## Algorithm for Avoiding Street Harassment by Using Secure Routes

Mariana García Universidad Eafit Colombia [mgarciaz2@eafit.edu.co](mailto:mgarciaz2@eafit.edu.co)

María Camila Raigosa Universidad Eafit Colombia [Mcraigosas@eafit.edu.co](mailto:Mcraigosas@eafit.edu.co)

Isabella Márquez Universidad Eafit Colombia [imarquezc@eafit.edu.co](mailto:imarquezc@eafit.edu.co)

Andrea Serna Universidad Eafit Colombia [asernac1@eafit.edu.co](mailto:asernac1@eafit.edu.co)

Mauricio Toro Universidad Eafit Colombia [mtorobe@eafit.edu.co](mailto:mtorobe@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.

The proposed algorithm is Dijkstra shortest path with custom costs, it is convenient for quickly testing the shortest path when the initial given conditions are changed for the same graph, it is able to get 3 shortest paths from Eafit university to National University with different cost configurations. First path was calculated in 42 seconds with a total distance of 2593 meters with a risk index of 26.67 Second path was calculated in 37 seconds with a total distance of 6419 meters with a risk index of 59.07. Third path was calculated in 35 seconds with a total distance of 2577 meters with a risk index of 29.37

## KEY WORDS

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

## 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.

## 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.

## Solution

Our solution to the problem is based on reducing the distance as the risk of harassment for pedestrians, the algorithm we use is the Dijkstra’s algorithm. We chose it because at the time of its implementation it was more efficient and better than the others to find the shortest route between two points on a map, it goes through the entire map (in this case we use the map of Medellin), where it makes sure to go to all its vertices, just try not to go through the busiest ones until there is none left untraveled. This will be better evidenced in the development of the following sections.

## 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.

## RELATED WORKS

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

## 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, considering the harassment rates of the area with the safety value of the region, considering 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 probability 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.

## 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 km².

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 "Bradenham’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

## 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, considering costs, routing obstacles, and the tests that were done with the data.

## 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 considering 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 were 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 several 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 direction’s function. Therefore, the algorithm is divided into 3 parts: Data processor, "K-mean" (latitude and longitude) and finally the "K nearest neighbor".

## 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.

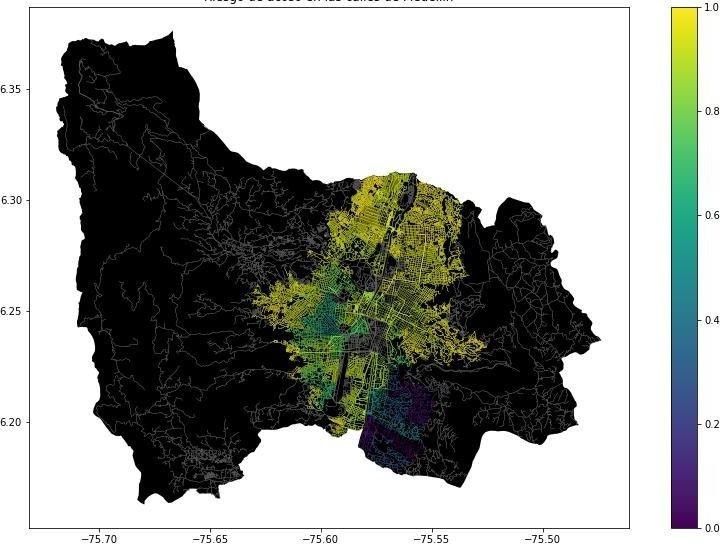
## Data collection and processing

The map of Medellín was obtained from *Open Street Maps* (OSM)1 and downloaded using the Python API2 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 Medellin 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 GitHub3 .

1 https://[www.openstreetmap.org/](http://www.openstreetmap.org/)

2 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



**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.

## 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|⋅(|E|+|V|⋅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|⋅|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

3https://github.com/mauriciotoro/ST0245Eafit/tree/master/

project/Datasets/

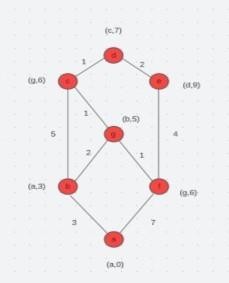
from node i to j is stored in the box dist[i][j]. Initially the value of all distances is infinite, except for those where there is a direct arc from i to j, in which case the value of dist[i][j] will be

the weight of the arc. Also, by conventio

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 ifpassing 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.

Gráfico, Gráfico de dispersión

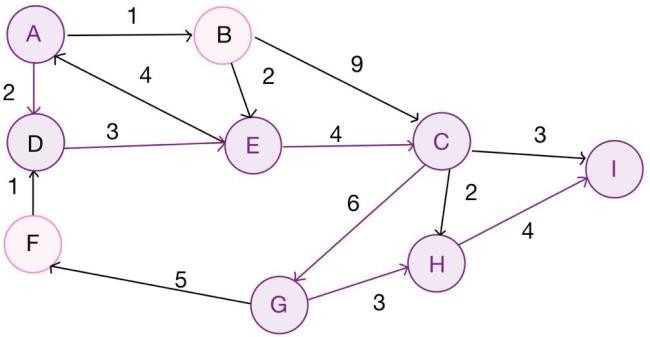
Descripción generada automáticamente

# 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 makeup the network is obtained, the algorithm stops.

Complexity o²



# 3.2.4 Fourth Algorithm

# Our last algorithm follows the patterns as well from the previous ones, starting as from the node “a” but with different values for the vertices. At some point, it can be found that two nodes have the same value, so the solution for it is picking up randomly one of the nodes previously mentioned.

# .Imagen que contiene mapa, texto, reloj, tabla Descripción generada automáticamente

.

ALGORITHM DESIGN ANDIMPLEMENTATION

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. 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. 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 GitHub 4 .

**4.1 Data Structures**

Explain the data structure that was used to represent the

map of the city of Medellín. Make a figure that explains it.

Do not use figures from the Internet. (In this semester,

examples of data structures are adjacency matrix,

adjacency list, adjacency list using a dictionary). The data

structure is presented in Figure 2. Diagrama, Dibujo de ingeniería

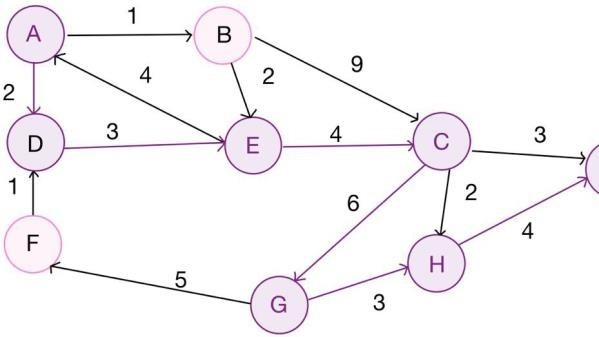
Descripción generada automáticamente

## Algorithms

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

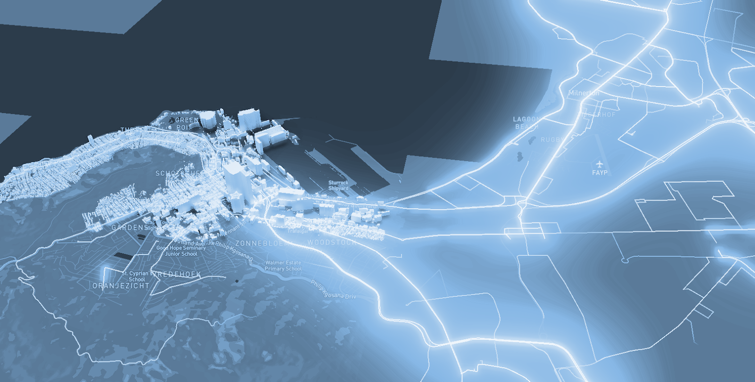
# Dijkstra's algorithm for finding a path that reduces both the distance and the risk of street sexual harassment

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.



**4.2.2 Calculation of two other paths to reduce both the distance and the risk of sexual street harassment**

The algorithm is exemplified in Figure 4.

****

**Figure 4:** Map of the city of Medellín showing three pedestrian paths that reduce both the risk of sexual harassment and the distance in meters between the EAFIT University and the National University.

**4.3 Algorithm complexity analysis**

If V is the number of nodes

In the worst case, all nodes are interconnected, the algorithm will go through all the nodes, which is O(V), and for each of those nodes, compute the minimum distance, which in the worst-case means going through all the nodes again, this will happen in O(V) time.

According to this:

O(V\*V) = O(V2)

|  |  |
| --- | --- |
| **Algorithm** | **Time complexity** |
| Dijkstra | O(V2) |
|  |

**For the algorithm, we must create a list of distance for every node to the initial node D**

|  |  |
| --- | --- |
| **Data Structure** | **Complexity of memory** |
| List | O(V) |

**4.4 Algorithm design criteria**

We are using a global Axes pyplot object in which we are drawing the map of Medellin, and the paths we are producing, for better readability of the code.

To represent the graph, we are using a dictionary of Vertex objects and each Vertex object has a list of its neighbors, so vertex access is O(1) which is more efficient.**5. RESULTS**

In this section, we present some quantitative results on the three pathways that reduce both the distance and the risk of sexual street harassment.

**5.1 Results of the paths that reduces both distance and risk of sexual street harassment**

Next, we present the results obtained from *three paths that reduce both distance and harassment,* in Table 3.

|  |  |  |  |
| --- | --- | --- | --- |
| **Origin** | **Destination** | **Distance** | **Risk** |
| Eafit | Unal | 2577 m | 29 |
| Eafit | Unal | 6419 m | 59 |
| Eafit | Unal | 2579 m | 28 |

Distance in meters and risk of sexual street harassment (between 0 and 1) to walk from EAFIT University to the National University.

**5.2 Algorithm execution times**

In Table 4, we explain the ratio of the average execution times of the queries presented in Table 3.

## 

|  |  |
| --- | --- |
| **Calculation of v** | **Average run times (s)** |
| v = r+d | 49 s |
| v = r^d | 39 s |
| v = r \* 3267 + d \* 1234 = | 45 s |

## Table 4: *execution times of Dijkstra algorithm with modifiable weight for each one of the three ways between EAFIT and Universidad nacional*

## 6.  CONCLUSIONS

The results can be interpreted as several ways one person can go from point a to point b in case, he doesn’t want to go through one of the paths, as expected, the first and third paths are nearly identical, but different from the second path. We would recommend the third path become is has less risk and only 2 meters longer than the first one. I think the execution times for Dijkstra algorithm are optimal because with enough time, we can calculate a shortest path for every pair of nodes, so for a user in a web or mobile app, we would have to give him the already calculated list of vertices that represent the path and graph them in a map.

**6.1 Future work**

In the future, we would like to go a step further with the algorithm and see it implemented on a real app, specifically on mobile devices. If this app gets successful, we expect that the statistics for kidnappings in Colombia become reduced exponentially since this crime, especially for women, has seen a rapid increase over the years. We would also like to implement machine learning into our project to get more accurate information and even predict certain dangers that may appear on the streets

# ACKNOWLEDGEMENTS

This research has been supported by University EAFIT.

We want to give our greatest gratitude to Students Juan and Diego Mesa, Universidad EAFIT, for helping us through the coding and execution of our research. We would also like to thank our friends who gave a constructive opinion on the project in order to achieve better results and our mascots that accompanied us all this time.

.

The authors thank Professor Juan Carlos Duque, Universidad EAFIT, for providing the data from the 2017 Medellín Quality of Life Survey, processed in a *Shapefile*.

# BIBLIOGRAPHY .

Shankar, V. and Sandeep, C., 2019. *Route- The Safe: ARobust Model for Safest Route Prediction Using Crime andAccidental Data*. [online] ResearchGate. Available at:

<https://[www.researchgate.net/publication/33](http://www.researchgate.net/publication/33)8096313\_Route- The\_Safe\_A\_Robust\_Model\_for\_Safest\_Rout e\_Prediction\_ Using\_Crime\_and\_Accidental\_Data> [Accessed 21 August

2022].

Ma, D., 2022. *Preventing Sexual Harassment Through aPath 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 routingfor motorised tourists based on open data and VGI*. [online].

ResearchGate. Available at:

<https://[www.researchgate.net/publication/30](http://www.researchgate.net/publication/30) 1566519\_Safe

ty- aware\_routing\_for\_motorised\_tourists\_based

\_on\_open\_dat a\_and\_VGI>.

Sarraf, R., 2016. *A Data Integration and Analysis Systemfor Safe Route Planning*. [online] ResearchGate. Available

at:

<https://[www.researchgate.net/publication/32](http://www.researchgate.net/publication/32) 7390893\_A\_

Data\_Integration\_and\_Analysis\_System\_for\_ Safe\_Route\_P

lanning>

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/algorithmia

-

dijkstra-bellman-ford-floyd-warshal