

## Algorithm for Avoiding Street Harassment by Using SecureRoutes

Mariana García  
Universidad Eafit  
Colombia  
mgarciaz2@eafit.edu.co

Isabella Márquez  
Universidad Eafit  
Colombia  
imarquezc@eafit.edu.co

Andrea Serna  
Universidad Eafit  
Colombia  
asernac1@eafit.edu.co

Mauricio Toro  
Universidad Eafit  
Colombia  
mtorobe@eafit.edu.co

María Camila Raigosa  
Universidad Eafit  
Colombia  
Mcraigosas@eafit.edu.co

### SUMMARY

Street harassment consists in unwanted comments, gestures, and actions forced on a stranger (mostly females) in a public place without their consent. This gender-based problem is important since everyone deserves to be treated with dignity and respect and to feel safe in public spaces regardless of gender. Many external problems that affect the victims arise, such as making the person worried about its physical safety and creating an environment of fear and intimidation from the abuser, causing the person to be afraid that when she goes out on the street this could happen again which makes it harder to do things in everyday life.

### KEY WORDS

shortest path, street harassment, identification, secure routes, prevention, crime

### 1. INTRODUCTION

The risk of street harassment is extremely high, especially for women. Paths with less distance that reduce sexual harassment are part of the solution to eradicate this problem, we need them since every day someone is affected either in a small or large magnitude. We demand that this does not continue to grow, to make this reality visible as a problem that requires an urgent solution.

#### 1.1. Problem

The impact of this problem on society consists of several points. Violence can be related to street sexual harassment since both come from people with whom they have never met. Another point of street sexual harassment is the harm that the aggressors cause, which can be physical or sexual, although it also involves verbal harassment. This occurs mostly in women because of all the stigmatization they have in society in relation to men. For this reason, the

latter seek to highlight this superiority that they have been given over women by exercising this type of violence in public spaces.

### 1.3 Structure of the Article

Following in section 2, we will present the related works with the problem presented on section 1.1. Subsequently, on section 3, we will show the gathering of data and methods used on the investigation. On section 4 exhibit the design of our algorithm. On section 5 there will be the display of the results. Finally on section 6, we will discuss said results and propose some directions to take on future works.

### 2. RELATED WORKS

Now, we will analyze and describe four related works with the search of pathways to prevent street harassment and delinquency in general.

#### 2.1 A data integration and data analysis system for route planning

The way to solve the problem of street harassment and robberies presented in the text is through the design of a safe route application that can be used to determine the safest and fastest route rather than just the fastest route.

The algorithm they used was to find out the conditions in which cases are likely to occur: clustering locations and then avoiding clusters of potentially dangerous people to find out how safe the location is, taking into account the harassment rates of the area with the safety value of the region, taking into account the conditions that women would feel unsafe with to compare this in a database.

To determine the safety of a location, factors such as the people in the area, time of day, lighting, monetary situation, cameras in the area must be taken into account,

etc. But as you can see it is impossible to identify the correspondence of the factors. Consequent to this it is necessary to define a safety function, which will take inputs based on the above mentioned factors to combine them and determine the relative safety of the region in relation to the expected location (that the probabilities of a case is minimal) and make route plannings more practical by changing the conditions according to the safety value. Based on this it was seen as a result that if only based on safety the route would be too long as the nodes would be far away from each other. but in some parts there is no good lighting or few people where the nodes are not so far away but without safety, so it would give the step to the combination of distance and safety by the maximum influence that safety can have on the cost of the route. So users will be provided with a set of questions to choose a route, and the decision making algorithm will analyze the answers and decide which route is the best for the user.

## **2.2 Preventing Sexual Harassment Through a Path Finding Algorithm Using Local Searches**

The problem that was solved within this project was, mainly, to have the ability to effectively predict the places that contain a high risk content around sexual harassment issues through "hot spots and "safe zones", so that the user can identify both the areas to move around and the areas to avoid.

The result obtained from the research was an IA implemented within a "heatmap" grid (which was mentioned above with its "hot spots" and "safe zones") with a 0-4 system, with 0 being the safest zone and 4 the most dangerous, having a distance from point to point of approximately 1.74 km<sup>2</sup>.

Thanks to the question "How to determine the safety of a route?" which the researchers ask, within their solution, 3 conclusions are reached, "determine in general the safety of a certain route only using the values associated with the destination" "calculate the average risk factors using the grid in a straight line to the destination" and finally "calculate the average risk factors using the grid step by step on the route to the destination". These last two involve a factor not previously considered, the line of sight, so their solution for the obstacles was the algorithm called "Bresenham's line".

segment of the grid, the next point within the grid determines whether it can be reached by evaluating where the segment line falls closest to the midpoint, up or down. And with the above explained, that "probability of safety" is applied within the explained heatmap

## **2.3 Safe routes for motorized tourists based on open data and VGI**

The problem presented is to avoid dangerous areas of a city, where through routing these can be avoided, taking into account the safety of tourists as some locals avoid them and only know them locally. The importance of crime in the means of transport is highlighted, and we want to detect their trajectories within a navigation solution that prevents them from entering dangerous areas.

Consistent with this, they included the use of freely available information in the form of open data and VGI, to provide local knowledge about places and be a source of information for tourists. Governmental open data such as local police was also included.

On the other hand, by obtaining data they wanted to detect crime hotspots using clustering techniques and open access information on the road network and in synthesis, combine open access data, Their test data were: lighting, police stations, roads and crime reports in the last month. With the above, a security-aware routing application was designed, taking into account costs, routing obstacles, and the tests that were done with the data.

## **2.4 Safe Route Model**

What this study seeks to address is the problems of sexual harassment of women identified in recent years, with the understanding that even with the help of widely implemented tools such as Google Maps or Waze, they do not have the capacity to identify dangerous or high-risk areas. This study is specifically focused on women tourists and New York City, understanding that the implementation of a tool which provides safety, especially in a metropolitan city so crowded with foreigners, is a concept that needs to be implemented as soon as possible.

Its solution focuses mainly on 3 aspects: First, a secure route is sought with the help of the data.

criminalistics of each city while using the shortest distance between the departure and arrival point; second, divide New York into smaller regions taking into account their hazard level; and third, calculate the risk value of the routes based on the values of nearby clusters.

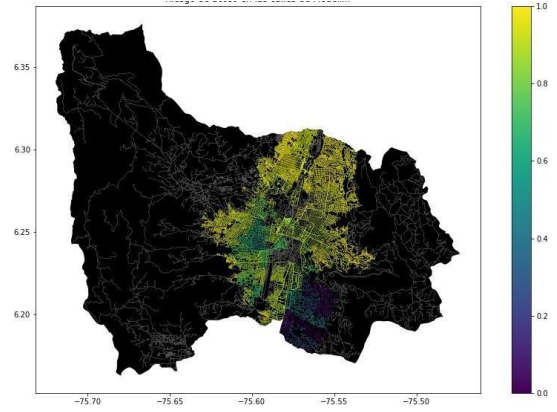
The following algorithm is used in this project; the crime and accident dataset was extracted directly from the NYPD databases, leaving out the columns that were not useful for the research, giving them the value of null, as well as renaming the columns to make them more understandable. The dataframes they created consisted of a number of clusters, latitude and longitude of the clusters and of "C" and "A" (crime dataset and accident dataset respectively). The last part consisted in using Google Maps to find the indicated roads by creating several commands within the application's directions function. Therefore the algorithm is divided into 3 parts; Data processor, "K-mean" (latitude and longitude) and finally the "K nearest neighbor".

### 3. MATERIALS AND METHODS

In this section, we explain how the data were collected and processed, and then different alternative pathway algorithms that reduce both the distance and the risk of street sexual harassment.

#### 3.1 Data collection and processing

The map of Medellín was obtained from *Open Street Maps* (OSM)<sup>1</sup> and downloaded using the Python API<sup>2</sup> OSMnx. The map includes (1) the length of each segment, in meters; (2) the indication of whether the segment is one-way or not, and (3) the known binary representations of the geometries obtained from the metadata provided by OSM. For this project, a linear combination (LC) was calculated that captures the maximum variance between (i) the fraction of households that feel unsafe and (ii) the fraction of households with income below one minimum wage. These data were obtained from the 2017 Medellín quality of life survey. The CL was normalized, using the maximum and minimum, to obtain values between 0 and 1. The CL was obtained using principal component analysis. The risk of harassment is defined as one minus the normalized CL. Figure 1 presents the calculated risk of harassment. The map is available on GitHub<sup>3</sup>.



**Figure 1.** Risk of sexual harassment calculated as a linear combination of the fraction of households that feel unsafe and the fraction of households with income below one minimum wage, obtained from the 2017 Medellín Quality of Life Survey.

#### 3.2 Alternatives of routes that reduce the risk of street harassment

Now, we will present different algorithms used to find a pathway that will reduce street harassment as well as the distances from point A to point B.

According to Dijkstra, an initial network F is constructed consisting only of the origin node. It must be ensured that all nodes that can be reached directly from F are added.

You must stop when the vertex to be included in F is the destination node. At each iteration, when you include a node to F you do so using the smallest path from the origin to the node. If another path exists, it must go through one of the other nodes that we can reach directly from F, which are greater or equal distance from the origin than node v, by construction. Since the weights are non-negative, any path that passes through any of these nodes to reach v cannot have a shorter distance than the one we were initially considering.

According to Floyd-Warshall, we must calculate the minimum distance between all pairs of vertices of a weighted network. It can be done in  $O(|V| \cdot (|E| + |V| \cdot \log(|V|)))$  by performing a Dijkstra for each node. But if the weights are negative, this method cannot be used, it must be done in  $O(|E| \cdot |V|^2)$  by performing n times the Bellman-Ford algorithm, one for each vertex. We perform n iterations; in each iteration we consider the optimal paths that have as intermediate nodes those with index less than i. So, after n

<sup>1</sup> <https://www.openstreetmap.org/>

<sup>2</sup> <https://osmnx.readthedocs.io/>

iterations we consider the optimal paths that include any node, i.e., the optimal paths in general. There is another matrix (dist) of n rows and n columns such that the minimum distance

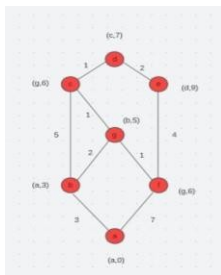
<sup>3</sup> <https://github.com/mauriciotoro/ST0245Eafit/tree/master/project/Datasets/>

from node i to j is stored in the box `dist[i][j]`. Initially the value of all distances is infinite, except for those where there is a direct arc from i to j, in which case the value of `dist[i][j]` will be the weight of the arc. Also, by convention

$\text{dist}[i][i] = 0$ . N iterations are performed, in each of which we will consider all pairs of nodes, so that at the k-th iteration if passing through vertex k on the path from i to j is optimal, we will update the value of the distance from i to j.

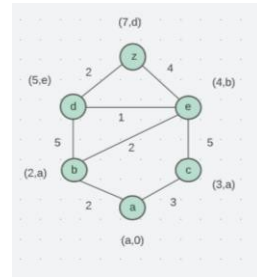
### 3.2.1 First Algorithm

In this algorithm, the circles with letters represent the nodes and lines the beard which represent the connections that exist (what they can run) and the distance is represented by the numbers that are above them. We will find the nearest route by starting from "a" which will be the initial node, by the parenthesis we will put the node from the left and on the right, the accumulated distance from the node of the past node to the next. It begins from "a" to 0 then it will move to either of the next beards having into account the distance. In this case it moved until b, so we are left with "a,3" (we will need to keep adding). We need to do the same on the other side and it runs the one which is less heavy (b in this case) and it follows this pattern until the end.



### 3.2.2 Second Algorithm

In this algorithm happens the same as the last one, only that the values we found on the parenthesis are crossed out in terms of the last one, but this will not affect the result. It is very similar to the last one mentioned, and it would be the same pattern, starting from "a" due to it being the initial node.

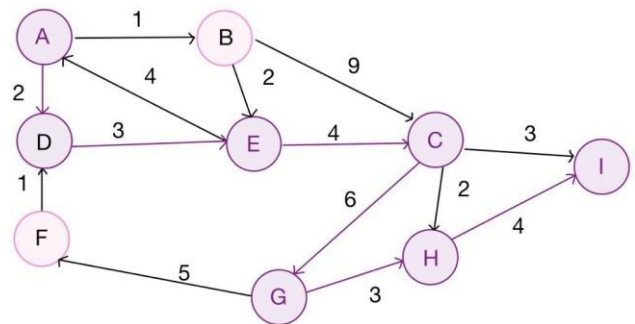


### 3.2.3 Third

#### Algorithm Dijkstra's

##### Algorithm

It tries to explore all the shortest paths that start from an origin vertex and goes through all the other vertices. When the shortest path from the origin vertex to the rest of the vertices that make up the network is obtained, the algorithm stops.



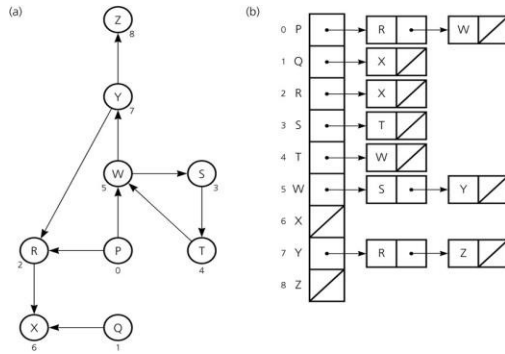
### Fourth Algorithm

#### 4. ALGORITHM DESIGN AND IMPLEMENTATION

In the following, we explain the data structures and algorithms used in this work. The implementations of the data structures and algorithms are available on Github1.

##### 4.1 Data structures

The data structure used to represent the map of the city of Medellin was the adjacent list, which by means of two arrangements, the name of each street was saved without repeating and the origin of the destination and its weight were stored, leaving at the end a single value that would be the best route.

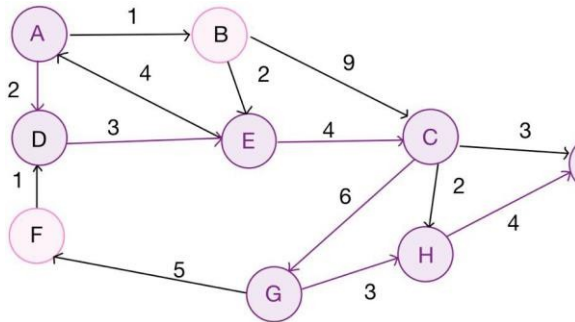


## 4.2 Algorithms

In this paper, we propose an algorithm for a path that minimizes both the distance and the risk of street sexual harassment.

### 4.2.1 Dijkstra's algorithm for finding a path that reduces both the distance and the risk of street sexual harassment

The design of the algorithm for calculating a path that reduces both the distance and the risk of harassment is that the algorithm finds the shortest distances from a point of origin to the other possible destinations in a network as an adjacency list. It must be emphasized that the routes must be safe in order to give the correct directions.



## BIBLIOGRAPHY .

Shankar, V. and Sandeep, C., 2019. *Route-The Safe: A Robust Model for Safest Route Prediction Using Crime and Accidental Data*. [online] ResearchGate. Available at: <[https://www.researchgate.net/publication/338096313\\_Route-The\\_Safe\\_A\\_Robust\\_Model\\_for\\_Safest\\_Route\\_Prediction\\_Using\\_Crime\\_and\\_Accidental\\_Data](https://www.researchgate.net/publication/338096313_Route-The_Safe_A_Robust_Model_for_Safest_Route_Prediction_Using_Crime_and_Accidental_Data)> [Accessed 21 August 2022].

Ma, D., 2022. *Preventing Sexual Harassment Through a Path Finding Algorithm Using Nearby Search*. [online] Omdena | Building AI Solutions for Real-World Problems. Available at: <<https://omdena.com/blog/path-finding-algorithm/>>

Keler, A. and Mazimpaka, J., 2016. *Safety-aware routing for motorised tourists based on open data and VGI*. [online]. ResearchGate. Available at: <[https://www.researchgate.net/publication/301566519\\_Safety-aware\\_routing\\_for\\_motorised\\_tourists\\_based\\_on\\_open\\_data\\_and\\_VGI](https://www.researchgate.net/publication/301566519_Safety-aware_routing_for_motorised_tourists_based_on_open_data_and_VGI)>

Sarraf, R., 2016. *A Data Integration and Analysis System for Safe Route Planning*. [online] ResearchGate. Available at: <[https://www.researchgate.net/publication/327390893\\_A\\_Data\\_Integration\\_and\\_Analysis\\_System\\_for\\_Safe\\_Route\\_Planning](https://www.researchgate.net/publication/327390893_A_Data_Integration_and_Analysis_System_for_Safe_Route_Planning)>

Aprende.olimpiada-informatica.org. 2019. *Shortest path between nodes*. [online] Available at:

<<https://nodo.ugto.mx/wp-content/uploads/2018/08/Dijkstra>>  
<<https://aprende.olimpiada-informatica.org/algoritmia>>

- dijkstra-bellman-ford-floyd-warshall

