**Street harassment, a problem to solve in every**


Miguel Vásquez Bojanini
Universidad Eafit
Colombia
mvasquezb@eafit.com

Juan Alberto Rodríguez
Universidad Eafit
Colombia
jarodriguo@eafit.edu.co

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

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


**For each version of this report: 1. Delete all text in red. 2. 2. Adjust the spaces between words and paragraphs. 3. Change the color of all text to black.**

**Red text =** Comments

**Text in black =** Contribution of Andrea and Mauricio

**Text in green** = To complete the first delivery

**Blue text** = To be completed for the 2nd deliverable

**Text in purple** = To be completed for the third deliverable


**ABSTRACT**

The problem is based on street harassment that anyone can constantly suffer in several ways; Attitudes such as whistles, sexist insults, and sexual actions are only a few examples of what people live every day. The truth is that women are the most affected ones in Medellin, considering it an unsafe city for them to travel alone based on danger perceptions or facts about violence. This problem is relevant because safety is a right that we all have regardless of our genre, we can consider verbal violence as all the comments that aren´t appropriate and cause discomfort.

After all the work we have put into this project, we can confirm that it is a very good way to improve mobility and safety in Medellin, the results are good and the implementation of the algorithm is very flexible, making it viable for other projects that contain a larger range of data. Also the use of Google's services is very convenient at the time of showing the routes in the map, making it easy and fast.

In terms of Quantitative results, we ended up with tree different routes taking in matter different forms of comparing both, sexual harassment, and length. Each, with its execution time and important data to use in this technical report.



**Key words**

Shortest route, street sexual harassment, identification of safe routes, crime prevention


## 1. INTRODUCTION

One of the hardest problems we as a community face is sexual harassment in public spaces, specifically in the most concurrent streets.

So, with this project, we are looking to improve our convivence, and most importantly, reduce the insecurity in the streets by creating safer routes for people to travel alone. After that, we can realize the degree of importance of talking about the problem.

### 1.1. The problem

Creating safer routes for people in a city is beneficial in many ways; it can make more tourists interested in visiting the city and investing in their local businesses, and it can make people who live around the area happier. Therefore, creating an algorithm that can help people feel safe on the streets will not only help the community but also improve the country's economy.

### 1.2 Solution

The solution that we are going to implement is Dijkstra´s algorithm, which determines the shortest path, given a source vertex, to the other nodes in a graph that has weights on each edge. In this case the weights are determined by the safety and length of the route.

We choose it because the algorithm has a complexity of O(n2), which makes it an efficient way to find distances, and that´s what we need to solve the main problem of this project.

### 1.3 Structure of the article

Next, in Section 2, we present work related to the problem. Then, in Section 3, we present the datasets and methods used in this research. In Section 4, we present the algorithm design. Then, in Section 5, we present the results. Finally, in Section 6, we discuss the results and propose some directions for future work.

## 2. RELATED WORK

Below, we explain four works related to finding ways to prevent street sexual harassment and crime in general.

1. **The path-finding algorithm to prevent sexual harassment**

The problem that was solved is based on identifying "safe spots" according to the user´s coordinates and directions generating routes to the nearest hospitals within 800 meters by nearby search, also there are some lists that create the coordinates for these locations and show walking directions to these places.

The algorithm that was used to solve the problems that we mentioned before is based on two important ideas. Determining safety-ness routes by the risk score associated with the destination, calculating the average of risk upon grid coverage of the line path from origin to destination, and evaluating each step of the route to get to the destination.

And everything that we mentioned can be finally solved with a Euclidian distance between points a and b. [7]

2. **A Data Integration and Analysis System for Safe Route Planning**

In the document, the authors want to influence safety by finding a collection of factors that could determine how safe a region is, this process starts by asking women in which conditions they feel unsafe, then, based on that they create a function that takes inputs and effectively combine them, describing it as an algorithm is used to determine relative safety concerning an ideal location where the chances of a case are minimum. Let´s assume that there are five factors and describe them as Ys, then we can simply define the value by (Y1 + Y2 + Y3 + …).

After making a graphic of how the function is, they think about two values in which the function is 0 and conclude that the first one is the minimum number of people that should be around a person, while the second is the maximum number of people which can be present without representing danger for the person.[8]

3. **Beyond the Shortest Route: A Survey on Quality-Aware Route Navigation for Pedestrians**

In this paper, the authors investigate different navigation systems that contribute to identifying the shortest or fastest routes in the user area. Taking as rating attributes what the authors call SWEEP (**S**afety-Health and **W**ell-being-**E**ffort-**E**xploration-**P**leasure), they also analyze the algorithms and data sources implemented in such systems.

Then, the authors describe the types of data these programs use, all in a summary table to compare different data sources and show a short description, feature types, and modalities. There are two features: **Dynamic** and **Simple** both based on the geographical characteristics of the streets of the selected route.

The authors show evaluation of route navigation systems, which is the main point of this paper, providing categorization, evaluation approaches, and making promises of future research. [1]

4. **Problem 4: Safety-aware routing for motorized tourists based on open data and VGI**

In this paper, the authors have the objective of creating safer and shorter routes for tourists in the city of Los Angeles, based on VGI information and open crime data sources from the government, additionally, they take information such as; historical crime in the area, crime hot spots, abandoned buildings, and police stations.

They explain crime mapping; using clustering techniques, Kernel Density Estimation, and nearest neighbor analysis. Although these are only examples, the authors go way deep in explanations, showing maps of the city with all of the different aspects, all of this to conclude different notes on how to approach all the things that a good algorithm of this kind should have.[2]

## 3. MATERIALS AND METHODS

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

### 3.1 Data collection and processing

The map of Medellín was obtained from *Open Street Maps* (OSM)[1] and downloaded using the Python API[2] 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 insecure and (ii) the fraction of households with incomes 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 components analysis. The risk of harassment is defined as one minus the normalized CL.

[1] https://www.openstreetmap.org/

[2] https://osmnx.readthedocs.io/

Figure 1 presents the calculated risk of bullying. The map is available on GitHub[3] .

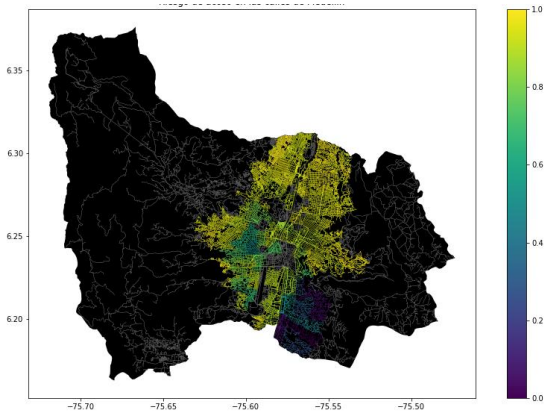


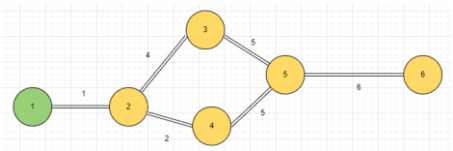

**Figure 1.** Risk of sexual harassment calculated as a linear combination of the fraction of households that feel unsafe and the fraction of households with income below one minimum wage, obtained from the 2017 Medellín Quality of Life Survey.

### 3.2 Algorithmic alternatives that reduce the risk of sexual street harassment and distance

In the following, we present different algorithms used for a path that reduces both street sexual harassment and distance.

#### 3.2.1 Dijkstra's algorithm

Dijkstra's algorithm is very notable when people are looking to find the shortest route, it works by implementing a graph of nodes, like this:

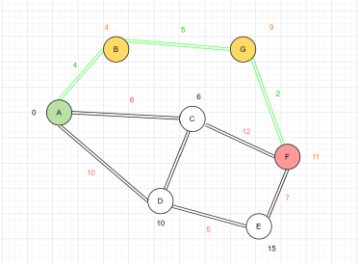


Taking a source node (Start point) and comparing the distances between other nodes. Each node is connected by a line that has a weight, these values are a crucial part of finding the shortest route, because they represent the distance between each point. The algorithm will start to find the routes to each node that cost less distance, marking the ones that he already visited until he has marked all nodes. [3]

#### 3.2.2 A* Algorithm

A* algorithm is one of the best path-finding methods and graph traversals, what makes it unique is that is considered an intelligent algorithm.

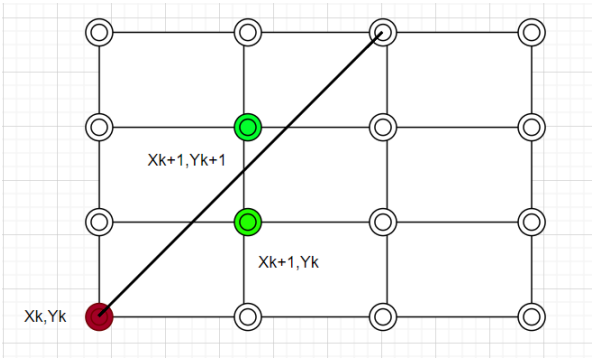
It works considering a graph with different nodes, having to reach a specific node as quickly as possible. The algorithm takes as a main parameter the value 'f' which represents the sum of two other values 'g' and 'h', then it picks the lowest f to decide which path it should take to reach the objective effectively. [4]




#### 3.2.3 Bresenham´s line algorithm

Drawing a line from an initial point that determines some points of an n-dimensional raster is what we use this algorithm for, something important about it is that is not a very difficult algorithm to implement because it works with pixels in relation with graphical operations.

It works with an endpoint of a line segment, so the next points it traverses to get to another endpoint is determined by evaluating where the segment crosses in relation with the midpoint of the two possible grid choices [5]

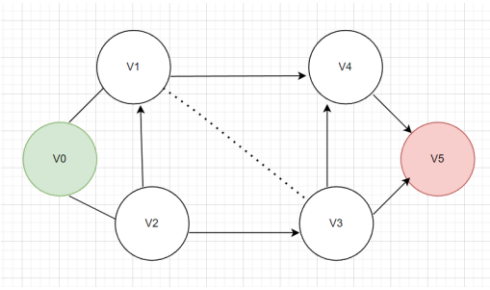



#### 3.2.4 The pulse algorithm

[3]https://github.com/bosh99/SH-proyect.git

The idea behind this algorithm is simple, propagating pulses from a start to an end node, as every pulse traverses the network it builds a partial path including the nodes that already visit the initial one, something important to emphasize is that each pulse that reaches the final node contains all the information for a path P from Vs to Ve.

Finally, the algorithm completely enumerates all the possible paths from Vs to Ve ensuring that the optimal P* is always found [6]



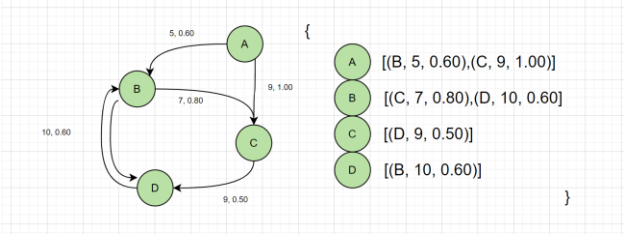

## 4. ALGORITHM DESIGN AND IMPLEMENTATION

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

To represent the map of Medellin, we used a Python dictionary in which every key is an origin, and it points to an array full of tuples. The format of each tuple is "(Destination, Sexual Harassment and length)" every tuple represents the different destinations connected to a specific origin. Also, we took the One-way value of the csv file in order to know if the destination could also be a origin and create a new key inside de dictionary.

The data structure is presented in this figure 10.




### 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 Algorithm for a pedestrian path that reduces both distance and risk of sexual street harassment

Dijkstra's algorithm basically finds the shortest path between a given node and the rest of the nodes. In our case, we need the shortest path only between two nodes. For that we made some changes to the algorithm, making it to stop when it has reached the destination, and returning the shortest path.

for finally finding the weight of the edges. The algorithm is exemplified in Figure 10.

**Figure 3:** Calculation of a path that reduces both distance and risk of harassment (please feel free to change this figure if you use a different algorithm).

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

The other two paths that are showed in the figure 4 are based in different forms of comparing both distance and sexual harassment. The red one represents the addition of distance and sexual harassment and the green one is the addition between $(\text{distance})^2$ and $(\text{sexual harassment})^2$. We must consider that the red path is like the original comparation, product between the two factors. The algorithm is exemplified in Figure 4.

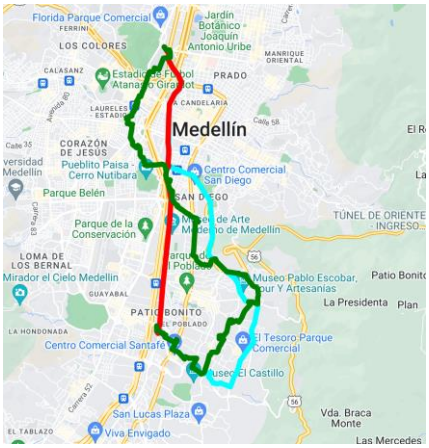



**Figure 4:** Map of the city of Medellín showing three pedestrian paths that reduce both the risk of sexual harassment and the distance in meters between the EAFIT University and the National University.

### 4.3 Algorithm complexity analysis

Explain, in your own words, the analysis, for the worst case, using the notation O. How did you calculate these complexities? Explain briefly.

| Algorithm | Time complexity |
|---|---|
| **Dijkstra's** | O (E*logV ) |

https://github.com/bosh99/SH-proyect.git

| | |
|---|---|
| N/A | |

**Table 1:** Vertex´s and Edges are represented with the letter V (Vertex) and E (Edge).

| Data Structure | Complexity of memory |
|---|---|
| Adjacency list | O(V+E) |
| N/A | |

**Table 2:** Memory complexity of the adjacency list filled with arrays of tuples |

### 4.4 Algorithm design criteria

Dijkstra's is one of the Worldwide known algorithms to use when solving a routes or shortest route related problem, that's why at the beginning of this project, after looking at other famous algorithms we encounter lots of information and reports that showed to us that Dijkstra was the best option for this project. We consider that the most important reasons we encountered are.

- Priority queue that supports the operations of extracting a minimum element. Letting us create a path more efficiently and make good use of other data structures.
- The usage of computer resources. This algorithm has a complexity of O (E*logV ) what makes it very useful and fast in order to generate the paths
- Access to information. We know that this project is based on autonomous search, there is many algorithms to make a solution for sexual harassment in Medellin, but the quality of information about Dijkstra is amazing.

## 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 | 5082.36 | 0.0019 |
| Eafit | Unal | 7894.27 | 0.0017 |
| Eafit | Unal | 642144.12 | 0.000012 |

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.

Calculate the execution time for the queries presented in Table 3.

| Calculation of v | Average run times (s) |
|---|---|
| v = SH*L | 0.786 seconds |
| v = SH+L | 0.714 seconds |
| $v = SH^2 * L^2$ | 0.999 seconds |

**Table 4:** *Dijkstra´s* execution times for each of the three calculator paths between EAFIT and Universidad Nacional.

## 6. CONCLUSIONS

The results of this project show tree different routes based on our criteria, but the code could be modified to analyze routes taking in matter other different aspects, like history of the area, geographic data, police reports and so many more. So, it could be implemented in many real situations, by anyone to implement in its software or app.

We are very happy with the result of this project, making it, in our opinion, very useful to the city and to our careers in EAFIT.

### 6.1 Future work

What we seek to improve for the future with this project is basically give it a more practical and effective approach making more easy for people the use of apps that contain the shortest route. Also, it would be interesting if the implementation of the safest route be seen in applications that are commonly used daily related to locations, addresses, etc. We think that, at the moment it is not in our plans to continue with the project but that doesn´t mean that we discard the option of continue with it in the future and focus on giving it adaptability in terms of virtual reality.

## ACKNOWLEDGEMENTS

This acknowledgment is directed towards the people that make possible the elaboration of this project, because for them, we could solve our doubts and improve what was done.

The monitors, people that are studying like us and who are always pending of questions, how can we make things better, what is the most appropriate way to solve things and much more. They deserve our most sincere thanks for being there and allowing us to learn more about programming, data

structures and stuff that we are sure will be useful for what is coming in our learning process.

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*.

**REFERENCES**

P. Siriaraya et al., "Beyond the Shortest Route: A Survey on Quality-Aware Route Navigation for Pedestrians," in IEEE Access, vol. 8, pp. 135569-135590, 2020, doi: 10.1109/ACCESS.2020.3011924.[1]

KELER, Andreas; MAZIMPAKA, Jean Damascene. Safety-aware routing for motorised tourists based on open data and VGI. *Journal of location Based services*, 2016, vol. 10, no 1, p. 64-77.[2]

Dijsktra's algorithm. *GeeksforGeeks*, 2022. https://www.geeksforgeeks.org/dijkstras-shortest-path-algorithm-greedy-algo-7/.[3]

A*SearchAlgorithm - GeeksforGeeks. *GeeksforGeeks*, 2022. https://www.geeksforgeeks.org/a-search-algorithm/.[4]

Brasenham´s algorithm. JavaTPoint, 2022. https://www.javatpoint.com/computer-graphics-bresenhams-line-algorithm[5]

Beasley JE, Christofides N. An algorithm for the resource constrained shortest path problem. Networks 1989;19:379–94 [6]

He, Z. and Qin, X. Incorporating a Safety Index into Pathfinding. *Trid.trb.org*, 2022. https://trid.trb.org/view/1501447.[7]

Preventing Sexual Harassment Through a Path Finding Algorithm Using Nearby Search. *Omdena | Building AI Solutions for Real-World Problems*, 2022. https://omdena.com/blog/path-finding-algorithm/.[8]

Shortest Path with Dijkstra's Algorithm. *Bradfieldcs.com*, 2022. https://bradfieldcs.com/algos/graphs/dijkstras-algorithm/.