., Sade, D., Lukuc, M., Simons, J., Wang, J., 2014. Vehicle-to-vehicle communications: Readiness of V2V technology for application. (No. DOT-HS-812-014). United States National Highway Traffic Safety Administration. https://rosap.ntl.bts.gov/view/dot/27999.
Hartenstein, H., Laberteaux, L., 2008. A tutorial survey on vehicular ad hoc networks. IEEE Commun. Mag. 46 (6), 164–171.
He, J., Cai, L., Pan, J., Cheng, P., 2017. Delay analysis and routing for two-dimensional VANETs using carry-and-forward mechanism. IEEE Trans. Mobile Comput. 16 (7), 1830–1841.
He, Z., Zheng, L., Chen, P., Guan, W., 2017. Mapping to cells: a simple method to extract traffic dynamics from probe vehicle data. Comput.-Aided Civil Infrastruct. Eng. 32 (3), 252–267.
Helbing, D., 2001. Traffic and related self-driven many-particle systems. Rev. Modern Phys. 73 (4), 1067–1141.
Hellinga, B., Izadpanah, P., Takada, H., Fu, L., 2008. Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments. Transp. Res. Part C 16 (6), 768–782.
Highway Department of the Hong Kong, 2018. Hong Kong road network. http://www.hyd.gov.hk/en/road_and_railway/existing/road_network/road.html (Accessed May 7, 2019).
Jiang, Z., Wu, J., Sabatino, P., 2014. GUI: GPS-less traffic congestion avoidance in urban areas with inter-vehicular communication. In: 2014 IEEE 11th International Conference on Mobile Ad Hoc and Sensor Systems. IEEE, pp. 19–27.
Kesting, A., Treiber, M., Helbing, D., 2010. Connectivity statistics of store-and-forward intervehicle communication. IEEE Trans. Intell. Transp. Syst. 11 (1), 172–181.
Kim, J., 2019. Vehicular messaging simulation, GitHub repository. https://github.com/jungyeol-kim/V2X-simulations.
Kim, Y.H., Peeta, S., He, X., 2016. An analytical model to characterize the spatiotemporal propagation of information under vehicle-to-vehicle communications. IEEE Trans. Intell. Transp. Syst. 19 (1), 3–12.
Knorr, F., Baselt, D., Schreckenberg, M., Mauve, M., 2012. Reducing traffic jams via VANETs. IEEE Trans. Veh. Technol. 61 (8), 3490–3498.
Kong, X., Xu, Z., Shen, G., Wang, J., Yang, Q., Zhang, B., 2016. Urban traffic congestion estimation and prediction based on floating car trajectory data. Future Gener. Comput. Syst. 61, 97–107.
Koscher, K., Czeskis, A., Roesner, F., Patel, S., Kohno, T., Checkoway, S., McCoy, D., Kantor, B., Anderson, D., Snachám, H., Savage, S., 2010. Experimental security analysis of a modern automobile. In: Security and Privacy (SP), 2010 IEEE Symposium on. IEEE, pp. 447–462.
Kühne, R.D., Rödiger, M.B., 1991. Macroscopic simulation model for freeway traffic with jams and stop-start waves. In: 1991 Winter Simulation Conference Proceedings. IEEE Computer Society, pp. 762–770.

Kurtz, T., 1981. Approximation of Population Processes (CBMS-NSF Regional Conference Series in Applied Mathematics). SIAM.
Kurtz, T.G., 1970. Solutions of ordinary differential equations as limits of pure jump markov processes. J. Appl. Probab. 7 (1), 49–58.
Laberteaux, K.P., Haas, J.J., Hu, Y.-C., 2008. Security certificate revocation list distribution for VANET. In: VANET' 08 - Proceedings of the 5th ACM International Workshop on VehiculAr Inter-NETworking. ACM, pp. 88–89.
Lakas, A., Chaqfeh, M., 2010. A novel method for reducing road traffic congestion using vehicular communication. In: Proceedings of the 6th International Wireless Communications and Mobile Computing Conference. ACM, pp. 16–20.
Lèbre, M.-A., Le Mouël, F., Ménard, E., 2015. On the importance of real data for microscopic urban vehicular mobility trace. In: Proceedings of the 14th International Conference on ITS Telecommunications (ITST'2015), Copenhagen, Denmark, pp. 22–26.
Liu, Y., Niu, J., Ma, J., Shu, L., Hara, T., Wang, W., 2013. The insights of message delivery delay in VANETs with a bidirectional traffic model. J. Netw. Comput. Appl. 36 (5), 1287–1294.
Mamun, M.S., Yin, Y., Srinivasan, S., 2010. Trip generation characteristics of special generators. University of Florida Transportation Research Center. https://cms.ce.ufl.edu/research/Trip_Generation_Characteristics_of_Special_Generators.pdf.
Mandal, K., Sen, A., Chakraborty, A., Roy, S., Batabyal, S., Bandyopadhyay, S., 2011. Road traffic congestion monitoring and measurement using active RFID and GSM technology. In: 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 1375–1379.
McGuckin, T., Lambert, J., Newton, D., Pearmine, A., Hubbard, E., Sheehan, R., 2017. Leveraging the Promise of Connected and Autonomous Vehicles to Improve Integrated Corridor Management and Operations: A primer. (No. FHWA-HOP-17-001). United States Federal Highway Administration. https://ops.fhwa.dot.gov/publications/fhwahop17001/index.htm.
Miller, C., Valasek, C., 2015. Remote exploitation of an unaltered passenger vehicle. Black Hat USA 2015.
Navas, J.C., Imielinski, T., 1997. GeoCast - geographic addressing and routing. In: MobiCom '97 - Proceedings of the Annual International Conference on Mobile Computing and Networking. ACM Press.
NHTSA, 2016. Federal Motor Vehicle Safety Standards; V2V Communications. (No. NHTSA-2016-0126). United States National Highway Traffic Safety Administration. http://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/v2v_nprm_web_version.pdf.
Ningning, Z., 2016. Number of cars on city's streets increases to 2.5m. Shanghai Daily. https://archive.shine.cn/metro/public-services/Number-of-cars-on-citys-streets-increases-to-25m/shdaily.shtml, (Accessed May 7, 2019).
Patire, A.D., Cassidy, M.J., 2011. Lane changing patterns of bane and benefit: observations of an uphill expressway. Transp. Res. Part B 45 (4), 656–666.
Petzold, L., 1983. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. SIAM J. Scientif. Stat. Comput. 4 (1), 136–148.
Pierce, G., Shoup, D., 2013. Getting the prices right: an evaluation of pricing parking by demand in San Francisco. J. Am. Plan. Assoc. 79 (1), 67–81.
Qian, Z.S., Li, J., Li, X., Zhang, M., Wang, H., 2017. Modeling heterogeneous traffic flow: a pragmatic approach. Transp. Res. Part B 99, 183–204.
Ramezani, M., Geroliminis, N., 2012. On the estimation of arterial route travel time distribution with Markov chains. Transp. Res. Part B 46 (10), 1576–1590.
Resource Systems Group, 2008. CCMPO Regional Transportation Model Documentation. Chittenden County Metropolitan Planning Organization.. https://studiesandreports.ccrpcvt.org/wp-content/uploads/2017/01/20080124_CCMPO-2.3.0_documentation.pdf.
Saleet, H., Langar, R., Naik, K., Boutaba, R., Nayak, A., Goel, N., 2011. Intersection-based geographical routing protocol for VANETs: a proposal and analysis. IEEE Trans. Veh. Technol. 60 (9), 4560–4574.
Scholliers, J., Van Noort, M., Johansson, C., Mans, D., Silla, A., Bell, D., Hancox, G., Leden, L., Giannelos, I., Bax, B., Malone, K., 2016. Impact assessment of ITS applications for vulnerable road users. Transp. Res. Procedia 14, 4515–4524.
de Souza, A.M., Yokoyama, R.S., Maia, G., Loureiro, A.A., Villas, L.A., 2015. Minimizing traffic jams in urban centers using vehicular ad hoc networks. In: 2015 7th international conference on new technologies, mobility and security (NTMS). IEEE, pp. 1–5.
Sugiyama, Y., Fukui, M., Kikuchi, M., Hasebe, K., Nakayama, A., Nishinari, K., ichi Tadaki, S., Yukawa, S., 2008. Traffic jams without bottlenecks—experimental evidence for the physical mechanism of the formation of a jam. N. J. Phys. 10 (3), 033001.
Sweatman, P., 2017. Heavy duty CAV – fast or slow?Presented at 2017 METRANS International Urban Freight Conference, Long Beach, CA.
Toyota, 2018. Toyota and Lexus to launch technology to connect vehicles and infrastructure in the U.S. in 2021. Toyota Newsroom. https://pressroom.toyota.com/toyota-and-lexus-to-launch-technology-connect-vehicles-infrastructure-in-u-s-2021/, (Accessed May 7, 2019).
U.S. Department of Transportation, 2017. Average age of automobiles and trucks in operation in the United States. http://www.bts.gov/content/average-age-automobiles-and-trucks-operation-united-states.
Wang, H., Li, J., Chen, Q.-Y., Ni, D., 2009. Speed-density relationship: from deterministic to stochastic. The 88th Transportation Research Board (TRB) Annual Meeting. Washington, DC.
Wang, M., Daamen, W., Hoogendoorn, S.P., Van Arem, B., 2016. Connected variable speed limits control and car-following control with vehicle-infrastructure communication to resolve stop-and-go waves. J. Intell. Transp. Syst. 20 (6), 559–572.
Wang, P., Hunter, T., Bayen, A.M., Schechtner, K., González, M.C., 2012. Understanding road usage patterns in urban areas. Scientif. Rep. 2, 1001.
Wang, R., Xu, Z., Zhao, X., Hu, J., 2019. V2V-based method for the detection of road traffic congestion. IET Intell. Transp. Syst. 13 (5), 880–885.
Whyte, W., Weimerskirch, A., Kumar, V., Hehn, T., 2013. A security credential management system for V2V communications. In: 2013 IEEE Vehicular Networking Conference. IEEE.
Won, M., Park, T., Son, S.H., 2017. Toward mitigating phantom jam using vehicle-to-vehicle communication. IEEE Trans. Intell. Transp. Syst. 18 (5), 1313–1324.
Wu, H., Fujimoto, R.M., Riley, G.F., Hunter, M., 2009. Spatial propagation of information in vehicular networks. IEEE Trans. Veh. Technol. 58 (1), 420–431.
Yang, P., Iyer, K., Frazier, P., 2018. Mean field equilibria for resource competition in spatial settings. Stochast. Syst. 8 (4), 307–334.
Yin, K., Wang, X.B., Zhang, Y., 2013. Vehicle-to-vehicle connectivity on two parallel roadways with a general headway distribution. Transp. Res. Part C 29, 84–96.
Yong-chuan, Z., Xiao-qing, Z., li ting, Z., Zhen-ting, C., 2011. Traffic congestion detection based on GPS floating-car data. Procedia Eng. 15, 5541–5546.
Yperman, I., Logghe, S., Immers, B., 2005. The link transmission model: an efficient implementation of the kinematic wave theory in traffic networks. In: Proceedings of the 10th EWGT Meeting. Poznan Poland, pp. 122–127.
Zhang, T., Antunes, H., Aggarwal, S., 2014. Defending connected vehicles against malware: challenges and a solution framework. IEEE Internet of Things J. 1 (1), 10–21.
Zhang, Z., Mao, G., Anderson, B.D., 2014. Stochastic characterization of information propagation process in vehicular ad hoc networks. IEEE Trans. Intell. Transp. Syst. 15 (1), 122–135.