International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

A Data Integration and Analysis System for Safe Route Planning

Aryan Guptaa[1], Bhavye Khetan[2]

Abstract: ***Crimes are rising day by day; thus, safety is becoming a major concern for people today. Even while travelling, people should be aware & choose the route which is safest to travel from. People who are new to the city have no idea about safe routes. Though people rely on Google maps for planning their routes; yet it only provides the shortest path & gives no consideration for the safety of the path. Although several other route planning apps exist which provide the safest route, these do not consider all the factors that account for the safety of the path. Apart from other navigation apps, this paper describes an innovative method to find the safest route having the highest safety value. We also realize that the safest route may turn out to be excessively long and time-consuming, thus, to make it practical, we effectively incorporate it along with the parameter of distance, thus making it useful in the real world.***

Keywords: Women Safety, Safe GPS, Safety Route, GPS, Data Integration, Analysis system, Women, Route Planning, Safety Metric

1. Introduction

Cases of sexual harassment against women are the 4th largest problem in India and have been increasing at an exponential rate since the last decade. Over 23, 000 cases were reported in 2013, and that number went over 37, 000 by 2018. Additionally, it has been reported that less than one-twentieth of the cases are reported, so one can accurately gauge that over 7 lakh cases could have taken place in 2018 and we estimate that number to grow to approximately 9 lakhs in 2020. Adding to the increase in crime rate caused by unemployment and loss of livelihood in the COVID-19 period, crimes against women are bound to cross all limits. This paper deals with a way to resolve this growing number of cases, which is by designing an application for safe routes that can be used to determine the safest and fastest route instead of just the fastest route.

An effective way to reduce the number of cases would be to find out the conditions in which such cases are prone to happen and then prevent those conditions from happening. But most of the time, these aren't things one can change or control by themselves. Thus, the next available option is to avoid the places where these conditions are present. We then need to find a way to implement this in practice. A possible way would be to cluster these locations and then avoid these clusters. This clustering can be done using various data science algorithms, the most prominent of which is k-means. However, although this may seem an effective solution, it does not tell us the relative safety of two locations, or in other words, we are not able to quantize the safety value of a location using this. Thus, a better way would be to develop a model that can tell us exactly how safe a location is. This will use various inputs, which would be the factors that can influence the relative rate of harassment in an area. However, it still lacks the way we would go from one place to another, that is we can't go to a place and then decide whether it is safe or not, we must decide in advance. We must incorporate this safety value in map routing, so one can refer to this safe route to get from one place to another safely. Another issue that comes up is if the safe route is too long to practically traverse. It can't be that to go from Delhi to Gurgaon, one has to go through Himachal Pradesh and then come to Gurgaon through Jaipur. Thus, we have to develop a way to make a route that is both safe, fast, and then present it appealingly.

With this, we aim to reduce the crimes against women in India, so that it becomes a much safer location for all and at the same time help the police to catch perpetrators by investigating the hotspots and encouraging the government to develop the unsafe zones to make them safe. Many women now are seeking futile ways to avoid these sorts of situations by not going out late at night or going through routes known to them or not travelling alone but in a group. But it can be safely concluded that these restrict the freedom of women and are unfair to them. Thus, researchers, police, and the government must work hand in hand to ensure the safety of women.

2. Methodology

This paper is intended at finding the safest route in terms of sexual violence against women, so we started by trying to think of all the factors we could which can influence the safety of a region or a point on the graph. This was done to find a collection of factors using which we can determine whether and how safe a region is. With the help of our team, we primarily conducted one-to-one surveys with women asking them which conditions would make them feel safe and which conditions would make them feel unsafe. By looking into and analyzing past cases in various cities in India, we set up a database using which we compared various situations to determine which factors would, directly and indirectly, affect the outcome in a case and how we can classify it. Most of the previous literature on this topic used the latitude and longitude of the cases to determine their location and cluster them but did not stop to consider the effectiveness of the program implementation. The problem with this is that it does not consider new areas that may become a hotspot in the future and it also completely disregards outliers. And if they start considering the outliers, then the accuracy of the model goes down drastically.

3. Influences on Safety of A Journey

The following factors can play a direct and indirect role in determining the safety of a location and can affect the outcome of a situation.

1) No. of cases in the area
2) People in the area

3) Police Stations / Chowkis
4) Time of day
5) Illumination
6) Network Speed
7) Literacy and Unemployment
8) Monetary status
9) Weather conditions
10) Holiday/Festival
11) Cameras in the area
12) Sound

As one can notice, it is not possible, humanly or mechanically, to determine the correspondence of each factor and then combine them all to get the safety value of a point.

Before we proceed to identify which factors, we need, let us define our Safety Function. The function will take in inputs based on the above factors and effectively combine them. It can be defined as an algorithm used to determine the relative safety of a region concerning an ideal location where the chances of a case occurring are the minimum.

For now, let us assume that there are five factors - y_1, y_2, y_3, y_4, and y_5.

Then we can simply define the Safety value (y_S) as

$$y_S = y_1 + y_2 + y_3 + y_4 + y_5$$

Here all the values will be from 0 to 100 and the safety value will be out of 500.

Now let us define our ideal situation for which we will calculate our safety value.

Assuming all our factors to be at their maximum value, we get a situation where there are a good number of people around, with no past cases in the area, time being around 3 pm, a lot of police station nearby, high network coverage, clear weather, etc. So now we set out to determine our most important factors. Taking the ideal situation, but removing the part with a lot of people, and now there will only be a woman alone with a stranger in the area. Would this be considered safe? But according to the safety function we had calculated, the minimum value of the situation would be 400 which is considered safe. We need to reform our safety function in a way that even when one critical factor is missing, its value should be considerably low. One of the common ways of doing so is to multiply the factors instead of adding them. A major drawback of this would be that non-critical factors would then have equal weightage as that of the critical ones; hence it is not possible to multiply the factors. The most fruitful solution here is to change the range of each of the factors. Instead of the factors ranging from 0 to 100, they can be from negative infinity to 100. Here, if the input value for each of the function falls below a certain level, the value of the function will fall drastically and so will the value of the Safety function, as the maximum value of each factor is capped at 100. We have identified the first one to be the number of people around. Using the same method for all the factors, we see that Illumination and No. of cases in the Area are two more of the essential factors. The non-essential factors will be discussed after we have made the function for each of the critical factors.

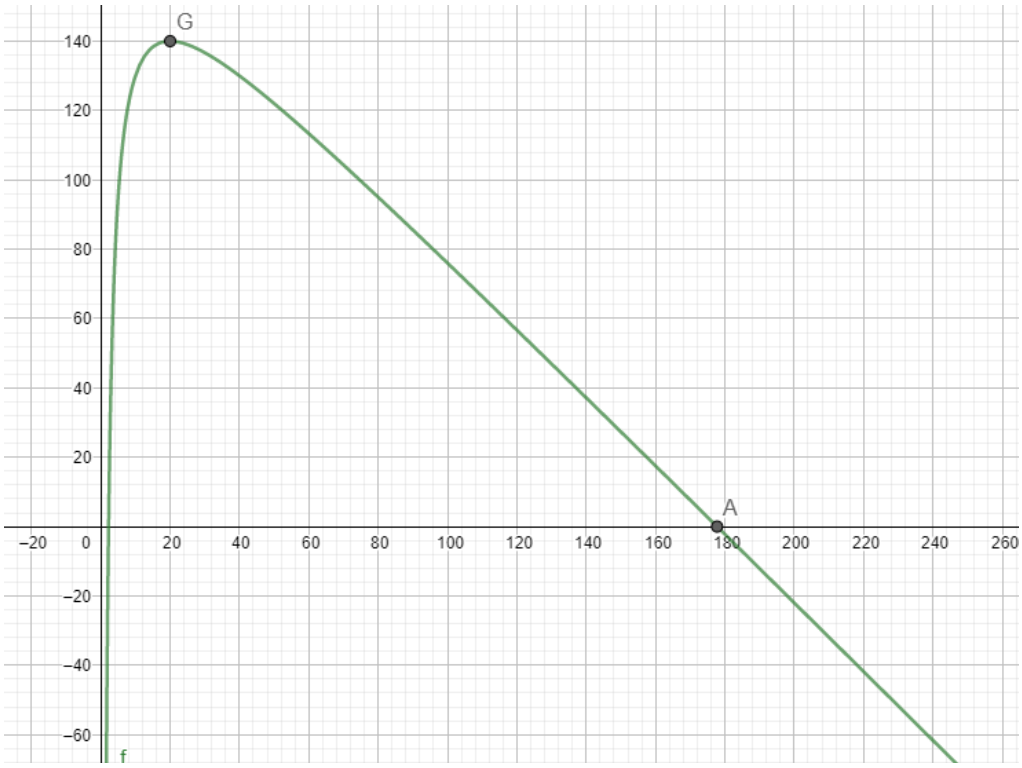
a) People in the Area

As discussed above, the number of people in the area plays a critical role and according to the corrections in the range, its value is from -∞ to 100, and its value should drop drastically when the number of people is very less. A key argument we need to consider is the range, that is, in how much area will we measure the number of people at that time. An appropriate range would be in which people would be able to hear you and come to your aid. Let's solve this problem practically. A person 500m away after hearing the commotion is more likely to turn and leave the scene of the crime. On the other hand, everyone in a 100m radius would be able to see the victim. Thus, the value of our range lies between 100 and 500 meters. A good estimate would be 200m because people in this range are likely to come and see what is happening instead of turning away.

But this alone is not sufficient enough to make a function that can accurately depict the dependence of the safety of women on the number of people. Therefore, before making the graph, we must visualize it and then use trial and error to perfect the values. A key factor in visualizing the graph will be the maximum of the graph. As we have decided, the graph will have 100 as its maximum value. So, what is the value on the x-axis when the function reaches that maximum value? That value has to be the ideal value when it is easy for her to blend into the crowd and for someone to notice and then call the emergency services and come for help. After doing a massive amount of research through reading research papers, consulting psychologists, and interviewing women, we found out that this value is 30, that is, it is ideal to have 30 people nearby. So, our function reaches its max value at 30.

Another important thing we require is the behaviour of the function after it has reached its maximum value, that is if it remains constant or it drops, and if does, how does it do so. Considering practically, if a region has a lot of people, the chance of the perpetrator hiding in the crowd and escaping becomes high, and as the number of people increases, the number of possible culprits also increases. It also takes the victim longer to reach her destination, so the duration of exposure also increases.

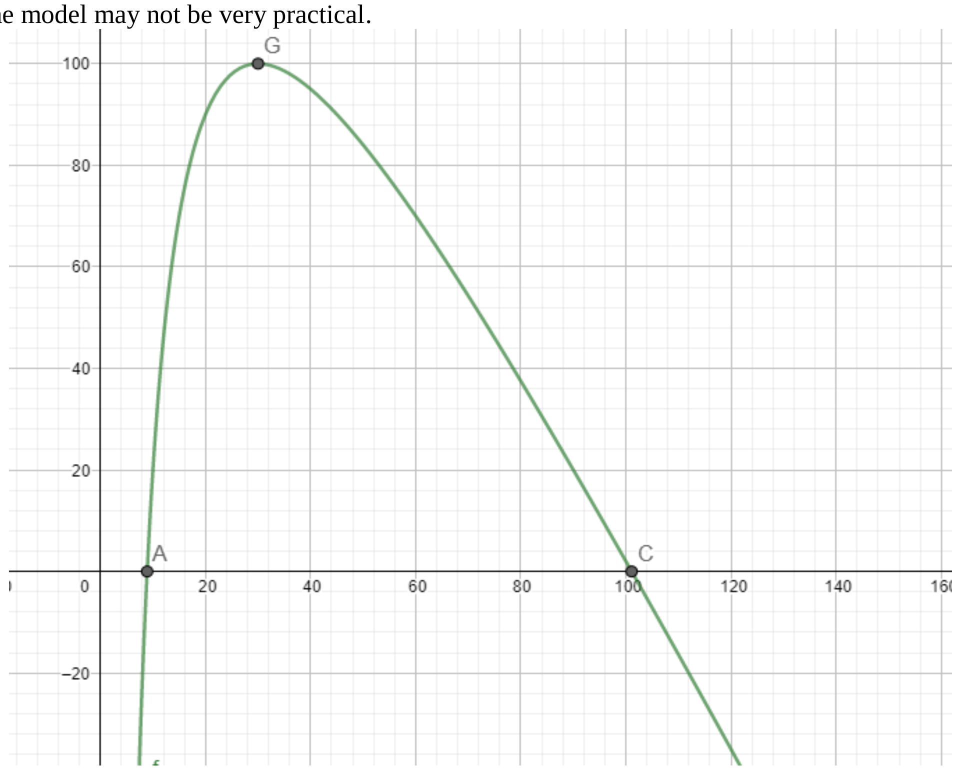
Hence, the graph will drop after it has reached its maximum value. However, this drop is not the same as that talked about earlier, as it is more gradual instead of drastic as the value goes down slowly. So now we can visualize our graph as something like this –



As we have decided on how the function will be like, we need two more parameters that are the two values at which the function will be 0.The first one symbolizes the minimum number of people that should be around a person while the second one signifies the maximum number of people which can be present without inducing danger for a person After various consultations and research, we figured that the first value should be around 10 and the second value should be around 100. Note that here we are not considering the exact values as to achieve the exact values, we may need to have decimal or irrational coefficients which not be a feasible test or calculate and the model may not be very practical.

Now we have the values we need to try and graph our function. After various tries, we came up with the function required.

Note that this method of developing a function or any other algorithm is called 'Bottom Up', which is to construct something from the very basics using data and trial and error.



Paper ID: SR201015234344 DOI: 10.21275/SR201015234344

$$y_P = 220 - \left(2x + \frac{1800}{x}\right)$$

Here is the table of important values of this function.

$f: y = 220 - \left(2x + \frac{1800}{x}\right)$

Roots(f, −6.4077575013012, 178.5965446816254)
→ A = (8.9022777120051, 0)
→ C = (101.0977222704978, 0)

Extremum(f, −6.4077575013012, 178.5965446816254)
→ B = (-0.0000000025542, 704733242586.19958496)

D = Intersect(f, yAxis, (0, 0))
→ undefined

E = (0, -∞)

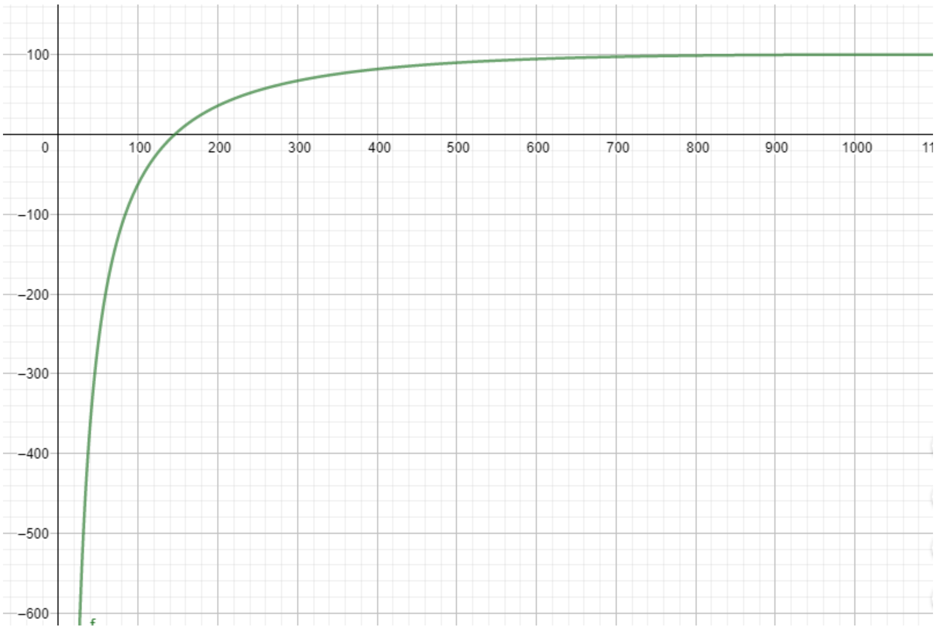
F = (30, 100)

b) Illumination

Like we did with people, let us consider a situation here. If in an area, there are an ample number of people, but it is night time, the chance of crime happening again are pretty high, as the perpetrator would be confident that the victim won't be able to see him and that he would be able to blend into the crowd. So, Illumination in the area is another key factor we have to consider. Again, the value for illumination ranges from -∞ to 100 and the value of the function drops drastically as the value of illumination reaches zero.

To visualize the graph, we must then consider the maximum value or the ideal value of illumination. The ideal value would be where one can see everything clearly and neatly, and the light shouldn't disturb their activities. Through our ground research and review of past cases, it was found that this value is close to that of the lighting on an overcast day, which is around 1000 lux.

We must decide how the function will behave after reaching this value. As greater illumination than this would not be harmful or be beneficial, the value of the function will remain constant after this. To determine the value of zero for this function, we went back to our database to find the minimum acceptable value of illumination, which came out to be 200 lux, which is that of a street light. Here we won't need the value of the second zero as the function does not go down. So, after various trials, we built our function as shown in the graph.



The function is

$$y_I = 140 - \left(\frac{x_I}{50} + \frac{20000}{x_I}\right), x_I < 1000$$
$$y_I = 100, x_I \geq 1000$$

The table of important values is

$f: y = 140 - \left(\frac{x}{50} + \frac{20000}{x}\right)$

Roots(f, −1029.6292430897056, 7642.5562972533326)
→ A = (145.8980337832146, 0)
→ G = (6854.1019662268172, 0)

Extremum(f, −1029.6292430897056, 7642.5562972533326)
→ B = (-0.0000000131448, 1521514177423.7412109)

E = Intersect(f, yAxis, (0, 0))
→ undefined

C = (0, -∞)

D = (1000.000017653153, 100)

Paper ID: SR201015234344 DOI: 10.21275/SR201015234344

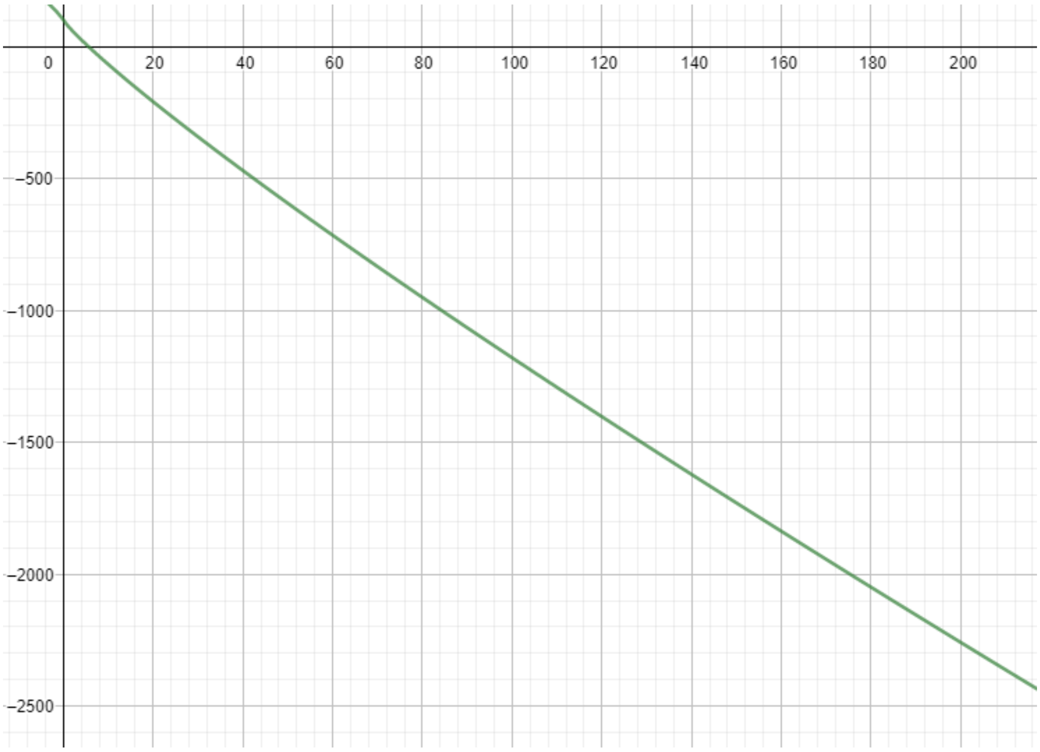
c) Number of Cases

A place with a large number of cases that has good illumination and an apt number of people is difficult to find. Crime happening here doesn't seem realistic at first, but that's the catch here. Even though we have illumination and people in the area, if cases are still coming from it, then it implies that some other factors are influencing that region. As it is not possible to find and then rationalize all those factors, we will take this as our key factor. Conversely, as we have taken this as our key factor, it covers all the other factors mentioned and can be taken as the sum of all of them. Now a question may arise that when we are taking this as the key factor, why don't we ignore the other two and consider them to be a part of this as well? The answer to that is simple – When we took mock situations for those factors, we did take the number of cases in that region to be zero, yet those locations were found to be unsafe, so they will be handled separately.

Like the number of people, here we need to consider the range in which we will count the number of cases. Since we had adopted the area of 200m radius around the location for the number of people, we will use the same range here. Again, the value for this function ranges from -∞ to 100 and decreases drastically as the number of cases increase.

An important factor we need to merge in here is time. It is possible that a location is safe during the day but unsafe at night, so there are a lot of cases at night, so we must classify it as safe during the day. Therefore, with a lot of consideration and evaluation, we have decided to take the cases within 3 hours of the current time, which is between 3 hours before and 3 hours after the time. Here determining the value at which the value of the function is maximum is not an issue – the maximum value of the function is when the number of cases is zero. To determine the value at which the function becomes zero, we have to determine the maximum number of acceptable cases in the area. Usually, this number would be around 15-20 as there are a lot of minor, unintentional, and misunderstood cases of sexual harassment in any area, but since a lot of them are not reported, the number we will use here is 5.

Since the value must drop drastically, the graph has to exponential in nature, so with our database and after many trials, we came up with the following graph and equation -



The equation is as follows –

$$y_C = 100 - 22x_C^{\frac{15}{17}}, x \geq 0$$

The table of important values is –

$f : y = 100 - 22\,x^{\frac{15}{17}}$

A = Roots(f, −14.6993400546052, 210.7469116226859)
→ (5.5622987362928, 0)

B = Extremum(f, −14.6993400546052, 210.7469116226859)
→ undefined

C = Intersect(f, yAxis, (0, 100))
→ (0, 100)

d) Other Minor Factors

As discussed earlier, several minor factors influence the safety of a journey. As one may think that we have already covered these factors earlier in the number of cases, these factors are more important than the rest of the factors and have an impact outside of the number of cases. As these are just auxiliary factors, they do not negatively affect the security of a location in any way. Even if we remove these keeping only the major factors, the region is still considered safe. The range of these factors is from 0 to 25.The first of these is the number of police stations in the neighbourhood. The presence of police stations can sometimes greatly discourage the perpetrators, but these may often be ignored so it is an auxiliary factor.

Considering the range for this, 200m is too small; hence here we are considering an area of 500m radius. This function gets its maximum value first for the number of police stations being three, and then the value remains constant. The equation for it is –

$$y_{PC} = \frac{25}{3} x_{PC}, for x_{PC} \leq 3$$
$$y_{PC} = 25, for x_{PC} > 3$$

The second factor we consider here is the network speed on one's device. This may seem to be unnecessary and unrelated, but the speed at which one can request for help is directly proportional to the network speed they have. Since the speed of 5MBps is enough for this purpose, that is when our function will reach its maximum value. The function is as follows

$$y_{NC} = 5x_{NC}, for\ x_{NC} \leq 5$$
$$y_{PC} = 25, for x_{PC} > 5$$

4. Route Planning

Now that we have our algorithm, we need to have to develop a mechanism for practical route planning. The most commonly used method for route planning is using the Dijkstra's Algorithm which would be most suited for us as well. We plan to modify the algorithm to suit our needs, so we will first explain the algorithm and how we modify it.

Dijkstra's Algorithm

Dijkstra's algorithm (or Dijkstra's Shortest Path First algorithm, SPF algorithm) is an algorithm for finding the shortest paths between nodes in a graph, which may represent, for example, road networks. It was conceived by computer scientist Edsger W. Dijkstra in 1956 and published three years later.

For a given source node in the graph, the algorithm finds the shortest path between that node and every other node. It can also be used for finding the shortest paths from a single node to a single destination node by stopping the algorithm once the shortest path to the destination node has been determined. For example, if the nodes of the graph represent cities and edge path costs represent driving distances between pairs of cities connected by a direct road (for simplicity, ignore red lights, stop signs, toll roads, and other obstructions), Dijkstra's algorithm can be used to find the shortest route between one city and all other cities. A widely used application of the shortest path algorithm is network routing protocols, most notably IS-IS (Intermediate System to Intermediate System) and Open Shortest Path First (OSPF). It is also employed as a subroutine in other algorithms such as Johnson's. The Dijkstra algorithm uses labels that are positive integers or real numbers, which are ordered. It can be generalized to use any labels that are partially ordered, provided the subsequent labels (a subsequent label is produced when traversing an edge) are monotonically non-decreasing. This generalization is called the generic Dijkstra shortest-path algorithm.

Let the node at which we are starting be called the initial node. Let the distance of node Y be the distance from the initial node to Y. Dijkstra's algorithm will assign some initial distance values and will try to improve them step by step.

1) Mark all nodes unvisited. Create a set of all the unvisited nodes called the unvisited set.
2) Assign to every node a tentative distance value: set it to zero for our initial node and to infinity for all other nodes. Set the initial node as current.[14]
3) For the current node, consider all of its unvisited neighbours and calculate their tentative distances through the current node. Compare the newly calculated tentative distance to the current assigned value and assign the smaller one. For example, if the current node A is marked with a distance of 6, and the edge connecting it with a neighbour B has length 2, then the distance to B through A will be 6 + 2 = 8. If B was previously marked with a distance greater than 8 then change it to 8. Otherwise, the current value will be kept.
4) When we are done considering all of the unvisited neighbours of the current node, mark the current node as visited and remove it from the unvisited set. A visited node will never be checked again.
5) If the destination node has been marked visited (when planning a route between two specific nodes) or if the smallest tentative distance among the nodes in the unvisited set is infinity (when planning a complete traversal; occurs when there is no connection between the initial node and remaining unvisited nodes), then stop. The algorithm has finished.
6) Otherwise, select the unvisited node that is marked with the smallest tentative distance, set it as the new "current node", and go back to step 3.

When planning a route, it is not necessary to wait until the destination node is "visited" as above: the algorithm can stop once the destination node has the smallest tentative distance among all "unvisited" nodes (and thus could be selected as the next "current").

Suppose you would like to find the shortest path between two intersections on a city map: a starting point and a destination. Dijkstra's algorithm initially marks the distance (from the starting point) to every other intersection on the map with infinity. This is done not to imply that there is an infinite distance, but to note that those intersections have not been visited yet. Some variants of this method leave the intersections' distances unlabeled. Now select the current intersection at each iteration. For the first iteration, the current intersection will be the starting point, and the distance to it (the intersection's label) will be zero. For subsequent iterations (after the first), the current intersection will be the closest unvisited intersection to the starting point (this will be easy to find).

From the current intersection, update the distance to every unvisited intersection that is directly connected to it. This is done by determining the sum of the distance between an unvisited intersection and the value of the current intersection and then relabeling the unvisited intersection with this value (the sum) if it is less than the unvisited intersection's current value. In effect, the intersection is relabelled if the path to it through the current intersection is shorter than the previously known paths. To facilitate shortest path identification, in pencil, mark the road with an arrow pointing to the relabelled intersection if you label/reliable it, and erase all others pointing to it. After you have updated the distances to each neighboring intersection, mark the current intersection as visited and select an unvisited intersection with minimal distance (from the starting point) – or the lowest label—as the current intersection. Intersections marked as visited are labelled with the shortest path from the starting point to it and will not be revisited or returned to.

Continue this process of updating the neighboring intersections with the shortest distances, marking the current intersection as visited, and moving onto the closest unvisited intersection until you have marked the destination as visited. Once you have marked the destination as visited (as is the case with any visited intersection), you have determined the shortest path to it from the starting point and can trace your way back following the arrows in reverse. In the algorithm's implementations, this is usually done (after the algorithm has reached the destination node) by following the nodes' parents from the destination node up to the starting node; that's why we also keep track of each node's parent. This algorithm makes no attempt of direct "exploration" towards the destination as one might expect. Rather, the sole consideration in determining the next "current" intersection is its distance from the starting point. This algorithm, therefore, expands outward from the starting point, interactively considering every node that is closer in terms of shortest path distance until it reaches the destination. When understood in this way, it is clear how the algorithm necessarily finds the shortest path. However, it may also reveal one of the algorithm's weaknesses: its relative slowness in some topologies.

Proof of Dijkstra's algorithm is constructed by induction on the number of visited nodes.

Invariant hypothesis: For each node v, dist[v] is the shortest distance from the source to v when travelling via visited nodes only or infinity if no such path exists. (Note: we do not assume dist[v] is the actual shortest distance for unvisited nodes.)

The base case is when there is just one visited node, namely the initial node source, in which case the hypothesis is trivial. Otherwise, assume the hypothesis for n-1 visited nodes. In which case, we choose an edge vu where u has the least dist[u] of any unvisited nodes and the edge vu is such that dist[u] = dist[v] + length[v, u]. dist[u] is considered to be the shortest distance from the source to u because if there were a shorter path, and if w was the first unvisited node on that path then by the original hypothesis dist[w] > dist[u] which creates a contradiction. Similarly, if there were a shorter path to u without using unvisited nodes, and if the last but one node on that path were w, then we would have had dist[u] = dist[w] + length[w, u], also a contradiction.

After processing u it will still be true that for each unvisited node w, dist[w] will be the shortest distance from the source to w using visited nodes only because if there were a shorter path that doesn't go by u we would have found it previously, and if there were a shorter path using u we would have updated it when processing u.

After all, nodes are visited, the shortest path from source to any node v consists only of visited nodes, therefore dist[v] is the shortest distance.

5. Modifications

As the cost primarily used in Dijkstra's algorithm is the square of the distance between two nodes, the most appropriate and accurate way to implement our algorithm would be to change the conditions of this cost according to our safety value. The new cost values will reflect the safety of the path instead of the distance. If we solely rely on safety to make the required route, then the route may obtain is too long to be traversable and thus is impractical, completely defeating our purpose.

Another factor we need to consider is that the safety value of which location should be incorporated in the cost. We can consider the safety value of specific nodes, or the area in general. Here let us consider that 2 nodes are far apart. Those two nodes are very safe themselves, close to ideal. However, in the route between the nodes, there is a part where there is no lighting or no people around. So, when we consider the nodes, we would find those to be ideal, but in actuality, the route is not very safe as of the unsafe part. As is it commonly said, ‘a chain is as strong as its weakest link’, thus we will consider the values of all the points on the route, and then the value which we will append with the distance will be the lowest, that is the safety value of the most unsafe place on the route. Now we move to how we combine the distance and safety to develop an algorithm that can transport someone safely and quickly to your destination. The main factor we need to decide here is how much our path cost will change according to our safety function that is what maximum influence the safety can have on the path cost of a route.

With this in mind, we must consider how important safety and distance are in this algorithm. As this is a safety-based route, higher preference will be given to the safety value of the route, without neglecting the distance factor in it. Let us consider the path cost to be p originally. As the cost should decrease with an increase in safety, therefore the safety value should be subtracted from this path cost to give the new cost. Note that here we will be more concerned about the least value a path can get, as compared to its maximum possible value, as the higher value would discourage the use of the path, which is desired in case of high negative safety value. The path value of the repute may drop into negative as in case of high safety value. Thus, the limit set here will be the negative of the original path cost p, i.e. -p will be the minimum value of the path. So, the max influence that the safety value can have on path cost is p – (-p) = 2p. Therefore, we must scale the safety value according to this metric. The maximum value of the safety function, which was 350 will now be scaled to 2p. Thus, the new path cost p_s will be

$$Ps = p - \frac{2p}{350} Ys$$

With this new path cost, we shall use the Dijkstra’s Algorithm as is and use that in our route planning. After the route has been planned, it will show the user the distance to the destination and also the safety along the route a style route progresses.

6. Analysis

a) Influences on Route Choice by Women

With this, women will now have an option of another route besides the fastest route commonly shown in route mapping applications. Overall, research on travel behaviour choice is still relatively young, although making progress in identifying the elements that are associated with higher levels of decision making are progressing rapidly both in the fields of math as well as psychology. Most research in this field has focused not with regards to the safer route, but in comparison to a more familiar route to the newly introduced fastest route. Though as the familiar route is comparatively much safer, we can use that as a parameter to determine how often that route is chosen, and the difference in distance that route is preferred. There appears to be some consensus on some of the primary influences on travel, which are

- Distance and ETA
- The closeness of Relatives/Friends/familiars on the route
- Interdependency of Route
- Familiarity with the route
- Preference of family and friends

b) Research Needs on Travel Mode Influences

More work is still needed to understand more about the influences on travel decisions and to tease out the multiple elements acting concurrently. For example, studies should identify and control for as many factors as possible, such as distance to destination, age, availability of a car, and other conditions, to obtain data on other influences, such as time constraints and trip-chaining priorities. Future studies also should examine the interrelationship between familial perceptions about aspects of the environment (safety, distance, etc), reality, and their decision to let women go out and use the proper travel convenience such as the public metro or bus. This could have implications for program design if the results show that fears are ungrounded in reality and this route could provide reassurance for households to allow women to walk or perhaps use public modes of transport. Also, the influences of these decisions, especially as they mature and are more independent the women are, should be studied. While they may not always choose this way, their experiences can influence their decisions, also that of those around them. More studies that compare male and female perceptions would provide valuable information to shape program elements aimed at the group.

c) Programs to Promote Women Safety

We know more about influences on travel behaviour to a school than we do about the effectiveness of programs, such as Safe Routes, which have been instituted in several communities across India in the form of Familiar routes to promote independent and safer travel. For the most part, these programs have focused on engineering solutions, enforcing government policies and other local rules in crime-prone areas, and programs that educate and encourage people to rise for the betterment of the society themselves. While these programs appear to be gaining popularity and funding, evaluation of results is sparse.

On the program side, the evaluations that have been conducted do not attempt to correlate specific program elements with behaviour change. From this, one can only conclude that efforts, taken together, made a slight difference indicating that there is potential for change given the right impetus and tools for the same. The lack of rigorous evaluation makes it difficult to conclude what aspects of these programs are effective in increasing the number of students who walk or bicycle to school. Evaluation of engineering and infrastructure improvements associated with Safe Routes has focused on the potential for infrastructure improvements to increase walking and cycling by addressing safety concerns, a primary deterrent to walking and cycling cited by interviewees. However, most of these programs and projects have not considered their actual effects on women which need to be documented to address safety concerns with confidence.

d) Research Needs on Program Evaluation

With limited research and implementation to promote the veracity of these routes, it is important to document the types of projects that have the most proven potential to get results. Specifically, we need to know what works and why – both to increase women's safety and the steps taken by the government and the NGOs. To date, evaluation of implementation and encouragement programs does not appear to have progressed beyond documenting travel mode change over time. In some cases, data on travel mode to and from has not been collected at program outset, but rather retro-actively through parent or surveys. Research should inform program design and ideally, evaluation should be built into all programs and projects. This will build the knowledge base by recording and disseminating information about the effectiveness of program elements. The results of the evaluation can provide a base from which to modify existing programs and design new programs to address unmet needs.

However, it is important to note that the parts of the research agenda proposed here may be difficult or infeasible for several reasons. Most of the suggestions for enhanced evaluation require more data, which translates into time and money. This means that funding at the federal level for program evaluation would need to be significantly increased to obtain additional and more detailed data that is already being collected. To isolate the effects of the program elements, communities would need to be willing and able to implement each program element separately. It may not be politically or programmatically feasible to disaggregate the program components in a way that provides the best research setting for evaluation. This could be somewhat overcome by identifying control sites for data collection, realizing that the pairings may not be exactly comparable in terms of the conditions required.

Following are recommendations for program evaluation that could inform and enhance local programs to encourage the safety of Women:

- Collect data before and after program implementation. While some studies obtain data, such as travel mode, before a program was initiated, many rely on recall during post-program data, which is less reliable.

- Include control sites. Much of the evaluation conducted to date has been only with schools receiving the program intervention. Without control sites, it is difficult to discern if other influences are at work.
- Design research that attempts to discern the relative influence of each program element on behaviour change. When several components of Safe Routes are started at the same time, it is difficult to determine which components effected a behaviour change and why. Staggering the implementation of program elements is one way to study the effects of each program element independently.

e) Implications on Design

Programs that promote transportation need to be tailored to address specific influences on travel behaviour and designed to evaluate their effectiveness in affecting those influences to change travel behaviour. Programs need to be informed by and incorporate evaluation into their implementation so funding can be directed to those programs that are proven to be most effective. To accomplish this, stronger working relationships should be forged between the program administrators at the state and local levels, and the scholars who are examining the effectiveness of these programs. Program funding from the central and state levels should require a more rigorous evaluation and provide additional funds to ensure that this happens.

A possible cause for tailored requirements would be local familiarity with the area people live in. Since locals are more familiar with the area, they are more likely to know the hidden characteristics of an area and can also assist with the same. For the same situation in cities, there are a lot of ways to reach your destination, and the user is used to commuting using a route with which she is comfortable, so that factor would be needed to be incorporated based on their data and other statistics, possibly including the data of others using the same route.

f) Implications on Policy and Planning

As stated earlier, many of the influences on travel are beyond the scope of Safe Routes. These programs can influence some aspects of the journey to school mode choice but are not enough to change all the influences on travel behaviour. While public agencies cannot directly affect household characteristics, such as employment or car ownership, they do have the responsibility for any decision that affects the environment. Local governments can enact planning guidelines that affect distance in several ways. They can require inter-connected street networks, sidewalks, and other elements that affect distance and route safety. Both local government and women need to consider policies that affect the spatial distribution of roads and other buildings.

7. Conclusion

In this paper, the authors presented an enhanced version of the Safe Routes for Women. One of the limitations of the former papers was the use of only past incidents, and the limited records made it difficult to calculate the safety level for all the roads. In this research, different angles and factors, namely Illumination, People in the area, Distance from police stations and other safe areas and network speed, and a new weighting method are used to increase the accuracy and precision of the Safe Routes.

A Safe Route Planner is very likely to suggest a safe path that is too long to possibly traverse, like one with highways, as they are safer with fewer incidents. However, this is not acceptable behaviour that threatens the usability of the routes. The authors presented modified Dijkstra's algorithm-based route planning avoidance feature to overcome this issue and users. In future studies, the authors are planning to add a decision-making algorithm to the system. Users will be provided with a set of questions for choosing a route, and then the decision-making algorithm will analyze the answers and decide what path is best for the user.

References

[1] Towers, S., Chen, S., Mailk, A. and Ebert D. 2018. Factors influencing temporal patterns in crime in a large American city: A predictive analytics perspective
[2] O'Hare, E.A. and O'Donohue, W. 1998. Sexual Harassment: Identifying Risk Factors
[3] Arjun, R.K., Shama, Reddy, P. and Yamuna, M. 2015. RESEARCH ON THE OPTIMIZATION OF DIJKSTRA'S ALGORITHM AND ITS APPLICATIONS
[4] Borger G. 2017. Safety First: Perceived Risk of Street Harassment and Educational Choices of Women
[5] Bajwa, M.U., Khan, A. and Nadeem, M. 2018. Empirical Study on Women Safety Concerns at Public Places: Case Study of Lahore City
[6] Weigand, L. 2008. The Effectiveness of Safe Routes to School and Other Programs to Promote Active Transportation to School
[7] Sarraf, R., McGuire, M.P. 2018. An Integrated Safe Route Planner