Optimization of New Pick-up and Drop-off Points for Public Transportation

Cristiano Martins Monteiro¹, Flávio Vinícius Cruzeiro Martins², Clodoveu Augusto Davis Junior¹

¹Computer Science Department – UFMG
²Computing Department – CEFET-MG

{cristianomartinsm,clodoveu}@ufmg.br, flaviocruzeiro@decom.cefetmg.br

Abstract. The expansion of cities, together with the advance of technological resources, has motivated the study of improvements for metropolitan dynamics. Among these improvements are those aiming to facilitate and to speed up the population’s routine activities. This work proposes and compares two methods to optimize the location of new pick-up and drop-off points in order to avoid long walks to get to a bus stop. Real datasets of the road network and bus stops in the city of Belo Horizonte were used. Results indicate the effort required by the city’s transit system to provide transit pick-up and drop-off points in a reasonable quantity and location, in order to improve the quality of the service rendered to the population.

1. Introduction

Commuting to school, work, shops and other points of interest is a common routine of metropolitan inhabitants (Logiodice et al., 2015). Due to the constant increase in urban population¹ and the greater difference between the urban and rural population², the study of improvements for urban transportation services is important to achieve efficiency and accessibility, mainly for residents in peripheral regions.

Commuting in urban areas is commonly performed by means of the public transit system. In the city of Belo Horizonte, Brazil, public transportation comprises bus, taxis and metro rail. The city’s metro rail has only one operating line, whose extension is 28 kilometers. This extension is relatively small, given the city has an area of 331 km² (Garrides et al., 2016) and maintains a road network of 9,047 kilometers. Therefore, in Belo Horizonte, buses and taxicabs (including shared ride services, such as Uber) stand out in relation to metro rail because they have greater coverage throughout the city.

Although bus fares are usually cheaper than the taxicab fares, some taxi services have attractive prices for taxipooling. In this option, passengers with different pick-up or drop-off points but with part of the route in common can share the same trip and pay lower fares. Such transport services based on pooling, as well as buses, could provide greater accessibility and attract new passengers if there are more embarkation and disembarkation points (also known as pick-up and drop-off points) distributed throughout the city (Loader and Stanley, 2009). This work aims to optimize the location of these new points in the city.

¹http://data.worldbank.org/indicator/SP.URB.TOTL
²http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS
of Belo Horizonte. An optimization algorithm was proposed to suggest these locations in order to benefit as many of the city’s neighborhoods as possible.

A real dataset of the Belo Horizonte’s road network was used to create a graph with all streets from the city, on which two optimization processes for new pick-up and drop-off points have been executed. Given the size of the dataset and the computational effort needed to perform an optimization process over such a large graph, the implementation has to use efficient data storage and retrieval methods. The optimization method proposed in this work is compared to an adapted version of the algorithm presented by Takakura et al. (2015). Results indicate the effectiveness of the proposed method in relation to the compared work, and the significant improvements that optimization can provide to the public transportation in Belo Horizonte. A broader distribution of pick-up and drop-off points throughout the city would especially benefit residents of neighborhoods far from downtown (Veras et al., 2016). The suggestion of these new pick-up and drop-off points can work as a basis for additional bus lines, or to propose the location of new taxi stands.

This paper is organized in five sections. Next section discusses related works. Section 3 presents the used datasets and the methodology. Section 4 explains the proposed optimization methods. Section 5 shows the results, and Section 6 concludes this paper.

2. Related Work

Recent works use open data to analyze the public transport operation or propose improvements in the location of bus stops and taxis. Some of these works apply meta-heuristics or stochastic processes to achieve their goals. Other works aim to explain city dynamics through public transportation data.

Logiodice et al. (2015) proposed an inaccessibility index to measure distance and time spent commuting between zones of a city. The authors applied this index on the São Paulo Metropolitan Area and contextualized results with the amount of trips made in each zone, and the average income of its population. The authors showed that areas with high inaccessibility have low income residents and are often peripheral.

Another exploratory analysis was presented by Kozievitch et al. (2016), who analyzed the bus service from Curitiba, Brazil, from the perspective of pattern discovery, statistical analysis, data integration, and the use of connected and open data. The density of routes of various buses lines in Curitiba was also discussed, and temporal patterns of bus fares paid with public transport cards were analyzed.

Spatial and temporal patterns from public transport were also analyzed by Monteiro et al. (2016). The authors have studied the San Francisco (USA) and Rome (Italy) regions with greater flow of taxicabs, presented the variations of taxi operation along the week, and determined the ten most common places for pick-up and drop-off in San Francisco. Although the downtown area in both cities show higher supply and demand of taxi services, taxi trips were also performed in the peripheral regions.

Silva Júnior et al. (2016) presented a spatial and temporal analysis of the taxi service in Belo Horizonte. Data were obtained from the WayTaxi³ app, and included one week of taxi calls, completed taxi trips, and cancelled calls. Among other results, the

³http://www.waytaxi.com
authors found out that 77% of the demanded routes were serviced by taxi drivers who were at 500 meters or less away from the passenger pick-up location. Besides, only 5% of the routes were serviced by taxi drivers whose locations were more than one kilometer away from the passenger. This indicates that, even though a passenger can call a taxi to pick him up exactly where he is, if there are no taxis nearby, it is unlikely that the passenger’s trip will be started.

The driving distance between the taxi driver and the passenger was the subject of research by Oliveira et al. (2015). The authors compared optimization techniques in order to define the best driver to respond to a call from a passenger. The algorithms aimed to minimize time and driving distance needed to move from the taxi driver’s location to the passenger’s location. The alternative that applies the Hungarian Algorithm for optimization, and considers the shortest path as its optimization distance (instead of Euclidean distance between the taxi driver and the passenger), achieved the best results.

In addition to taxis, walking distance to a bus stop in Belo Horizonte was analyzed by Veras et al. (2016). Data from an origin-destination survey from 2012 were used to analyze the Accessibility Index of the city. The lower the walking distance to a bus stop and the waiting time until the bus arrives, the greater (and better) is the Accessibility Index for a given region. According to the literature analyzed by Veras et al. (2016), the accessibility of a region is not considered to be bad if the walking distance to a bus stop is, on average, less than or equal to 500 meters. For the waiting time, the threshold is 12 minutes. The analysis indicates that the 500 meters threshold is exceeded in most of Belo Horizonte, and there are discrepancies on this distance even for neighboring regions.

A Genetic Algorithm was proposed by Takakura et al. (2015) to define the location of ten new bus stops for the city of Nonoichi, in Japan. The authors’ goal was to minimize the walking distance between students dormitories and the nearest bus stop with connection to the Kanazawa Institute of Technology. The best solution achieved by the authors would reduce walking up to 702 meters.

Nalawade et al. (2016) proposed an optimization method for bus stops spacing in the city of Aurangabad, India. The proposed method runs in a specific way for each category of bus stops. Three categories of bus stops were defined: Connection, Key, and Ordinary. Connection bus stops are those near airports, railway stations or other means of transportation, they are maintained in this process. The optimization of location of Key bus stops reallocates the other bus stops aiming to minimize the global distance among them. The optimization of the location of Ordinary bus stops allocates them aiming to maximize the covered population. The optimization of Key and Ordinary bus stops use the Random Walk technique to explore the nodes and edges from the city’s road network.

This paper differs from the related work by proposing a method to optimize the location of new pick-up and drop-off points (NPDPs), in order to improve the accessibility in the city of Belo Horizonte. The term Pick-up or Drop-off Point (PDP) will be used in this work to abstract the concept of bus stops, taxi stands, stops for taxipooling, as well as other places destined for people to get in or out of a public transport vehicle. The results of this study can be useful for public transit companies, taxi services (including taxipooling and competitors like Uber), and other private initiatives such as Buser. The dataset used and the proposed optimization method are described in the next section.
3. Dataset and Methodology

This section presents the dataset used in this study and the applied methodology. The Subsection 3.1 describes the road network data and the bus stops location dataset from Belo Horizonte used in this work, as well their treatment and integration. And the Subsection 3.2 explains the neighborhood selection to the optimization.

3.1. Data Treatment and Integration

In this work, two datasets from Belo Horizonte were integrated: the road network\textsuperscript{4}, and the bus stops locations, maintained by the public transport company BHTrans\textsuperscript{5}. Figure 1 illustrates two different regions from the city using these datasets. Figure 1 (a) shows the road network (points and lines in yellow) and the bus stops (purple points) near the Pampulha lake. Figure 1 (b) illustrates the edges and bus stops in the downtown region of Belo Horizonte. The circular edges near the image’s center surround Raul Soares square, a well-known local landmark.

![Figure 1. Illustration of the datasets](image-url)

Figure 1 (a) shows that the location of bus stops usually does not touch the edges from the road network dataset, and that the distance between a bus stop and an edge varies. In some cases (especially for edges at a corner), is unclear to define in which edge the bus stop is located. In many of these cases, both the street and bus stop location can be correct, because a bus stop in Belo Horizonte can vary from a simple sign fixed in a pole, to a Bus Rapid Transit (BRT) station. But, if the location is incorrect, it is impossible to assert in an automatic way whether the error is on the road network data, or on the bus stops dataset. According to Monteiro et al. (2017), errors while matching a point to a street (provided by different datasets) can happen due to factors such as: (i) GPS system inaccuracy when calculating the coordinates of the point, street or both; (ii) misunderstandings when recording or proving the data; (iii) missing streets or streets recorded with the wrong direction.

Since the georeferenced data integration is subject to different sources of error, the optimization method proposed in this work simplifies the location of bus stops and PDPs

\textsuperscript{4}https://geodadosbh.pbh.gov.br/
\textsuperscript{5}http://servicosbhtrans.pbh.gov.br/bhtrans/e-servicos/S43F01-extracao.asp
by simply matching them to the nearest edge. With this matching, some edges had more than one bus stop. About 12% of the bus stops were located at an edge with two or more bus stops. This overlapping of bus stops is considered reasonable, because even tough a small street fragment can have more than one bus stop, usually each bus stop in the street fragment will support different bus lines, providing after all only one bus stop for each bus line. The next Subsection presents the selection of neighborhoods for the optimization.

3.2. Selection of Neighborhoods
As analyzed by Veras et al. (2016), the number of bus stops by region in Belo Horizonte has a high variation. Nowadays, there are neighborhoods with as few as one bus stop for every two kilometers of streets on average, and neighborhoods with one bus stop at every 200 meters on average. Neighborhoods with a high number of bus stops are not the optimization focus in this work. Therefore, neighborhoods with more than one bus stop for every 800 meters of streets were not selected to be optimized.

One bus stop for every 800 meters of streets implies that, for each address contained in that neighborhood, there will be, on average, a bus stop 400 meters away to the left and another bus stop at same distance to the right. That threshold was chosen because, as mentioned by Veras et al. (2016) and Nalawade et al. (2016), 400 meters of walking to get to a bus stop is considered reasonable. Considering this threshold, 299 neighborhoods were selected to be optimized. Among the neighborhoods not selected are those located in the central region of the city, and neighborhoods such as “Gameleira”, “Campus UFMG”, “São Gabriel” and “Venda Nova”, which act as regional centers away from downtown. The following section presents the proposed optimization method.

4. Optimization Method
This section presents the proposed method to optimize the location of NPDPs. The optimization is based on the Simulated Annealing algorithm (Kirkpatrick et al., 1983). This algorithm is a meta-heuristic that explores a search space looking for the optimal solution for a given problem.

Initially, the Simulated Annealing generates a random solution for the problem. From this generated solution, the algorithm evaluates neighbor solutions walking on the search space towards an optimal solution. In order to prevent the algorithm from getting stuck at a local maximum or minimum, the optimization process also allows (momentarily) solutions worse than the best one found so far. In the Simulated Annealing, the decision to allow a worse solution is given based on a probability. This probability varies along the optimization, leading to larger movements through the search space at the beginning of the optimization (to speed up the process), and smaller movements through the search space at the end of the optimization (to refine the best solution found). This algorithm is inspired in the annealing materials process, motivating the name Simulated Annealing (Kirkpatrick et al., 1983).

Meta-heuristics are useful to optimize the location of bus stops, for example, when the optimizing area is extensive. An exact method for combinatorial optimization of new bus stops location would be to perform Breadth First Search on the Belo Horizonte’s road network graph, restarting the search from every existing bus stop. This procedure would have time complexity order of $O(|P| \times (|V| + |E|))$, where $|P|$ is the amount of
existing bus stops, \(|V|\) represents the number of nodes in the road network, and \(|E|\) is the number of edges. Regarding the 299 selected neighborhoods to be optimized, \(|P| = 9,428\), \(|V| = 127,196\) and \(|E| = 201,816\). Keeping one instance of this graph for each individual of a meta-heuristic like a Genetic Algorithm would make the method impracticable due to memory issues. The Simulated Annealing algorithm was chosen because it is a non-population-based meta-heuristic, well suited for this problem. The proposed optimization method that uses Simulated Annealing was implemented in two ways, described in Sections 4.1 and 4.2

4.1. Random Walk

This optimization method aims to maximize the reach of NPDPs, i.e., the length of street segments served by each NPDP. Reach maximization was performed using Random Walk on the road network’s graph, following Nalawade et al. (2016), to optimize the spacing between bus stops.

A neighbor solution for Simulated Annealing would consist in adding or removing a NPDP. However, in our approach existing bus stops are not updated or removed. The task to add a NPDP consists on performing a Random Walk (following the direction of graph edges) starting near a existing bus stop, and defining a NPDP for every 400 meters that are walked without coming by any existing bus stop, and without crossing edges or nodes previously walked to add another NPDP. Figure 2 illustrates this procedure.

![Random Walk suggesting a NPDP](image.png)

**Figure 2. Random Walk suggesting a NPDP**
Figure 2 (a) displays existing bus stops using purple points, red lines indicate the street segments matched to a bus stop (as described in Section 3.1), and the red point is the starting node of the Random Walk. This starting node is chosen randomly among the nodes close to an existing bus stop. Hence, there will be at least one existing bus stop connecting to the suggested NPDP. This enables using NPDPs to create new bus stops, for example. To prevent a NPDP from being located too close to an existing bus stop or previous NPDP, the Random Walk is restarted whenever an edge with a bus stop or NPDP is crossed, as illustrated in Figure 2 (b). A NPDP is suggested after at least 400 meters were randomly walked without finding bus stops, NPDPs, or repeating edges already walked when suggesting a previous NPDP. Figure 2 (c) illustrates this procedure. The method also verifies if the NPDP is being located right before other NPDP or bus stop. Therefore, it checks whether the Random Walk’s next step will not have an existing bus stop or NPDP. If the next step is also clear, the NPDP is added as illustrated by the green point in Figure 2 (d).

The parameter of 400 meters was chosen because Nalawade et al. (2016) also used it in their optimization process applying Random Walk. When looking for a location for the NPDP, if the Random Walk reaches a sink node (node without outgoing edges) or becomes stuck in a cycle (walking more than 100 steps without suggesting a NPDP), the NPDP addition is canceled. The objective function is to maximize the sum of walked distances without crossing a existing bus stop, or reaching a NPDP previously suggested, or visiting an edge or node already visited while suggesting a previous NPDP. The fitness function used is the same as the objective function. The following Subsection presents a version of this method based on the work of Takakura et al. (2015).

4.2. Grid-Based Random Walk
This method is based on the procedure proposed by Takakura et al. (2015) to define new bus stop locations. The method uses a grid with a predefined number of cells, where each cell represents a city region uniformly divided. For each region, is suggested only one bus stop. This technique ensures that the new bus stops are distributed throughout the city, avoiding bus stop concentrations in small areas.

For the city of Belo Horizonte and the problem of suggesting new pick-up and drop-off points, would be necessary 8,050 NPDPs in order to achieve the average of 400 meters walked to get to a bus stop in the selected neighborhoods. This number of 8,050 NPDPs was got by dividing the streets length by the desired walking distance. Therefore, a grid with 8,050 cells (one for each NPDP) must be created. According to the grid-based approach proposed by Takakura et al. (2015), each cell must have four candidate points to become a NPDP. The optimizing process consists in defining which candidate point will be selected as the suggested NPDP.

However, cells located on lakes or buildings can be away from streets. This reduces the diversity of bus stops, because probably there will be another candidate point closer to the nearest street. Figure 3 illustrates this situation near Pampulha lake. Each point indicates a candidate point, but only the green points are near a street segment (edge on the graph). The red points would be avoided along the optimization process, because there is a green point closer to a street segment.

This method was adapted to the Random Walk as follows: beyond the conditions
of walked distance without NPDPs (mentioned in Section 4.1), the suggested NPDP must be close to a candidate point in the grid. The NPDP is considered close to a candidate point in the grid if both are matched to the same edge, as mentioned in Section 3.1. Therefore, although the grid-based approach ensures a more spaced distribution of NPDPs throughout the city, it also reduces the possible locations for NPDPs.

The following section presents a comparison of the results obtained using the Random Walk and the grid-based implementations to the NPDP location problem.

5. Results

This section presents the results found by the optimization process. Empirical tests were made to define a set of parameters that enable the methods to converge. The defined parameters were: initial temperature = 1,000; iteration number for each temperature = 1,000; temperature decreasing rate = 0.9; and minimum temperature = $10^{-10}$.

Figure 4 shows the convergence curve after 40 executions of the optimization methods Random Walk and Grid-Based Random Walk. Figure 4 (a), presents the evolution of the solution with best fitness score in each execution of both optimization methods. Figure 4 (b) presents the number of NPDPs suggested. The “Temperature” axis of both graphs is in logarithmic scale. The curve in both graphs shows saturation when the temperature reaches a value of about 10. After this temperature, the fitness score and the number of NPDPs keep increasing, but more slowly.

The Random Walk method presented in Section 4.1 achieved better fitness scores. The Grid-Based Random Walk method achieved fitness scores close to those from the Random Walk method (only 7% of difference between the best results from both) but using much fewer NPDPs (23.8% less). This large difference for the number of NPDPs is due the location constraint imposed by the grid, in which the suggested NPDP must coincide with a point in the grid. This suggests that the Grid-Based method may be
suitable for multi-objective versions of this optimization problem, in which the goals are to maximize the reach of NPDPs and minimize the number of NPDPs required.

Table 1 presents the basic statistics of both optimization methods. The Grid-Based Random Walk had a lower standard deviation for the fitness scores and number of NPDPs. This lower variation is probably due to the the grid-imposed reduction of candidate locations for NPDPs. The values on the table varied linearly from the lowest to the highest. The absence of strong variations among the 40 executions of each method indicates that the execution samples generalize well the solutions generated by both proposed methods.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Best fitness score</th>
<th>Number of NPDPs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Random Walk</td>
<td>Grid-Based R.W.</td>
</tr>
<tr>
<td>Lowest</td>
<td>3,165.92</td>
<td>2,952.09</td>
</tr>
<tr>
<td>1st Quartile</td>
<td>3,202.07</td>
<td>2,982.1</td>
</tr>
<tr>
<td>Average</td>
<td>3,221.12</td>
<td>2,999.33</td>
</tr>
<tr>
<td>Median</td>
<td>3,221.95</td>
<td>3,003.16</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>3,239.43</td>
<td>3,011.36</td>
</tr>
<tr>
<td>Highest</td>
<td>3,282.84</td>
<td>3,053.45</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>25.93</td>
<td>23.9</td>
</tr>
</tbody>
</table>

Figure 5 compares the boxplots from the results of both optimization methods presented in Table 1. Figure 5 (a) presents the best fitness scores achieved and Figure 5 (b) shows the number of NPDPs suggested. In each figure, the boxes from the boxplots do not overlap. This indicates that there is significant statistical difference between the results (Krzywinski and Altman, 2014). Therefore, it can be asserted that the Random Walk method achieved better fitness scores than the Grid-Based Random Walk, and that the Grid-Based Random Walk generates less NPDPs than the Random Walk method.

Figure 6 illustrates the optimized Belo Horizonte neighborhoods using the best solution found. The following analysis is based on the solution with a fitness score of 3,282.84, found by the Random Walk method. Figure 6 (a) presents the neighborhoods selected to be optimized, as described in the Section 3.2. The red neighborhoods have
participated of the optimization process, and the green neighborhoods already have one bus stop for every 800 meters of streets or less. This measure is represented in the figure as Average Distance Between Bus Stops (ADBBS). After the optimization process, in 71 of the 299 neighborhoods the ADBBS stayed above 800 meters. Together, these 71 neighborhoods cover an area of 37 km$^2$, the equivalent to 11.16% from the 331.4 km$^2$ of Belo Horizonte. Nevertheless, the length of streets in these 71 neighborhoods represents only 6.5% of the total length of streets in Belo Horizonte. The lower street length in these 71 neighborhoods hampered the access and generation of NPDPs by the Random Walk.
In spite of this limitation, some of these regions with ADBBS ≥ 800 meters are slums whose streets can even not be passable by bus, and not being useful for this mean of transport. The solution with best fitness score would provide to the public transportation of Belo Horizonte the contributions listed on Table 2. Therefore, considering that the NPDPs can work as bus stops (for example), the creation of 7,076 NPDPs (equivalent to 75.1% of the existent bus stops) would reduce the average walking distance to reach a bus stop, improving the accessibility of 76.3% of the neighborhoods to a regular level, as defined by Veras et al. (2016) and Nalawade et al. (2016). In relation to the whole city, these NPDPs would reduce the ADBBS in 39%. But, in relation to the street length from neighborhoods that had ADBBS ≥ 800 meters, there was a reduction of 92.8%.

Table 2. Impact of the generation of additional NPDPs

<table>
<thead>
<tr>
<th>Measures</th>
<th>Without R.W.</th>
<th>With R.W.</th>
<th>Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of bus stops + NPDPs</td>
<td>9,428</td>
<td>16,504</td>
<td>+75.1%</td>
</tr>
<tr>
<td>Average distance for bus stop or NPDP</td>
<td>959.57 m</td>
<td>585.59 m</td>
<td>-39%</td>
</tr>
<tr>
<td>Number of neighborhoods to be improved</td>
<td>299</td>
<td>71</td>
<td>-76.3%</td>
</tr>
<tr>
<td>Total area of improved neighborhoods</td>
<td>280.1 km²</td>
<td>37 km²</td>
<td>-86.8%</td>
</tr>
<tr>
<td>Street length to be improved</td>
<td>8,195 km</td>
<td>587.9 km</td>
<td>-92.8%</td>
</tr>
</tbody>
</table>

Therefore, the proposed optimization methods achieved their goal to optimize the locations for New Pick-up and Drop-off Points. The best solution found can improve the accessibility of the public transportation in Belo Horizonte.

6. Conclusion

Optimization algorithms for public transportation compose a recent and promising target of research. Simulated Annealing was effective in performing the optimization without need to keep in memory a population to evolve, as Genetic Algorithms do. In this approach, it would be necessary to maintain in memory one copy of the city’s road network for each individual of the population. Therefore, given Belo Horizonte’s road network size, population-based meta-heuristics wouldn’t scale up well due to memory issues.

The Random Walk optimization method achieved better fitness scores than the Grid-Based Random Walk. However, the location restrictions for NPDPs in the grid-based method enabled the optimization to reach fitness scores close to the ones obtained by the Random Walk method, but using much fewer NPDPs. This indicates that the Grid-Based Random Walk can be attractive for a multi-objective version of this problem.

Although the Random Walk has been effective in suggesting NPDPs for Belo Horizonte, its dependency on following the road network hampered the access to neighborhoods with fewer registered streets. As future work, we suggested proposing and comparing a method that is not dependent on the random walk to reach peripheral regions of the city. Another future work is to use a measure based on the local average of demographic density, instead of the fixed threshold of 400 meters walked to a bus stop. We also propose implementing a multi-objective version for this optimization with a goal to also minimize the number of NPDPs defined. Finally, future work should evaluate questions on performance and scalability for the optimization process, especially for more extensive regions, such as the Metropolitan Region of Belo Horizonte, and not only the municipality.
7. Acknowledgements

The authors acknowledge the support from CNPq and FAPEMIG, Brazilian agencies in charge of fostering research and development.

References


