Comparison of route-finding algorithms in urban graphs: A case study in Alto Selva Alegre, Arequipa, Peru

Gian Franco Condori-Luna
PhD student
Federal University of Sao Paulo (UNIFESP)
Sao Paulo, Brasil
gian.luna@unifesp.br

***Abstract*—Rabies vaccination campaigns in densely populated urban areas require precise planning to maximize coverage and minimize operational costs. This paper presents an analysis and comparison of the Dijkstra, Bellman-Ford, and A* algorithms applied to the road graph of the Alto Selva Alegre district in Arequipa, Peru, to determine optimal walking distances from homes to vaccination points. Based on the article: "Optimizing the location of vaccination sites to stop a zoonotic epidemic", we replicated and optimized part of their methodology using open source tools, achieving a significant reduction in calculation times. This study demonstrates that the use of efficient algorithms can overcome the limitations imposed by calculation platforms such as Leaflet Routing Machine and Mapbox Directions API, offering a more agile and economical alternative for planning public health campaigns.**

I. Introduction

Rabies is a zoonotic disease that affects thousands of people and animals every year, particularly in densely populated urban areas where the control of the dog population is a significant challenge. Annual rabies vaccination campaigns are essential to prevent outbreaks, but their planning presents logistical and economic limitations.

One of the key strategies in these campaigns is to select optimal vaccination sites, both fixed and mobile. Fixed sites remain operational throughout the day in strategic locations, while mobile sites move through high-density canine areas where fixed vaccination sites do not have reach. However, establishing an excessive number of fixed sites significantly increases costs, which requires a methodological approach to identify an optimal number of locations that maximize coverage and minimize walking distances for residents.

In this context, the paper by Castillo-Neyra et al. [1] introduces a method based on calculating walking distances from each home to potential vaccination points. Using the Mapbox Directions API, the authors calculated these distances for an initial set of 70 candidate points, ultimately selecting the 20 most efficient points. However, restrictions of their tool, such as the limit of 20 queries per minute, hampered the speed of the analysis, requiring several days to complete their calculations in certain scenarios.

Identify applicable funding agency here. If none, delete this.

In the present study, we replicate and improve part of this methodology by using open source libraries such as OSMnx and NetworkX, along with clustering techniques using K-Means to simulate potential vaccination points since we did not have that information. In addition, we evaluate and compare the performance of three classic short-route algorithms: Dijkstra, Bellman-Ford, and A*. Our main objective is to demonstrate the efficiency of these tools in terms of execution time, offering a viable and faster alternative for planning vaccination campaigns.

II. Methodology

The methodology used in this study combines open source tools to model urban graphs, clustering techniques to identify key points, and short-route algorithms to calculate optimal distances. The approach allows the performance of the algorithms to be evaluated by considering the execution time of the results. The main stages of the process are detailed below.

A. Generation of the Urban Graph

The Alto Selva Alegre district, located in Arequipa, Peru, was modeled using the OSMnx library, which allows extracting road data from OpenStreetMap (OSM). This tool is widely recognized for its ability to generate highly accurate urban graphs, representing intersections as nodes and connections between them as arcs weighted by geographic distance.

Since GPS coordinate data for the original dwellings in the district used in the base study are not available, it will be assumed that each node in the graph represents a dwelling. Although this approximation reduces the total number of dwellings modeled compared to the actual number, it is adequate as a reference for the purposes of this analysis. With this approximation, it will be possible to calculate the execution time of the method proposed by the authors of the original study. This will allow a direct comparison between this method and the three alternative algorithms proposed in this work.

Figure 1 shows the graph of the Alto Selva Alegre district with its nodes and vertices.

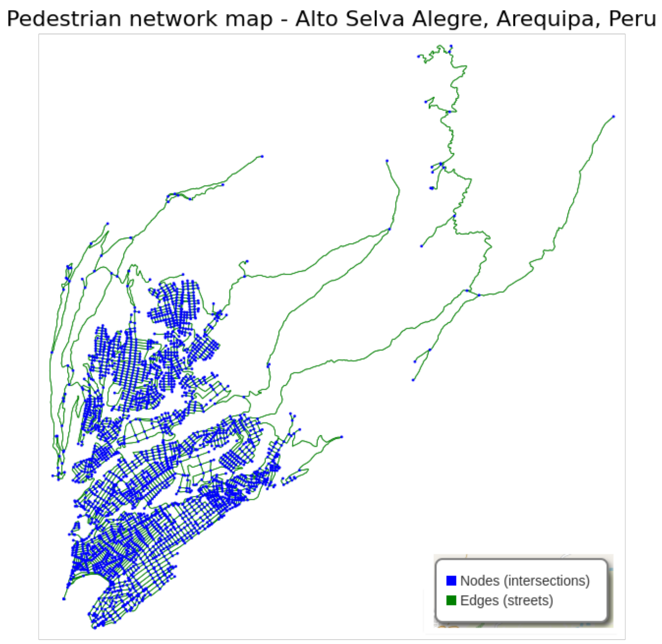

Fig. 1. Graph obtained from the OSMnx library, which allows extracting road data from OpenStreetMap (OSM) with nodes at each intersection and edges containing the weight of the distance walked between each node.

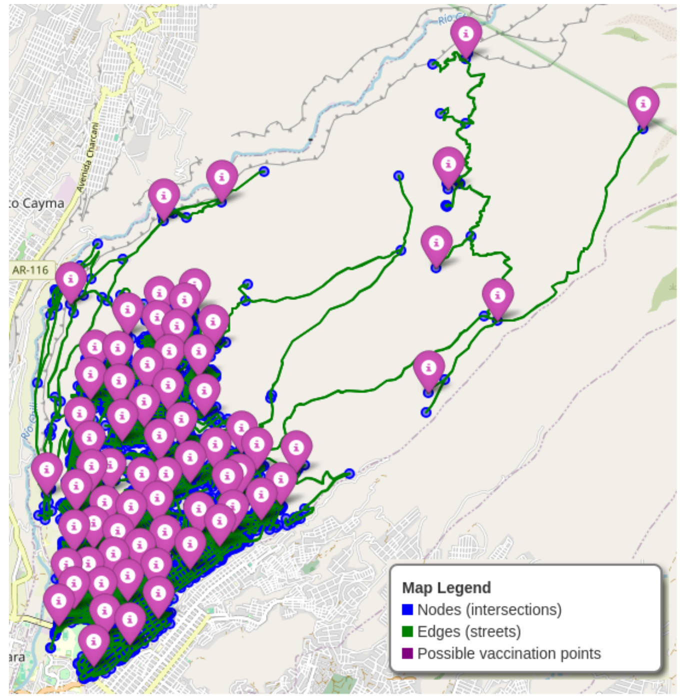

Fig. 2. This map shows the centroids of each cluster obtained by the Kmeans algorithm throughout the Alto Selva Alegre district.

B. Clustering Identification

In the base article, 70 potential vaccination points were selected and determined by Ministry of Health personnel based on their experience and prior knowledge in vaccination campaigns. From these 70 initial points, the article concluded with the identification of 20 fixed vaccination points that optimize the use of resources and maximize coverage.

To simulate the 70 initial points mentioned in the base article, this study used the K-Means clustering algorithm, available in the Scikit-learn library, to divide all nodes in the graph into 70 groups. This approach allowed for clustering based on geographic proximity, providing a reproducible and objective method for identifying potential strategic vaccination points.

The process was carried out as follows:

- **Determining the Number of Clusters:** A total of 70 clusters were selected, corresponding to the initial points described in the base article. This decision ensures consistency in the comparison of results.
- **Application of the K-Means Algorithm:** The geographic coordinates of the nodes in the graph were used as input data for the algorithm. This allowed K-Means to form clusters representing geographically coherent communities within the district.
- **Centroid Identification:** The centroids of the generated clusters were considered as the 70 possible initial vaccination points. These centroids were subsequently used to calculate the shortest walking distances from each node to its corresponding point.

Figure 2 shows the 70 centroids used.

C. Implementation of Short Route Algorithms

To calculate the walking distances from each house (represented as a node) to the potential fixed vaccination points (the previously generated centroids), three classical short-route algorithms were evaluated. Each was selected for its specific characteristics, allowing a comprehensive comparison of their performance in the context of the modeled urban graph.

- **Dijkstra [2]:** This algorithm calculates the shortest paths from a source node to all other nodes in a weighted graph.
  - Advantages: Its simplicity and robustness make it a widely used standard approach in route optimization problems.
  - Limitations: Although efficient on small or moderately sized graphs, its performance may degrade on dense or large graphs due to the exhaustive search in all possible directions. This is particularly relevant in the context of the modeled graph, which includes numerous road connections.
- **Bellman-Ford [3], [4]:** Similar to Dijkstra, this algorithm also finds shortest paths, but it stands out for its ability to handle arcs with negative weights, which makes it especially useful in scenarios where these may exist.
  - Advantages: Its iterative design guarantees the accuracy of the results even in complex graphs with negative arcs.
  - Limitations: Although our graph does not include negative weights, Bellman-Ford was used to evaluate its comparative efficiency. Its main disadvantage lies in its higher computational cost, since it checks all

the edges of the graph in each iteration, which can be less efficient in dense structures.

- **A* [5]:** This algorithm combines Dijkstra's exhaustive search with a heuristic function that estimates the remaining distance to the destination, making it a more efficient choice in large and complex graphs.
  - Advantages: The heuristic applied in this case was the direct Euclidean distance between nodes, which allowed a significant acceleration in the calculations. This is especially useful in dense urban graphs such as the one modeled in this study.
  - Limitations: Although its performance is superior in terms of speed, the quality of the results may depend greatly on the choice and accuracy of the heuristics.

D. Performance Analysis and Evaluation

To assess the efficiency and consistency of the evaluated algorithms, a detailed analysis was developed covering the following key stages:

- **Multiple Iterations:** Each algorithm was run three times, ensuring that the results were consistent and reproducible. During each iteration, the distances from all nodes in the graph to potential vaccination points (centroids) were calculated. This approach allowed us to identify variations in the calculations and ensure that performance was not influenced by external factors, such as system conditions or memory fluctuations.
- **Execution Time Measurement:** Total execution times were recorded for each algorithm when processing the entire urban graph. This data was used to perform a direct performance comparison between the implemented methods.

III. Results

The case study analysis focused on a graph with a total of 3,031 nodes, representing the households in the Alto Selva Alegre district. For each household, the shortest distance to the 70 possible simulated fixed vaccination points was calculated using the method described previously. This resulted in a total of 212,170 individual queries to determine the optimal routes.

The base article approach, which performs a maximum of 20 queries per minute, would require approximately 176.81 hours (7.37 days) to complete all queries. This time was calculated based on the parameters described in that study.

A. Execution Time

The execution times obtained for the algorithms implemented in this work were those shown in Table I, which shows the result of each iteration performed with each of the algorithms:

TABLE I
THIS TABLE SHOWS THE RESULTS OF THE ITERATIONS PERFORMED WITH EACH ALGORITHM.

	Dijkstra	A*	Bellman-Ford
Iteration 1	3834.75 s	3926.30 s	15263.10 s
Iteration 2	3854.02 s	3925.74 s	15000.08 s
Iteration 3	3908.80 s	4013.35 s	14510.25 s

Compared to the method used in the base paper, which takes more than 7 days, the results obtained show a drastic improvement in execution time. This highlights the advantage of using optimized and open source algorithms to tackle similar problems. Table II shows the average results of the three algorithms used together with the result that the authors of the base paper would obtain if they had performed their calculations with our data.

TABLE II
THIS TABLE SHOWS THE AVERAGE RESULTS OBTAINED FROM THE THREE ALGORITHMS USED IN THIS STUDY AND ALSO THE TIME THAT WOULD BE OBTAINED WITH THE METHOD USED IN THE BASE ARTICLE.

	Seconds	Hours
base article	NA	176.81
Dijkstra	3865.85 ± 31.37	1.07
A*	3955.14 ± 41.17	1.14
Bellman-Ford	14924.47 ± 311.96	4.15

IV. Discussion

The implementation of open source tools such as OSMnx and NetworkX in this study has proven to be an economical and efficient alternative to address optimization problems in urban graphs. These libraries not only facilitate the extraction and modeling of spatial data from platforms such as OpenStreetMap (OSM), but also offer flexibility that is especially valuable in research projects and large-scale practical applications.

In contrast, commercial tools such as Mapbox Directions API, although equipped with advanced functionalities and technical support, present significant limitations in scenarios with high computational demand. The costs associated with their use and quota restrictions can be prohibitive for studies that require processing a large volume of data, such as the analysis of optimal routes for thousands of homes.

The evaluation of the performance of the algorithms reveals that, although the A* algorithm presents an outstanding combination of accuracy and speed thanks to its heuristic approach, recent analyses indicate that Dijkstra's algorithm is slightly superior in terms of efficiency for the case studied. Dijkstra demonstrates greater consistency when calculating optimal routes in dense graphs, which positions it as a slightly more robust alternative for applications such as planning vaccination campaigns and other community-level logistics interventions. This finding underlines the importance of evaluating multiple approaches to identify the most suitable solution based on the specific characteristics of the problem and the available resources.

Finally, the combination of reproducible approaches, open-source tools, and advanced algorithms underlines the importance of democratizing access to technological solutions. These methodologies can be easily adapted to other urban contexts, broadening their impact on critical problems such as resource distribution in vulnerable areas.

V. Conclusion

This study demonstrates the effectiveness of short-route algorithms applied to urban graphs as a strategy to optimize the selection of vaccination points in complex urban contexts. The integration of open source tools, such as OSMnx and NetworkX, together with reproducible methodologies, allowed not only to model the road infrastructure accurately, but also to significantly reduce computation times compared to other methods.

Although the A* algorithm is theoretically presented as an ideal solution due to its combination of accuracy and speed through the use of heuristics, the results obtained in this study revealed that Dijkstra's algorithm offers slightly superior performance in the specific case analyzed. This finding highlights the importance of rigorously evaluating different approaches to identify the most appropriate solution according to the characteristics of the problem and the urban environment.

The results obtained confirm that the proposed approach is not only viable, but also scalable, offering a practical and economical solution for planning public health campaigns. The flexibility of these methodologies allows their adaptation to other cities or regions, expanding their applicability to different geographic and logistical contexts.

Furthermore, the approach focused on the use of open data and efficient algorithms underlines the potential of accessible technologies to address complex urban challenges. This study represents a step towards the development of data-driven tools that can improve decision-making in public policies, especially in critical areas such as health, transportation and the distribution of essential resources.

VI. Future Research

The development of this work opens up multiple opportunities for future research. One promising line is to explore hybrid models that integrate short-route algorithms with multi-objective optimization techniques, which would allow incorporating additional factors such as population density, accessibility to essential services, and socioeconomic characteristics of the areas studied. This approach could offer more comprehensive solutions, especially in contexts where location decisions must balance multiple criteria.

It is also suggested to perform a more detailed analysis on the impact of different heuristics on the performance of the A* algorithm, especially in larger-scale and more complex urban graphs. Evaluating alternative heuristics, such as those based on historical traffic data or human mobility patterns, could improve the efficiency of the algorithm and better adapt it to dynamic scenarios.

Another relevant direction would be to establish a framework to compare the results obtained with those reported in the base article, which would allow evaluating the relative accuracy of the proposed solutions. Although this analysis was not possible in the present study due to the lack of access to accurate data and the considerable time that such an effort would require, its future implementation would provide a more rigorous validation and strengthen confidence in the developed methods.

Finally, the implementation of these methodologies in real environments, such as cities with more complex infrastructures or in health emergency situations, would provide valuable data to validate and refine the proposed models, expanding their applicability and robustness.

References

[1] R. Castillo-Neyra, S. Xie, B. R. Bellotti, E. W. Diaz, A. Saxena, A. M. Toledo, G. F. Condori-Luna, M. Rieders, B. B. Bhattacharya, and M. Z. Levy, "Optimizing the location of vaccination sites to stop a zoonotic epidemic," *Scientific Reports 2024 14:1*, vol. 14, pp. 1–11, 7 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-66674-x

[2] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269–271, 12 1959. [Online]. Available: https://dl.acm.org/doi/10.1007/BF01386390

[3] L. R. Ford, "Network flow theory." 1956. [Online]. Available: https://www.rand.org/pubs/papers/P923.html

[4] R. Bellman, "On a routing problem," *Quarterly of Applied Mathematics*, vol. 16, pp. 87–90, 4 1958. [Online]. Available: https://www.ams.org/qam/1958-16-01/S0033-569X-1958-0102435-2/

[5] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, pp. 100–107, 1968.