

RESEARCH ARTICLE

Network-constrained spatio-temporal clustering analysis of traffic collisions in Jiangnan District of Wuhan, China

Yaxin Fan¹, Xinyan Zhu^{1*}, Bing She^{2*}, Wei Guo¹, Tao Guo³

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, **2** China Data Center, University of Michigan, Ann Arbor, United States of America, **3** Wuhan Digital Engineering Research Institute, Wuhan, China

* geozxy@263.net (XZ); coolnanjizhou@163.com (BZ)



OPEN ACCESS

Citation: Fan Y, Zhu X, She B, Guo W, Guo T (2018) Network-constrained spatio-temporal clustering analysis of traffic collisions in Jiangnan District of Wuhan, China. PLoS ONE 13(4): e0195093. <https://doi.org/10.1371/journal.pone.0195093>

Editor: Li Daqing, Beihang University, CHINA

Received: September 28, 2017

Accepted: March 18, 2018

Published: April 19, 2018

Copyright: © 2018 Fan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data used in this work are all third party data that are owned by the Wuhan Traffic Management Bureau in China (<http://www.whjg.gov.cn>). The Wuhan Traffic Management Bureau in China cannot make these collision data publicly available due to legal restrictions. However, other researchers would be able to access these data in the same manner as the authors. Wuhan Traffic Management Bureau in China are willing to share the data with researchers upon request. Wuhan Traffic Management Bureau in China can be contacted by phone or mail. The specific division in the Wuhan Traffic Management

Abstract

The analysis of traffic collisions is essential for urban safety and the sustainable development of the urban environment. Reducing the road traffic injuries and the financial losses caused by collisions is the most important goal of traffic management. In addition, traffic collisions are a major cause of traffic congestion, which is a serious issue that affects everyone in the society. Therefore, traffic collision analysis is essential for all parties, including drivers, pedestrians, and traffic officers, to understand the road risks at a finer spatio-temporal scale. However, traffic collisions in the urban context are dynamic and complex. Thus, it is important to detect how the collision hotspots evolve over time through spatio-temporal clustering analysis. In addition, traffic collisions are not isolated events in space. The characteristics of the traffic collisions and their surrounding locations also present an influence of the clusters. This work tries to explore the spatio-temporal clustering patterns of traffic collisions by combining a set of network-constrained methods. These methods were tested using the traffic collision data in Jiangnan District of Wuhan, China. The results demonstrated that these methods offer different perspectives of the spatio-temporal clustering patterns. The weighted network kernel density estimation provides an intuitive way to incorporate attribute information. The network cross K-function shows that there are varying clustering tendencies between traffic collisions and different types of POIs. The proposed network differential Local Moran's I and network local indicators of mobility association provide straightforward and quantitative measures of the hotspot changes. This case study shows that these methods could help researchers, practitioners, and policy-makers to better understand the spatio-temporal clustering patterns of traffic collisions.

Introduction

Streets are one of the most common types of physical networks. Human activities in the urban space, although dynamic in nature, are largely constrained by the street networks [1]. Such constraints are an important force for pushing geographical entities to distribute along streets.

Bureau in China for research data inquiries and collaborations is the Technology Management Division. Their phone number is: +86-02785395787. Their mailing address is: No 2 Youyi Road, Qiaokou District, Wuhan, Hubei, China. A note is needed in the envelope to indicate the purpose, for example: 'Research Data Inquiries to the Technology Management Division, Wuhan Traffic Management Bureau in China.

Funding: This work is supported by National key R & D program (2016YFB0502204, http://www.most.gov.cn/eng/programmes1/200610/t20061009_36224.htm), LIESMARS (State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University) Special Research Funding (<http://www.lmars.whu.edu.cn/en/>), and Research on Key techniques of dynamic map (Key Open Fund 4201420100041, LIESMARS, Wuhan University, <http://www.lmars.whu.edu.cn/en/>). Xinyan Zhu received these fundings to support this work. The funders had no role in study design, data collection and analysis, decision and publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Events in the urban space are also more likely to be located on, or near, street segments. Traffic collisions are one of the most common types of such events. The World Health Organization reported that road traffic injuries have become the number one cause of death among the age group 15–29, and an estimated 3% of GDP is lost to road traffic deaths and injuries globally [2]. A sustainable transportation system allows the access needs of individuals to be met safely and consistent with human health [3]. Therefore, ensuring traffic safety is a major goal of local Traffic Management Bureaus for the sustainable urban transportation [4–6]. The analysis of traffic collisions is key to reducing traffic injuries [7]. However, maintaining traffic safety is an extremely complex task that involves drivers, vehicles, pedestrians, cyclists, motorcyclists, and road environment, etc. Measuring the road risks at the segment level is critical for traffic officers for resource allocation and policy-making. Traffic officers could also use the information for more targeted management practices. On the other hand, if drivers and pedestrians are aware of locations and time of collision hotspots on the roads, they are more likely to avoid them or adopt more defensive ways when approaching them. Traffic collision analysis provides approaches to visualize the spatial distribution and patterns of road risks, and quantify the risks through a set of spatial statistics. Therefore, the identification of the spatio-temporal clustering patterns of traffic collisions would help practitioners and policy-makers in better understanding the dynamics of collision hot-spots at a finer spatial scale. In recent years, GIS and spatio-temporal analysis methods have increasingly been used to study the characteristics of traffic collisions. These methods include both exploratory ones such as mapping and geovisualization techniques and confirmatory models from spatial statistics [8].

Recent years have seen a growing interest in applying network-constrained spatial analysis methods to study the characteristics of urban events. These exploratory models can quantitatively measure the spatial patterns and interactions of traffic collisions, using the spatial location and time stamp of the events. However, traffic collisions are not isolated events in space. The semantic information and temporal dimension of traffic collisions are also important for the comprehensive understanding of the spatio-temporal clustering patterns.

The semantic information of traffic collisions, including both the inherent attributes and the environmental factors of the location where the collision happens, is also related to the collisions. The inherent attributes include the types of collisions, financial loss, the cause of collisions, the number of vehicles involved, etc. These attributes are often incorporated into a regression model for confirmatory analysis, but they can also be incorporated into exploratory methods. For example, the inherent attributes of collisions could be integrated into the kernel density estimation method, thus, practitioners can observe the spatial pattern from different perspectives. The environmental factors include the characteristic of the location where the event happens, and also the surrounding locations. The characteristics of the location where the event happens are directly related to the event, including the road type, traffic volumes, and the weather condition. The characteristics of the surrounding locations are mostly the characteristics of the surrounding points of interest (POIs), such as grocery stores, parking lots, and hospitals. These POIs might not directly relate to individual collisions, but their spatial distribution might correlate with spatial distributions of collisions collectively.

The temporal dimension of the traffic collisions could be used to detect the changes in the spatial distributions of events. The collision hotspots might emerge, disappear, or move over time. The resulting changes could be presented by showing the spatial analysis results from different time points, with a common comparable scale. For example, we can use kernel density estimations to see how the spatial distribution changes across different times of the day. Another approach would be to first compute a measure from two time points, and then use a hotspot detection method to analyze the changes quantitatively. In addition, although the analysis can be done between two arbitrary time points or periods, a meaningful organization of

time would be more useful for practitioners when implementing regulating policies. This includes the monthly or seasonal cycles, day of the week, and time of the day. Researchers have studied other types of events using a variety of temporal organizations, such as crime patterns [9].

This paper attempts to explore the spatio-temporal clustering patterns of traffic collisions with semantic information. The weighted network kernel density estimation is extended to explicitly incorporate attribute information with normalized weights. The resulting visualization provides practitioners with multiple perspectives of the spatial distribution of traffic collisions with a certain normalized attribute. The network cross K-function is applied to investigate the relationship between traffic collisions and different types of POIs. To investigate the temporal change of collision clustering, this work proposed the network-based differential local Moran's I and Local Indicators of Mobility Associations (LIMA) [10] by substituting the planar weight matrix with the network weight matrix. The network differential Moran's I method quantifies the clustering of changes on the street network, while the network LIMA measures the degree of local concordance and discordance over time. The remainder of this paper is organized as follows: The next section introduces the relevant works from literature. Section 3 describes the study area. Section 4 discusses the network-constrained methods. Section 5 presents a case study of Jiangnan District, Wuhan, China. Section 6 gives the discussions and outlines the future work.

Sustainable traffic safety and spatio-temporal clustering analysis of traffic collisions

A sustainable transport system must be able to provide different modes of transport to people in a safe way [11]. The 2030 Agenda for Sustainable Development of the United Nations has set a target of halving the total number of deaths and injuries from traffic crashes by 2020 [12]. Researchers have also used traffic safety as indicators for the sustainable transport systems [13, 14]. The goal of sustainable safety is to prevent traffic collisions. Researchers have proposed ways to achieve sustainable traffic safety by investigating new road patterns [15] or new hardware, such as eye mark recorders [16]. Yet the urban mobilities in our daily experiences are dynamic and complex. Achieving traffic safety needs more than transportation engineering measures [17]. Thus, it is key to involve all parties, including drivers, pedestrians, and traffic officers, in the process of achieving better traffic safety. In practice, policy-makers tend to impose various security policies. However, safety policies sometimes encourage riskier behaviors, which expose greater risks to the transportation system [18]. Drivers and pedestrians might not always willing to follow rules [19]. Individuals, in general, have difficulties making decisions regarding modes of transport when safety is considered an objective [11]. The spatial analytics of traffic collisions will provide all parties with a toolset to measure road risks at finer spatio-temporal scale.

The spatial analytic methods for traffic collision analysis can be generally classified into three types [7]: mapping and topological analysis, identification of clustering patterns, and analysis of contributory factors. This work focus on the second type. Since traffic collisions are constrained by the street network, traditional planar spatial methods introduce a systematic bias in the analysis result [20]. The Ripley's K-function [21], a widely used method to detect spatial aggregation, might possibly over-detect or underestimate the aggregation in network-constrained phenomena [22, 23].

Network-constrained methods have been extensively studied for finer-scale urban analyses in recent years. These methods can be broadly classified into event-based approaches and link-based approaches [24]. The event-based approach includes the network kernel density

estimation (KDE) [25], and the K-function [26], and moving-segment approach [27]. In particular, the network KDE and K-function have been applied widely in recent years for exploring the spatial patterns of traffic collisions and other types of events [22, 28–33]. Researchers have then started to incorporate semantic information into the analysis pipeline. Ni et al. [34] proposed a weighted kernel density estimation method for studying the spatial distribution characteristics of healthcare facilities. Rui et al. [35] used the network cross K-function to study the spatial correlation between Suguo hypermarkets with other hypermarkets and commercial centers. Our work built upon the weighted kernel density estimation method with normalized weights to study the spatial patterns of traffic collisions. The network cross K-function is used to investigate the spatial aggregation patterns between traffic collisions and different types of POIs.

The link-based approaches tried to identify the hot spots by using local spatial autocorrelation tests [36, 37]. Yamada and Thill [20] used the local Moran's I, Local Getis, and Ord G statistics in the local indicators of network-constrained clusters (LINCS) for detecting local-scale clustering of highway accidents. These methods can detect street segments with statistically significant patterns by use of Monte Carlo simulation procedures. Recent years have also seen an increasing interest in using time geography to measure traffic collision risk [38, 39]. In other areas, such as crime pattern analysis, the time dimension has been used in Markov analysis and combined with other methods, such as Moran's I [40–42]. Researchers have proposed other types of extensions in the network space. Xianrui and Zhongren [43] proposed a spatio-temporal K-function over the network to analyze taxi load-unload data. Eckley and Curtin [44] used the spatio-temporal interaction Knox test on traffic collision data. Other statistical models are often used for analysis of contributory factors, such as spatial lag model [45], Bayes models [46–48], and Geographically weighted regression [49–51]. This work extended the differential Local Moran's I and LIMA into the network space, which provides a set of new local measures to quantify the spatio-temporal clustering at the road segment level.

Study area and data

Wuhan is a rapidly growing city in central China and the capital of Hubei province. According to the Statistical Communiqué of Wuhan on the 2016 National Economic and Social Development [52], the number of permanent residents was 10,766,200 people at the end of 2016. The number of cars has reached 2.31 million, and there are 489 bus routes in operation. Jiangnan District is one of the seven major urban districts in Wuhan. It is also the most densely populated, while also the most prosperous, district. Jiangnan District has a total area of 33.43 square kilometers and lies on the north shore of the Yangtze River. The data used in this work include traffic collisions, POIs, and the road network. All these data were fully anonymized before we accessed them.

The Wuhan Traffic Management Bureau started to equip traffic police with personal data assistants from mid-2016. The traffic collision data collected after that time are better in quality and coverage. This work uses the traffic collision data in Jiangnan District from 1 July 2016 to 31 December 2016. There are a total of 11,445 collisions with locations recorded during this period. Fig 1 displays the overall distribution of these events with the street network. It shows a clear clustering tendency of the traffic collisions in certain road segments.

Fig 2 shows the number of collisions under two temporal organizations: the hour of the day, and weekday/weekend. The horizontal axis represents the hour of the day is categorized into six periods, plus an extra period that represented the average count. The vertical axis represents the daily average of the cumulative count by weekdays, weekends, and overall, respectively. Unexpectedly, the collisions are mostly aggregated in the daytime. It also suggested that although no

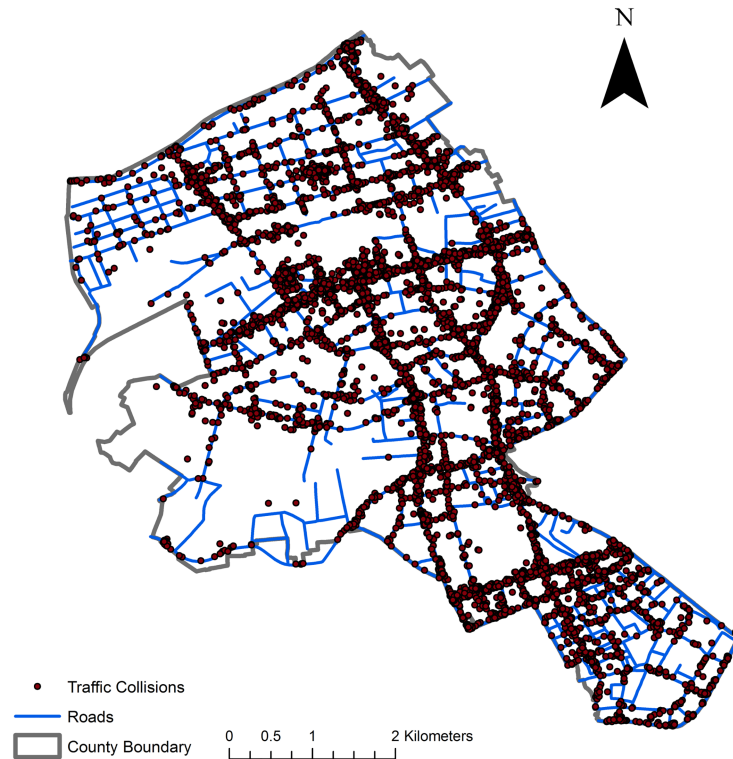


Fig 1. Traffic collision distribution in Jiangnan District, Wuhan, China.

<https://doi.org/10.1371/journal.pone.0195093.g001>

significant difference exists between weekdays and weekends, overall, their structure does differ when taking into account the hour of the day factor. This is particularly evident in the morning (6–10). [Table 1](#) illustrates the count of collisions by the ranges of the direct financial loss. Most of the collisions are minor incidents and do not incur any direct financial losses.

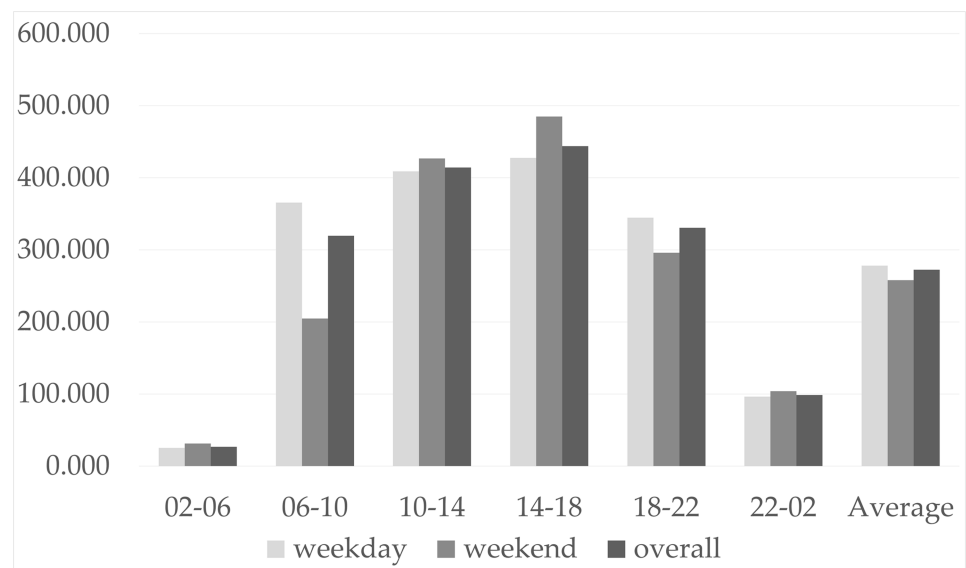


Fig 2. Traffic collision statistics by the hour of the day and weekday/weekends.

<https://doi.org/10.1371/journal.pone.0195093.g002>

Table 1. The count of collisions by the ranges of direct financial loss.

Direct financial loss (Yuan)	Count
0	10,616
1–100	50
101–500	46
501–1,000	665
1,001–5,000	64
5,000+	4

<https://doi.org/10.1371/journal.pone.0195093.t001>

Six types of POIs are integrated into this study for network cross K-function analysis. The counts for each type of POI points are given in Table 2. The Transportation Services refer to parking lots and public transit stations. The overall spatial distribution of these POI points is displayed in Fig 3.

Network-constrained analysis

The street network in this work is formulated as an undirected and planar network $N = (V, L)$, with a set of nodes V and edges L . The event points $E = \{e_1, \dots, e_n\}$ refers to the traffic collisions. Each event point e is represented by its location, time, and a set of attributes. The event points in E are assumed to follow a probabilistic point process that distributed on the network N . The other type of points are the POI points $P = \{p_1, \dots, p_m\}$, and this work represent each POI point by its location and type. For simplicity, P is assumed to be a fixed set of points snapped to the network.

Weighted network kernel density estimation

The network-constrained KDE is a direct extension of the planar KDE into the network space. The core idea is to divide the network space into linear pixels [25] and uses the shortest path between event points to calculate the distance, instead of using Euclidian distance. For an arbitrary point q on N , the weighted kernel estimator is represented as:

$$\lambda(q) = \frac{1}{n} \sum_{i=1}^n w_i * K_i(q) \tag{1}$$

where w_i is the weight for the event point i . This work used a quantile mapping method to map an original event attribute into a weighted vector. This is derived from [53], which computes the weights of links for network Voronoi diagrams. The weight vector falls into a normalization interval $R = [r_{min}, r_{max}]$, a user-defined parameter that controls the influence of the weights on the kernel estimator. Suppose the original event attribute vector is c . The process starts with the sorting of c , and then the ranks of values are used to compute the normalized

Table 2. The count of POIs by types in Jiangnan District, Wuhan, China.

POI Type	Count
Transportation Services	107
Hotels	87
Sports and Recreation	52
Residential Communities	127
Vehicle Maintenance	49
Food	280

<https://doi.org/10.1371/journal.pone.0195093.t002>

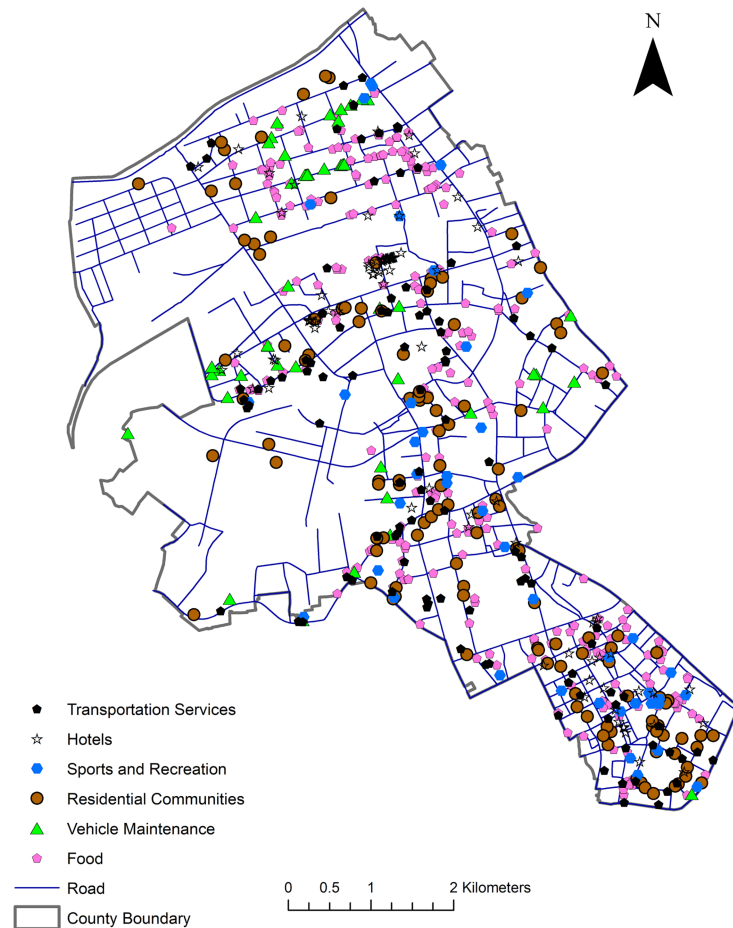


Fig 3. POI distribution in Jiangnan District, Wuhan, China.

<https://doi.org/10.1371/journal.pone.0195093.g003>

value. The number of distinct ranks is $n_d = n - n_e$, where n_e is the number of equal values in c . The event with the k th largest value will be normalized to:

$$c' = \frac{(k - 1) * (r_{max} - r_{min})}{(n_d - 1)} + r_{min} \tag{2}$$

where $K_i(q)$ is the kernel function at the kernel center i . The equal-split kernel function is used to prevent biased estimates at road intersections [54]. Suppose the shortest path from i to q contains p nodes: v_1, \dots, v_p , and let n_i represent the degree of the node v_i . $K_i(q)$ is given by:

$$K_i(q) = \begin{cases} \frac{k(d(q, i), h)}{(n_1 - 1)(n_2 - 1) \dots (n_p - 1)} & \text{for } 0 \leq d(q, i) \leq h \\ 0 & \text{for } d(q, i) > h \end{cases} \tag{3}$$

where h is the bandwidth, while $k(d(q, i), h)$ is the base kernel function. The idea is to divide the kernel values at each node along the path from i to q , and distribute them to adjacent edges. It is accepted by the research community that the form of the kernel function is less important than the choice of bandwidth [55, 56]. The basic units of computation in the network KDE relies on lixels, and the density values are output for each lixel [33]. The lixel length determines the computational intensity.

Network cross K-function

The K-function method is considered as an approach to investigate the second-order characteristics of a spatial point process, which is widely used to measure spatial dependence. The network cross K-function extends the measure to take into account two different types of points [35]. In other words, the cross K-function quantifies the spatial interrelationships between two types of point sets. The theoretical form of the cross K-function could be written in the following form:

$$K_{ba}(d) = \frac{1}{\rho_b} \frac{\sum_{i=1}^n n(d, b|a_i)}{n} \tag{4}$$

where ρ_b is the density of point type b on the network, while $n(d, b|a_i)$ is the number of points of type b that are within distance d from point i of type a . The distance is also calculated by the shortest path method. In the context of this work, the traffic collisions are the type- b points, and POI points are the type- a points. $K_{ba}(d)$ could be written as:

$$K_{ba}(d) = \frac{1}{\rho_b} \frac{\sum_i \sum_j I(|s(a_i, b_j)| < d)}{n} \tag{5}$$

where $|s(a_i, b_j)|$ denote the distance of shortest path from POI a_i to the traffic collision b_j , and $I(|s(a_i, b_j)| < d)$ is the indicator function with the value 1 if the distance is smaller than d and 0 otherwise. Okabe also proposed a transformation method to transform a non-uniform network into a uniform network [57]. Similar to the planar K-function, the Monte Carlo simulation method is used to test the distribution pattern of point events. This can be done by generating simulated point patterns on the network repeatedly according to the completely spatial random assumption. Then, the observed K-function curve is compared with the simulated K-function curve. Judging from the relations of the curves, we can then tell whether the traffic collisions are clustered around, dispersed from, or unrelated to certain types of POIs.

Network differential local Moran’s I method

The local Moran’s I statistic developed by Anselin [58] is a widely accepted measure of spatial autocorrelation. For region i , the local I statistic for an attribute v is defined as:

$$I_i = z_i \sum_j w_{ij} z_j \tag{6}$$

where z_i and z_j are the normalized value of v , and w_{ij} is a binary indicator of whether areas i and j are adjacent. This adjacency relationship is represented as the spatial weight matrix W . Positive values of the local Moran’s I statistic suggests a clustering tendency, while negative values indicate spatial dispersiveness of the distribution.

The differential local Moran’s I method is a natural extension of the Moran’s I statistics. It measures the spatial patterns of the changes of the same attribute between two different times [59]. The form of the differential local Moran’s I is as follows:

$$I_i = z(v_{it_2} - v_{it_1}) \sum_j w_{ij} z(v_{jt_2} - v_{jt_1}) \tag{7}$$

where $z(v_{it_2} - v_{it_1})$ and $z(v_{jt_2} - v_{jt_1})$ are the normalized values of the changes in v from time t_1 to t_2 . The simulation process is done by using conditional random permutations. In some circumstances where the analyst wants to compare two time periods, the attribute could be

normalized by the length of two periods, which gives the following:

$$I_i = z\left(\frac{v_{it_2}}{L_{t_2}} - \frac{v_{it_1}}{L_{t_1}}\right) \sum_j w_{ij} z\left(\frac{v_{jt_2}}{L_{t_2}} - \frac{v_{jt_1}}{L_{t_1}}\right) \tag{8}$$

where L_{t_2} and L_{t_1} are the length of the two periods. For example, if we would like to compare the event distribution between weekdays and weekends, the length of the periods could be measured in days. Then L_{t_2} and L_{t_1} would be 5 and 2, respectively. In this work, we use the number of traffic collisions as the attribute v . The periods used in the analysis are of a certain temporal organization, such as weekday/weekend, or the hour of the day.

This work extends the differential Moran’s I into the network space by substituting the planar weight matrix W with the network weight matrix W_N . Therefore, w_{ij} defines the neighboring relationships between two network segments. Researchers have previously used W_N to compute the local Moran’s I, Local Getis, and Ord G statistics in the network space [20, 32]. Two types of network weight matrices exist: the node-based matrix and the distance-based matrix. The node-based matrix will only treat network segments as neighbors when they are directly connected. The distance-based matrix determines the neighboring relationships based on whether the distance between the centers of two segments is less than a distance threshold or not. We adopted the distance-based matrix because it can represent the segment relationships more flexibly by controlling the threshold parameter and it is commonly used in the literature [20, 33].

Network local indicators of mobility association

The local indicators of mobility association measures are derived from the global indicators of mobility association, which is further derived from the general rank correlation coefficient proposed by Kendall [60]. Here we consider the two observation vectors \mathbf{v}_{t_1} and \mathbf{v}_{t_2} that represent the same variable in two periods. The coefficient $\tau(\mathbf{v}_{t_1}, \mathbf{v}_{t_2})$ is given by:

$$\tau(\mathbf{v}_{t_1}, \mathbf{v}_{t_2}) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(v_{it_1} - v_{jt_1}) \text{sgn}(v_{it_2} - v_{jt_2})}{n(n-1)/2} = \frac{C - D}{n(n-1)/2} \tag{9}$$

where the sgn function extracts the sign of the difference between two units, thus taking values 1 or -1. If $\text{sgn}(v_{it_1} - v_{jt_1}) \text{sgn}(v_{it_2} - v_{jt_2}) = 1$, the pair of observation between unit i and j is concordant across two periods t_1 and t_2 . If $\text{sgn}(v_{it_1} - v_{jt_1}) \text{sgn}(v_{it_2} - v_{jt_2}) = -1$, the pair is discordant. C and D represent the number of concordant and discordant pairs. The use of ranks makes Kendall’s τ robust to departure from bivariate normality [61].

To consider ties in the observations which would lead $\text{sgn}(v_{it_1} - v_{jt_1}) \text{sgn}(v_{it_2} - v_{jt_2}) = 0$, extra pairs could be accounted for in the denominator [61], which gives:

$$\tau'(\mathbf{v}_{t_1}, \mathbf{v}_{t_2}) = \frac{C - D}{\sqrt{C + D + E_{t_1}} \sqrt{C + D + E_{t_2}}} \tag{10}$$

where E_{t_1} represents the number of extra pairs introduced when $\text{sgn}(v_{it_1} - v_{jt_1}) \neq 0$ and $\text{sgn}(v_{it_2} - v_{jt_2}) = 0$, while E_{t_2} represents the number of extra pairs introduced when $\text{sgn}(v_{it_1} - v_{jt_1}) = 0$ and $\text{sgn}(v_{it_2} - v_{jt_2}) \neq 0$. The values of τ' falls on the range [-1,1]. A value of 1 indicates that all pairs are concordant. This means that larger values of τ' implies less distributional mixing from period t_1 to t_2 . Detailed explanations for handling ties are given in [60].

Rey proposed a spatial concordance measure based on Kendall’s τ measure [62]. The spatial measure is based on the decomposition of the pairs of observation into those that are neighbors and those that are not. Suppose a binary spatial weight matrix W is constructed to

represent whether units i and j are neighbors, define matrix $\bar{W} = J - W - I$, where J is a matrix of ones and I is an identity matrix. The measure τ can thus be decomposed into:

$$\tau(\mathbf{v}_{t_1}, \mathbf{v}_{t_2}) = \psi \tau_w(\mathbf{v}_{t_1}, \mathbf{v}_{t_2}) + (1 - \psi) \tau_{\bar{w}}(\mathbf{v}_{t_1}, \mathbf{v}_{t_2}) \tag{11}$$

where $\psi = \sum_i \sum_j w_{i,j} / n(n - 1)$, and τ_w and $\tau_{\bar{w}}$ are the decomposed concordance measures for the neighboring pairs and the non-neighboring pairs. τ_w is then considered as a type of Global Indicators of Mobility Association (GIMA), given by:

$$\tau_w(\mathbf{v}_{t_1}, \mathbf{v}_{t_2}) = \frac{\sum_i \sum_j w_{i,j} \text{sgn}(v_{it_1} - v_{jt_1}) \text{sgn}(v_{it_2} - v_{jt_2})}{\sum_i \sum_j w_{i,j}} \tag{12}$$

The approach of handling ties in Eq (10) could be used to incorporate extra pairs in computing τ_w . Rey further constructs three types of Local Indicators of Mobility Association (LIMA) [10]: the local concordance τ_i , the neighbor set LIMA $\tilde{\tau}_i$, and the neighborhood set LIMA $\tilde{\tilde{\tau}}_i$. Let $\text{concordance}(i, j) = \text{sgn}(v_{it_1} - v_{jt_1}) \text{sgn}(v_{it_2} - v_{jt_2})$, then be:

$$\tau_i = \frac{\sum_{j \neq i} \text{concordance}(i, j)}{n - 1} \tag{13}$$

$$\tilde{\tau}_i = \frac{\sum_j w_{i,j} \text{concordance}(i, j)}{\sum_j w_{i,j}} \tag{14}$$

$$\tilde{\tilde{\tau}}_i = \frac{\sum_{m \in NS_i} \sum_{n \in NS_i, n \neq i} \text{concordance}(m, n)}{|NS_i|(|NS_i| - 1)} \tag{15}$$

where NS_i is the neighborhood set of i plus i . This work will consider $\tilde{\tau}_i$ and $\tilde{\tilde{\tau}}_i$ in the case study which takes the local spatial context into account. The $\tilde{\tau}_i$ measure investigate the local concordance between a unit and its neighbors, while $\tilde{\tilde{\tau}}_i$ extends $\tilde{\tau}_i$ by conducting the computations between all pairs of observations in a unit's neighborhood set. The inference is done by using conditional random permutations.

Similar to the network differential local Moran's I, this work extends the LIMA measures into the network space by using the network weight matrix W_N . The LIMA measures use binary weight matrices, thus $w_{i,j}$ defines whether two network segments are neighbors, while NS_i is the neighboring segments of the segment i .

Network computations

The input to the network-constrained analysis methods are the shapefiles of the streets, events, and POIs. Fig 4 gives a synthesized workflow of these methods. During the preprocessing phase, the street network is first constructed from the streets and segmented into network segments. It is a common practice to split the network edges into equal sizes approximately [20, 25]. The segment size is predefined by analysts. The street network used in this work is a generalization of the real-world roads, which do not consider lanes and complex intersection structures. This would produce an offset between the event points and the street network. The offset is also subject to errors in GPS readings when recording the events. Therefore, the event points need to be projected into the network. The snapping process of event points and POIs is for finding the nearest edge for an event or POI point. This process can be accelerated by first constructing a spatial index (e.g., R tree) for the network N. For the network KDE and cross K-function analysis, the events points are inserted as endpoints in N. This insertion

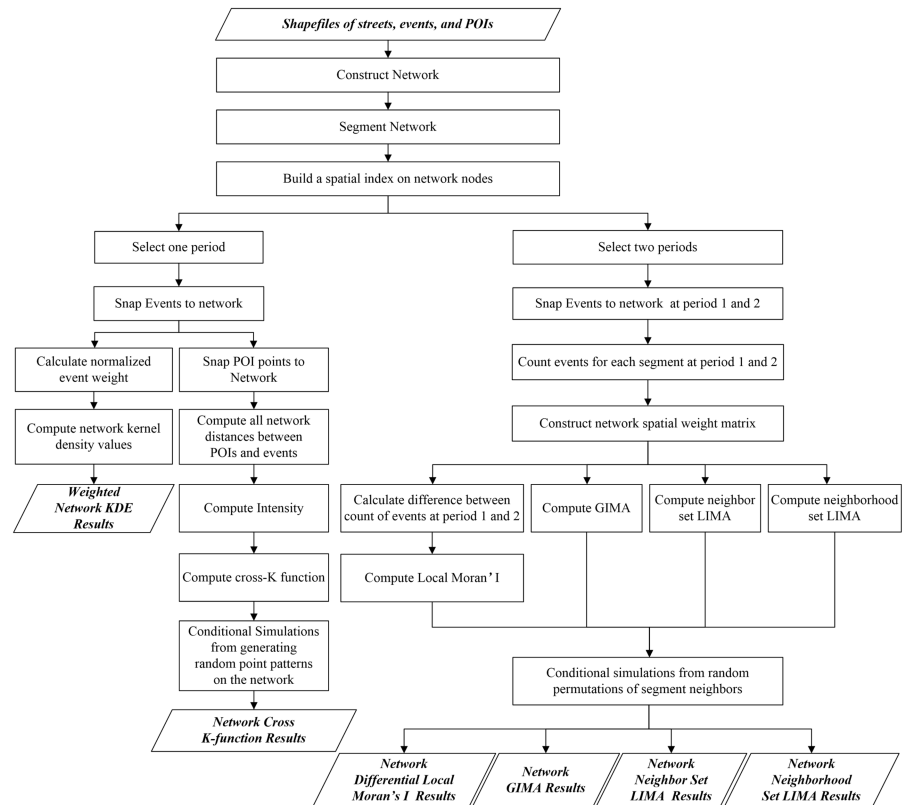


Fig 4. The workflow of network-constrained analysis.

<https://doi.org/10.1371/journal.pone.0195093.g004>

process will transform N into a new network N' with its original segments split by events points. For the network differential Moran's I and GIMA/LIMA analysis, the numbers of events on each edge are counted for computing the indicators.

The distance computation on the network is a basic operation for all the analysis methods in this work, but in slightly different ways. The weighted network KDE method finds the distances of other events to an event point within a given bandwidth. The network cross K-function method computes all network distances between all POIs and all events. For the network differential Moran's I and GIMA/LIMA analysis, a distance-based network spatial weight matrix is constructed based on the distances between network segments. The simulations are done by random permutations of segment neighbors.

Results

The computations of the methods were implemented in Python. PySAL is leveraged for Local Moran's I and network computations [63]. The network local indicators of mobility association measures are implemented based on the PySAL-giddy package, which originates from the spatial dynamic module of PySAL for analyzing the dynamics of longitudinal spatial data [64]. The case studies are conducted on a machine with an i7 Intel CPU and 16 GB DDR3 memory.

Weighted network KDE analysis

The bandwidth parameter is an important issue in the network KDE analysis. Porta et al. [65] proposed a 100–300 m bandwidth in urban applications. This work chose a bandwidth of 200

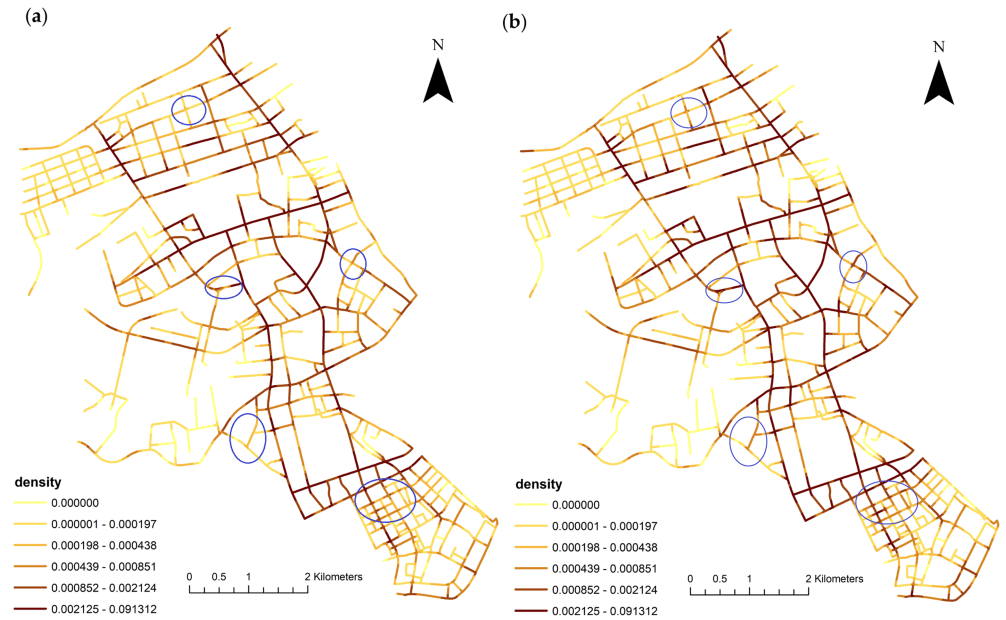


Fig 5. Comparison of unweighted and weighted network KDE for all traffic collisions in Jiangnan, Wuhan, China. (a) Unweighted KDE; and (b) weighted KDE.

<https://doi.org/10.1371/journal.pone.0195093.g005>

meters. The length of lixel is set as 40 meters, as suggested in Xie and Yan [25]. The normalization interval for the weighted attribute of direct financial loss is set to [1, 10]. Fig 5 compares the distribution of the unweighted and weighted KDE for all traffic collisions in the experimental data. The blue oval-shaped markers indicate some clear differences and they appear mostly in road intersections.

Fig 6 displays the comparison of six weighted network KDE results of traffic collisions by the hour of the day. The map classifications are all completed using the quantile method. The Figs 5 and 6 clearly show that the spatial structures of traffic collisions are quite different between daytime and nighttime. Traffic collisions on the main road have a relatively higher frequency at all times in a day. During 2–6 in the morning, there are fewer accidents because there are fewer cars, and most of these accidents are distributed around road intersections and main roads. The Jiangnan district is the major economic and business center of Wuhan, the traffic flow stays relatively high from 6:00 to 22:00. Most accidents are minor incidents caused by traffic violations such as overtaking, failing to yield, and cut-in. The spatial distribution of traffic collisions remains largely stable from 6:00 to 22:00, while small variations do exist.

Network cross K-function analysis

The network cross K-function is used to analyze the relationships between traffic collisions and different types of POIs. The results are used to measure quantitatively the degree of network aggregation between traffic collisions and surrounding POIs. The results were plotted in R with outputs from the Python program, and shown in Fig 7. The plots show clearly the relationships between traffic collisions and POIs varies considerably for different POI types.

Among all the POI types, traffic collisions show the strongest degree of network aggregation around POI points of Hotels and Vehicle Maintenance. However, they tend to follow a random distribution around POI points of Sports and Recreation. Mild network aggregation exists for POI points of Transportation Services, Residential Communities, and Food. It is possible to

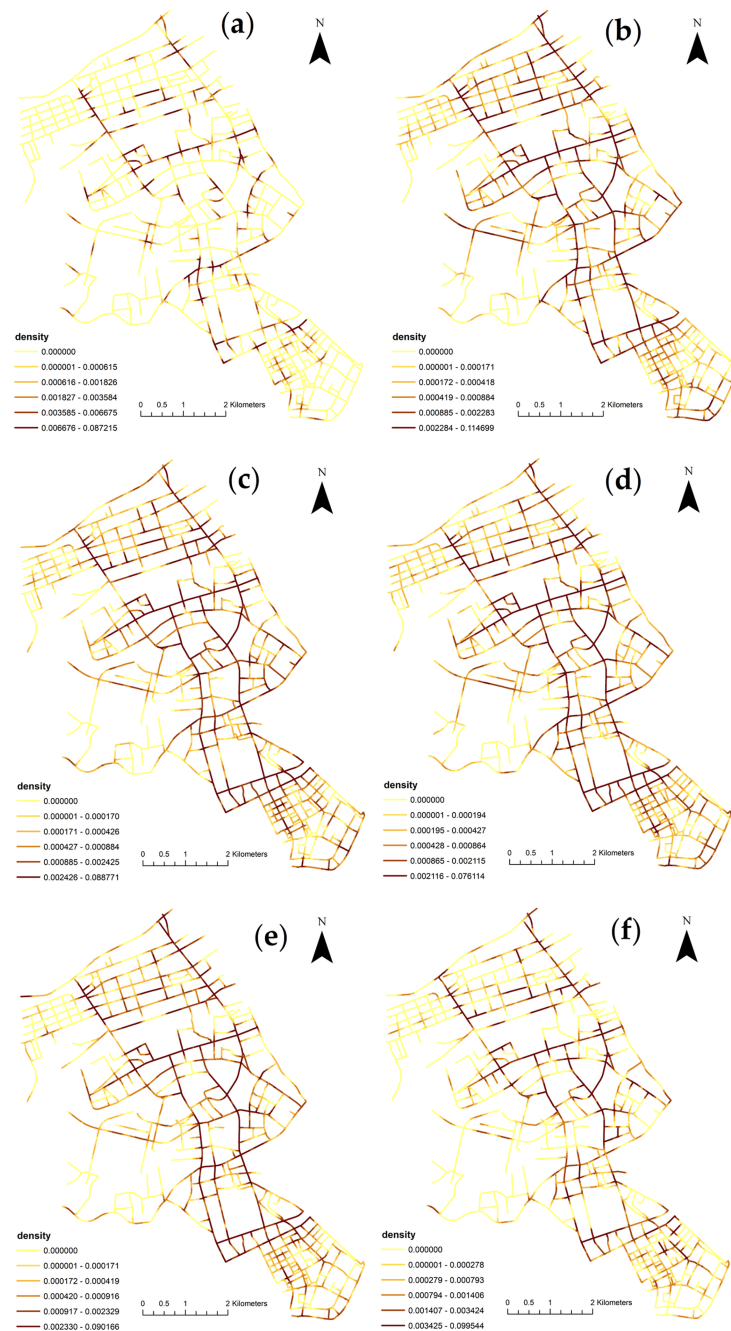


Fig 6. Comparison of weighted network KDE of traffic collisions by the hour of the day in Jiangnan, Wuhan, China: (a) 2–6; (b) 6–10; (c) 10–14; (d) 14–18; (e) 18–22; and (f) 22–2.

<https://doi.org/10.1371/journal.pone.0195093.g006>

form assumptions of the network aggregations. For example, the aggregation of traffic collisions around Vehicle Maintenance POIs might be due to the fact that the vehicles going to these POIs might have issues at that time. The collisions around hotels might be due to several factors, including a large volume of incoming and outgoing traffic, driving under the influence, and other traffic violations such as overtaking and cut-in. The mild aggregation of collisions around Transportation Services, Residential Communities, and Food might be largely because

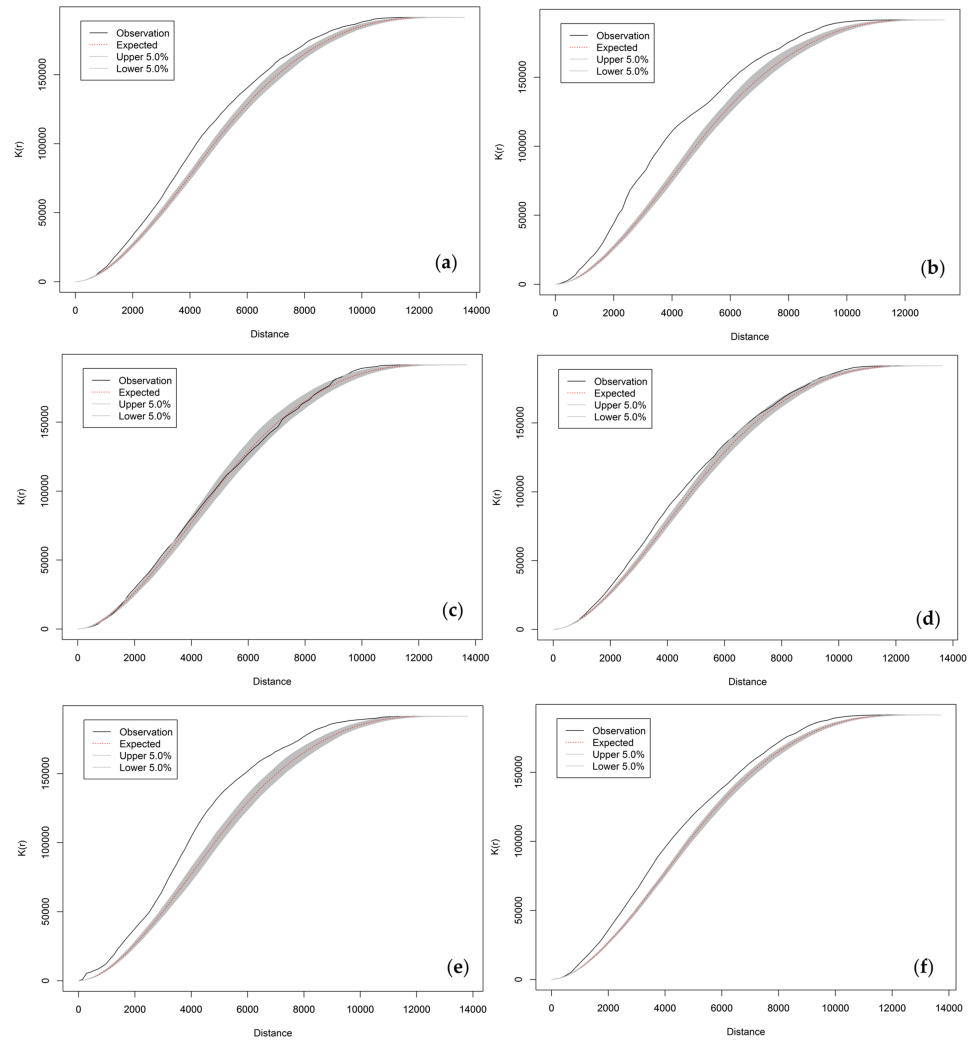


Fig 7. Network cross K-function analysis between traffic collisions and different types of POIs in Jianghan, Wuhan, China: (a) Transportation Services; (b) Hotels; (c) Sports and Recreation; (d) Residential Communities; (e) Vehicle Maintenance; and (f) Food.

<https://doi.org/10.1371/journal.pone.0195093.g007>

of scratch incidents related to parking violations. The Sports and Recreation POIs mostly have direct access to public transportation and less traffic flow, and thus there is no significant network aggregation of traffic collisions around them. However, it is important to note that the network cross K-function analysis is still a descriptive measure of the network aggregation. Therefore, it is more suitable to be used in the exploratory phase. The assumptions formed in this phase needs to be validated in further statistical regression analysis with supplementary data.

Network differential local Moran’s I analysis

The network differential Local Moran’s I method is used to quantify the changes in space between two periods. The segmentation width is set to 100 meters in this analysis as this is the standard distance in management practice for segmenting roads in the Wuhan Traffic Management Bureau. The bandwidth threshold for computing the spatial weight matrix is set to

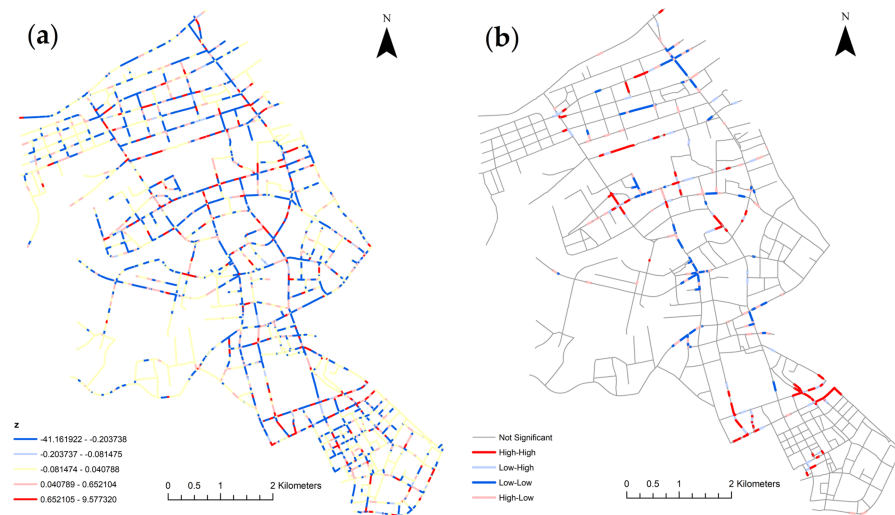


Fig 8. The network differential local Moran's I analysis from weekdays to weekends in Jiangnan, Wuhan, China: (a) Z values; and (b) pattern maps.

<https://doi.org/10.1371/journal.pone.0195093.g008>

300 meters. This means that a road segment will be neighbors to all segments that are within 300 meters in network distance. The significance level is set to 0.05 and the number of iterations in the Monte Carlo simulation is set to 999 times. In this study, we first grouped all collisions into weekdays and weekends and compared these two periods. Fig 8 shows the Z value distributions and patterns of the changes from weekdays to weekends. The Z value distribution map indicates the normalized values of changes. The negative values suggest a drop in the number of traffic collisions from weekdays to weekends, while positive values suggest an increase in the number. The pattern map gives a clear view of where the cluster segments locate. The patterns correspond to the changes in the number of traffic collisions from weekdays to weekends. Therefore, it does not directly reflect the large or small number of traffic incidents on road segments. Particularly, the high-high segments indicate clusters with high and significant increases in the number of traffic incidents. These places are mostly around large commercial areas where citizens frequently go on the weekends. Oppositely, the low-low segments indicate clusters with large and significant drops in the number of traffic incidents. Many of these places are near residential communities and industrial areas with less traffic flow on the weekends.

To demonstrate the use of finer temporal organizations, we first grouped the events by weekdays and weekends, then under each group, classified the events further by the hour of the day. Then the two same hours of the day periods from weekdays and weekends can be compared. Fig 9 shows the patterns of the changes of these six hours of the day periods from weekdays to weekends. During the 2–6 and 22–2 periods, the low-low segment clusters show that there are significant drops in the number of collisions from weekdays to weekends. This is because collisions on weekdays spread across the whole area, while collisions on weekends are more concentrated in certain areas. Starting at 6:00 AM, some road segments emerge as high-high clusters. This indicates that relatively more collisions happen on those clustered segments on weekends than weekdays, even though the total number of collisions on weekdays is larger than weekends. The results indicate that temporal analysis alone (as in Table 2) cannot capture the spatial dynamics of traffic collisions. The differential Local Moran's I provides an effective tool to quantify and map the micro-level change of collisions in the spatial dimension.

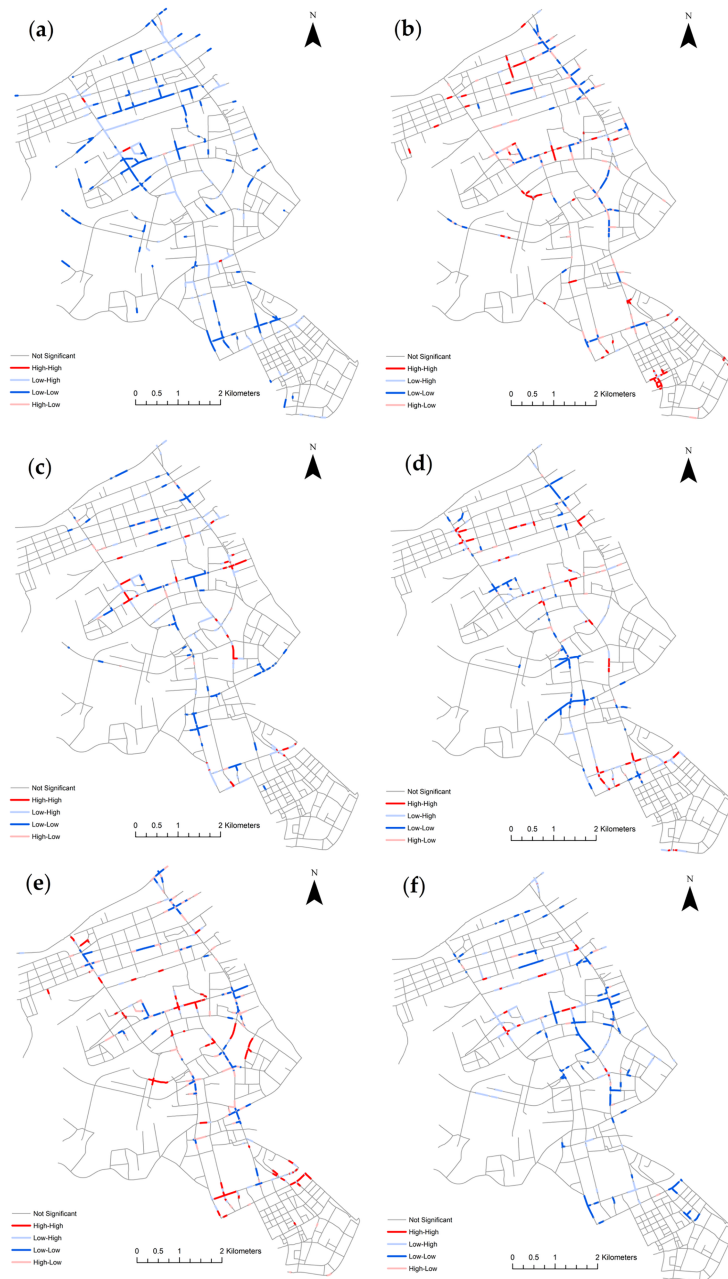


Fig 9. Network differential Local Moran's I pattern maps of the changes of six hours of day periods from weekdays to weekends in Jiangan, Wuhan, China: (a) 2–6; (b) 6–10; (c) 10–14; (d) 14–18; (e) 18–22; and (f) 22–2.

<https://doi.org/10.1371/journal.pone.0195093.g009>

Network LIMA

The network LIMA analysis offers another angle of the spatio-temporal clustering of traffic collisions. The neighbor set LIMA $\tilde{\tau}_i$ investigates the concordance relationship between a focal segment and its neighbors. The neighborhood set LIMA $\tilde{\tau}_i$ expands $\tilde{\tau}_i$ by taking all pairs of segments belonging to the neighborhood set of a segment into the computation. As the LIMA statistics are based on the ranks of the variables, there is a potential loss of statistical power. However, they are robust to outliers and have better generality by relaxation of the underlying

Table 3. The matrix of global concordance (τ_w) between the six hours of day periods.

	2–6	6–10	10–14	14–18	18–22	22–2
2–6	-	0.232	0.233	0.201	0.251	0.235
6–10	0.232	-	0.46*	0.464	0.467*	0.355
10–14	0.233	0.46*	-	0.487	0.504*	0.344
14–18	0.201	0.464	0.487	-	0.496	0.337
18–22	0.251	0.467*	0.504*	0.496	-	0.385
22–2	0.235	0.355	0.344	0.337	0.385	-

<https://doi.org/10.1371/journal.pone.0195093.t003>

correlation statistics [10]. Thus, the network differential Local Moran’s I Analysis and network LIMA are complementary to each other. The network differential Local Moran’s I Analysis detects the significant clusters of changes, while the network LIMA detects segments with significant rank changes relative to its neighbors or segments with significant rank changes among its neighbors.

In this work, we use the network LIMA to quantify the spatio-temporal patterns of rank concordance of the traffic collisions between six hours of day periods. In accordance with the network differential Local Moran’s I analysis, the segmentation width is set to 100 meters, and the bandwidth threshold for computing the spatial weight matrix is set to 300 meters. The GIMA is used first to investigate the global concordance between the six hours of day periods. Table 3 displays the symmetric concordance matrix, with significant values ($p < 0.05$) indicated in bold and with asterisks. The matrix shows that the period 2–6 has the largest rank changes with all other periods, but the τ_w values are not significant, possibly because that the majority of segments have no collisions during late at night. A similar trend exists for the period 22–2. The only significant changes happen between the pairs of 6–10 and 10–14, 6–10 and 18–22, as well as 10–14 and 18–22. Overall, the positive values in the matrix shows that concordance is the dominant pattern.

Figs 10 and 11 show the distribution of $\tilde{\tau}_i$ and $\tilde{\tau}_i$ values of six consecutive pairs of the six hour of day periods. Although the majority of the roads show a tendency of concordance, the maps show complex local interactions among the segments with significant LIMA measurements. This is reflected by segments with both positive and negative values of $\tilde{\tau}_i$ and $\tilde{\tau}_i$ spread throughout the study area. The yellow segments indicate places that tend to retain their ranks among their neighbors or neighborhood set, while the blue segments indicate places that tend to switch ranks with its neighbors or neighborhood set. The results will help practitioners identify roads that are stable in terms of risks, and places with risks that change in certain periods. The maps also capture the most changes in the four period pairs: 6–10 to 10–14, 10–14 to 14–18, 14–18 to 18–22, and 18–22 to 22–2. This is due to the fact that most of the segments have zero collisions during the 22–2 and 2–6 periods. In practice, the network LIMA in this work could be used to identify road segments that have sustained risks across different periods (i.e. the yellow segments), or periodical risks in certain periods (i.e. the blue segments). The periods 6–10, 10–14, 18–22 are periods with vibrant urban activities and largest traffic flows. From 6–10 to 10–14, there are a lot of yellow segments, signaling the traffic patterns are similar from early in the morning until noon. From 10–14 to 14–18, and 14–18 to 18–22, the number of yellow segments has reduced, which indicates more diversified traffic patterns. From 18–22 to 22–2, there are more blue segments than yellow segments, indicating an overall shift of traffic activities. To investigate further the spatial distribution and causes of these sustained or periodical risks at certain road segments, additional data such as the traffic flow, pedestrian traffic, and road characteristics needs to be integrated into the statistical regression analysis.

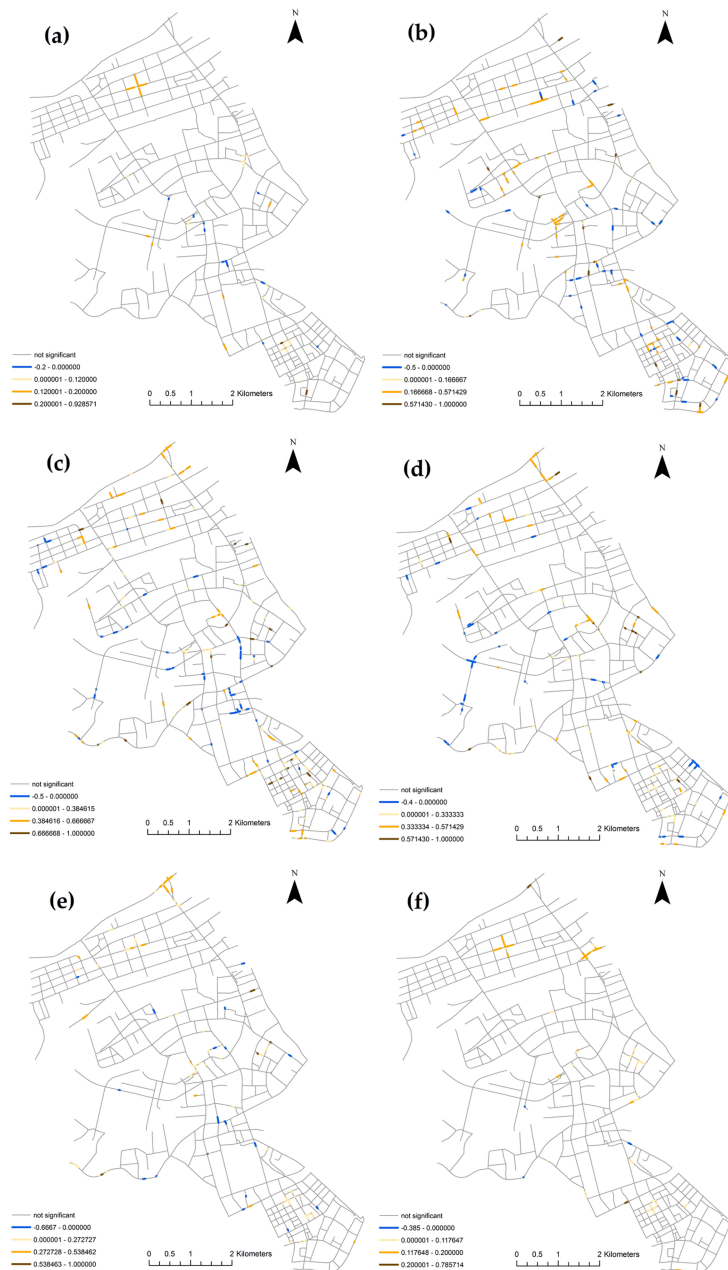


Fig 10. Spatial distribution of the neighbor set $\tilde{\tau}_i$ of six consecutive pairs of the six hours per day periods in Jiangnan, Wuhan, China: (a) 2–6 to 6–10; (b) 6–10 to 10–14; (c) 10–14 to 14–18; (d) 14–18 to 18–22; (e) 18–22 to 22–2; and (f) 22–2 to 2–6.

<https://doi.org/10.1371/journal.pone.0195093.g010>

Compared with the network KDE and network differential Local Moran's I methods, the network LIMA analysis captures the local interactions of rank changes at the segment level. Figs 10 and 11 convey essentially same information, yet more clusters appear in the $\tilde{\tau}_i$ distribution visually. However, these clusters are located at roughly the same places as the $\tilde{\tau}_i$ clusters. In other words, the clusters are more expanded in the $\tilde{\tau}_i$ maps because all of the neighborhood sets are included in the computations.

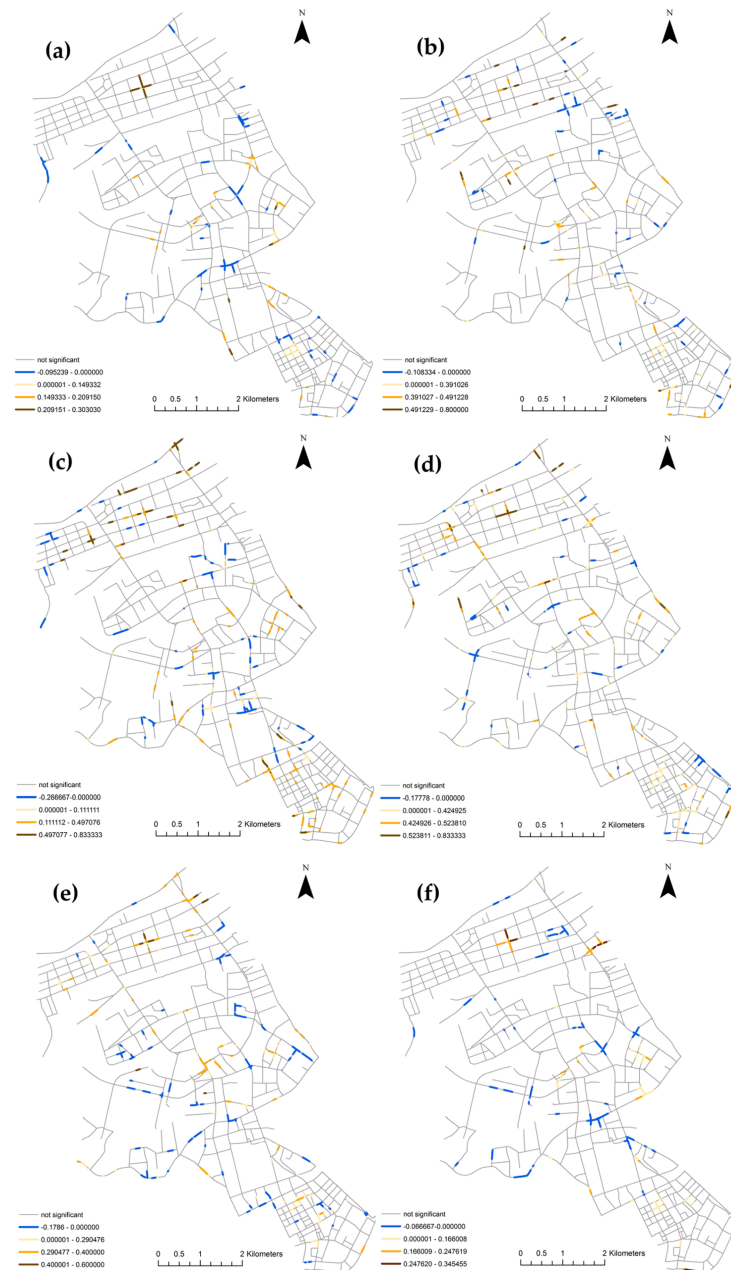


Fig 11. Spatial distribution of the neighborhood set $LIMA \tilde{z}_i$ of six consecutive pairs of the six hours per day periods in Jiangnan, Wuhan, China: (a) 2–6 to 6–10; (b) 6–10 to 10–14; (c) 10–14 to 14–18; (d) 14–18 to 18–22; (e) 18–22 to 22–2; and (f) 22–2 to 2–6.

<https://doi.org/10.1371/journal.pone.0195093.g011>

Discussion and conclusions

GIS-based methods have nowadays been used extensively for mapping traffic collisions, which helps practitioners identify hazardous road locations. On the other hand, the demand for understanding the spatio-temporal clustering patterns of traffic collisions, and network-constrained phenomenon in general, have promoted the developments of new models and algorithms within the research community. The advances in handling large datasets empower

analysts greatly by providing more effective and efficient ways to integrate multi-source, heterogeneous data. In this context, this work attempted to analyze the spatio-temporal patterns of traffic collisions that integrate the semantic information of the events and surrounding POIs. This work developed a series of network spatial analysis methods for traffic collision analysis. The weighted network KDE method is extended to allow flexible normalization of attribute weights. The differential Moran's I method and LIMA are extended to the network space in order to detect and quantify the changes of the collision hotspots over time.

The case study proved that these methods could help practitioners to better understand the spatio-temporal patterns and changes in traffic collisions in two ways. First, semantic information is integrated into the analysis methods. Specifically, the weighted network KDE is a simple and straightforward way to integrate the semantics of traffic collisions. The idea of weights could also be embedded in other methods, such as the Knox test, to analyze the spatio-temporal interaction of traffic collisions. The network cross K-function is utilized to analyze the correlation between the dynamic spatial point process of traffic collisions and the fixed spatial locations of POIs. This offers policymakers a new perspective of how collisions interact with the surrounding locations. The second is to analyze data under a meaningful organization of time, such as the hour of the day. This could be done by either conducting analysis at different time points or quantifying the change between time points. The proposed network differential Local Moran's I and LIMA are such measures to quantify and visualize the hotspot changes over time.

The methods in this work are essentially exploratory, which helps practitioners and policymakers discover the spatio-temporal clustering patterns from various perspectives at the road segment level. Additionally, the results from these methods are easy to interpret, thus providing a channel for announcing road safety information to the public. These methods could be applied to the analysis of other network-constrained phenomena, such as events of city management. The analysis in this work could be combined with percolation analysis in a traffic network, which offers an innovative perspective to identify bottleneck links considering real-time traffic dynamics [66]. The spatial distribution and patterns of bottleneck links and traffic collisions could be compared to reveal interesting relationships between traffic flow and collisions. Future work will also integrate more data in the analysis pipeline. This includes the attributes related to the roads such as the traffic volumes, road types and weather condition. Furthermore, those attributes related to surrounding POIs could also be integrated, such as the population density, pedestrian traffic, and building structures. These data are also useful for confirmatory analysis in spatio-temporal regressions. In terms of time, the analysis could be done on how the patterns of collisions change around different POIs on certain dates, including holidays, construction, or important events. From a modeling standpoint, the network cross K-function could be extended to the spatio-temporal version for measuring the spatio-temporal aggregations, and it is also critical to analyze the properties of network cross K-function results using proper statistical methods. Additionally, the edge correction issue should be considered in the network space.

Author Contributions

Conceptualization: Yaxin Fan, Xinyan Zhu.

Formal analysis: Bing She, Wei Guo.

Investigation: Bing She, Wei Guo, Tao Guo.

Methodology: Yaxin Fan, Xinyan Zhu.

Writing – original draft: Bing She, Wei Guo.

Writing – review & editing: Bing She, Tao Guo.

References

1. Jiang B, Claramunt C. Topological analysis of urban street networks. *Environment and Planning B*. 2004; 31(1):151–62.
2. Organization WH. Global status report on road safety 2015: World Health Organization; 2015.
3. Haghshenas H, Vaziri M. Urban sustainable transportation indicators for global comparison. *Ecological Indicators*. 2012; 15(1):115–21. <https://doi.org/10.1016/j.ecolind.2011.09.010>
4. Wang Z, Yue Y, Li Q, Nie K, Tu W, Liang S. Analyzing Risk Factors for Fatality in Urban Traffic Crashes: A Case Study of Wuhan, China. *Sustainability*. 2017; 9(6):897. <https://doi.org/10.3390/su9060897>
5. Sun DJ, Zhao Y, Lu Q-C. Vulnerability analysis of urban rail transit networks: a case study of Shanghai, China. *Sustainability*. 2015; 7(6):6919–36.
6. Pojani D, Stead D. Sustainable urban transport in the developing world: beyond megacities. *Sustainability*. 2015; 7(6):7784–805.
7. Yao S, Loo BPY, Yang BZ. Traffic collisions in space: four decades of advancement in applied GIS. *Annals of GIS*. 2016; 22(1):1–14. <https://doi.org/10.1080/19475683.2015.1085440>
8. Loo BP, Anderson TK. *Spatial Analysis Methods of Road Traffic Collisions*: CRC Press; 2015.
9. Glasner P, Leitner M. Evaluating the Impact the Weekday Has on Near-Repeat Victimization: A Spatio-Temporal Analysis of Street Robberies in the City of Vienna, Austria. *ISPRS International Journal of Geo-Information*. 2017; 6(1):3. <https://doi.org/10.3390/ijgi6010003>
10. Rey SJ. Space–Time Patterns of Rank Concordance: Local Indicators of Mobility Association with Application to Spatial Income Inequality Dynamics. *Annals of the American Association of Geographers*. 2016; 106(4):788–803.
11. Mohan D, Tiwari G. Sustainable Transport Systems: Linkages between Environmental Issues, Public Transport, Non-Motorised Transport and Safety. *Economic and Political Weekly*. 1999; 34(25):1589–96.
12. UN GA. Transforming our world: The 2030 agenda for sustainable development. A/RES/70/1, 21 October, 2015.
13. Persia L, Cipriani E, Sgarra V, Meta E. Strategies and Measures for Sustainable Urban Transport Systems. *Transportation Research Procedia*. 2016; 14:955–64. <https://doi.org/10.1016/j.trpro.2016.05.075>
14. Hidalgo D, Huizenga C. Implementation of sustainable urban transport in Latin America. *Research in Transportation Economics*. 2013; 40(1):66–77. <https://doi.org/10.1016/j.retrec.2012.06.034>
15. Wei VF, Lovegrove G. Sustainable road safety: a new (?) neighbourhood road pattern that saves VRU lives. *Accident; analysis and prevention*. 2012; 44(1):140–8. <https://doi.org/10.1016/j.aap.2010.12.005> PMID: 22062348.
16. Wakabayashi H, Ando M, Kawaguchi T, Horie Y, Hotta K-h. Sustainable Traffic Safety Management at Accident Black Spots Combined with Drivers' Psychology and Vehicle Engineering Using Eye Mark Recorder. *Transportation Research Procedia*. 2014; 3:90–9. <https://doi.org/10.1016/j.trpro.2014.10.094>
17. Loo BPY. Role of Stated Preference Methods in Planning for Sustainable Urban Transportation: State of Practice and Future Prospects. *Journal of Urban Planning and Development*. 2002; 128(4):210–24.
18. Goldman T, Gorham R. Sustainable urban transport: Four innovative directions. *Technology in Society*. 2006; 28(1–2):261–73. <https://doi.org/10.1016/j.techsoc.2005.10.007>
19. Wegman F, Aarts L, Bax C. Advancing sustainable safety. *Safety Science*. 2008; 46(2):323–43.
20. Yamada I, Thill J-C. Local Indicators of Network-Constrained Clusters in Spatial Patterns Represented by a Link Attribute. *Annals of the Association of American Geographers*. 2010; 100(2):269–85. <https://doi.org/10.1080/00045600903550337>
21. Ripley BD. The Second-Order Analysis of Stationary Point Processes. *Journal of Applied Probability*. 1976; 13(2):255–66. <https://doi.org/10.2307/3212829>
22. Yamada I, Thill J-C. Comparison of planar and network K-functions in traffic accident analysis. *Journal of Transport Geography*. 2004; 12(2):149–58. <http://dx.doi.org/10.1016/j.jtrangeo.2003.10.006>
23. Lu Y, Chen X. On the false alarm of planar K-function when analyzing urban crime distributed along streets. *Social Science Research*. 2007; 36(2):611–32. <https://doi.org/10.1016/j.ssresearch.2006.05.003>

24. Loo BP, Yao S. The identification of traffic crash hot zones under the link-attribute and event-based approaches in a network-constrained environment. *Computers, Environment and Urban Systems*. 2013; 41:249–61.
25. Xie Z, Yan J. Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*. 2008; 32(5):396–406. <https://doi.org/10.1016/j.compenvurbsys.2008.05.001>
26. Okabe A, Yamada I. The K-Function Method on a Network and Its Computational Implementation. *Geographical Analysis*. 2001; 33(3):271–90.
27. Steenberghen T, Aerts K, Thomas I. Spatial clustering of events on a network. *Journal of Transport Geography*. 2010; 18(3):411–8.
28. Sugihara K, Satoh T, Okabe A, editors. Simple and unbiased kernel function for network analysis. 10th International Symposium on Communications and Information Technologies; 2010; Meiji University, Meiji, Japan: IEEE Circuits and Systems Society.
29. Mohaymany AS, Shahri M, Mirbagheri B. GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. *Geo-spatial Information Science*. 2013; 16(2):113–9. <https://doi.org/10.1080/10095020.2013.766396>
30. Xie Z, Yan J. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of transport geography*. 2013; 31:64–71.
31. Loo BP, Yao S, Wu J, editors. Spatial point analysis of road crashes in Shanghai: A GIS-based network kernel density method. *Geoinformatics, 2011 19th International Conference on*; 2011: IEEE.
32. Yamada I, Thill J-C. Local Indicators of Network-Constrained Clusters in Spatial Point Patterns. *Geographical Analysis*. 2007; 39(3):268–92. <https://doi.org/10.1111/j.1538-4632.2007.00704.x>
33. Nie K, Wang Z, Du Q, Ren F, Tian Q. A Network-Constrained Integrated Method for Detecting Spatial Cluster and Risk Location of Traffic Crash: A Case Study from Wuhan, China. *Sustainability*. 2015; 7(3):2662. <https://doi.org/10.3390/su7032662>
34. Ni J, Qian T, Xi C, Rui Y, Wang J. Spatial Distribution Characteristics of Healthcare Facilities in Nanjing: Network Point Pattern Analysis and Correlation Analysis. *International Journal of Environmental Research and Public Health*. 2016; 13(8):833. <https://doi.org/10.3390/ijerph13080833> PMID: 27548197
35. Rui Y, Yang Z, Qian T, Khalid S, Xia N, Wang J. Network-constrained and category-based point pattern analysis for Suguo retail stores in Nanjing, China. *International Journal of Geographical Information Science*. 2016; 30(2):186–99. <https://doi.org/10.1080/13658816.2015.1080829>
36. Flahaut Bt. Impact of infrastructure and local environment on road unsafety: Logistic modeling with spatial autocorrelation. *Accident Analysis & Prevention*. 2004; 36(6):1055–66.
37. Flahaut B, Mouchart M, San Martin E, Thomas I. The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach. *Accident Analysis & Prevention*. 2003; 35(6):991–1004.
38. Yao S, Loo BP, Lam WW. Measures of activity-based pedestrian exposure to the risk of vehicle-pedestrian collisions: space-time path vs. potential path tree methods. *Accident Analysis & Prevention*. 2015; 75:320–32.
39. Lam WW, Loo BP, Yao S. Towards exposure-based time-space pedestrian crash analysis in facing the challenges of ageing societies in Asia. *Asian Geographer*. 2013; 30(2):105–25.
40. Ye X, Xu X, Lee J, Zhu X, Wu L. Space–time interaction of residential burglaries in Wuhan, China. *Applied Geography*. 2015; 60:210–6. <http://dx.doi.org/10.1016/j.apgeog.2014.11.022>.
41. Wu L, Xu X, Ye X, Zhu X. Repeat and near-repeat burglaries and offender involvement in a large Chinese city. *Cartography and Geographic Information Science*. 2015; 42(2):178–89. <https://doi.org/10.1080/15230406.2014.991426>
42. Rey SJ, Mack EA, Koschinsky J. Exploratory Space–Time Analysis of Burglary Patterns. *Journal of Quantitative Criminology*. 2012; 28(3):509–31. <https://doi.org/10.1007/s10940-011-9151-9>
43. Xianrui X, Zhongren P, editors. The K-function analysis of space-time point pattern on road network. *Geoinformatics, 2011 19th International Conference on*; 2011 24–26 June 2011.
44. Eckley DC, Curtin KM. Evaluating the spatiotemporal clustering of traffic incidents. *Computers, Environment and Urban Systems*. 2013; 37(0):70–81. <http://dx.doi.org/10.1016/j.compenvurbsys.2012.06.004>.
45. Levine N, Kim KE, Nitz LH. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accident Analysis & Prevention*. 1995; 27(5):675–85.
46. Wang Z, Yue Y, Li Q, Nie K, Yu C. Analysis of the Spatial Variation of Network-Constrained Phenomena Represented by a Link Attribute Using a Hierarchical Bayesian Model. *ISPRS International Journal of Geo-Information*. 2017; 6(2):44. <https://doi.org/10.3390/ijgi6020044>
47. Miaou S-P, Song JJ, Mallick BK. Roadway traffic crash mapping: a space-time modeling approach. *Journal of Transportation and Statistics*. 2003; 6:33–58.

48. Agüero-Valverde J, Jovanis PP. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention*. 2006; 38(3):618–25. <http://dx.doi.org/10.1016/j.aap.2005.12.006>.
49. Li Z, Wang W, Liu P, Bigham JM, Ragland DR. Using Geographically Weighted Poisson Regression for county-level crash modeling in California. *Safety science*. 2013; 58:89–97.
50. Pirdavani A, Bellemans T, Brijs T, Wets G. Application of geographically weighted regression technique in spatial analysis of fatal and injury crashes. *Journal of Transportation Engineering*. 2014; 140(8):04014032.
51. Erdogan S. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. *Journal of safety research*. 2009; 40(5):341–51. <https://doi.org/10.1016/j.jsr.2009.07.006> PMID: 19932314
52. Wuhan Municipal Bureau of Statistics., Statistical Communiqué of Wuhan on the 2016 National Economic and Social Development 2016 [8/30/2017]. <http://www.whjt.gov.cn/details.aspx?id=3439>.
53. She B, Zhu X, Ye X, Guo W, Su K, Lee J. Weighted network Voronoi Diagrams for local spatial analysis. *Computers, Environment and Urban Systems*. 2015; 52:70–80.
54. Okabe A, Satoh T, Sugihara K. A kernel density estimation method for networks, its computational method and a GIS-based tool. *International Journal of Geographical Information Science*. 2009; 23(1):7–32. <https://doi.org/10.1080/13658810802475491>
55. Bailey TC, Gatrell AC. *Interactive spatial data analysis*: Longman Scientific & Technical Essex; 1995.
56. O'Sullivan D, Wong DWS. A Surface-Based Approach to Measuring Spatial Segregation. *Geographical Analysis*. 2007; 39(2):147–68. <https://doi.org/10.1111/j.1538-4632.2007.00699.x>
57. Okabe A, Satoh T. Uniform network transformation for points pattern analysis on a non-uniform network. *J Geogr Syst*. 2006; 8(1):25–37.
58. Anselin L. Local Indicators of Spatial Association—LISA. *Geographical Analysis*. 1995; 27(2):93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
59. Anselin L. GeoDa Workshop 2016 [cited 2017 8/30/2017]. https://s3.amazonaws.com/geoda/software/docs/geoda_1.8_2.pdf.
60. Kendall MG. *Rank correlation methods*. 3rd ed ed. London: Griffin; 1970.
61. Rey SJ. Fast algorithms for a space-time concordance measure. *Computational Statistics*. 2014; 29(3–4):799–811.
62. Rey SJ. *Spatial dependence in the evolution of regional income distributions*. Spatial econometrics and spatial statistics Palgrave, Hampshire. 2004:193–213.
63. Rey SJ, Anselin L. PySAL: A Python library of spatial analytical methods. *Handbook of applied spatial analysis*. 2010:175–93.
64. Rey S. Geospatial Distribution Dynamics (giddy) in PySAL 2017 [8/30/2017]. <https://github.com/pysal/giddy>.
65. Porta S, Latora V, Wang F, Strano E, Cardillo A, Scellato S, et al. Street centrality and densities of retail and services in Bologna, Italy. *Environment and Planning B: Planning and Design*. 2009; 36(3):450–65.
66. Li D, Fu B, Wang Y, Lu G, Berezin Y, Stanley HE, et al. Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proceedings of the National Academy of Sciences*. 2015; 112(3):669–72.