IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

CRIME HOTSPOT DETECTION AND SAFE PATH SCHEMER USING STATISTICAL AND GEO-SPATIAL METHODS

[1]Mrs. Jyothi Prasad, [2]Mohammed Aimon, [3]Rashmitha Hegde, [4]Shadman, [5]Shraddha Bhat

[1]Assitant Professor, Mangalore Institute of Technology and Engineering, Moodabidre, Karnataka, India[2] Student, Mangalore Institute of Technology and Engineering, Moodabidre, Karnataka, India [3]Student, Mangalore Institute of Technology and Engineering, Moodabidre, Karnataka, India [4]Student, Mangalore Institute of Technology and Engineering, Moodabidre, Karnataka, India [5]Student, Mangalore Institute of Technology and Engineering, Moodabidre, Karnataka, India

Abstract: The safe path schemer and crime hotspot identification methods used in this research study combine statistical and GIS techniques. The suggested methodology makes use of crime data from numerous sources, such as police reports, social media, and citizen reporting systems, to pinpoint urban crime hotspots. The statistical techniques used include regression analysis, hypothesis testing, and clustering based on density. The results of the statistical analysis are visualised using geospatial techniques like geographic information systems (GIS), which are also used to find spatial patterns and connections between crime hotspots and other urban aspects like land use and transportation systems. Additionally, a safe path schemer is created using a network analysis algorithm that makes use of the data from crime hotspots to determine the safest and most practical paths for both vehicles and pedestrians. To determine the safest paths, the algorithm takes into account a number of variables, such as crime rates, road connections, and automobile and pedestrian traffic. The suggested methodology is evaluated using actual crime statistics from a large metropolis, and the findings demonstrate that the system is successful in locating crime hotspots and offering safe route planners for metropolitan areas.

Overall, this research study makes a significant contribution to the subject of urban crime prevention and public safety and offers insights into the application of statistical and geospatial methodologies to enhance crime prevention efforts and foster safer urban settings.

Keywords - K-Means Clustering, Crime hotspot detection, Safest path, Crime count

1. INTRODUCTION

Cities around the world continue to struggle with crime, and one of the biggest obstacles to lowering crime rates is locating and addressing crime hotspots—areas with high crime rates that call for focused interventions. Researchers may now more easily detect crime hotspots and create safe path schemes that improve public safety thanks to recent advancements in statistical and geospatial methodologies.

Using a combination of statistical and geospatial tools, the project suggests a novel method for identifying crime hotspots and safe travel routes. To pinpoint crime hotspots in urban regions, our methodology draws on a variety of data sources, including police reports, social media, and citizen reporting systems. We analyse the data and locate crime hotspots using statistical methods like clustering, regression, and hypothesis testing. Following this, patterns and connections between crime hotspots and other urban elements, such as transport systems and land use, are found by using geospatial techniques like geographic information systems (GIS) to visualise the findings of the statistical study. Additionally, we create a safe path schemer utilising a network analysis algorithm that makes use of the data from crime hotspots to determine the safest and most practical paths for both vehicles and pedestrians. The algorithm determines the safest routes for various user groups by taking into account a number of variables, such as crime rates, road connections, and pedestrian and vehicle traffic. Using Google Maps API as a mapping tool, we also implement data visualization algorithms to display the results of our statistical analysis in a clear and concise manner. By utilizing data visualization techniques such as heat maps, we are able to present crime hotspots in a visually appealing way that is easily interpretable for both researchers and stakeholders.

The results show the effectiveness of our methodology in locating crime hotspots and creating safe path schemes for urban regions. The methodology is tested using real-world crime data from a major metropolis. Our research offers insights into how statistical and geospatial tools can be utilised to strengthen crime prevention efforts and adds to the ongoing endeavour to increase urban safety.

2. NEED OF THE STUDY

In many cities, crime is a huge issue with serious social repercussions. The use of data and analytics to address this issue is becoming more popular as technology advances. Geospatial and statistical techniques have become effective tools for analysing crime trends and pinpointing crime hotspots in urban regions.

In order to build a more comprehensive picture of crime patterns and trends in urban regions, web scraping is utilised in this research article to get information for data visualisation. This technique enables researchers to collect data from a variety of sources, including social media and citizen reporting systems. It is simpler for policymakers and law enforcement organisations to comprehend and apply the findings of this research to improve public safety when using a mapping tool like Google Maps API to visualise crime hotspots and safe path schemes in a user-friendly and interactive way. KMeans clustering is a crucial statistical technique used in this study to pinpoint urban crime hotspots. Researchers may identify regions with high crime rates and create more specialised crime prevention tactics using this technology, which enables the clustering of crime data based on its spatial closeness.

Additionally, the creation of a safe path schemer algorithm is essential to ensuring that the research goes beyond simply identifying crime hotspots and offers practical ways to reduce crime risks. The algorithm makes use of the crime hotspot data to determine the most effective and safest routes for vehicles and pedestrians, hence creating safer urban settings.

The main objective of this study is to advance the fields of urban public safety and crime prevention. The goal of the research is to create a technique that combines many data sources and analytical tools, such as statistical and geospatial methodologies, to provide a thorough understanding of crime patterns and trends.

3. LITERATURE REVIEW

1. SN Singh and J Kumar, "Spatial visualization approach for detecting criminal hotspots: An analysis of total cognizable crimes in the state of Haryana," 2017 International Conference on Intelligent Computing and Control (I2C2),doi: 10.1109/I2C2.2017.8226459.

In the research, a strategy is suggested for identifying crime hotspots in Haryana using spatial visualisation methods. To pinpoint high-crime locations and trends, it makes use of hotspot analysis, choropleth maps, and spatial autocorrelation analysis. The findings point out a number of hotspots, mostly in urban areas, and emphasise the need of spatial visualisation in creating crime prevention initiatives.

2. E. Galbrun, K. Pelechrinis, and E. Terzi, "Urban navigation beyond shortest route: The case of safe paths," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 1444-1453. doi: 10.1109/BigData.2017.8258074.

The study paper goes beyond the shortest route to examine the idea of safe paths in urban navigation. The authors provide a technique for creating safe paths based on information on crime, connectivity of the road network, and traffic from both vehicles and pedestrians. The suggested approach computes the safest and most practical routes for both people and cars after accounting for all relevant variables. The authors evaluate their methodology using actual data from a large city, and the results show how well the system works at recommending safe routes. The study stresses the need of taking crime data into account when building routes and offers insightful information on the creation of secure channels for urban mobility.

3. J. J. Corcoran, I. D. Wilson, and J. A. Ware, "Predicting the geo-temporal variations of crime and disorder," in 2013 IEEE International Conference on Pervasive Computing and Communications Workshops pp. 442-447. doi: 10.1109/PerComW.2013.6529511.

In order to improve urban navigation by taking the quickest path, the research article suggests a way to forecast the geotemporal changes of crime and disorder in metropolitan regions. The suggested method makes use of statistical models, geographic information systems, and historical crime data to identify crime hotspots and create safer travel options. The outcomes show that the strategy is successful in identifying crime hotspots and creating safer travel routes. The study emphasises the potential of statistical and geospatial tools to improve urban navigation and public safety beyond the quickest path.

4. F. Wang and Y. Xu, "Estimating O-D travel time matrix by Google Maps API: implementation, advantages, and implications," Journal of Transport Geography, vol. 24, pp. 142-153, 2012.

The method for predicting an origin-destination (O-D) journey time matrix using the Google Maps API is presented in this work. The suggested approach entails entering the origin and destination addresses into the API, which then determines trip times based on current traffic information and the state of the road network. The authors found that the Google Maps API does a good job of estimating travel times for both driving and taking public transport when they compare the predicted travel times with ground truth data obtained through a survey. The study explores the ramifications of using such data for transport planning and policy-making while highlighting the benefits of using the Google Maps API, including its accessibility, accuracy, and timeliness.

5. J. Wang, J. Hu, S. Shen, J. Zhuang, and S. Ni, "Crime risk analysis through big data algorithm with urban metrics," Phys. A, Stat. Mech. Appl., vol. 545, May 2020, Art. no. 123627.

A approach for crime risk assessments using big data algorithms and urban metrics is proposed in the study paper. To pinpoint crime hotspots and forecast crime risk levels in urban regions, it blends a variety of data sources, statistical approaches, and machine learning techniques. The approach uses decision trees, clustering, hypothesis testing, regression analysis, and urban and social media data analysis. The outcomes show how the suggested approach is effective at locating high-risk regions and giving details about the underlying causes of crime. The study highlights how big data and urban metrics may be used to create efficient crime prevention plans..

6. I. Ayala, L. Mandow, M. Amor, and L. Fuentes, "An evaluation of multi-objective urban tourist route planning with mobile devices," in LNCS Ubiquitous Computing and Ambient Intelligence, vol. 7656, 2012, pp. 387-394.

When recommending tourist routes, the study took into account accessibility, user preferences, distance, and time. A user poll was employed in the study to assess the app's utility and usability. The study emphasises how user preferences and a variety of objectives should be taken into account when developing applications for organising urban tourist routes. The study's methodology comprised creating a mobile application, and its methods included user research and multi-objective optimisation.

4. METHODOLOGY

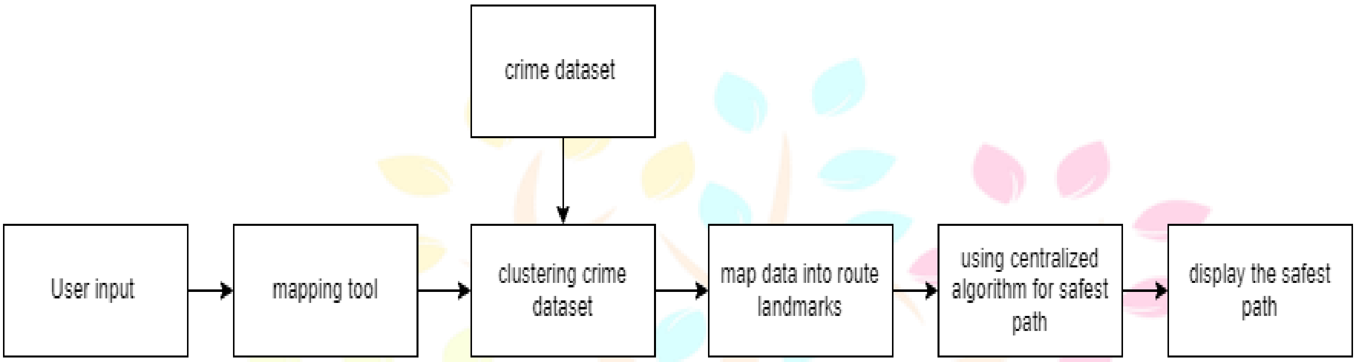

Fig 1. Methodology of safest path schemer

4.1. Data

The dataset is based upon different types of crime committed in Delhi which were extracted from different government websites The dataset consists of 14 attributes in the crime dataset, Nm_pol is the name of the area, number of people murdered, number of people raped, number of people gangraped, number of people robbed, number of theft occurred , number of assault murders happened, sexual harassment , the size of the area , total number of crimes in the region , latitude and longitude of the area , the area where the crime occured and the area where the crime happened.

Data Preprocessing

The next step is to perform Extraction, Load and Transformation of the dataset. It is significantly important to clean the data when contains null values, skewed, biased and unbalanced. The preprocessing will help us improve the end result by removing the noise in the data. This step is important because it takes around 80% of the whole analytical pipeline in a standard clustering project.

Using of Google Maps API (mapping tool)

The use of google maps api is generally used for displaying the crime prone area using its Response API and Places API which provide details about the place which includes the address of the place, duration, distance and latitude & longitude. This helps in calculating the danger Index of the area by counting the intensity of the crime. There are few standard algorithms which helps Google Map API calculate the distance between the source and destination. They are as follows:

Dijkstra algorithm

A* algorithm

4.1.1. Dijkstra algorithm

Dijkstra's algorithm is a graph search algorithm used to find the shortest path between nodes in a weighted graph. It was conceived by computer scientist Edsger W. Dijkstra in 1956 and is one of the most popular algorithms used in graph theory and network routing.

The algorithm starts by initialising the starting node with a distance of 0 and all other nodes with a distance of infinity. It then repeatedly selects the node with the smallest distance that has not yet been visited and updates the distances of its neighbours, if a shorter path is found. This process continues until the destination node is reached or all reachable nodes have been visited.

To keep track of the distances and visited nodes, Dijkstra's algorithm typically uses a priority queue or heap. The algorithm has a time complexity of $O(|E|+|V|\log|V|)$, where $|V|$ is the number of vertices in the graph and $|E|$ is the number of edges.

Dijkstra's algorithm is guaranteed to find the shortest path in a graph with non-negative edge weights, but it may not work correctly in graphs with negative weights. For such graphs, other algorithms like the Bellman-Ford algorithm or the Floyd-Warshall algorithm should be used instead.

4.1.2. A* algorithm

The A* algorithm is a heuristic search algorithm used to find the shortest path between nodes in a graph. It is an extension of Dijkstra's algorithm, which uses a heuristic function to guide the search towards the goal node more efficiently.

The algorithm works by maintaining two lists: an open list and a closed list. The open list contains nodes that still need to be explored, while the closed list contains nodes that have already been explored. The algorithm starts by adding the starting node to the open list with a heuristic value equal to the estimated distance to the goal node. The heuristic function is used to estimate the remaining distance to the goal node, and it is often an admissible heuristic that underestimates the true distance.At each iteration, the algorithm selects the node with the lowest f-value (the sum of the distance from the starting node and the estimated remaining distance to the goal node) from the open list and expands its neighbors. For each neighbor, the algorithm calculates the new g-value (the actual distance from the starting node to the neighbor) and the new f-value, and it adds the neighbor to the open list if it is not already on the closed list or the open list with a lower f-value.

The algorithm continues until the goal node is reached or the open list is empty. If the goal node is reached, the algorithm returns the shortest path. Otherwise, the algorithm fails to find a path.

The A* algorithm is often faster than Dijkstra's algorithm because it uses the heuristic function to prioritize the search towards the goal node. However, the quality of the heuristic function greatly affects the efficiency of the algorithm. If the heuristic is not admissible, the algorithm may not find the optimal path. If the heuristic is consistent (satisfies the triangle inequality), the algorithm is guaranteed to find the optimal path. The time complexity of the A* algorithm is O(b^d), where b is the branching factor (the maximum number of successors of any node) and d is the depth of the solution.

4.2. Using K-means algorithm for clustering

K-means clustering algorithm can be used in the prediction of crime dataset to identify patterns and groupings within the data that can help in the identification of crime hotspots. Crime data can be grouped into clusters based on their similarities and differences, and this can be used to predict future crime occurrences in specific locations.

Here are the steps in using the K-means clustering algorithm for crime prediction:

1. Data preparation: The first step is to prepare the crime data for clustering. This involves collecting the data and cleaning it to remove any missing values or errors that could affect the analysis.

2. Feature selection: The next step is to select the relevant features that will be used in the clustering analysis. Features such as location, time of day, type of crime, and demographics of the population in the area can be used as input variables.

3. Data normalization: It is important to normalize the data to ensure that all input variables are on the same scale. This helps to ensure that the clustering algorithm gives equal weight to all features.

4. K-means clustering: The next step is to apply the K-means clustering algorithm to the crime data. This involves selecting the number of clusters (K) to be used in the analysis and running the algorithm to group the data into K clusters.

5. Cluster analysis: The final step is to analyze the clusters that have been identified. This involves examining the characteristics of each cluster to identify crime hotspots and potential risk factors. It also helps to identify areas that may require additional resources for crime prevention.

In summary, K-means clustering algorithm can be used in the prediction of crime dataset to identify patterns and groupings within the data that can help in the identification of crime hotspots and risk factors. This can be useful in the development of crime prevention strategies and resource allocation for law enforcement agencies.

4.3. Using Deterministic ranking Algorithm for sorting the safest path

In the context of a crime prediction dataset, a deterministic ranking algorithm could be used to assign a danger index to each location or neighbourhood based on the likelihood of crime occurrence. The danger index would range from 0 to 4, with 0 indicating a low likelihood of crime and 4 indicating a high likelihood of crime.

The algorithm could use a set of predetermined criteria to assign the scores based on various factors that are known to affect crime rates. For example, the algorithm could take into account variables such as past crime rates, population density, poverty levels, unemployment rates, and proximity to high crime areas.

The algorithm could use statistical modeling techniques to calculate the danger index based on the chosen variables. For example, it could use logistic regression to predict the likelihood of crime occurrence in a given area based on historical data and other factors. The use of a deterministic ranking algorithm ensures that the danger index assigned to each location will be consistent and reproducible for every run of the algorithm. This can be useful for making comparisons between locations or tracking changes in danger levels over time.

However, it is important to note that the criteria used by the algorithm to assign the danger index may not capture all the factors that contribute to crime. Moreover, the algorithm may be biased if the variables used are not representative of the population. Therefore, it is important to carefully consider the variables used by the algorithm and to regularly review and update them as necessary to ensure the accuracy and fairness of the danger index.

5. RESULTS AND DISCUSSION

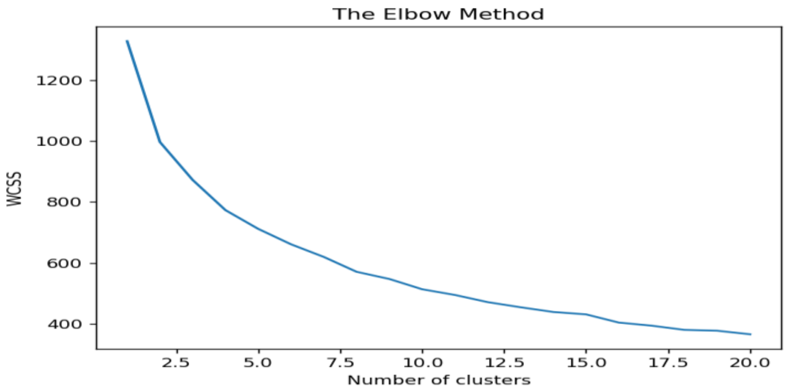

Fig 2. The Elbow method

The elbow method is a technique used to determine the optimal number of clusters in the K-means clustering algorithm. The method involves plotting the within-cluster sum of squares (WCSS) against the number of clusters, and selecting the number of clusters at the "elbow" point of the curve.

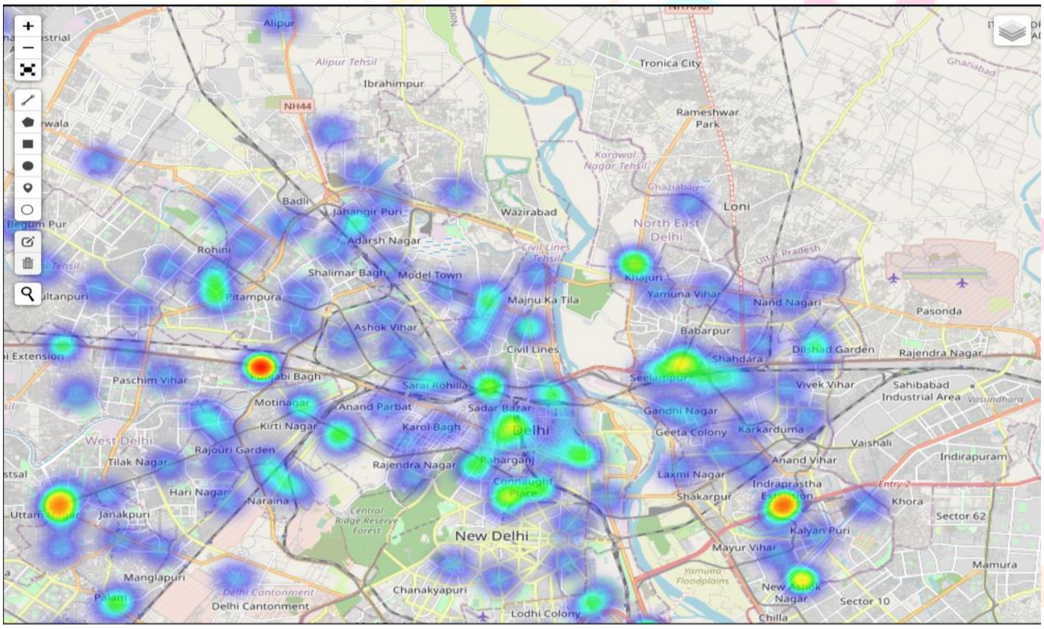

Fig 3. Heatmap showing the crime prone area

A heatmap is depicted according to the intensity of the crime in the area using the leaflet javascript library.

Fig 4. Safest path from a source to destination

Using the google maps API we find the number of ways to travel from source to destination. Here we can see the safest path from source to destination using the danger index.

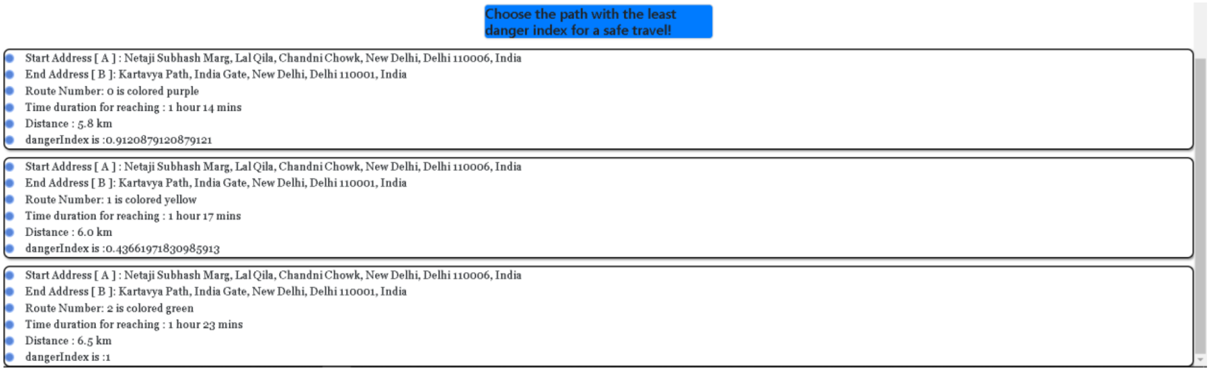

Fig 5. Finding the safest path by comparing the danger indices

As there are several paths contributing from source to destination , here the user has to compare and find the safest path according to the least amount of danger index.

6. CONCLUSION

The research study proposes a strategy for identifying crime hotspots and devising safe travel routes using statistical and geospatial techniques. The 14 variables in the dataset that the authors used were connected to various sorts of crimes in Delhi, India. They performed data preprocessing, mapped crime-prone locations using the Google Maps API, and used the Dijkstra and A* algorithms to determine safe routes. In order to anticipate future crime occurrences in particular regions, they also employed the K-means clustering method to find patterns and groupings within the crime data. The findings demonstrated that the suggested strategy was successful in identifying crime hotspots and locating secure routes. The K-means clustering technique helped with future crime prediction by successfully identifying trends in the crime data. Dijkstra and A* algorithms were used to aid people travelling through high-risk zones locate the quickest and safest route.

A deterministic ranking algorithm can be utilised for crime hotspot detection and safe path schemer in addition to the approaches stated above.. The ranking can be used to identify high-crime regions and give them priority for interventional efforts. The implementation of deterministic ranking algorithms is straightforward and might serve as a good foundation for pinpointing crime hotspots. However, if the underlying data is skewed or lacking, they may not account for all pertinent factors, such as the demographics of the area or the efficacy of current interventions, and may produce biassed results.

Overall, crime hotspot detection and safe path schemer can be accomplished using a combination of statistical and geospatial techniques, including clustering algorithms, mapping applications, and deterministic ranking algorithms. These techniques can assist police departments and city planners in locating high-risk regions and developing practical solutions to lower crime rates and enhance public safety. The above strategy can be made even better by including further components like weather, lighting, and population density to improve the accuracy of crime predictions.

7. REFERENCES

[1] SN Singh and J Kumar, "Spatial visualization approach for detecting criminal hotspots: An analysis of total cognizable crimes in the state of Haryana," 2017 International Conference on Intelligent Computing and Control (I2C2),doi: 10.1109/I2C2.2017.8226459.

[2] E. Galbrun, K. Pelechrinis, and E. Terzi, "Urban navigation beyond shortest route: The case of safe paths," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 1444-1453. doi: 10.1109/BigData.2017.8258074.

[3] J. J. Corcoran, I. D. Wilson, and J. A. Ware, "Predicting the geo-temporal variations of crime and disorder," in 2013 IEEE International Conference on Pervasive Computing and Communications Workshops pp. 442-447. doi: 10.1109/PerComW.2013.6529511.

[4] F. Wang and Y. Xu, "Estimating O-D travel time matrix by Google Maps API: implementation, advantages, and implications," Journal of Transport Geography, vol. 24, pp. 142-153, 2012.

[5] J. Wang, J. Hu, S. Shen, J. Zhuang, and S. Ni, "Crime risk analysis through big data algorithm with urban metrics," Phys. A, Stat. Mech. Appl., vol. 545, May 2020, Art. no. 123627.

[6] Ayala, L. Mandow, M. Amor, and L. Fuentes, "An evaluation of multi-objective urban tourist route planning with mobile devices," in LNCS Ubiquitous Computing and Ambient Intelligence, vol. 7656, 2012, pp. 387-394.