Network Constrained Spatio-temporal Analysis Tool for Traffic Accidents in Leeds, United Kingdom

ANG JiaYing
School of Information Systems
Singapore Management University
Singapore
jiaying.ang.2015@sis.smu.edu.sg

Sheryl CHONG Man Er
School of Information Systems
Singapore Management University
Singapore
mechong.2015@sis.smu.edu.sg

TAN Yan Lin
LKC School of Business
Singapore Management University
Singapore
yanlin.tan.2015@business.smu.edu.sg

ABSTRACT

Application of geospatial analytics on traffic accidents is imperative for urban development design to improve road safety. Targeting road traffic accidents specific to certain casualty groups is also essential for effective traffic management. However, clusters of traffic accidents on networks are likely to evolve overtime. This characteristic shows that traffic accidents are not isolated in time and space. Therefore, traffic accident analyses have to incorporate both time and space elements so that temporary resources by the traffic police could be allocated efficiently, or future road planning efforts by transport authorities could be implemented in the right direction. However, currently available web-based traffic collisions applications mainly focus on visualising traffic accident point patterns. A web-based geospatial analytics tool, SIGNAL, is then developed to address the need for an interactive network constrained spatio-temporal dashboard on traffic accidents. This paper aims to explore the use of network constrained spatio-temporal statistical techniques on traffic accidents data in Leeds, United Kingdom. Four key methods are employed to conduct analyses in SIGNAL. Firstly, Network Constrained Kernel Density Estimation is used to derive insights on traffic collision intensity patterns. To enable identification of statistically significant clusters, Network Constrained K-Function is incorporated. Lastly, Network Constrained Cross K-Function and Network Constrained Cross Pair Correlation are adopted for investigating correlation between traffic collision points and variables of interests, such as pedestrian crossings, motorway junctions and schools. The results obtained from our demonstration highlight key insights that could help transport authorities, traffic police and even business users better understand spatio-temporal clustering patterns and correlations of traffic accidents.

KEYWORDS

Network Constrained, Spatio-temporal Analysis, Traffic Accidents, Leeds, Geospatial Analytics, R, R Shiny, Kernel Density Estimation, K-Function, Cross K-Function, Cross Pair Correlation.

1 INTRODUCTION

In order to attempt to reduce traffic accidents, it is important to understand where, when and who is involved in traffic accidents. A better understanding of spatio-temporal patterns specific to casualty groups aids in developing appropriate preventive measures by the authorities. For example, a road segment could have high intensity of traffic accidents for the elderly, but those traffic accidents could possibly occur only at certain time periods, such as at night. In fact, spatio-temporal analysis of traffic accidents is well-known to be necessary due to the tendency of occurrence along certain road networks and also at the same time, be distributed in certain time periods, such as years and months [1].

The use of spatio-temporal analysis involves two dimensions - spatial dimension and temporal dimension. The temporal dimension describes the evolution of an object overtime while the spatial dimension highlights the movement of an object over a topography. Other factors such as environmental characteristics, involving weather conditions and road surface, are also known to be associated with distinct spatial patterns [1].

In view of the benefits of applying spatio-temporal analyses to traffic accidents, our initial focus was on developing a web-based geospatial application for Singapore as there is a general lack of spatio-temporal geospatial applications developed specific to this area. The Singapore Police Force is also concerned about accidents involving motorcyclists and elderly jaywalkers as motorcyclist accidents still account for more than half of the traffic accidents in 2017, and the number of elderly road fatalities have been on the rise [3]. Although this serves as grounds for adapting our application to Singapore's context, our team has used data from Leeds, United Kingdom instead, due to its easily accessible rich diversity of information. Specifically, Leeds' traffic accident data includes coordinates of traffic accidents, time, weather and details on casualties, all of which are essential for an in-depth spatio-temporal analysis for a certain casualty group. While behavioral patterns of drivers and pedestrians in United Kingdom may not be reflective of that of Singapore's, insights still provide some indicative directions for investigations by Singapore Land Transport Authority and Singapore Police Force.

This paper reports on our development efforts in designing and implementing a geospatial analytics tool for use by transport authorities in analysing traffic accidents. The first few sections of the paper provides a general overview of our motivation and objectives. This is followed by a quick review on current research papers specific to applying geospatial techniques on traffic accidents. After which, the approach used in developing our application, SIGNAL, including the statistical methods used, would be explained before moving on to provide an overview of the application interface. The results obtained from analysing Leeds' traffic accidents data would then be discussed. This paper concludes by highlighting the areas for improvement and future development work for extending our geospatial application.

2 MOTIVATION & OBJECIVES

Our application development efforts were motivated by the gap between currently available road traffic geospatial tools and analysts' needs. Current geospatial road traffic applications mainly portray traffic accidents as point events, with little analyses generated to provide in-depth insights. As such, a web-enabled geospatial analytics tool, SIGNAL, was developed with the purpose of allowing users to conduct statistical analyses on road networks for selected target groups or environmental conditions. It aims to provide transport authorities and traffic police with an analytical tool for discovering network-constrained spatio-temporal patterns of traffic accidents. Specifically, it focuses on the following objectives:

- To visualise the intensity of traffic accidents on road networks cartographically on an internet-based map such as ESRI
- To conduct statistical simulations on road segments to reveal evidence of clusters or correlation patterns
- To provide a user-friendly interface to for practitioners to apply relevant filters for different time periods selected

3 RELATED WORKS

Spatial analysis of traffic accidents has been conducted by researchers for many years. Ya Xin et al. [5] have applied various spatial statistical analysis methods to study traffic accidents in Wuhan, China. Four techniques were used to explore spatio-temporal clustering patterns - Weighted Network Kernel Density Estimation, Network Cross K-Function, Network differential Local Moran's I and Network Local Indicators of Mobility Association. The weighted Network Kernel Density Estimation was used to identify traffic accidents hotspot while Network Cross K-Function was used to explore whether is there any clustering tendencies between traffic collisions and different variable. Network differential Local Moran's I and Network Local Indicators of Mobility Association provides straightforward and quantitative measures of changes in traffic accidents hotspots. These statistical techniques are comprehensive in nature and some elements should be incorporated into our SIGNAL application.

A similar study was conducted to analyse the traffic accidents in New York [6]. Spatio-temporal Network Kernel Density Estimation (STNKDE) was used to explore hotspots of traffic accidents at different time periods. The importance of using space-time was emphasised in this paper as more efficient allocation of resources was possible by incorporating the second-dimension of time. However, due to its comprehensive discussion on implementing space-time element, other geospatial techniques were not discussed.

As seen, Network Constrained Kernel Density Estimation appears to be a commonly used technique to determine the traffic accident hotspots. Often, other spatial statistical analytics methods are accompanied with Network Constrained Kernel Density Estimation to offer in-depth analysis of traffic collisions. Our application was thus inspired to incorporate both elements.

4 ANALYSIS METHODS

As mentioned in the previous section, a combination of different spatial statistical analysis methods is often used by geospatial analysts to conduct spatial analysis of traffic accidents. For our SIGNAL application, both first-order and second-order methods in examining accident point processes are used. First-order method includes Network Constrained Kernel Density Estimation while second-order methods used are Network Constrained K-Function, Network Constrained Cross K-Function and Network Constrained Cross Pair Correlation Function.

4.1 NETWORK CONSTRAINED KERNEL DENSITY ESTIMATION

Kernel Density Estimation (KDE) is one of the most popular methods for analyzing the first order properties of a point event distribution due to its easy and simple implementation [1]. It involves estimating the probability density function of a variable, or in geospatial terms, the density of features in a neighborhood. Investigating the average density of points along the network provides a quick insight on which segments of roads have higher intensity of traffic accidents. There are two main types of intensities when estimating kernel densities in networks. The first type is homogeneous intensity function, where all points are independent and uniformly distributed in any given set and has randomness that is characterized by complete spatial randomness. The second type is inhomogeneous intensity function, where points are also independence between disjoint sets but unevenly distributed according to their spatially varying intensity functions.

Since traffic accidents almost always happen on roadways and inside a network, homogeneity intensity function is applied. The kernel estimate of intensity is formally defined as,

$$\widetilde{\lambda}(u) = \sum_{i} \kappa(x_i, u)$$

where x is a given set of point patterns, {x1, … xn}, and κ (v, u) is the smoothing kernel to smooth out points on the network, as shown in Figure 1, in a book by Adrian et al (2016). The smoothing of the point is done with a Gaussian kernel and thickness of line is proportional to the kernel value. This can be performed using the density.lpp function from the spatstat package in R.

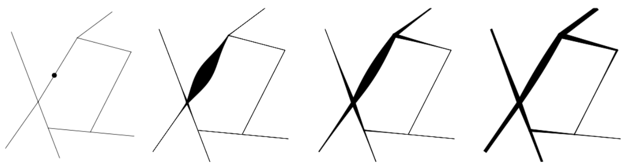
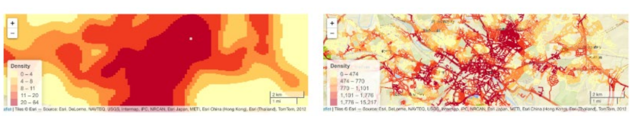


Figure 1. Smoothing kernel on a network, with increasing bandwidth sigma from the left to right.

Due to the nature of the data points studied in this paper, that is traffic accident and casualty points, it is important to understand the impact of and difference between the conventional KDE method and the network constrained KDE method.

In the conventional KDE (non-network constrained), calculation of intensity is based on Euclidean distance search bandwidth and does not take into consideration the presence of road network structures. This hinders the ability of pinpointing exact locations that has high or low intensities of traffic accidents as the density values are measured per area unit over a 2-D space.

Comparatively, network constrained KDE estimates the intensity of traffic accidents strictly over a network space. This allows for a clear distinction of road networks that has high intensity of traffic accidents versus those that are not as density values are measured per linear unit over a network instead.



(a) (b)

Figure 2. Comparison of the (a) conventional KDE and the (b) network constrained KDE. Performing analysis on the same data points using Gaussian kernel and 500 meters bandwidth.

Figure 2 illustrates the difference between the conventional KDE and network constrained KDE on the same set of data and area. Evidently, it is expected that the conventional KDE will have the highest density in the middle and lower density near the edges, as shown in Figure 2 (a). This could be attributed to the fact that most accidents tend occur near the city or central area, as compared to other areas. However, it is inappropriate to conclude as such.

Although Figure 2 (b) also showed high intensity in the central area, there are observable differing levels of intensities within it. This would be helpful for the authorities to distinguish areas that they should focus on in attempts to reduce the occurrence of traffic accidents.

Thus, network constrained KDE is chosen as the appropriate method to analyze the intensity of traffic accident and casualty points.

4.2 NETWORK CONSTRAINED K-FUNCTION

The Okabe-Yamada Network Constrained K-Function defines the Network Constrained K-Function, by adapting the Ripley's K Function through replacing the Euclidean distance with the shortest path distance. In this method, given a point v on the network, all locations in the network that can be reached from v by a path of length shorter than or equal to a radius r, defined by the user, would be considered. It is defined by the function below, with λ(L) denoting the total length of the linear network:

$$\widehat{K}_{\text{net}}(r) = \frac{\ell(L)}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \mathbf{1}\left\{d_L(x_i, x_j) \leq r\right\},$$

While the above Network Constrained K-Function constrains points to networks, a second-order stationary point process is required for our analyses. The Okabe-Yamada Network Constrained K-Function assumes that the network itself is homogeneous, which is not the case as different locations in the network is surrounded by different configurations of line segments. Network Constrained K-Functions obtained from different networks are not directly compatible in this case.

The second-order Network Constrained K-Function is proposed by Ang et al (2012), known as the 'geometrically corrected K-Function', will be used in our application. It is an extension of the Ripley's K-Function's benefits of enabling comparison between different point processes with different intensities, observed in different windows, combined with Okabe-Yamada network Constrained K-function. The geometrically corrected K-Function is defined by, for all r ≤ R, where u is any location on the network:

$$K^L(r) = \frac{1}{\lambda}\mathbb{E}\left[\sum_j \frac{1\{0 < d_L(u, x_j) \leq r\}}{m(u, d_L(u, x_j))} \;\middle|\; u \in \mathbf{X}\right]$$

The above analysis is computed in spatstat by the function linearK, and by default uses the second-order Network Constrained K-Function, which assumes homogeneity. The function linearK requires the input of accidents constrained to road network, as seen in below Figure 3. Network Constrained K-Function is then computed based on the linnet captured in the area of analysis.

Figure 3: Linnet of Traffic Accidents in Leeds' city centre

4.3 NETWORK CONSTRAINED CROSS K-FUNCTION

Network Constrained K-Function handles points of the same type while Network Constrained Cross-K Function is used for two different sets of points. Estimation is based on measuring pairwise distances from all points of type i to all points of type j. Thus, for any pair of types i and j, the function calculates the expected number of points of type j lying within a distance r of a typical point of type i, standardised by dividing the intensity of points of type j, for r >= 0. The Cross-K Function is as shown below, and it is constrained to a network in our application.

$$K_{ij}(r) = \frac{1}{\lambda_j}\mathbb{E}\left[t(u,r,\mathbf{X}^{(j)}) \mid u \in \mathbf{X}^{(i)}\right]$$

This analysis is computed in spatstat by the function linearKcross. Similarly, linearKcross requires the input data to be in the form of linnet, with accidents and pedestrian crossings as marked point pattern processes so that the function can differentiate between the two-point patterns (Figure 4 below).

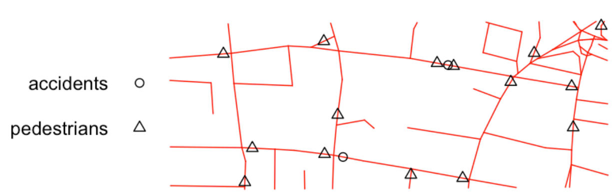


Figure 4: Linnet of Traffic Accidents & Pedestrian Crossings in Leeds' city centre

4.4 NETWORK CONSTRAINED CROSS PAIR CORRELATION FUNCTION

Similar to Network Constrained Cross K-Function, Network Constrained Cross Pair Correlation Function measures pairwise distances from all points of type i to all points of type j. However, this function calculates the expected number of points of type j lying at a distance equal to distance r of a typical point of type i, standardised by dividing the intensity of points of type j, for r >= 0. The cross pair correlation function is as shown below, and it is constrained to a network in our application.

$$\widehat{g}(r) = \frac{|W|}{2\pi rn(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}\kappa_h(r-d_{ij})e_{ij}(r)$$

This analysis is computed in spatstat by the function linearpcfcross. The preparation and processing of data for Network Constrained Cross Pair Correlation follows closely that of Network Constrained Cross K-Function.

5 APPROACH

5.1 DATA COLLECTION

Data of Leeds Traffic Accident Data (2013 to 2017), Schools and Pedestrian Crossings data were collected from the UK Open Database in CSV file format, while Leeds' District Boundary Map, Road Network and Motorway Junctions were obtained from UK Consumer Data Research Centre (CDRC) in Shapefile (SHP) format.

5.2 DATA CLEANING

Traffic accident point events were separated from casualty point events before removing duplicates. Unique accident points allow for visualization of intensity and spatial distribution of traffic accidents. Standardization of data, such as, ensuring columns of each CSV files and their data types are the same before reclassifying selected columns, is conducted.

5.3 DATA TRANSFORMATION

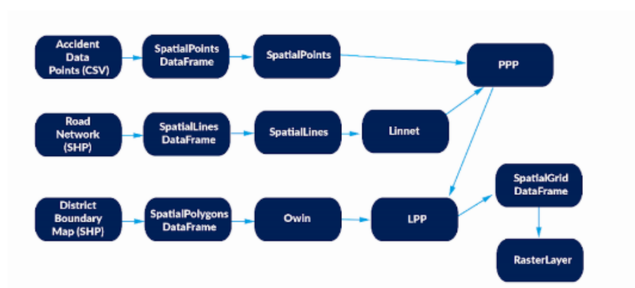


Figure 5: Overview of Data Transformation

As our geospatial application involves the use of Network Constrained analyses, data has to be transformed to appropriate formats before relevant functions could be used. Specifically, accident traffic points have to be converted to SpatialPoints and then to Point Pattern Processes. Roads would have to be converted to SpatialLines and then to Linnet before they could be combined with Point Patterns Processes to form Linear Point Patterns. In order to constrain the Linear Point Patterns to a target area of interest, it has to be intersected with the Owin of District Boundary Map. The below figure summarises the key data transformation that takes place.

6 SYSTEM ARCHITECTURE

Our application is built using R programming language due to its numerous spatial analysis package such as RGDAL, Spatstat and Leaflet and the ease and flexibility of creating an interactive web application. Our application is then deployed to Shinyapp.io so that users are able to view and interact with it online using their preferred browser.

The following R packages are used during the development of our application:

shiny	shinydashboard	tidyverse	dplyr
DT	Lubridate	Leaflet	Rgdal
Sf	Sp	Spdep	maptools
Spatstat	Raster	Rgeos	shinycssloaders
shinyjs	classInt	V8	Rsconnect

7 APPLICATION

7.1 APPLICATION USER INTERFACE DESIGN

Our application consists of 6 tabs which will lead the user to our overview tab, different network constrained analysis tabs and our data set.

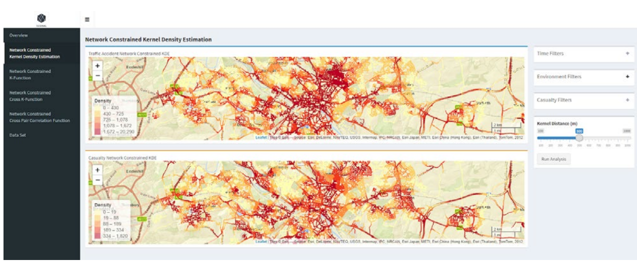


Figure 6: Application User Interface for Network Constrained Kernel Density Estimation

For Network Constrained KDE tab, there are 2 maps showing the result for traffic accident and casualty. The user is able to zoom in or out and move the map to an area of interest When the user moves or zooms into the top map (traffic accident KDE), the bottom map (casualty KDE) will be updated automatically as both maps are synchronised. This ensures that analyses on both maps are at the same geographical area for fair comparison. On right side of the application, there are different types of filters and controls for the user to select.

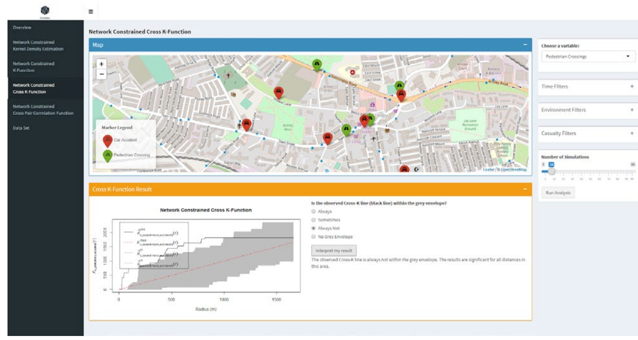
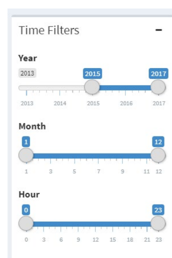


Figure 7 Application User Interface for Network Constrained Cross K-Function

Analyses in other tabs also make use of the same user interface for easy navigation. The map now contains markers to indicate the geographical positions of traffic accidents events and other variables (Pedestrian Crossing, Motorway Junctions and Schools). The user can zoom in or out and move the map to an area that they wish to run the analysis on. There is a result box at the bottom of the map will display the result of the analysis. On the right side of the application, there are also different types of filters and controls for the user to select.

7.2 FILTERS AND COMPONENTS OF THE APPLICATION

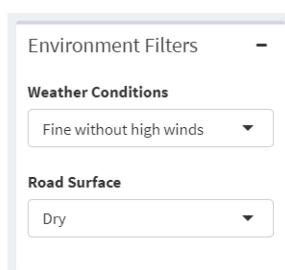
7.2.1 TIME FILTERS



The user will be able to filter the data by year, month and hours.

7.2.2 ENVIRONMENT FILTERS

The user will be able to filter the data by:

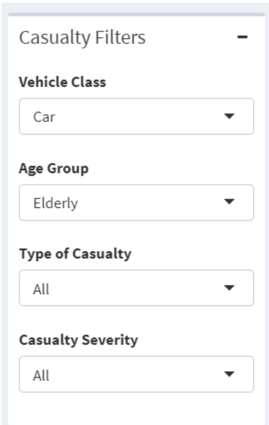


Weather conditions – All, Fine without high winds, Fine with high winds, Snowing without high winds, Snowing with high winds, Raining without high winds, Raining with high winds, Fog or mist, unknown and others

Road Surface – All, Dry, Frost / Ice, Wet / Damp, Snow, Others and Flood (surface water over 3cm deep)

7.2.3 CASUALTY FILTERS

The user will be able to filter the data by:



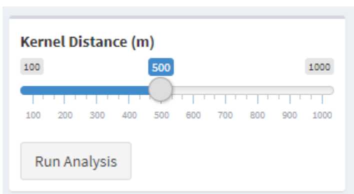
Vehicle Class – All, Car, Motorcycle, Bus / Coach, Bicycle, Goods Vehicle, Taxi / Private Hire, Mini Bus, Agricultural Vehicle, Mobility Scooter, Horse and Tram

Age Group – All, Adult, Elderly, Children and Young Adults

Type of Casualty – All, Driver or rider, Passenger and Pedestrian

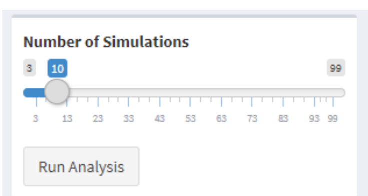
Casualty Severity – All, Slight, Serious and Fatal

7.2.4 KERNEL DISTANCE SLIDER



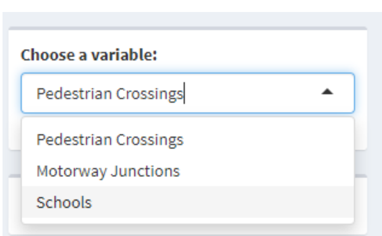
The bandwidth of kernel density plot. The user is able to drag the slider to state the kernel distance (in metres) which they want to run for the analysis.

7.2.5 SIMULATION SLIDER



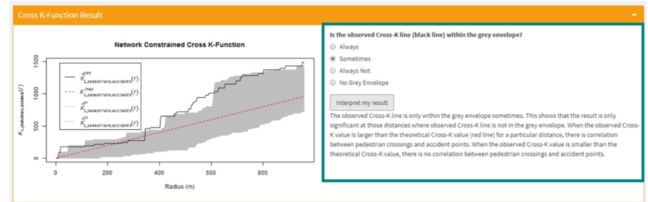
The user is able to drag the slider to state the number of simulations which they want to run for the analysis.

7.2.6 VARIABLE SELECTION DROPDOWN



The user is able to choose which variable they want to run the analysis with the traffic accidents event(s). This variable selection is only for Network Constrained Cross K-Function and Network Constrained Cross Pair Correlation Function analysis.

7.2.7 GRAPH INTERPRETATION FUNCTION



The user will choose an option according to the graph output and a general interpretation will be shown on the application. The graph interpretation function aids the user in interpreting the graph output. This graph interpretation function will be placed in all analysis except Network Constrained KDE.

8 RESULT

Network Constrained KDE method detects hot spots by calculating the density of traffic accidents for each raster cell size (23 by 23 meters) along the network within a specified bandwidth (e.g. 500 meters). Density values shown in the legend is measured in square kilometers, and it can be converted to expected count for easier interpretation. Since density is count divided by area, multiplying density by area will give an expected count (Krause, 2013). For instance, suppose the analysis is on accidents in a city over one month, a raster cell with a density value of 15,000, and the raster cell size area of 0.000529 (0.023 * 0.023) square kilometers, will give an expected count of approximately 8 accidents (calculated by 15,000 * 0.000529). This can be interpreted as, given the same accident conditions month to month, it can be expected to see about 8 accidents in the following month.

Following are three network constrained KDE use cases to explain how the results of the models can be analyzed and how it can be useful to potential users. In analyzing these use cases, the following filters are assumed to be constant throughout:

- Month: 1 to 12
- Road Surface: Dry
- Weather Conditions: Fine without high wind
- Casualty Severity: All (Slight, Severe, Fatal)
- Kernel Distance (m): 500

Table 1 summarizes the three use cases, highlighting the key differences in the filters applied.

Filters	Use Case 1 (Elderly Casualties)	Use Case 2 (Motorcycle Accidents over a three-year period)	Use Case 3 (Pedestrian Accidents, comparing Time)
Year	2015 to 2017	2015 vs 2016 vs 2017	2015 to 2017
Hours	0 - 23	0 - 23	9 - 17, 17 - 0
Vehicle Class	All	Motorcycle	All
Age Group	Elderly	All	All
Type of Casualty	All	All	Pedestrian

Table 1: Summary of Network Constrained use cases

8.1 USE CASE 1 – ELDERLY CASUALTIES

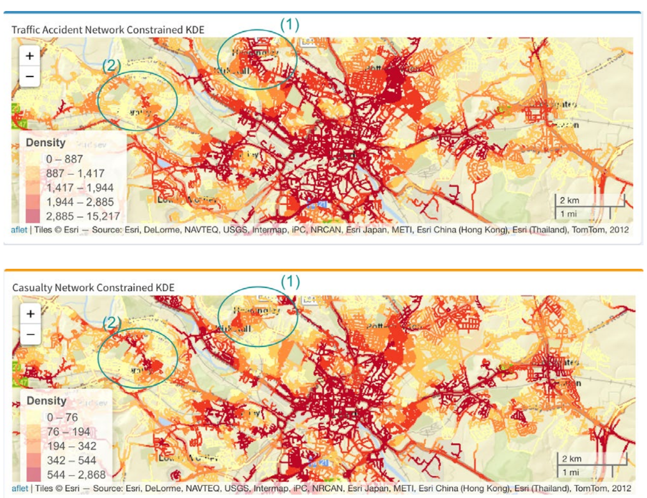


Figure 8: Elderly Casualties

When it comes to traffic accidents, it is instinctive for most people to only focus on the traffic accidents KDE to identify areas with high intensity of accidents and overlook the concentration of casualty points. For instance, if the authorities are looking into reducing the number of elderly casualties, naturally, they will focus on areas with high intensity of traffic accidents. However, that may not be representative of the age group that they are targeting. Thus, the solution provided two different network constrained KDE maps, one focusing on traffic accidents as a whole and the other focusing on the casualties.

In Figure 8, although the intensities of the two maps are not strikingly different, there are notable areas that can be focused on. In the first circled area (denoted as (1) in Figure 8), the intensity of traffic accidents is relatively high, however, the intensity of casualties is the opposite. This indicates that if the authorities are focusing on reducing elderly casualties, they should not be focusing all their resources in this area. Instead, they should be focusing in areas with higher intensity of casualty but were not so apparent when looking at the intensity of traffic accident, as shown in the second circled area (denoted as (2) in Figure 8). In Figure 8 (2), it is shown that there are relatively low traffic accidents happening in at area, however, the intensity of elderly casualties are quite high. This serves to show that the authorities should be focusing on this area, should their primary goal be reducing the number of accidents involving elderly casualties.

8.2 USE CASE 2 – MOTORCYCLE ACCIDENTS OVER THREE YEAR PERIOD

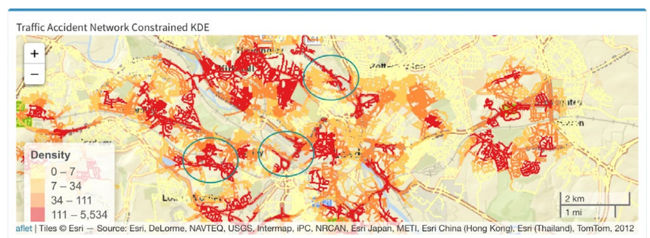


Figure 9 (a) Traffic Accident Network Constrained KDE on Motorcycle Accidents in 2015

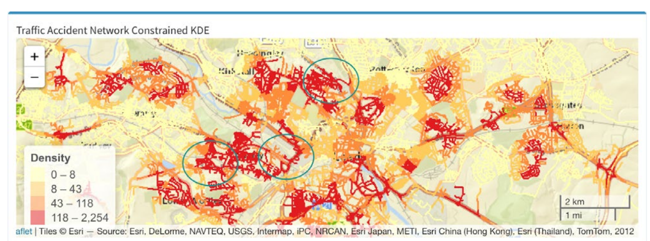


Figure 9 (b) Traffic Accident Network Constrained KDE on Motorcycle Accidents in 2016

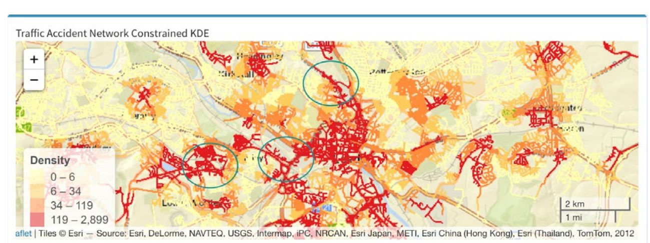


Figure 9 (c) Traffic Accident Network Constrained KDE on Motorcycle Accidents in 2017

With the increasing concerns on motorcycle accidents, it is important to understand areas with constant high intensities. In Figure 9, traffic accident network constrained KDE is plotted over three consecutive years, from 2015 to 2017, respectively. Three areas have been circled to highlight the constant high intensity of traffic accidents at the same area over the years. It is an important signal for the authorities to uncover reasons why these areas are persistently having high intensity of traffic accidents and implement countermeasures to lower motorcycle accidents.

8.3 USE CASE 3 – ACCIDENTS OVER DIFFERENT TIME PERIODS

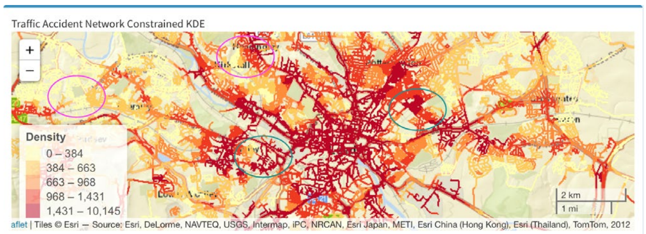


Figure 10 (a) Traffic Accident Network Constrained KDE – Day Time (9am to 5pm)

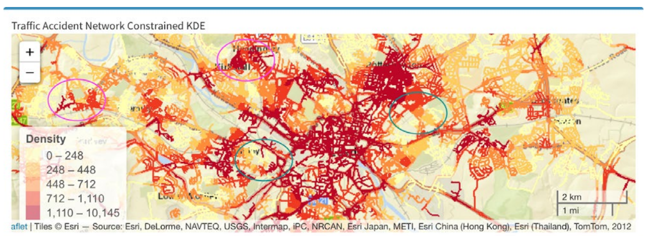


Figure 10 (b) Traffic Accident Network Constrained KDE – Day Time (5pm to 12am)

In analyzing the temporal aspect of the traffic accidents, we defined day time as 9am to 5pm and night time as 5pm to midnight (12am). As seen in Figure 10, circled in teal color are areas where there are higher intensity of accidents in the day time as compared to night time. On the contrary, circled in pink color are areas where there are higher intensity of accidents in the night time as compared to the day time. With this information, it is worth investigating the possible reasons of such distinction. One likely explanation for having areas with higher intensity at night as compared to the day time, could be due to the reduced lighting, where the number street lights present at those areas of concern might not be sufficient, which inherently resulted in more accidents happening at that area. Another possible reason could be due to alcohol-impaired driving, where drivers are driving under influence. The authorities can look into increasing the frequency of traffic patrolling at those area, in an attempt to reduce potential drink driving from happening.

After analysing the intensity of traffic accidents, authorities can proceed to investigate if there are signs of traffic accident clusters or correlation patterns between variables such as pedestrian crossings and traffic accident points in a selected road segment. These results are drawn from Network Constrained K-Function, Network Constrained Cross K-Function and Network Constrained Cross Pair Correlation in our application. Subsequent discussions on results is focused on the city centre of Leeds, where Network Constrained Kernel Density Estimation has pointed out as having consistently high intensity of traffic accidents.

8.4 USE CASE 1 – ELDERLY CASUALTIES

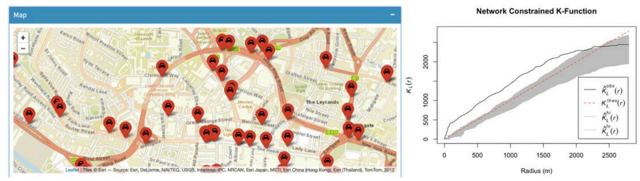


Figure 11: Network Constrained K-Function on Elderly Casualties

Network Constrained K-Function has proved that there is evidence of statistically significant clustering of traffic accidents involving all types of elderly casualties in the north of city centre for most distances along Inner Ring Road and The Headrow Road, as shown in above Figure 11.

A similar analysis was generated for accidents involving motorcyclists as shown in below Figure 12.

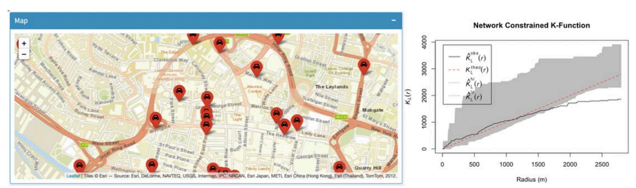


Figure 12: Network Constrained K-Function on Accidents Involving Motorcyclists

Unlike traffic accidents involving elderly casualties, traffic accidents involving motorcyclists appear to be show signs of being statistically significant dispersion, only at distances of 2 km and above along Inner Ring Road and The Headrow Road. Thus, while Network Constrained Kernel Density Map shows high intensity of accidents involving motorcyclists, it does not mean that the accidents are clustered. These analyses made were possible with the use of Monte Carlo simulation, without which the use of Kernel Density Estimation would be unable to prove whether there is a sign of statistically significant clustering or dispersion.

Network Constrained Cross K-Function and Network Constrained Cross Pair Correlation are used to determine if there is evidence of correlation between accident points and variables selected. Below analyses are generated between pedestrian crossings and elderly, as well as with accidents involving motorcycles.

8.5 USE CASE 1 – ELDERLY CASUALTIES

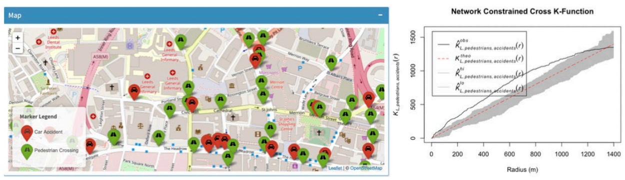


Figure 13: Network Constrained Cross K-Function Between Pedestrian Crossings and Elderly Casualties

Network Constrained Cross K-Function revealed evidence of correlation between pedestrian crossings and elderly for most of the distances along The Headrow Road and Woodhouse Lane, peaking at 800m (seen in above Figure 13). This means that accidents involving elderly tend to occur near pedestrian crossings more often than random. Authorities aiming to investigate traffic accidents involving elderly could look into this road segment, paying special attention to pedestrian crossings. The same analysis was conducted for accidents involving motorcyclists and below results (Figure 14) shows no significant correlation. Such accidents occur randomly at pedestrian crossings and it would be more relevant for the user to investigate correlation with other variables, such as motorway junctions.

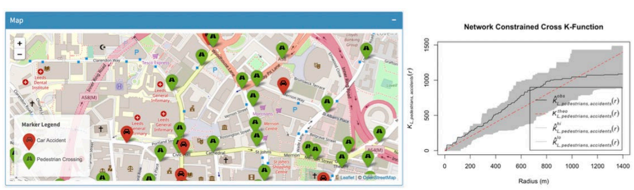


Figure 14: Network Constrained Cross K-Function Between Pedestrian Crossings and Accidents Involving Motorcyclists

While Network Constrained Cross Pair Correlation offers an alternative to computing correlation, there are instances where its results contradicts that of Network Constrained Cross K-Function. As Network Constrained Cross Pair Correlation only include points equal to the distance from pedestrian crossings, it is recommended for users to evaluate the importance of including points less than the radius from pedestrian crossings when choosing between the two functions.

8.6 USE CASE 1 – ELDERLY CASUALTIES

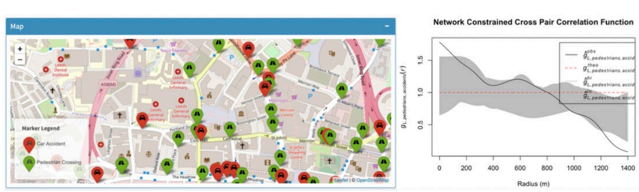


Figure 15: Network Constrained Cross Pair Correlation Between Pedestrian Crossings and Elderly Casualties

As shown above, pedestrian crossings and elderly casualties appear to correlate significantly at smaller distances (about 200m and below), while they do not correlate significantly at large distances (about 1200m and above). This differs from our findings with Network Constrained Cross K-Function, which shows that pedestrian crossings and elderly tends to correlate at larger distances. Similarly, Network Constrained Cross Pair Correlation shows that pedestrian crossings and accidents involving motorcyclists (as shown in below Figure 16) are mostly random, and do not correlate significantly at large distances (above 800m and above), which partially supports our findings from Network Constrained Cross K-Function.

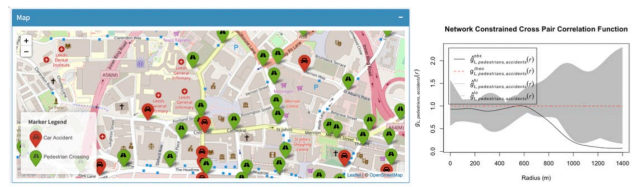


Figure 16: Network Constrained Cross Pair Correlation Between Pedestrian Crossings and Accidents Involving Motorcyclists

9 DISCUSSION

The demonstration of SIGNAL at GeoWorks, Singapore Land Authority, on 8th April 2019 has allowed our team to conduct user studies to obtain feedback. The user studies proved that the statistical analyses implemented in our application can help practitioners better understand spatial-temporal patterns and spatial relationships involving road accidents. The separation of traffic accident and casualty network-constrained kernel density estimation maps offers an alternative view on which road segments to focus on, depending on the transport authorities' target group. Different intensities revealed by both maps for the same set of filters points to different road segments to investigate on. Additionally, results from statistical simulations are made easier to interpret using the multiple-choice selection based on users' observations, thus allowing business users to be more confident in their statistical conclusions. Moreover, the methods used in SIGNAL are exploratory, which helps practitioners to uncover spatio-temporal patterns using relevant filters at a road segment level. The spatial correlations of variables such as schools could be compared to traffic accident points involving different age groups or weather conditions easily to reveal interesting insights.

10 LIMITATIONS

While SIGNAL incorporates a few key network-constrained analyses, it is currently focused on a city in England. Although the analyses may not be relevant to Singapore, the presence of certain types of roads in Leeds, such as roundabouts, is useful as we have obtained feedback that Singapore is currently considering roundabouts in its road upgrading plans. Analysing if accident clusters tend to form around roundabouts could influence

Singapore's plans in building roundabouts. Another limitation is that uploading of SHP files and CSV files by users have yet to be incorporated due to complications.

11 FUTURE WORKS

SIGNAL has the potential to be extended and further refined. The application can be adapted for use in Singapore after collecting appropriate traffic accident data. Accessibility of emergency services, such as SCDF stations, from accident-prone areas could then be calculated using Hansen's Accessibility with Network Distance, to determine if there are sufficient emergency vehicles located near accident-prone areas. Animated visualisations to show a quick overview of trends could be considered while Network Constrained Getis-Ord Gi* [2] could be used to detect which statistically significant high-risk road segments.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Kam Tin Seong, Associate Professor of Information Systems at School of Information Systems, Singapore Management University for his kind support and guidance.

REFERENCES

[1] Ivan, Haidu (2012). The spatio-temporal distribution of road accidents in Cluj-Napoca. Geographica Technica 16(2), pp. 32-38. Available at https://www.researchgate.net/publication/292514453_The_spatio-temporal_distribution_of_road_accidents_in_Cluj-Napoca.

[2] Xie, Zhixiao and Yan, Jun. (2008). Kernel Density Estimation of Traffic Accidents in a Network Space. Computers, Environment, and Urban Systems, 35 (5), 396-406.

[3] Singapore Police Force (2019). Annual Road Traffic Accident Situation 2018 and Launch of "Use Your Roadsense Campaign 2019". Available at https://www.police.gov.sg/news-and-publications/statistics?category=Road%20Traffic%20Situation.

[4] Ke, Wang, Du, Fu, Qin (2015). A Network-Constrained Integrated Method for Detecting Spatial Cluster and Risk Location of Traffic Crash: A Case Study from Wuhan, China, pp. Sustainability 2015, Vol 7, 2662-2677. Available at https://pdfs.semanticscholar.org/2f28/7df273eb01ef04c72e10564919006fde77dc.pdf.

[5] Krause, Eric. (2013). How should I interpret the output of density tools?. Retrieved from https://www.esri.com/arcgis-blog/products/product/analytics/how-should-i-interpret-the-output-of-density-tools/.

[6] Fan, Y., Zhu, X., She, B., Guo, W. & Guo, T. (2018) Network-constrained spatio-temporal clustering analysis of traffic collisions in Jianghan District of Wuhan, China. PLoS ONE 13(4): e0195093. Retrieved from https://doi.org/10.1371/journal.pone.0195093

[6] Romano, Benjamin & Jiang, Zhe. (2017). Visualizing Traffic Accident Hotspots Based on Spatial-Temporal Network Kernel Density Estimation. 1-4. Retrieved from https://dx.doi.org/10.1145/3139958.3139981