

PREVENTION OF STREET HARRASSMENT THROUGH CONSTRAINED SHORTEST PATH ALGORITHMS

Isabel Mora
Universidad Eafit
Colombia
imorar@eafit.edu.co

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

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

Blue text = To complete for the 2nd deliverable

Violet text = To complete for the 3rd deliverable

ABSTRACT

Street sexual harassment is an issue impacting the overall safety of cities and the well-being of citizens, particularly women. Such a problem is prominent in Medellin, where women repeatedly say they feel unsafe while walking. Despite the negative side effects of street harassment, the lack of data and algorithms that warn citizens of areas where harassment is likely to occur perpetuates people's exposure to these types of situations. In this paper we present an algorithm that calculates routes based on both distance and safety in order to improve security perception, encourage walking and prevent cases of harassment.

Which is the algorithm you proposed to solve the problem? What QUANTITATIVE results did you achieve? What are the conclusions of this work? The abstract should have at most 200 words. (In this semester, you should summarize here execution times, and results of lowest risk path and shortest path).

Keywords

Constrained shortest path, street sexual harassment, secure-path identification, crime prevention.

1. INTRODUCTION

Gender-based street harassment occurs when a person is subject to unwanted sexual comments, gestures and actions by a stranger in a public space. These types of actions are identified because they are directed at the victim due to their gender or sex. Actions which classify as street harassment include catcalling, sexist slurs and names, public sexual demands, public flashing, masturbation and rape [5].

Gender-based street harassment is an issue worldwide, with a disproportionately greater effect on women. Although it's a problem described as transnational, developing countries, which often have weak infrastructure, poor lighting conditions in public places and unsafe transport systems, often have worse cases of street sexual harassment. For instance, a study conducted by YouGov found that Bogota, Mexico City and Lima were among the top five worst cities in terms of verbal harassment [5].

Evidently, street harassment is an issue affecting safety perception in cities, on top of contributing to negative

feelings in victims, particularly women. According to BMJ [6], street harassment makes victims feel humiliated (53.7%) and disturbed about their body (29.3%), and causes sleep and appetite problems. In Medellin specifically, 34.4% of teenage girls say they are victims of street harassment several times a day, while 60% of women feel Medellin is not a safe city due to its persistent patriarchal culture [7]. Thus, the constant threat of harassment when walking around a city, including Medellin, leads to fear, and can even stop people from walking or using the city's transportation network. Most importantly, the lack of data and algorithms that warn citizens of areas where harassment is likely to occur perpetuates people's exposure to these types of situations.

1.1. Problem

As seen by the figures presented in the introduction, street sexual harassment is an issue deeply affecting safety in cities, including Medellin. However, people do not have access to data that can help them avoid areas of the city where sexual harassment incidents are likely to occur. Even worse, citizens and tourists usually rely on navigation platforms, such as Waze or Google Maps to obtain routes for traveling around the city. These types of applications calculate routes based solely on distance, which not only fails to inform users of unsafe areas, but can also potentially lead people to take riskier routes just because they are shorter. Thus, there's a great need to develop an algorithm that calculates routes within the city based on both distance and safety. In the case of sexual harassment, this means developing a constrained shortest path algorithm that provides users with routes that are short and have a low risk of sexual harassment. Through the implementation of this algorithm, security perception in the city would improve greatly, more people, especially women, would be confident about walking and taking public transportation, and cases of street harassment would decrease significantly.

1.2 Solution

Explain, briefly, your solution to the problem (In this semester, the solution are algorithms for constrained shortest paths. Which algorithms did you choose? Why?)

1.3 Article structure

In what follows, in Section 2, we present related work to the problem. Later, in Section 3, we present the data sets and methods used in this research. In Section 4, we present the algorithm design. After, in Section 5, we present the

results. Finally, in Section 6, we discuss the results and we propose some future work directions.

2. RELATED WORK

In what follows, we explain four related works to path finding to prevent street sexual harassment and crime in general.

2.1 Urban Navigation Using SafePaths

Galbrun, Pelechris and Terzi [1] were the first to define safe routes with the use of criminal data. Specifically, concerned with growing insecurity in cities, and aiming to take advantage of publicly available datasets, the investigators sought to improve the quality of life of those traveling around cities by developing an algorithm that uses criminal data to yield navigation options based on both distance and safety. The preliminary experiment was done using data from Chicago and Philadelphia, which they used to create a city risk model that included the probability of crime on any given road segment. Overall, their central objective was to find a short and low-risk route between two given locations. However, since these two variables cannot be computed together as a single problem, the investigators developed a solution based on a bi-objective shortest path problem that outputs a set of paths that have varying degrees of safety and distance [1]. Despite their success, this SafePaths problem is based on crime in general, and fails to address specific types of crime, such as street harassment, along a certain path.

The investigators working on the SafePaths problem used a deterministic algorithm to determine the best route, which is done by computing all possible routes between two given locations. Overall, given a fixed pair of origin and destination locations, the algorithm computes two important paths: the safest path and the shortest path. When these two initial paths differ, as it is often the case, a recursive algorithm is then used to consider, in each iteration, an intermediate non-dominated path between the original intervals. This process is repeated until the difference between the shortest and the safest path is minimal [1].

2.2 Safe Routing with Crowdsourcing

A group of investigators in Mexico City, much like the authors of the work described in section 2.1, were concerned with security issues that arise in big cities and urban centers, especially in developing countries where crime is more widespread. Popular mobile systems compute paths based entirely on distance, and don't consider safety hazards that may arise along a recommended route. Thus, the researchers, developed an approach that integrates crowd-reported crime data with official government data in order to obtain safer routes in Mexico City for locals and tourists alike. The academic paper states that a preliminary experiment was implemented with 75% of effectiveness [2].

The approach proposed by the researchers consists on collecting Tweets that are related to crime and integrating them with crime data from official governmental institutions through an automatic system that considers descriptions and attributes in the data. After this, the Bayes algorithm is used to classify data that couldn't be integrated automatically. Most importantly, this algorithm is used to assign probabilistic crime rates to specific parts of the city. After collecting, sorting and classifying the data, a mobile application was developed to display the safest route between two given locations. The safe route is obtained through the Dijkstra algorithm, in which the weight of the nodes is derived from the average number of crimes in the given location. Thus, the algorithm outputs a route that avoids places with higher crime rates [2].

2.3 Personalized Safest Route

Analysis of criminal activity reveals that a significant number of offences towards civilians occur when people are in transit throughout a city. This, coupled with increasing rates of criminal reports throughout the world, motivated researchers Tarlekar, Bhat, Pandhe, and Halamkar [3] to determine the safest route between two given locations in order to improve people's safety while traveling. The researchers developed a mobile application that outputs a personalized safest route based on the user's gender and age. To develop their application and algorithm the researchers used official data based on geographical locations that includes reports of criminal incidents that have occurred over the past 12 years in San Francisco.

The algorithm used by the investigators to obtain their desired result was the Iterative Dichotomiser 3 or ID3 algorithm. Initially, the application receives an input with a starting location and a destination. With these two locations a route is calculated, and all the possible streets along that route are analyzed based on the dataset relevant to the specific user. In other words, risk is calculated based on data that matches the user's gender and age. The safety route between the two locations can be calculated by the safety of the streets that make up the route. Thus, the ID3 classifier, which generates a decision tree, checks the safety of all the streets and outputs a value of "Yes" or "No". The streets marked with a "No" are analyzed further and are the ones that determine the overall risk of the route. At the end of the process, the route with minimum risk value is selected as the safest and is returned to the user. The implementation of the algorithm detailed in the academic article was successful, with a reported accuracy of 86% [3].

2.4 Preventing Harassment with Anonymous Reports

As the final related work, we present the SafeStreet application created by researchers Ali, Rishta, Ansari, Hashem and Khan [4]. Troubled by the lack of attention given to widespread forms of public harassment, such as catcalling and groping, along with the negative psychological effects of these experiences on women, the researchers aimed to create

an application to help victims avoid public harassment incidents. Specifically, through the application, victims can anonymously self-report occurrences of street harassment with the ultimate goal of creating a large database of assault-related statistics. With this large and dynamic data base created through crowdsourcing, other application users can get live data of areas of the city prone to harassment and, subsequently, opt to take routes that avoid these areas [4]. Compared to the related works presented in sections 2.1, 2.2, and 2.3, the SafeStreet application is the most specific, as it was designed distinctively to reduce sexual harassment, instead of crime in general.

The application’s main feature is a map where users can see combined records of harassment which were previously reported in the servers. The harassment statistics are shown through hierarchical clustering, which allows reports to be shown for both large areas, like counties and cities, and small regions, such as distinct roads. If a user wants to travel to a certain location, SafeStreet yields both a safe path and an unsafe path. In addition, the application can provide users with the safest time of the day to travel to the desired destination [4]. The exact algorithm used by the application to output the safest path is not detailed in the article, but it’s clear that SafeStreet work only with the harassment statistics and does not consider distance, which is a great deficiency.

3. MATERIALS AND METHODS

In this section, we explain how data was collected and processed and, after, different constrained shortest-path algorithm alternatives to tackle street sexual-harassment.

3.1 Data Collection and Processing

The map of Medellín was obtained from Open Street Maps (OSM)¹ and downloaded using Python OSMnx API². The (i) length of each segment, in meters; (2) indication whether the segment is one way or not, and (3) well-known binary representation of geometries were obtained from metadata provided by OSM.

For this project, we calculated the linear combination that captures the maximum variance between (i) the fraction of households that feel insecure and (ii) the fraction of households with income below one minimum wage. This data was obtained from the quality of life survey, Medellín, 2017. The linear combination was normalized, using the maximum and minimum, to obtain values between 0 to 1. The linear combination was obtained using principal components analysis. The risk of harassment is defined as one minus the normalized linear combination. Figure 1 presents

the risk of harassment calculated. Map is available on Github³.

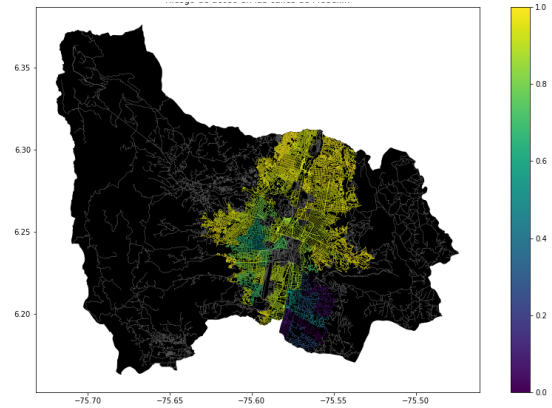


Figure 1. Risk of sexual harassment calculated as a linear combination of the fraction of households that feel insecure and the fraction of households with income below one minimum wage, obtained from Medellín’s 2017 Life Quality Survey.

3.2 Constrained Shortest-Path Alternatives

In what follows, we present different algorithms used for constrained shortest path.

3.2.1 Dijkstra’s Algorithm

Dijkstra’s algorithm, based on graph theory, is used to find the shortest path between two nodes on a graph. The graph used in the algorithm is said to be weighted, since each edge connecting nodes has certain value, or weight, to indicate how “costly” it is to travel through the edge. The most common variant of the algorithm sets a single node as the source and finds the shortest path from the source to every other node within the graph. It must be noted that not all nodes are connected to each other, so the algorithm finds paths from one node to another by passing through other nodes [8].

Dijkstra’s algorithm starts by “labeling” each node with the known distance from the source. Before any calculations are made, the source node is assigned a value of 0, and all the other nodes are assigned a value of ∞ . The values of the nodes distinct from the source are updated as the algorithm is executed, and the value will always represent the shortest distance found until that point. As the first step, the algorithm “travels” through all the edges connected to the source node, and updates the value of the distance taken to get to each of the “secondary” nodes. After this, the algorithm chooses the secondary node with the shortest dis-

¹ <https://www.openstreetmap.org/>

² <https://osmnx.readthedocs.io/>

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

tance, and calculates the new distances through the edges that depart from the chosen secondary node. If a new distance is shorter than one found in the previous step, the distance assigned to that node is updated. In addition, the algorithm keeps track of the path followed to obtain the assigned shortest distance. Overall, Dijkstra's algorithm follows 3 main steps: updates node distances, keeps track of paths followed, and chooses the next vertex. Once all vertices on the graph have been chosen or "visited", the value assigned to each node, with its corresponding route, is the shortest distance and path from the source node [9].

Algorithm complexity

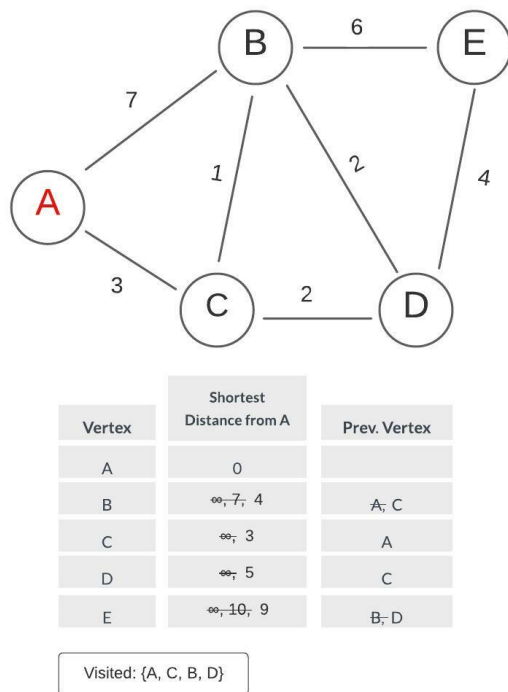


Figure 2. Dijkstra's algorithm diagram

3.2.2 A* Algorithm

A* is an algorithm based on graph theory that is widely used to find the shortest path between two given locations. The algorithm is particularly useful for finding paths in graphs that include obstacles or unwalkable segments. This is ideal for constrained shortest path problems, as areas where a certain condition is exceeded (in the case of this article, the risk r of being sexually assaulted) can be taken as obstacles through which the path cannot pass. Overall, A* works by maintaining a tree of paths originating at the starting node and extending those paths one edge at a time until the desired condition is obtained. Most importantly, the A* algorithm is based on heuristics, as it uses a heuristic function to determine a (probabilistic) cost of reaching the destination from each node [10].

A* starts its execution by calculating the value of the function $f(n) = g(n) + h(n)$ for each of the neighboring nodes of the starting node. In this function, $g(n)$ refers to the cost of the path from the origin, and $h(n)$ refers to the cost from the destination. The heuristic part of the function is $h(n)$, as this value cannot always be calculated deterministically due to obstacles [10]. In the traditional version of A*, where each node is a square on a graph, neighboring nodes vertically or horizontally have a cost of 10, and diagonal nodes have a cost of 14. Nodes that are part of obstacles are not assigned a value. After calculating $f(n)$ for all the neighboring nodes, the algorithm decides to extend the path of the neighboring node with the lowest $f(n)$ value. $F(n)$ values are calculated for the neighboring nodes of the new node, and this process is repeated. The algorithm also keeps track of the node where a given $f(n)$ value came from, as $f(n)$ values for certain nodes are updated repeatedly when a shortest path is found. The algorithm finishes its execution when the path it chooses to extend is a path from start to destination or if there are no more paths to be extended [11].

Algorithm complexity

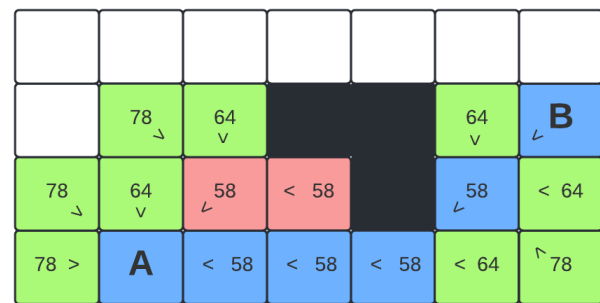


Figure 3. A* Algorithm example

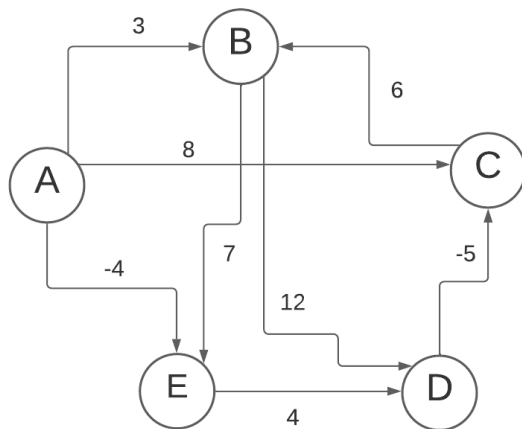
3.2.3 Floyd-Warshall Algorithm

The Floyd-Warshall algorithm is commonly used for finding the shortest path between each pair of nodes in a graph. The graphs used with Floyd-Warshall are weighed and directed, meaning that each edge between vertices is assigned a value and a direction [12]. This algorithm is particularly useful for graphs that have edges with negative values, as these don't push the algorithm into infinite searches, as it is often the case with other algorithms like Dijkstra's.

Primarily, the algorithm is executed with the use of a matrix, which is constructed from the given graph. Each entry $[i][j]$ of the matrix represents the value of the edge between nodes i and j . If there's no edge connecting i and j , the matrix entry is given a value of ∞ . Furthermore, the distance of a node to itself is 0, so the matrix's diagonal will always be equal to 0. After the matrix is constructed from the given graph, the algorithm calculates the distance between two nodes. In this process, it checks if there's any intermediate nodes between the two chosen originally. If this is the case,

and if the distance when passing through the intermediate node is less than the direct distance between the two original nodes, the shortest distance value is updated in the matrix. This process is repeated for all the intermediate vertices, and then for every other pair of nodes in the graph. At the end of its execution, the algorithm outputs an updated matrix with the shortest paths between each pair of nodes in the graph. An updated version of the Floyd-Warshall algorithm includes an extra step that calculates the shortest-path tree for every node in the graph, which allows for the reconstruction of the path between two nodes [12].

Algorithm complexity



	A	B	C	D	E
A	0	3	8	∞	-4
B	∞	0	∞	12	7
C	∞	6	0	∞	∞
D	∞	∞	-5	0	∞
E	∞	∞	∞	4	0

Figure 4. Floyd-Warshall algorithm input example

3.2.4 Breadth-First Search Algorithm

The breadth-first search algorithm is used for searching tree data structures until a node with a certain condition is found. The algorithm starts at the root node and explores each level of the tree completely before moving on to the next level. This differs from the depth-first search algorithm, which explores a tree branch until its last level before backtracking and exploring other branches. Among many other things, breadth-first search is used for finding the shortest path from a source vertex to all the other nodes on the graph. Unlike other algorithms, like Dijkstra's and Floyd-Warshall, breadth-first search is used on graphs that are unweighted. This means that edges between nodes are not assigned a specific length. Instead, the distance from a

certain node to the source node is determined by how many levels beneath the source node it is found, with each level being assigned a value of one [13].

BFS works by assigning two values to each node in the tree: a distance referring to the minimum number of edges needed to get to the source, and the name of the previous vertex in the shortest path. Before the algorithm is executed, the two values of every vertex in the graph are set to null and the source vertex is assigned a distance value of 0. Then, all the neighbors of the source node are visited first and assigned a distance value of 1, since they are one level below the source. Afterwards, the algorithm explores all the neighbors of the nodes whose distance is 1 and assigns them a distance of 2. This continues through all levels of the tree until all nodes reachable from the source are visited. As the algorithm gets deeper into the tree, it must be careful not to visit nodes that have already been visited, so it can only explore nodes whose distance value is still null. Furthermore, to keep track of which neighbor node to explore from next, the algorithm uses a queue. Once a node is visited it is added to the queue, and the node that has been in the queue the longest is the one chosen to explore from next [14].

Algorithm complexity

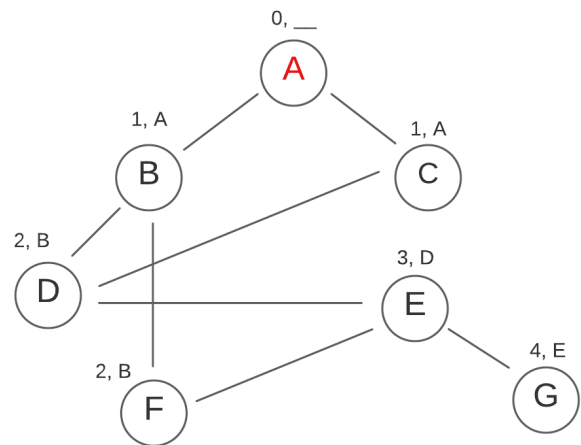


Figure 5. Breadth-first search

REFERENCES

1. Galbrun, E., Pelechris, K. and Terzi, E., 2015. Urban navigation beyond shortest route: The case of safe paths. [online] Science Direct. Available at: <https://www.sciencedirect.com/science/article/pii/S0306437915001854>.
2. Mata, F., Torres-Ruiz, M., Guzmán, G., Quintero, R., Zagal-Flores, R., Moreno-Ibarra, M. and Loza, E., 2016. A Mobile Information System Based on Crowd-Sensed and Official Crime Data for Finding Safe Routes: A Case Study of Mexico City. [online] Mobile Information Systems.

Available at:

<https://www.hindawi.com/journals/misy/2016/8068209/>

3. Tarlekar, S., Bhat, A., Pandhe, S. and Halarnkar, T., 2016. Algorithm to Determine the Safest Route. International Journal of Computer Science and Information Technologies, Vol. 7 (3), 1536-1540. Available at: <http://ijcsit.com/docs/Volume%207/vol7issue3/ijcsit20160703106.pdf>

4. Ali, M., Rishta, S., Ansari, L., Hashem, T. and Khan, A., 2015. SafeStreet: Empowering Women Against Street Harassment using a Privacy-Aware Location Based Application. [online] ResearchGate. Available at: https://www.researchgate.net/publication/300656203_SafeStreet

5. Stop Street Harassment, 2022. What Is Street Harassment? | Stop Street Harassment. [online] Available at: <https://stopstreetharassment.org/about/what-is-street-harassment/>

6. Neeti, R. and Pamela, O., 2012. Street harassment: an unaddressed form of gender-based violence. [online] BMJ Journals. Available at: https://injury prevention.bmj.com/content/18/Suppl_1/A145.2.info

7. Ruta N Medellín, 2022. Reducir el acoso callejero: el reto de la Secretaría de Mujeres - Ruta N. [online] Available at: <https://www.rutanmedellin.org/es/programas-vigentes/2-uncategorised/592-reto-de-mujeres>

8. Wikipedia, 2022. Dijkstra's algorithm. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Dijkstra%27s_algorithm#Algorithm

9. Spanning Tree, 2020. How Dijkstra's Algorithm Works. [video] Available at: https://www.youtube.com/watch?v=EFg3u_E6eHU

10. Abiy, T., Pang, H. and Tiliksew, B., 2022. A* Search. [online] Brilliant Math & Science Wiki. Available at: <https://brilliant.org/wiki/a-star-search/#heuristics>

11. Lague, S., 2014. A* Pathfinding (E01: algorithm explanation). [video] Available at: <https://www.youtube.com/watch?v=-L-WgKMFuhE>

12. Datta, S., 2020. Floyd-Warshall Algorithm: Shortest Path Finding. [online] Baeldung. Available at: <https://www.baeldung.com/cs/floyd-warshall-shortest-path>

13. Wikipedia, 2022. Breadth-first search. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Breadth-first_search

14. Cormen, T. and Balkcom, D., 2015. The breadth-first search algorithm (BFS). [online] Khan Academy. Available at: <https://www.khanacademy.org/computing/computer-science/algorithms/breadth-first-search/a/the-breadth-first-search-algorithm>