Role of Road Network Features in the Evaluation of Incident Impacts on Urban Traffic Mobility

Chen-Shuo Sun, Xin Pei*, Junheng Hao, Yewen Wang, Zuo Zhang, and S.C. Wong

Highlights

1) Network features are introduced to study incident-induced impacts on road networks mobility.

2) Four network features are extracted to code the distinctive functionality of urban intersections.

3) Incidents impacts are measured in both temporal and spatial dimension.

4) Temporally, accident delay is significantly correlated with the Betweenness Centrality and K-shell.

5) Spatially, micro impact and macro impact are found to be strongly associated with the four network features.

Abstract

In this paper, we seek to investigate the spatiotemporal impacts of traffic incident on urban road networks. The theoretical lens of a complex network leads us to expect that such effects are associated with the functionality that an intersection acts in network, and also, the location of incident sites. Incident impacts are measured in both temporal and spatial dimension through collaboratively mining the rich data that account for traffic flow and incident detailed information. In complex network context, the urban road network can be converted into a weighted direct graph

[1] Chen-Shuo Sun, Xin Pei, Junheng Hao, Yewen Wang and Zuo Zhang are with the Department of Automation, Tsinghua University, Beijing, China; S.C. Wong is with the Department of Civil Engineering, The University of Hong Kong, Hong Kong, China. *Corresponding author: Fax: +86 10 62795043, E-mail address: peixin@mail.tsinghua.edu.cn, (X. Pei). E-mail address: scuseei@vip.163.com (C.S. Sun), haojh_thu13@163.com (J. Hao), wyw10804@sina.com (Y. Wang), zuozhang@tsinghua.edu.cn (Z. Zhang), hhecwsc@hku.hk (S.C. Wong).

with intersections as nodes and road segments as edges with their geographic information. Four network features of Betweenness Centrality, weighted PageRank, Hub, and K-shell are then assigned to each intersection to assess its functionality. Temporally, we find out significant correlations between incident delay and several network features by applying hazard-based models. Spatially, micro impact and macro impact are found to be strongly associated with the four network features through estimating a Bayesian Negative-binomial conditional autoregressive model and a generalized linear model, respectively. Our study provides a basis of leveraging urban road network context to evaluate incident impacts, with some explanations, useful insights and possible extensions that would assist traffic administrations to guide the post-incident resilience and emergency management, and help road users to avoid potential congestion.

Keywords

Traffic mobility; Incident impacts; Network features; Hazard-based model; Bayesian Negative-binomial CAR model; Generalized linear model

1. Introduction

Intersections are the busiest but dangerous locations in road networks due to the number of turning movements and the resultant conflict points (Chen and Xie, 2016). The hazards at intersections not only lead to more injuries and property losses, as showcased by previous studies (Chandler, 2013; Barua et al., 2010), they also lead to a decline in traffic efficiency, i.e., incidents at or near intersections induce traffic congestion and thus reduce the connectivity of the road network. To improve the reliability of road networks and facilitate emergency rescue, traffic management agencies need to acquire the incident impacts on road networks (Weng et al., 2015; Konduri et al., 2003). Specifically, if drivers can be informed in advance via online systems of potential congestion, they may actively change their routes; if management authorities are able to grasp an incident's spatiotemporal impact, they can mitigate traffic congestion more efficiently.

Previous research on the estimation and evaluation of the spatiotemporal impacts mainly focused on road segments of freeways. Deterministic queuing diagrams (Erera and Garrick, 1998) and shock waves (Wirasinghe, 1978; Wang et al., 2016) were the conventional methods adopted by these studies. Sheu et al. (2001) developed a stochastic estimation approach to real-time prediction of incident congestions. Recently, Chung et al. (2010, 2012, 2015) developed a binary integer programming method to estimate the spatiotemporal impacts of freeway incidents. Unlike freeway segments, urban road networks are interconnected and interdependent, and thus accord with complex networks. However, the impact of incidents on traffic mobility in urban areas, specifically the role of road network features at an incident location in the impact analysis, has not been carefully examined.

Road network features have been drawing increasing attention in urban transport studies, especially, usually involved as risk factors in road safety analysis. Marshall et al. (2011) and Rifaat et al. (2011) found that road network structure has a significant impact on traffic safety. Wang et al. (2013) used Closeness Centrality, Betweenness Centrality, and Meshedness coefficients to measure road network properties within traffic analysis zones (TAZs) and found they were closely related to crash frequencies. Analogously, Zhang et al. (2015), adopted Betweenness Centrality and overall clustering coefficients to quantify road network structures, and proved that they are associated with the frequency of non-motorist accidents. In most studies, road network features have been used to represent the general profile and properties of the road network in an entire zone or area. The role of local network features, nevertheless, in the evaluation of incident impact on road networks has not been fully considered in previous analyses.

As a typical kind of incidents with negative impact on mobility, traffic accidents have always been analyzed by previous research in terms of accident frequency (Lord and Mannering, 2010; Abdel-Aty and Radwan, 2000) and injury severity (Savolainen et al., 2011). These studies investigated the roles of different risk factors in road safety, but rarely considered the accident impacts on the reliability of road networks. Accidents, in fact, not only cause injuries that lower the safety performance of roads, but also give rise to congestion that deteriorates the mobility of the surrounding roads and even the whole road networks through the malfunction or removal of those key road segments and intersections (Li et al., 2015). Hence, it is promising to analyze the spatiotemporal impacts of accidents in mitigating the negative influence of accidents, thus providing useful information for incident management. Moreover, the understanding of incident impacts on road network reliability would be beneficial in improving network design and developing incident management strategies (Lo and Tung, 2003).

In this paper, we seek to explore the impact of incidents, especially, injury accidents, on an urban road network in both temporal and spatial dimensions. Primarily, road network features are extracted in terms of Betweenness Centrality, weighted PageRank, Hub, and K-shell, which constitute the key independent variable set to tap into the incident impacts analysis. Next, we proceed to examine the incident impacts on traffic mobility from three perspectives: one in the time dimension and two in the spatial dimension. Specifically, our paper explores three concrete topics: temporally, as incident can lead to delay, we explore the association between the incident-induced delay at nearby intersections with the network features; spatially, at micro or local level, an intersection's mobility is affected by incidents, and this local impact should be associated with the network features of its location, as different intersections have disparate roles within a road network; at macro or global level, the connectivity or mobility of the entire road network would also be affected by an incident, and the reduction in network mobility should also be related to the incident site and its network features. This study contributes to the literature streams by innovatively combining complex network theory with incident impact analysis, that is, the proposed framework and inference approach can be applied to study the impacts of non-recurrent traffic events, with the purpose of identifying the network features that contribute to these impacts. Also, our data driven approach can be implemented practically and could yield managerial implications for post-event response actions and incident management.

The rest of this paper is organized as follows. Section 2 provides an overview of the contextual setting and data preparation, and Section 3 describes the empirical models used in our research. Section 4 presents and discusses the findings. Finally, conclusion and managerial implications for traffic safety and mobility are outlined in Section 5.

2. Data Preparation

2.1 Data Overview

The data in this study were collected in a medium-sized city in northern China, where 69 major urban intersections (denoted as $V, |V| = 69$) were laid out on a chessboard-style network (Fig. 1 small panel). Three datasets are organized to extract the accident impacts as dependent variables and network features as key independent variables.

1) The traffic flow dataset is used to derive the benchmark of traffic flow. It contains 584 million records, each of which represents a vehicle passing a given intersection at a certain timestamp.

2) The accident information dataset accounts for 299 injury-causing accidents that happened in this region (denoted as $\mathrm{ACC}, |\mathrm{ACC}| = 299$) and includes the accident type, severity, property loss, illumination conditions, and truck involvement.

3) The geographic information dataset contains the longitude and latitude of the intersections. The road information is harnessed to build the graph model, from which the network features are extracted.

The following two subsections first consider the network features according to the complex network theory with the road network geographic information. Accident impacts are then measured by incorporating the traffic flow and accident information datasets.

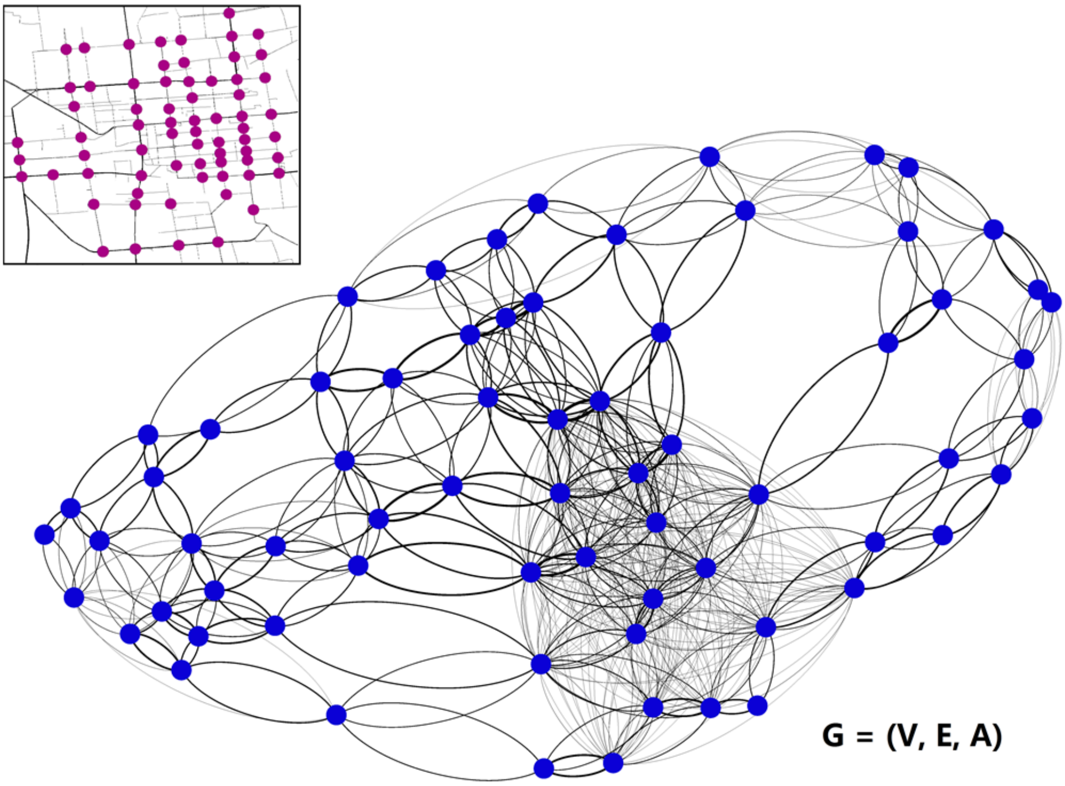

Fig. 1. Road network in reality (small panel) and its corresponding graph model (main panel). (*Nodes in both figures denote intersections. Edges in the small panel denote the real road segments while edges in the main panel denote the logical edges that link neighboring intersections.*)

2.2 Intersection Network Feature Measures

To procure the network features for each intersection, the road network in reality (see Fig. 1 small panel) is converted to a weighted directed graph $G = (V, E, A)$, where intersections are regarded as nodes V and links between them as edges E (see Fig. 1 main panel). A is the adjacency matrix with element a_{ij} denotes the weight of the edge connecting intersections i and j, which should be defined according to specific network features.

In the literature, there are three definitions of weight. First, the element of A is set to 1 if intersections i and j are connected and 0, otherwise (Ben-Tal et al., 2011), denoted as:

$$a_{ij}^L = a_{ji}^L = \begin{cases} 1, i \in \{j\} \\ 0, i \notin \{j\} \end{cases},$$

Second, weight can be defined as traffic commuting distance (Wisetjindawat et al., 2006), where we set a_{ij} as the length of the shortest path (calculated through Dijkstra algorithm) between the intersections i and j if they are adjacent:

$$a_{ij}^{D} = \text{Dijkstra}(i, j),$$

Third, weight can be defined as the intensity of the connection (Newman 2004). The annual average daily traffic (AADT) of i at the corresponding direction during midday period (i.e., 10:00am-16:00pm) is used to represent weights in this study, as a greater average traffic volume embodies a more intense connection between two places:

$$a_{ij}^{Q} = \text{AADT}(j, \text{Dir}: i \rightarrow j),$$

Applying complex network theory to the constructed weighted graph, we consider the following four network features: Betweenness Centrality, weighted PageRank, Hub, and K-shell. We exclude in-degree and closeness Centrality that are widely used in complex network study here because they are highly correlated to weighted PageRank. Each selected feature represents a particular aspect of network behavior that does not overlap with the other selected features. The graphs with distance-weighted edges, connection-intensity-weighted edges and unit edges are chosen to assess Betweenness Centrality, weighted PageRank, Hub, and K-shell, respectively.

2.2.1 Betweenness Centrality

Betweenness Centrality (BC) is used to quantify the frequency with which a node acts as a bridge along the shortest path of any node pairs in the network (Freeman, 1979; Crucitti et al., 2006; Bell et al., 2017). A graph with distance-weighted edges is used to calculate the

Betweenness Centrality because BC is determined by the physical distance between nodes in the road network:

$$\mathrm{BC}(i) = \frac{1}{(|V|-1)(|V|-2)} \sum_{\substack{s \neq i \neq t \in V \\ s \neq t}} \frac{\sigma_{st}(i)}{\sigma_{st}},$$

where σ_{st} is the number of shortest geodesic paths from s to t, and $\sigma_{st}(i)$ is the number of shortest geodesic paths from s to t that pass through node i. As Fig. 2(a) illustrates, BC measures nodes' transfer ability.

2.2.2 Weighted PageRank

Weighted PageRank (WPR) measures the probability that a vehicle will traverse a certain intersection by random walk. A graph with connection-intensity–weighted edges is used to calculate the WPR because WPR reflects the node's importance in terms of connection strength:

$$\mathrm{WPR}(i) = \frac{(1-d)}{|V|} + d \sum_{j \in \mathrm{IN}(i)} \frac{\mathrm{WPR}(j) a_{ij}^{Q}}{|\mathrm{OUT}(j)|},$$

where $\mathrm{IN}(i)$ are the in-neighbors of node i, i.e. the nodes incoming to node i, $|\mathrm{OUT}(j)|$ is the out-degree of j, i.e.the number of nodes outgoing from node j, $d \in [0,1]$ is a damping factor, which is normally set to 0.85 (Brin and Page, 2012), and a_{ij}^{Q} are the weights of the out-going edges from j. Schematic of WPR's definition can be found in Fig. 2(b).

2.2.3 Hub

Hub is also adapted from a webpage modeling approach, specifically as applied to the analysis of social networks. To find a hub and its corresponding index authorities in the network, we use the HITS algorithm that by solving the following simultaneous equations (Kleinberg, 1999):

$$\begin{cases} \text{Auth} = \vartheta A \text{ Hub} \\ \text{Hub} = \rho A^{\mathrm{T}} \text{Auth} \end{cases},$$

where A is the connection-intensity–weighted adjacency matrix, as Hub is defined based upon connection strength; $\vartheta, \rho \in \mathbf{R}$, $\vartheta\rho = \lambda^{-1}$ (ρ is set to 1 to avoid loss of generality); and λ is the largest eigenvalue of the co-citation matrix, AA^{T} or $A^{\mathrm{T}}A$. A previous study of microblog communities (Java et al., 2007) found that some blog users with very high Authority have relatively low Hub because they have many followers but do not follow many other bloggers, or vice versa. In the road network, an intersection with a higher Hub is one that points to more nodes, whereas a high Authority represents a node that is linked by hubs (as Fig. 2(c) demonstrates).

2.2.4 K-Shell

K-shell (KS), was developed by scholars to study how epidemics within networks spread. According to Kitsak et al. (2010), the topology of a network organization plays a vital role in the spread of viruses. In some cases, a high out-degree node that is strategically placed in the core of the network can make significant effects and induce infection throughout a huge fraction of the network. However, if a high out-degree node is situated at the periphery of a network, it will have a smaller impact on the spread of the virus. Hence, to study the spreading process, it is necessary to distinguish a network's core and periphery; KS is an effective tool for this purpose, and is obtained via the K-shell decomposition algorithm. Specifically, in the graph with unit edges, we first remove all nodes (including their edges) with a degree of one, and assign them to the 1-shell. Then, we recursively repeat the same procedure until all of the nodes in the network have been assigned to a corresponding KS (Carmi et al., 2007), as Fig. 2(d) shows.

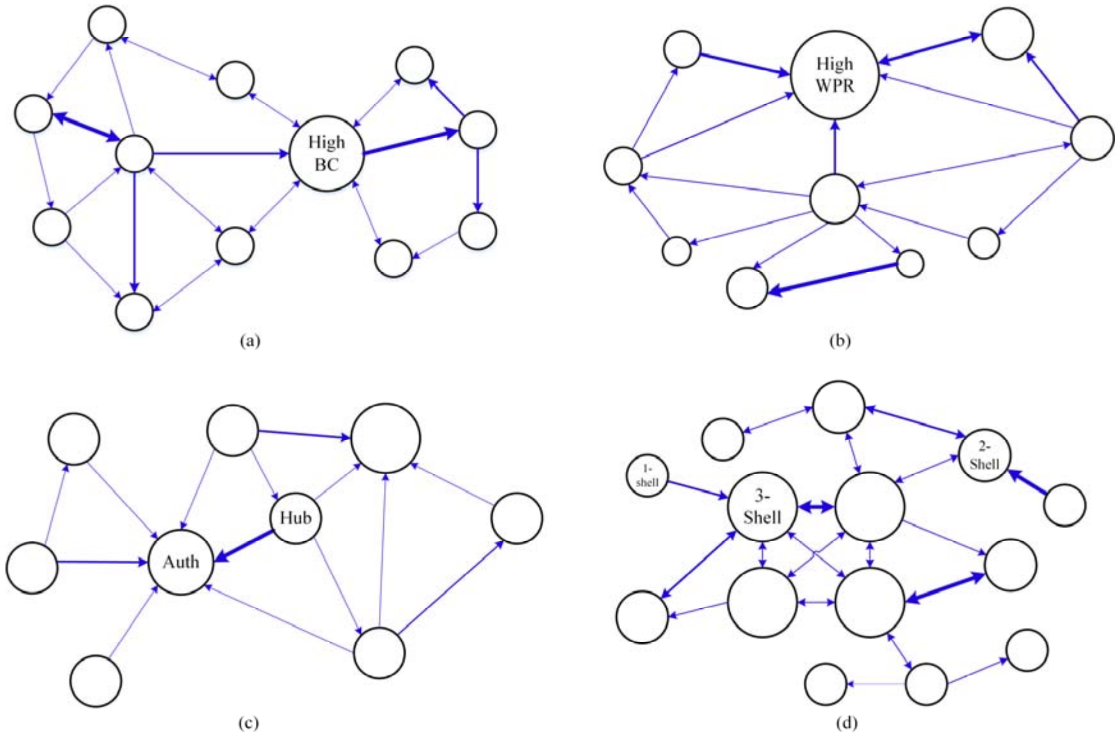

Fig. 2. Schematic diagram of the network features. A greater radius indicates a larger value of the corresponding feature.

2.2.5 Interpretations of Network Features

The color maps in Fig. 3 intuitively illustrate how different network features can reflect the functionality of intersections within a road network.

As Fig. 3(a) shows, the BC feature measures the importance of an intersection in terms of the shortest path. Intersections with a higher BC are more likely to be the shortest paths, and they may be more prone to be traversed by road users who like to take shortcuts.

The distribution of intersections with WPR values is demonstrated in Fig. 3(b). Originally, WPR was proposed to assign importance to webpages in a Web graph. By analogy, as we are interested in the relative importance of nodes in a road network (i.e., road intersections), a node's WPR value is expected to reflect its importance in terms of connection strength: if two intersections have similar WPR values, the probability that a driver will visit either of them should be similar.

"Hub" acts as a resource list that directs road users to recommended authorities. As we have already defined other features that function as authorities (i.e., PageRank, which is highly correlated with the Authority index in our case), only Hub is considered in the following steps to measure an intersection's "guiding" role. As can be seen in Fig. 3(c), high-Hub intersections are nodes that point to other nearby important intersections in terms of connection strength.

In Fig. 3(d), the actual distribution of intersections' KS values is in line with its definition: high-KS intersections are located in the core area of the road network, and low-KS intersections are located in the periphery.

Each of the above network features represents one specific facet of an intersection's role (or functionality) in the road network. They together constitute a prime covariate set that leverages the urban road network context to study the accident impacts.

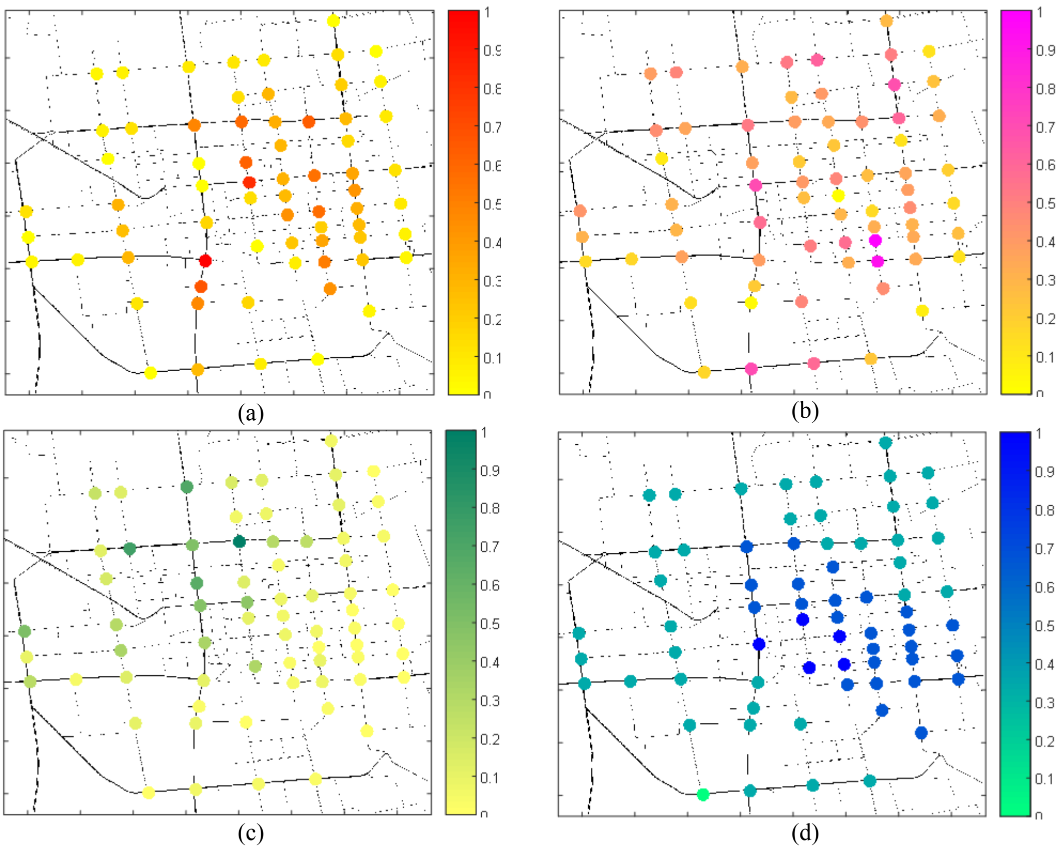

Fig. 3 Distributions of values for various intersection network features within the urban road network: (a) Betweenness Centrality; (b) Weighted PageRank; (c) Hub; (d) K-shell.

2.3 Accident Impacts Measurement

This study investigates how an accident's impact is associated with road network features. The multifaceted impacts induced by traffic accidents are involved in the models as dependent variables, which are measured both temporally and spatially. On the time dimension, we select the accident-induced delay at nearby intersections as the metric. Spatially, we consider the microscopic and the macroscopic effect. For the former, we measure the frequency that an intersection's mobility is affected by accidents at individual-level. For the latter, we capitalize on network mobility reduction to quantify an accident's overall impact on the road network.

Accident impacts measurement in this study is based on one reasonable assumption that are also supported by several prior research (Skabardonis et al., 2003; Williams and Guin, 2007). As traffic accidents are nonrecurring events that would cause congestion or delays at or near urban intersections, it is assumed that the traffic flow under such abnormal situation should be significantly different from that under the recurrent situations. In this regard, we need to firstly find out the benchmark, based on which the outlier from normal region can be regarded as impacts by accidents. Following pertinent literatures (Sun et al., 2016; Hojati et al., 2016), we acquire this benchmark, namely, the traffic flow profile under recurrent situations, from historical data.

A stepwise data driven procedure is taken to obtain the recurrent traffic flow profile. To begin with, let $i = 1, \ldots, 69$ denotes the index of all urban intersections; $j = \{\mathrm{S} \rightarrow \mathrm{N}, \mathrm{E} \rightarrow \mathrm{W}, \mathrm{N} \rightarrow \mathrm{S}, \mathrm{W} \rightarrow \mathrm{E}\}$ represents four directions at which the traffic flow enters an intersection; $d = 1, \ldots, 365$ represents all days in one year; $t = 1, \ldots, k, \ldots, K$ denotes the index of time interval at which the traffic volume is aggregated (i.e., $K = 144$ for 10-min interval), the aggregated traffic flow at direction j of intersection i at day d therefore is denoted as $f_{i,j,d,t}$, and the annual aerage then becomes $\bar{f}_{i,j,t}$.

Next, since the traffic demand is subject to temporal effects, i.e., seasonal variations and day-of-the-week effect (Rakha and Van Aerde, 1995; Thomas et al., 2010), $\bar{f}_{i,j,t}$ may not be a good indicator of recurrent traffic flow profile during a short span of time and therefore it is necessary to specify the temporal term d according to the reported accident starting time. In this study, we manually separate workdays (Monday to Friday), weekends (Saturday and Sunday) and special days (festivals and special events). We find that the flow profiles of these three categories of days show significant heterogeneities: the flow profile of workdays has two salient peaks (i.e., morning peak and evening peak) while that of weekends and special days does not; the overall traffic demand of special days is 60-70% as much as that of weekends. In addition, the seasonal variation is found to be significant in this city: the traffic demand during summer (Jun. to Sept.) is 80% as much as that during winter (Nov. to Feb.). Hence, as we compare the traffic flow during accident and the recurrent flow profile, we adjust the temporal term d proportionally. To be more specific, as we study a given accident, denoted as $acc \in ACC$, which took place on day d^{acc}, we collect 15 days from the same category both before and after d^{acc}, denoted as d^A $(A = 1, \ldots, 30)$ and then calculate the moving average $\bar{f}_{i,j,d^{acc},t} = \frac{1}{30}\sum_A \bar{f}_{i,j,d^A,t}$. The obtained $\bar{f}_{i,j,d^{acc},t}$ can be regarded as the recurrent flow profile or core benchmark specified for a given accident.

Besides the core benchmark, we also define the band. The aggregated traffic flow in the set of d^A at k-th interval is represented by $f_{i,j,d^{acc},k}$, which contains 30 points. Due to the stochastic nature of traffic flow, $f_{i,j,d^{acc},k}$ is assumed to follow $\mathcal{N}\left(\bar{f}_{i,j,d^{acc},k}, \sigma_{i,j,d^{acc},k}\right)$, where:

$$\sigma_{i,j,d^{acc},k} = \sqrt{\frac{1}{30}\sum_A \left(f_{i,j,d^A,k} - \bar{f}_{i,j,d^{acc},k}\right)^2},$$

is the standard deviation of the traffic flow of at direction j of intersection i at the k-th time interval. For the normal distribution, the values less than two standard deviations from the mean account for 95.45% of the data. As non-recurrent cases such as traffic accidents can be viewed as rare events, we thus apply the idea of outlier detection to find the impacts of accident (Guo et al., 2015). The normal region in this study is defined as $[\bar{f}_{i,j,d^{acc},t} + 2\sigma_{i,j,d^{acc},t}, \bar{f}_{i,j,d^{acc},t} - 2\sigma_{i,j,d^{acc},t}]$. Once we have detected more than two clustered outliers (i.e., points that lay outside the normal region) in current traffic flow around the reported accident starting moment, we then consider this abnormality to be the impact of a non-recurrent traffic incident. Moreover, since the reported time was subject to artifacts, both spatial boundary and temporal boundary are set to avoid misidentification. Hojati et al. (2016) used 60 min before and after the specified incident time as temporal boundary, and Chung and Recker (2015) defined the accident impact spatiotemporal boundary as maximum freeway section by 90 minutes. In this study, we use 30 min before and after the reported time and 1.8km as spatiotemporal boundary for outlier detection.

Through the above procedure that collaboratively mine the traffic data and traffic accident data, it becomes feasible to quantify some basic impact indicators for each accident record acc $\in$ ACC. Specifically, for each accident, we can detect a set of intersections whose traffic flow is affected by this accident, along with the count of clustered outliers on the traffic flow of these intersections, upon which we are able to derive three sorts of accident impacts with concrete meaning. Fig. 4 illustrates the schematic diagram of the benchmark and the outliers detected.

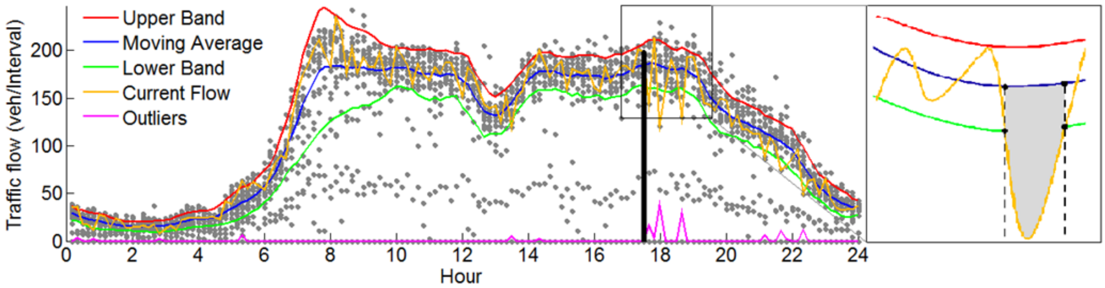

Fig. 4. Plot for detecting whether the traffic flow at an intersection is affected by a nearby accident. (*The blue line denotes the average flow profile of a given direction of an intersection; red and green lines denote the upper and lower bands of the normal interval, respectively. The yellow line indicates the current traffic flow, and the pink line indicates the outliers detected. The grey area represents the flow reduction caused by the accident.*)

2.3.1 Temporal Impact

"Accident delay" refers to the time from the moment the incident took place to the minute the traffic flow returned to normal (Garib et al., 1997; Nam and Mannering, 2000; Chung, 2010). The delay can thus be calculated by the time interval during which the current traffic flow is significantly different from the recurrent flow profile. The delay for a given accident acc, in direction j at intersection i can be calculated as follows:

$$T_{i,j,d^{acc}} = N_{i,j,d^{acc}} \times T^{\mathrm{A}},$$

where $N_{i,j,d^{acc}}$ is the number of clustered outliers detected by the aforementioned method and T^{A} is the aggregated time, which in this study is 10 min. At the intersection level, the accident delay of a given accident acc is determined by the maximum delay among the different directions, which is expressed as $T_{i,d^{acc}} = \max\{T_{i,j,d^{acc}}\}$. For all 299 accidents, we detect 846 intersection-level delay samples, which are used as a dependent variable in the subsequent analysis.

2.3.2 Microscopic Spatial Impact

Microscopic spatial impact focuses on the intersection-level accident impact with regard to the network features, i.e., we are curious whether intersections with different functionalities (embodied by their distinctive network features) would demonstrate different propensities to be affected by accidents. An understanding of how the intersections' network features are associated with its susceptibility to accidents would help transportation administrators to determine which part of the road network is more likely to be affected by accidents. At the same time, as suggested by Li et al. (2015), improvements in traffic on those key road segments or intersections can significantly improve global traffic. In the same vein to the accident occurrence, we define the annual affected times (AAT) for each intersection $i \in V$ to depict the local spatial impact of accident, by counting the number of times that i is affected by accidents. For instance, for intersection #1, this value is 19, indicating that intersection #1 was affected by accidents 19 times in the focal year. We believe that such a straightforward measure is a suitable proxy for the local impact of accidents on each intersection i, as the larger the value of $\mathrm{AAT}(i)$, the higher the likelihood that the traffic flow of i will show non-recurrent outliers. For all 69 intersections, we acquire tantamount intersection-level impact samples as the second dependent variable.

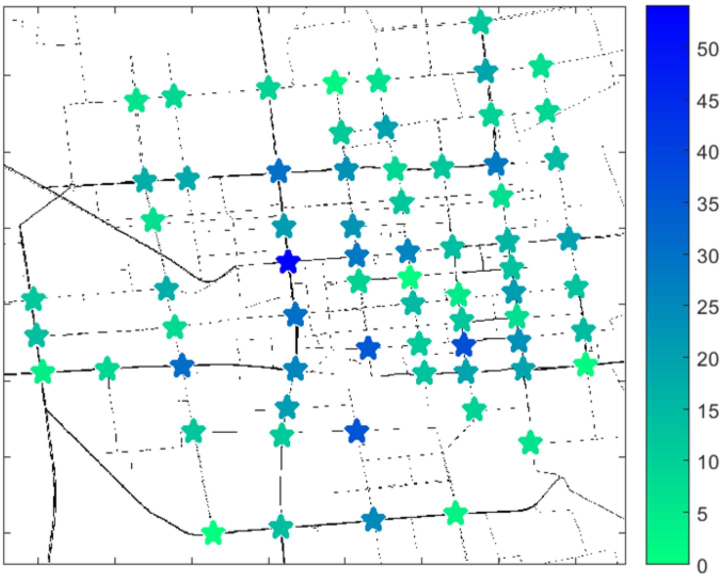

Fig. 5. Spatial distribution of each intersection's annual affected times

It is worth mentioning that the local impact may subject to spatial correlation. As seen in Fig. 5, intersections that are affected seriously seem to cluster along the main roads. According to Goodchild (1992) and Zhou et al. (2016), spatial correlation exists in most spatial dataset, since adjacent regions may share similar socioeconomics or demographic characteristics. More relevantly, Li et al. (2014) showed that jams in city traffic have long-range spatial correlations that decay slowly with distance. To specify an appropriate model, we need to conduct the spatial correlation test. We draw upon Moran' I statistic to judge if we should adopt spatial model:

$$\text{Moran}'\text{s I} = \frac{\sum_{i=1}^{|V|}\sum_{j=1}^{|V|} a_{ij}\cdot(\text{AAT}(i)-\overline{\text{AAT}})(\text{AAT}(j)-\overline{\text{AAT}})}{\sum_{i=1}^{|V|}\sum_{j=1}^{|V|} a_{ij}\cdot\sum_{i=1}^{|V|}(\text{AAT}(i)-\overline{\text{AAT}})^2/n},$$

where $\text{AAT}(i)$ is the annual affected times of intersection i, $\overline{\text{AAT}} = 15.49$ is the mean of AAT value of all intersections, a_{ij} is the corresponding element of the adjacency matrix. Here, we adopt all three adjacency matrices (i.e., distance-weighted, connection-intensity-weighted and unit) to calculate the Moran' I with its Z test statistic (Moran, 1950). The results of Moran's I test are shown in Table 1, in which we confirm that the AAT data is subject to the spatial correlation, given the significant level 0.05. Accordingly, the spatial correlation should be taken into account in the following modelling.

Table 1

Results for the Moran' I spatial correlation test

	Unit	Distance	Connection intensity
Moran' I	0.196**	0.289**	0.229**
Z(I)	3.518	3.135	3.137

** p<0.05

2.3.3 Macroscopic Spatial Impact

Network-level impact can be imputed to the aggregated manifestation of its local counterparts: as road segments or intersections are blocked due to accident-induced effects, it would cause the road users to pay additional cost in time or travel a longer distance, which means the network becomes less connected with reduced mobility(Bell, 2000). To measure this negative impact, we apply a notion that is widely used in the complex network analysis: the average diameter increment. According to Albert et al., (1999), the average diameter (D) of a network is defined as its interconnectedness, namely, the average value of the shortest paths of any two-node pair. A change in average diameter, ΔD, reflects the extent of the impact when a certain set of V is removed or blocked (Albert et al., 2000). In a road network, once an accident takes place, it will exert a negative impact on the traffic flow of nearby intersections, which can be seen as errors. For simplicity, the negatively affected intersections are assumed to be total malfunctions, i.e., treated as deadlocks. Under this assumption, the average diameter of the affected network with its interior errors must be greater than that of the original network, and its difference is considered as the accident damage to the mobility of the road network, as it represents the average increase in traveling distance. For each accident $acc \in \mathrm{ACC}$, its average diameter increment, ΔD, is expressed as

$$\Delta D(acc) = \frac{1}{|V \backslash V_{acc}|} \sum_{i',j' \in V \backslash V_{acc}} d(i',j') - \frac{1}{|V|} \sum_{i,j \in V} d(i,j),$$

where V_{acc} denotes the set of all of the affected intersections and $d(i,j)$ is the shortest distance from node i to j. The larger the values of $\Delta D(acc)$, the heavier the damage to the road network. For all 299 accidents, we derive network-level impact samples as the third dependent variable with its distribution shown in Fig. 6, from which an exponential distribution seems to fit the data well.

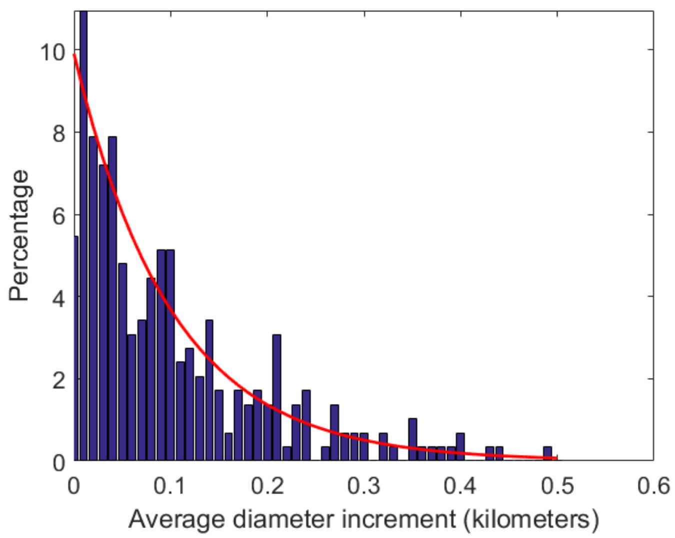

Fig. 6. Distribution of the average diameter increment

2.4 Summary Statistics

Table 2 presents the summary statistics of the dependent and independent variables. To evaluate the accident impacts more accurately, we include other control variables. These intersection-specific variables (i.e., AADT, V/C ratio and number of lanes) and accident-specific variables (i.e., truck involvement, fatality, property loss, illumination, and accident type) are in line with what had been incorporated in the relevant previous studies (Chung and Recker, 2015; Sun et al., 2016). Before estimating the association measures for accident impacts, correlation tests were conducted for each pair of independent variables. Thus, the independent variables in Table 2 are free of multicollinearity risks.

Table 2

Descriptive statistics for the independent and dependent variables

Variables	Min	Max	Mean	S.D.
Independent variables				
Intersection network feature [a]				
Betweenness Centrality	0	1	0.229	0.217
Weighted PageRank	0	1	0.448	0.220
Hub	0	1	0.216	0.244
K-shell	0	1	0.498	0.180
Intersection-specific variables				
AADT (veh./15min)	96	291	217.226	46.228
V/C ratio (volume/capacity ratio)	0	2.220	0.711	0.083
Number of lanes	4	27	13.653	3.601
Accident-specific variables				
Truck involvement	0	1	0.204	0.404
Fatality	1-Property loss only			
	2-Injury and/or fatality			
Property loss level	1-0			
(RMB)	2-(0,500]			
	3-(501,2000]			
	4-(2001,10000]			
	5->10000			
Illumination	1-daytime			
	2-night with illumination			
	3-night without illumination.			
Accident type	1-vehicle crash			
	2-other			
Dependent variables				
Accident delay (minute)	10	115	34	26
Impact propensity	1	54	15.411	10.010
Average diameter increment (km)	0.004	0.492	0.313	0.097

[a] Network features are scaled to [0,1].

3. Statistical Modeling Methods

In this study, accident impacts are measured in two dimensions. Temporally, accident delay is gauged at intersections in the vicinity of the accident spot. Spatially, 1) the local impact is measured by the frequency at which each intersection is affected and 2) the network impact of each accident is measured by the average diameter increment of the entire network. As these three dependent variables are subject to different categories (i.e., duration time, count data and positive real numbers), we apply three distinct statistical models that can exactly capture the nature of the data to determine the relationship between these dependent variables and the corresponding network features and other control variables.

3.1 Hazard-based Duration Model

A hazard-based model to study incident duration was first introduced by Jones et al. (1991), and has then been applied by many researchers (Nam and Mannering, 2000; Chung, 2010; Hojati et al., 2013). In the literature, the three main hazard-based duration models in wide use are the semiparametric, parametric and frailty models. These three models are all applied and compared in terms of their likelihood ratio statistics in our study of accident delay.

The following Cox model, developed by Cox and Oakes (1984), is a semiparametric method aimed at evaluating the influence of covariates with little or no prior knowledge of the hazard function (Hou et al., 2014):

$$h(t) = \frac{f(t)}{S(t)} = \lim_{\Delta t \to 0} \frac{\Pr(t \le T \le t + \Delta t | T \ge t)}{\Delta t},$$

where $h(t)$ is the hazard rate at which the incident delay will end at time t, given this delay has lasted for t min, $f(t)$ is the probability density function of the event duration (i.e., accident delay) and $S(t) = 1 - \int_0^t f(u)du$ is the survival function that accounts for the probability that the event will last until t min. Given the hazard function, the Cox proportional hazard model is described as follows:

$$h_C(t|\lambda) = \lambda h_0(t),$$

where $h_0(t)$ represents the unspecified baseline hazard function (hazard function $h_{\mathrm{C}}(t|\lambda)$ reduces to the baseline hazard function $h_0(t)$ when all independent variables are equal to zero. In other words, $h_0(t)$ is the hazard function in the absence of covariates) and $\lambda = \exp(-\boldsymbol{\beta X}) > 0$ describes how the hazard response to m covariates of $\boldsymbol{X} = (1, X_1, \dots, X_m)$; $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_m)^{\mathrm{T}}$ is a vector of m parameters to be estimated. To evaluate the intersection-level accident delay, the intersection network features, intersection-specific variables and accident-specific variables are all included as covariates, among which the network features are the primary concern.

The proportional hazard (PH) model specifies a theoretical ground of the hazard function and considers the influence of a multiplicative factor derived from independent variables (Breslow, 1975). Four distributions—exponential, Weibull, log-logistic and gamma—are usually chosen to fit the hazard function. The Weibull hazard function was selected in this study. The assessment of accident delay data is expressed as follows:

$$h_{\mathrm{PH}}(t|\lambda, p) = \lambda p(\lambda t)^{p-1},$$

where $p > 0$ is known as the Weibull scale parameter.

A frailty model further introduces a heterogeneity term to evaluate the influence of unobserved factors that may not be captured by independent variables, and thus outperforms the original PH model (Chung et al., 2015). The Weibull hazard function with the heterogeneity of Inverse-Gaussian distributions with unit mean and variance θ is given by:

$$h_{\mathrm{F}}(t|\lambda, p, \theta) = \frac{\lambda p(\lambda t)^{p-1}}{\sqrt{1+2\theta(\lambda t)^p}},$$

whose mathematical derivation is available in Appendix A.

Likelihood ratio statistics are calculated to compare the fitness of the above models:

$$\chi^2 = -2[LL(0) - LL(\beta)],$$

where $LL(0)$ is the initial log-likelihood, and $LL(\beta)$ is the log-likelihood value at the convergence of this model. Higher levels of significance for χ^2 indicate superior goodness of fit (Nam and Mannering, 2000).

3.2 Bayesian Negative-binomial CAR Model

As per the Moran' I test results, a statistical model that incorporate the spatial correlation between adjacent intersections is used to evaluate the accident-impact propensity of each intersection in the road network. The Bayesian Negative-binomial CAR model derives from simple count data situation. Due to the stochastic nature of the occurrence of events, the relationship between the number of occurrences and the dependent variable y_i usually follows the Poisson distribution:

$$y_i \sim Possion(\lambda_i),$$

where λ_i is the expectation of y_i.

In this study, the dependent variable y_i refers to the number of times an intersection is affected by accidents in a year (annual affected times; $\mathrm{AAT}(i)$). As $\mathrm{AAT}(i)$ is subject to over-dispersion, a Negative-binomial (NB) model (Poch and Mannering, 1996) is applied by linking the expectation of y_i to the independent variables with a random error term such that

$$\lambda_i = \exp(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X} + \varepsilon_i),$$

where $\boldsymbol{\beta}^{\mathrm{T}}$ is the vector of the estimable parameters and $\exp(\varepsilon_i)$ follows a gamma distribution. Covariate $\boldsymbol{X}$ in this model includes the intersection network features as well as other intersection-specific variables.

According to Wang and Kockelman (2013), Buddhavarapu et al. (2016), Guo et al. (2017), and Xu et al. (2017), the heterogeneity brought by spatial correlation can be accounted for by applying a Bayesian conditional autoregressive (CAR) model, in which the random effect term ϕ_i is involved as follows:

$$\lambda_i = \exp(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X} + \varepsilon_i + \phi_i),$$

The effect of spatial correlation among intersections, ϕ_i, has the following conditional distribution:

$$\phi_i|\phi_{(-i)} \sim N\left(\frac{\sum_{j,i\neq j} a_{ij}\phi_j}{\sum_{j,i\neq j} a_{ij}}, \frac{\tau_c}{\sum_{j,i\neq j} a_{ij}}\right),$$

where $\phi_{(-i)}$ is the set of all ϕ without ϕ_i; a_{ij} is the spatial relationship between i and j, which in our case equals 1 if i and j are adjacent, and 0 otherwise; τ_c is the precision parameter that Wakefield et al. (2000) suggest follows the gamma prior distribution Gamma (0.5,0.0005).

3.3 Generalized Linear Model

To estimate the overall impact of each accident on network mobility, the average diameter increment is evaluated by an estimation model with the network features of the accident site as covariates. As the average diameter increment follows an exponential distribution, as shown in Fig. 6, the Generalized Linear Model (GLM) of the gamma family with a log-linear link function is implemented to reveal the relationships of the non-negative dependent variable Y with the independent variables $\boldsymbol{X}$, which can be expressed as the following form:

$$\ln(Y) = \beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X} + \varepsilon,$$

where the random error term ε follow a normal distribution. Covariates vector $\boldsymbol{X}$ in this model includes the network features of the intersection that is closest to the accident site as well as other accident-specific variables.

4. Results and Discussion

4.1 Temporal Impact

Three hazard models (i.e., Cox, PH and frailty models) were developed and estimated using Stata® software. Table 3 summarizes the parameter estimates of all of the network features and selected control variables that are statistically significant at 90% confidence level for at least one of these models.

Table 3

Summary of estimation results of three hazard models for accident delay

Covariate		Hazard Ratio [a] (Std. Err.)		
		Cox model	PH model	Frailty model
Betweenness Centrality (BC)		1.269* (0.109)	1.316*** (0.122)	1.535*** (0.221)
Weighted PageRank (WPR)		1.028 (0.144)	1.053 (0.148)	1.023 (0.222)
Hub		0.986 (0.089)	0.988 (0.093)	0.984 (0.087)
K-shell (KS)		1.144 (0.113)	1.217** (0.123)	1.360** (0.206)
V/C ratio		0.995*** (0.001)	0.994*** (0.001)	0.989*** (0.001)
Number of lanes		0.955** (0.022)	0.922*** (0.021)	0.897*** (0.032)
Property loss	0		*Control*	
	1-500	1.195 (0.069)	1.351* (0.216)	1.577* (0.385)
	501-2000	1.181 (0.197)	1.309 (0.219)	1.532* (0.393)
	2001-10000	1.101 (0.225)	1.011 (0.226)	1.072 (0.366)
	>10000	1.423 (0.360)	1.869** (0.474)	2.317* (0.916)
Heterogeneity term θ		--	--	2.321 (0.274)
$LL(0)$		-13614	-3056	-2889
$LL(\beta)$		-13592	-2999	-2807

Likelihood ratio statistics	44	114	**164**

* p<0.1, ** p<0.05, *** p<0.01.

From Table 3, it can be seen that the frailty model has the highest likelihood ratio statistic (i.e., χ^2 value) and performs better than the semiparametric and parametric models without heterogeneity. Incorporating the heterogeneity term θ, which allows parameters to vary across observations, enables the exploration of additional significant factors affecting incident delay (Hojati et al., 2013). Moreover, the coefficient of the inverse-Gaussian heterogeneity term θ is estimated to be 2.321 with a significant likelihood ratio test of $H_0: \theta = 0$ (p-value = 0.000), which proves its existence. The findings with respect to control variables are in line with our previous work (Sun et al., 2016). Intuitively, accidents with heavier property loss lead to longer delay because clearing the debris of such accidents may need more time. Additionally, with more lanes, the dispersion capacity of an intersection should be stronger, and thus the accident-induced congestion can be mitigated more easily and faster.

The empirical result indicates that high-BC intersections suffer from significantly longer accident impact duration, as the accident delay will be 53.5% (1.535 − 1) greater if the BC increases from the minimum to the maximum accordingly. A high-BC intersection may be more likely to be considered as an optimal shortcut. Even if an incident occurs nearby, road users may still choose a high-BC intersection due to the shorter travel distance and no advance warning of the incident impact.

The estimation results reveal that the KS value of an intersection also has a statistically significant, yet less influential effect compared with BC, on its accident delay. Accident delay will be 36% (1.360 − 1) greater if the KS increases from the minimum to the maximum. High-KS nodes are typically located within the core of the network (i.e., in an urban area), whereas low-KS nodes are normally distributed at the periphery (near suburbs). In the city where our dataset was collected, the urban area usually suffers from heavier traffic and a higher risk of congestion; once an accident occurs, vehicles at central urban intersections (as Fig. 3(d) shows) must wait longer for congestion to clear.

4.2 Microscopic Spatial Impact

Bayesian Markov Chain Monte Carlo (MCMC) method is used for the Negative-binomial CAR model estimation (Zhou et al., 2016). We use WinBUGS® software to run the algorithm. The posterior distribution of estimators are obtained with three MCMC chains of 30,000 iterations each, of which the first 5,000 iterations are considered as burn-in periods. Two-tailed tests at the 5% confidence interval (C.I.) are used to reach significance at outside the areas of the 2.5% and 97.5% percentiles. To be deemed significant at the 5% level, a variable must exclude zero in its C.I. range. The posterior results of the parameters for the impact propensity evaluation are listed in Table 4.

Table 4
Results of Bayesian CAR model

Covariate	Posterior Mean (Posterior S.D.)	(95% C.I.)
Betweenness Centrality	20.71* (9.06)	(5.09, 40.15)
Weighted PageRank	33.46* (15.11)	(6.23, 63.58)
Hub	-32.15* (5.59)	(-43.80, -21.29)
K-shell	2.92 (7.77)	(-13.09, 16.46)
AADT	9.98 (12.52)	(-14.41, 32.49)
Number of lanes	-6.66 (5.68)	(-17.52, 4.86)
Intercept	-28.44* (12.41)	(-46.99, -1.66)

*variables are significant at the 5% level, as their C.I. values exclude zero.

We select AADT and the total number of lanes of an intersection as control variables, as it is reasonable to deduce that higher traffic volume and capacity may lead to a higher accident impact.

However, estimates show that both AADT and the number of lanes are not significantly related to the local impact. On the contrary, three network characteristics are found to be significantly associated with the dependent variable.

With respect to Betweenness Centrality, low-BC intersections in the road network have lower risks of being affected by traffic accidents than high-BC intersections, perhaps because a low-BC intersection is less likely to be considered an optimal shortcut and therefore has less chance of being selected by drivers. Once an accident occurs, road users may still be unwilling to choose a low-BC intersection as an alternative due to the longer travel distance, which lowers the likelihood that it will be affected.

Weighted PageRank plays a similar role as BC does. One plausible explanation could be that as nearby intersections depend on one high-WPR intersection, it merges the traffic flows from neighbors, as Fig. 3(b) shows. Suppose that an accident causes abnormal flows at four nearby intersections (V_1, V_2, V_3, V_4), the traffic flows from which will all merge into another point V_5; then V_5 may suffer from a higher risk of showing an outlier due to the joint probabilities. In other words, V_5, as a high-WPR intersection, which may attract much more traffic flow once an accident occurs nearby, has a higher susceptibility to accident impacts. Moreover, among the three network features that demonstrate significant effect on AAT, the effect of WPR is considered to be the most influential, as its estimated coefficient is larger than that of the rest two features.

The association between Hub and AAT is noticeable, as it is the only network feature that shows a counter-intuitive negative correlation with AAT. According to the information theory, high-Hub nodes in a network often serve as large directories to authorities, but do not actually hold authoritative information in themselves. In road networks, high-Hub intersections may direct road users to intersections of great importance, thus possess a relatively stronger evacuation ability, which lowers the risk of being congested. Such result is in accordance with what Fig. 3(c) conveys: high-Hub intersections distribute distinctively from those important intersections in terms of the shortest path and connection strength.

4.3 Macroscopic Spatial Impact

The GLM model is developed and estimated using Stata® software; the results are shown in Table 5. As the likelihood ratio (LR) statistic is significant, the proposed model has a good fit with the observations. It is found that almost all of the significant network features of accident site (i.e., the nearest intersection in this study) are positively associated with the diameter increment caused by the focal accident. As the diameter characterizes the connectivity of the road network, a larger D indicates a shorter expected path between any two intersections. Once an accident takes place in the urban area, it begins to exert a negative influence on the traffic flow, i.e., it causes delays and congestion. As a result, the connectivity and mobility of the road network declines.

Table 5
Results of the GLM model

Covariate	Coef. (Std. Err.)
Betweenness Centrality (BC)	0.867*** (0.238)
Weighted PageRank (WPR)	-0.224 (0.425)
Hub score	1.046*** (0.230)
K-shell (KS)	1.245*** (0.285)
Intercept	-3.319 (0.214)
R-squared	0.459
Prob. > F	<0.000

* p<0.1, ** p<0.05, *** p<0.01.

These results indicate that if an accident takes place near a high-BC intersection, the overall road network mobility is highly reduced. Intuitively, this means that within road networks, high-BC intersections/nodes are those nodes where the shortest paths between all of the pairs of nodes gather, and thus they have a huge influence on the network's connectivity. Once they are blocked or congested, the overall connectivity of the network is reduced (Wasserman and Faust, 1994; Newman, 2001).

Accident network damage is shown to be positively associated with the Hub because the removal of high-Hub nodes eliminates the shortest pathways connecting the relevant authorities. In addition, compared to vehicle crashes, other types of accidents (stationary, off-road, rollover) lead to less mobility reduction, presumably because of lower road occupancy.

Furthermore, the results also indicate that if an accident occurs at high-KS nodes, the mobility of the road network is significantly damaged. In addition, as demonstrated by the absolute value of the estimated coefficient, the effect of KS could be regarded as the most influential network feature that correlates with the macroscopic spatial impact. In the road network, an ongoing accident can be considered a single spreading origin, which behaves like an epidemic. Our results show that the most efficient spreaders are located in the inner core of a network (as Fig. 3(d) shows). There are many pathways that transmit effects to other nodes, thus disabling more nearby intersections and causing serious mobility problems. Kitsak et al. (2010) present similar results.

5. Conclusions

This study investigates how network features are taken into account for the evaluation of incident impacts on traffic mobility in urban road networks. Geographic information of 69 intersections with their linking roads within an urban road network is used to build up three directed graphs with three kinds of weighted edge. Based on these graphs, four network features (Betweenness Centrality, weighted PageRank, Hub, and K-shell) are measured. The incident impacts are measured as temporal and spatial impacts by accidents in terms of: one in the time dimension (accident delay), two in the spatial dimension (local impact propensity as micro spatial impact and network mobility reduction as macro spatial impact). The main findings on the association of incident impacts with network features are listed in Table 6.

Table 6

Summary of results (only significant network features are presented)

Network features	Accident Delay	Impact propensity	Mobility reduction
Betweenness Centrality (BC)	positive	positive	positive
Weighted PageRank (WPR)	--	positive	--
Hub	--	negative	positive
K-shell (KS)	positive	--	positive

Hazard-based models are harnessed to study how the accident delay at intersections is affected by the intersection’s network features. The estimation results unravel that intersections with larger BC or KS values tend to suffer from longer delay. A Negative-binomial Bayesian CAR model is developed to analyze the intersection-level spatial impact. The results show that the annual affected times of an intersection are significantly associated with its Betweenness Centrality, weighted PageRank, and Hub, but are not related to its traffic features. A gamma-family linear model is utilized to investigate the incident-induced network mobility reduction. Modeling results reveal that this overall impact is strongly correlated with the K-shell, Betweenness Centrality, and Hub of the intersection that is closest to the accident site.

The findings shown in Table 6 could yield several managerial implications, as they provide a particular perspective to evaluate the impact of incidents, especially accidents in this study, on traffic safety and mobility. As safety issues invariably draw our attention, the analysis and prevention of post-incident consequences have become essential for both road administrations and users. Theoretically, our study contributes to the literature streams by proposing an efficient framework and proper methods that combine complex network theory with incident impact evaluation. Managerially, traffic administration could draw upon our idea to conduct more efficient post-incident resilience enhancing actions in the following ways. First, results in the paper provide useful insights to evaluate incident-induced consequences and thus take actions based on the network's spatial characteristics. For instance, traffic administration should rationally prevent more traffic from entering high-risk intersections (e.g., high-BC intersections) after an incident happens to avoid heavier congestion and longer delay; meanwhile, traffic can be guided to intersections with high Hub. Second, our findings on the impact of incidents on traffic mobility with regard to network features can also help the traffic administrations to identify those susceptible intersections within a road network and those accidents that could potentially cause heavier network damage. Last but not least, for road users, if they are able to acquire timely traffic status and equitable warnings (e.g., "please detour around, though this is the shortest path, your passing may worsen the traffic"), they may strategically alter the route.

Our research is subject to the limitation of data availability, that as real distance and accurate traffic flow distributions at the edges are not accessible, we have to use the calculated distance and a proxy measure as the edge weight to construct the graph. Future studies may exploit more fine-grained data to measure network features in a more precise manner, and thus more accurately understand incident impacts. Also, since the traffic flow data is collected only at the major intersections, we are not able to model those minor intersections. If flow data at more microscopic level is available, then the entire road network can be modeled.

Finally, our work is only a first step. It unravels the importance of studying incident impacts in the context of network features. By leveraging this new dimension into incident impacts analysis, we hope to inspire more research into related questions, such as the causal effects of network features on incident occurrences and other spatiotemporal impacts, and how we can measure these network features in a more accurate, dynamic manner. We believe that the answers to these questions will further deepen our understanding of the impact of incidents on urban road networks.

Acknowledgments

The work described in this paper was supported by grants from the National Natural Science Foundation of China (Grant No. 71301083), the Joint Research Scheme of the National Natural Science Foundation of China/Research Grants Council of Hong Kong (Project No. 71561167001 & N_HKU707/15), and the Research Funds of Tsinghua University (No. 20151080412). The last author was also supported by the Francis S Y Bong Endowed Professorship in Engineering.

References

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accid. Anal. Prev., 32(5), 633-642.

Albert, R., Jeong, H., Barabási, A.L., 1999. Internet: Diameter of the world-wide web. Nature, 401(6749), 130-131.

Albert, R., Jeong, H., Barabási, A.L., 2000. Error and attack tolerance of complex networks. Nature, 406(6794), 378-382.

Barua, U., Azad, A., Tay, R., 2010. Fatality risks of intersection crashes in rural undivided highways of Alberta, Canada. Transp. Res. Rec., 2148, 107–115.

Bell, M.G.H., Kurauchi, F., Perera,S., Wong,W., 2017. Investigating transport network vulnerability by capacity weighted spectral analysis. Transp. Res. B, 99, 251-266.

Ben-Tal, A., Do Chung, B., Mandala, S.R., Yao, T., 2011. Robust optimization for emergency logistics planning: Risk mitigation in humanitarian relief supply chains. Transp. Res. B, 45(8), 1177-1189.

Breslow, N.E., 2011. Analysis of survival data under the proportional hazards model. Int. Stat. Rev., 1975, 45-57.

Brin, S., Page, L., 2012. Reprint of: The anatomy of a large-scale hypertextual web search engine. Comput. Netw., 56(18), 3825-3833.

Buddhavarapu, P., Scott, J.G., Prozzi, J.A., 2016. Modeling unobserved heterogeneity using finite mixture random parameters for spatially correlated discrete count data. Transp. Res. B, 91, 492-510.

Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y., Shir, E., 2007. A model of Internet topology using K-shell decomposition. Proc. Natl. Acad. Sci., 104(27), 11150-11154.

Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. Transp. Res. B, 46(1), 253-272.

Chandler, B.E., 2013. Signalized Intersections Informational Guide, 2nd edition. Federal Highway Administration (FHWA), FHWA-SA-13-027.

Chen, C., Xie, Y., 2016. Modeling the effects of AADT on predicting multiple-vehicle crashes at urban and suburban signalized intersections. Accid. Anal. Prev., 91, 72-83.

Chung, Y., 2010. Development of an accident duration prediction model on the Korean Freeway Systems. Accid. Anal. Prev., 42(1), 282-289.

Chung, Y., Recker, W.W., 2012. A methodological approach for estimating temporal and spatial extent of delays caused by freeway accidents. IEEE Trans. Intell. Transp. Syst., 13(3), 1454-1461.

Chung, Y., Recker, W.W., 2015. Frailty models for the estimation of spatiotemporally maximum congested impact information on freeway accidents. IEEE Trans. Intell. Transp. Syst., 16(4), 2104-2112.

Cox, D.R., Oakes, D., 1984. Analysis of survival data (Vol. 21). Boca Raton, FL, USA, CRC Press.

Crucitti, P., Latora, V., Porta, S., 2006. Centrality measures in spatial networks of urban streets. Phys. Rev. E, 73 (3), 35–39.

Erera, A., Lawson, T., Daganzo, C., 1998. Simple, generalized method for analysis of traffic queue upstream of a bottleneck. Transp. Res. Rec., 1646, 132-140.

Freeman, L.C., 1978/1979. Centrality in social networks conceptual clarification. Soc. Networks, 1(3), 215–239.

Garib, A., Radwan, A.E., Al-Deek, H., 1997. J. Transp. Eng., 123(6), 459-466.

Goodchild, M.F., 1992. Geographical data modeling. Comput. Geosci., 18(4), 401-408.

Guo, J., Huang, W., Williams, B.M., 2015. Real time traffic flow outlier detection using short-term traffic conditional variance prediction. Transp. Res. C, 50, 160-172.

Guo, Q., Xu, P., Pei, X., Wong, S.C., Yao, D., 2017. The effect of road network patterns on pedestrian safety: A zone-based Bayesian spatial modeling approach. Accid. Anal. Prev., 99(A), 114-124.

Hojati, A. T., Ferreira, L., Washington, S., Charles, P., Shobeirinejad, A., 2016. Reprint of: Modelling the impact of traffic incidents on travel time reliability. Transp. Res. C, 70, 86-97.

Hojati, A.T., Ferreira, L., Washington, S., Charles, P., 2013. Hazard based models for freeway traffic incident duration. Accid. Anal. Prev., 52, 171-181.

Hou, L., Lao, Y., Wang, Y., Zhang, Z., Zhang, Y., Li, Z., 2014. Time-varying effects of influential factors on incident clearance time using a non-proportional hazard-based model. Transp. Res. A, 63, 12-24.

Java, A., Song, X., Finin, T., Tseng, B., 2007. Why we twitter: Understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, ACM, 56-65.

Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. Accid. Anal. Prev., 23(4), 239-255.

Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., Makse, H.A., 2010. Identification of influential spreaders in complex networks. Nat. Phys., 6(11), 888-893.

Kleinberg, J.M., 1999. Authoritative sources in a hyperlinked environment. J. ACM, 46(5), 604-632.

Konduri, S., Labi, S., Sinha, K., 2003. Incident occurrence models for freeway incident management. Transp. Res. Rec., (1856), 125-135.

Li, D., Fu, B., Wang, Y., Lu, G., Berezin, Y., Stanley, H.E., Havlin, S., 2015. Percolation transition in dynamical traffic network with evolving critical bottlenecks. Proceedings of the National Academy of Sciences, 112(3), pp.669-672.

Li, D., Jiang, Y., Kang, R., Havlin, S., 2014. Spatial correlation analysis of cascading failures: congestions and blackouts. Scientific reports, 4.

Lo, H.K., Tung, Y.K., 2003. Network with degradable links, capacity analysis and design. Transp. Res. B, 37(4), 345-363.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transp. Res. A, 44(5), 291-305.

Marshall, W.E., Garrick, N.W., 2011. Does street network design affect traffic safety? Accid. Anal. Prev., 43(3), 769-781.

Moran, P., 1950. A test for the serial dependence of residuals. Biometrika, 37, 178-181.

Nam, D., Mannering, F., 2000. An exploratory hazard-based analysis of highway incident duration. Transp. Res. B, 34(2), 85-102.

Newman, M.E., 2001. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys. Rev. E, 64(1), 016132.

Newman, M.E., 2004. Analysis of weighted networks. Phys. Rev. E, 70(5), 056131.

Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. J. Transp. Eng. 122(2), 105-113.

Rakha, H., Van, Aerde M., 1995. Statistical analysis of day-to-day variations in real-time traffic flow data. Transp. Res. Rec., 26-34.

Rifaat, S.M., Tay, R., De Barros A., 2011. Effect of street pattern on the severity of crashes involving vulnerable road users. Accid. Anal. Prev., 43(1), 276-283.

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accid. Anal. Prev., 43(5), 1666-1676.

Sheu, J.B., Chou, Y.H., Shen, L.J., 2001. A stochastic estimation approach to real-time prediction of incident effects on freeway traffic congestion. Transp. Res. B, 45(8), 35(6), 575-592.

Skabardonis, A., Varaiya, P., Petty, K., 2003. Measuring recurrent and nonrecurrent traffic congestion. Transp. Res. Rec., (1856), 118-124.

Sun, C.S., Hao, J., Pei, X., Zhang, Z., Zhang, Y., 2016. A data-driven approach for duration evaluation of accident impacts on urban intersection traffic flow. IEEE 19th Int. Conf. Intell. Transp. Sys., 354-1359.

Thomas, T., Weijermars, W., Van Berkum, E., 2010. Predictions of urban volumes in single time series. IEEE Trans. Intell. Transp. Syst, 11(1), 71-80.

Wakefield, J.C., Best, N.G., Wallgen, L., 2000. Bayesian approaches to disease mapping Spatial Epidemiology: Methods and Applications. Oxford University Press, UK.

Wang, J., Liu B., Zhang, L., Ragland, D.R., 2016. Modeling secondary accidents identified by traffic shock waves. Accid. Anal. Prev., 87, 141-147.

Wang, X., Wu, X., Abdel-Aty, M., Tremont, P.J., 2013. Investigation of road network features and safety performance. Accid. Anal. Prev., 56, 22-31.

Wang, Y., Kockelman, K.M., 2013. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods, Accid. Anal. Prev., 60, 71-84.

Wasserman, S., Faust, K., 1994. Social network analysis: Methods and applications, Vol. 8. Cambridge, Cambridge University Press.

Weng, J., Qiao, W., Qu, X., Yan, X., 2015. Cluster-based lognormal distribution model for accident duration. Transportmetrica A, 11(4), 345-363.

Williams, B.M., and Guin, A., 2007. Traffic management center use of incident detection algorithms: Findings of a nationwide survey. IEEE Trans. Intell. Transp. Syst., 8(2), 351–358.

Wirasinghe, S.C., 1978. Determination of traffic delays from shock-wave analysis. Transp. Res., 12(5), 343-348.

Wisetjindawat, W., Sano, K., Matsumoto, S., 2006. Commodity distribution model incorporating spatial interactions for urban freight movement. Transp. Res. Rec. (1966), 41-50.

Xu, P., Huang, H., Dong, N., S.C. Wong, 2017. Revisiting crash spatial heterogeneity: A Bayesian spatially varying coefficients approach. Accid. Anal. Prev., 98, 330-337.

Zhang, Y., Bigham, J., Ragland, D., Chen, X., 2015. Investigating the associations between road network structure and non-motorist accidents. J. Transp. Geogr., 42, 34-47.

Zhou, Y.,Wang. X, Holguin-Veras, J., 2016. Discrete choice with spatial correlation: A spatial autoregressive binary probit model with endogenous weight matrix (SARBP-EWM). Transp. Res. B, 94, 440-455.

Appendix A

For the event duration T that follows Weibull distribution, its probability density function (pdf) $f(t) = \lim_{\Delta t \to 0} \frac{P(t \le T \le t + \Delta t)}{\Delta t}$ is given by

$$f_{\mathrm{PH}}(t|\lambda, p) = \lambda p(\lambda t)^{p-1}\exp[-(\lambda t)^p],$$

where $\lambda = \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_m X_m)$ is the covariates influence, $\boldsymbol{\beta}^{\mathrm{T}} = (\beta_0, \beta_1, \ldots, \beta_m)$ is the vector of hazard ratios that has to be estimated, and p is the scale parameter. Its corresponding survival function has the form:

$$S_{\mathrm{PH}}(t) = 1 - \int_0^t f_{\mathrm{PH}}(u)du, \tag{a1}$$

Based on PH model, after the heterogeneity term ω is introduced, the unconditional survival function of the Frailty model could be expressed as:

$$S_{\mathrm{F}}(t) = \int S_{\mathrm{PH}}(t)^{\omega}\, g(\omega)d\omega, \tag{a2}$$

where $g(\omega)$ is the pdf for ω:

$$g(\omega) = \left(\frac{\lambda_{\mathrm{IG}}}{2\pi\omega^3}\right)^{1/2} \exp\left\{\frac{-\lambda_{\mathrm{IG}}(\omega-\mu)^2}{2\mu^2\omega}\right\},$$

where $\mu(\mu > 0)$ is the mean and $\lambda_{\mathrm{IG}}(\lambda_{\mathrm{IG}} > 0)$ is the shape parameter. For mathematical tractability, here we set the mean $\mu = 1$ and variance $\theta = \frac{\mu^3}{\lambda_{\mathrm{IG}}}$, thus the pdf of ω becomes:

$$g(\omega) = \left(\frac{1}{2\pi\theta\omega^3}\right)^{1/2} \exp\left\{\frac{-(\omega-1)^2}{2\omega\theta}\right\}, \tag{a3}$$

The survival function of the Frailty model is obtained by substituting Eq. (a3) into Eq. (a2):

$$S_{\mathrm{F}}(t) = \exp\{\frac{1}{\theta}(1 - [1 - 2\theta\ln\{S(t)]^{\frac{1}{2}})\}, \tag{a4}$$

Last, by substituting Eq. (a1) into Eq. (a4), and applying $S(t) = \exp\left(-\int_0^t h(t)dt\right)$, it leads to the expression of Weibull hazard function with Inverse-Gaussian heterogeneity:

$$h_{\mathrm{F}}(t|\lambda, p, \theta) = \frac{\lambda p(\lambda t)^{p-1}}{\sqrt{1+2\theta(\lambda t)^p}},$$

as we described in the main text.