Federal University of Ouro Preto
Institute of Exact and Applied Sciences
Department of Computer and Systems

Students clustering approaches aiming to minimize the bus stop number: a comparative study in real problems

Henrique Queiroz do Amaral

TRABALHO DE CONCLUSÃO DE CURSO

ORIENTAÇÃO:

Rafael Frederico Alexandre

COORIENTAÇÃO:

Fernando Bernardes de Oliveira

Julho, 2019

João Monlevade–MG

Henrique Queiroz do Amaral

Students clustering approaches aiming to minimize the bus stop number: a comparative study in real problems

Orientador: Rafael Frederico Alexandre

Coorientador: Fernando Bernardes de Oliveira

Monografia apresentada ao curso de Sistemas de Informação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina "Trabalho de Conclusão de Curso II".

Universidade Federal de Ouro Preto

João Monlevade

Julho, 2019

A485s Amaral, Henrique Queiroz.
Students clustering approaches aiming to minimize the bus stop number: [manuscrito]: a comparative study in real problems / Henrique Queiroz Amaral. - 2019.

35f.:

Orientador: Prof. Dr. Rafael Frederico Alexandre.
Coorientador: Prof. Dr. Fernando Bernardes de Oliveira.

Monografia (Graduação). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Aplicadas. Departamento de Computação e Sistemas de Informação.

1. Algoritmos . 2. Estudantes - Análise por agrupamento. 3. Levantamentos de rotas - Transporte escolar . I. Alexandre, Rafael Frederico. II. Oliveira, Fernando Bernardes de. III. Universidade Federal de Ouro Preto. IV. Titulo.

Catalogação: ficha.sisbin@ufop.edu.br

CDU: 004.775

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIENCIAS EXATAS E APLICADAS
DEPARTAMENTO DE COMPUTACAO E SISTEMAS

FOLHA DE APROVAÇÃO

Henrique Queiroz do Amaral

Students clustering approaches aiming to minimize the bus stop number: a comparative study in real problems

Membros da banca

Rafael Frederico Alexandre - Doutor - Ufop
Fernando Bernardes de Oliveira - Doutor - Ufop
Mateus Ferreira Satler - Doutor - Ufop
George Henrique Godim da Fonseca - Doutor - Ufop

Versão final
Aprovado em 12 de Julho de 2019

De acordo

Professor (a) Orientador (a)

Documento assinado eletronicamente por **Rafael Frederico Alexandre**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 19/10/2020, às 15:41, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 , informando o código verificador **0093854** e o código CRC **E5344286**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.007922/2020-97 SEI nº 0093854

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: - www.ufop.br

It is mine genuine gratefulness and warmest regard that I dedicate this work to my mother Ana and my father José, who made everything possible, from the very beginning until this unforgettable momment in my life. I am also thankful to my supervisor, advisor, professor and friend Rafael, who trusted and guided me througth the most important moments in my graduation, with patience, severity, attentiveness and motivational support.

Acknowledgements

To perform the planning, development and execution of this work, I am thankful to the Federal University of Ouro Preto's Department of Computing and Systems, which promoted and supported the research and provided optimization tools that lead to its execution. In special to my advisor Rafael Alexandre, who started and leaded the project. Also to my co advisor Fernando Oliveira, who guided us through the execution. And last but not least, to all my family that always supported me and made me the person I have become. Special thanks to mother and father.

"The great enemy of knowledge is not ignorance, it is the illusion of knowledge."

— Stephen Hawking (1942 – 2018),
in: Interview - 2001

Abstract

The Bus Stop Selection Procedure (BSSP) is a School Bus Routing Problem (SBRP) subproblem whose objective is to cluster students at bus stops. This work introduces some BSSP strategies, applied in real georeferenced data of Brazilian cities. The most successful BSSP strategy presented in this work, the Isochrone Approach, returns from a single point in the map all the known points reachable inside the maximum distance allowed for walking. The comparison with another recent strategy in the literature presented relevant results reducing the number of visitation points. Reducing the visitation points, lower the Routing subproblem's complexity tends to be.

Key-words: School Bus Routing Problem, Clustering students, Bus Stops.

Resumo

O Problema de Seleção de Pontos de Ônibus (PSSO) é um subproblema do Problema de Roteamento de Veículos Escolares (PRVE), cujo objetivo é agrupar estudantes em pontos de ônibus. Este trabalho busca introduzir estratégias para o PSSO, aplicadas em uma base de dados georeferenciados de cidades Brasileiras. A estratégia mais bem sucedida para o PSSO a ser apresentada neste trabalho é a Estratégia Isócrono. Ela retorna de um único ponto referencial no mapa, todos os pontos conhecidos e alcançáveis, dentro de uma distância máxima que restringe o quanto um aluno pode caminhar, de acordo com sua capacidade locomotiva. A comparação com outro trabalho recente na literatura apresentou resultados relevantes na redução do número de pontos de ônibus do problema. Quanto menos pontos para visitação, menor tende a ser a complexidade do subproblema de Roteamento sequente.

Palavras-chaves: School Bus Routing Problem, Clustering students, Bus Stops.

List of Figures

Figure 1 – Rural environment. Students in green triangles and roads in brown. . . 15

Figure 2 – Urban environment. Students in green triangles and roads in brown. . . 16

Figure 3 – Isochrone Approach. The reference Student (orange triangle) and the reachable vertices inside the distance constraint (blue lines) 21

Figure 4 – Percentual proportion (Y axis) between approaches per city (X axis). The further from 1 (Y axis), the greater the difference between approaches. 29

Figure 5 – One of the executions before the bus stops selection. Rural environment characteristics. 30

Figure 6 – The execution after the bus stop selection. Students are associated to bus stops. Lines on red represents the students walked path between their homes and the bus stops. 31

List of Tables

Table 1 – Sets, parameters and variables 18
Table 2 – Combination of Selection and Allocation Approaches used on each instance 26
Table 3 – Results Table - For each city, following the Objective Function presented, the bus stop number is highlighted in blue. 28

Contents

1 INTRODUCTION 12
1.1 Objectives 13
1.2 Structure of this project 13

2 LITERATURE REVIEW 14
2.1 Related Work 14
2.2 Literature definition of the Problem 14
2.3 Solving the BSS problem 17

3 PROBLEM DEFINITION 18

4 DEVELOPMENT 20
4.1 Bus Stops Selection Approaches 20
4.1.1 Isochrone Approach 20
4.1.2 Two Students Approach 21
4.2 Bus Stops Allocation 23

5 RESULTS 25
5.1 Methodology 25
5.2 Parameters 26
5.3 Results and Comparison 27
5.3.1 Results Overview 27
5.3.2 Detailed Comparison 29

6 CONCLUSION 32
6.1 Final Considerations 32
6.2 Future work 32

BIBLIOGRAPHY 33

1 Introduction

The SBRP is a known problem in the literature and considered as an important application in the real-world context (SARUBBI et al., 2016)(LIMA et al., 2016)(SCHITTEKAT et al., 2013), (JOZEFOWIEZ; SEMET; TALBI, 2008). It can be assigned as an important tool for defining routes that belongs to the set of best solutions. The SBRP derives from Vehicles Routing Problem (VRP), which is a classic problem of literature whose objective is to target vehicles on efficient routes, from starting points (source/warehouse) to destination points (sink/consumers), collecting or delivering products (TOTH; VIGO, 2002), (RIERA-LEDESMA; SALAZAR-GONZáLEZ, 2012), (MIRANDA, 2018), (LI; FU, 2002). In SBRP, the source is equivalent to a garage or starting point, and the sink (or destination) are the schools (PARK; KIM, 2010). The students can be considered as loads, and their gathering and delivery must be considered simultaneously in the problem.

As an applicable problem in the real-world, the SBRP can be assign as an important tool at the Brazilian context, where the government strives to provide assistance to public school students (LIMA et al., 2016). The Brazilian reality in the context of rural locations involves among several complications, the daily displacement of long distances to educational centers with more adequate teaching support and infrastructure. The routes management for these students transportation is an enormous challenge, considering the lack of adequate tools and the covered territory's great extension (CARVALHO et al., 2010). Besides these factors, the SBRP approach in rural environments is relevant because Brazil has a large number of students located in rural areas (5.3 million in 2017 according to (INEP, 2017)). Furthermore, rural locations usually do not have pre-set bus stops, justifying the use of the BSSP subproblem, which is the focus of this work.

According to Park and Kim (2010), the SBRP consists of smaller subproblems and can be solved by five steps: data preparation, bus stop selection, bus route generation, school bell time adjustment and route scheduling. As a SBRP subproblem, the BSSP's objective is to define points that gather as many students as possible, respecting constraints such as maximum walking distance per student and maximum number of students per vehicle. The selection and definition of bus stops must generate consistent locations, especially in rural environments, outside properties, in viable road sections and especially covering the largest number of students as possible, respecting each students' maximum walking distance. In this work, it is presented new strategies and its comparison with another recent work in literature (SARUBBI et al., 2016) for the Bus Stop Selection, along with the Data Preparation step. The Data Preparation step involves the students, schools, vehicles and the roads network representation.

1.1 Objectives

This study aims to introduce new strategies in the BSS problem, the Isochrone Strategy and the Two Students Strategy. The new strategies can be splitted in two stages: The bus stop candidates generation, in which new points are introduced to the initial problem's set, and the bus stop selection, the step that focuses on filtering the candidate points, on a selection method based on a GRASP (Greedy Randomized Adaptive Search Procedure) (ROA; SUÁREZ, 2015) procedure. Isochrone Strategy uses the concept of isochrone maps. Isochrone concept (KRISMER et al., 2017) allows by a shortest path tree built under a (MUROTA; SHIOURA, 2014)-like algorithm, these points set manipulation in order to achieve useful and efficient results. In our strategy, the use of *Postgre pgRouting* library's Isochrone Function allows the Isochrone strategy to manage, by one point, all the reachable points inside the distance contraint default set in the problem. Isochrone strategy is computationally more efficient than Two Students strategy, since Two Students generate its reachable vertices by the massive comparison pair by pair.

It is also presented in this work, how the strategies are capable of generating solutions for instances with different characteristics, respecting the problem's restrictions. The more students are grouped together, the lower the problem's complexity.

1.2 Structure of this project

The topics covered are arranged as follows: literary review of the SBRP is shown in Chapter 2. The definition of the problem and the concepts discussed in the monography, along with the proposed algorithm and the difference between the baseline algorithm are shown in Chapter 4. The results exibition and its analyses are shown in Chapter 5. The final conisderations and future directions are presented in Chapter 6.

2 Literature Review

2.1 Related Work

Vehicle allocation to attend a load transportation demand consists of a daily modern society problem. It is frequently executed without the aid of automated tools, based just in technical knowledge and a deeper study of the road network layout. This type of work can be efficient when considered in a low-scale context, with fewer collect/delivery points or routes available. However, it is a considerable characteristic of this problem the fact that its complexity is directly related to the number of points considered in the graph that represents the roads network. The more visitation points, the harder it becomes to find the best route. From this need arises the VRP, a NP-Hard problem (PARK; KIM, 2010). Therefore, its complexity overflows the capacity of known polynomial programming (LAPORTE; NOBERT; TAILLEFER, 1988) and relies on metaheuristic optimizations, aiming to find a solution which although not the best, is part of the solutions set close to optimal.

2.2 Literature definition of the Problem

On the SBRP, the load of students must be collected and delivered, where at first, each student is considered a visitation point. It means it will be necessary to make a comparison between all students in a combinatorial problem. As an alternative to the excessive visitation points, one approach is to try clustering the students so whenever possible, they leave their home and walk to a not so far common point. Many studies consider that these clusters points are already given (PARK; KIM, 2010), (ELLEGOOD et al., 2019), especially in urban regions, where exists a regular public transportation fleet that demands this type of organization. In rural environments, these points are usually not considered, due to the low density of students in a given area. However, the differentiation of rural and urban environments has a subjective character. Rural environments can have small towns or villages that may justify the creation and analyses of bus stops as clusters. In this work, the BSSP is present as a way to reduce the number of visitation points and consequently, the combinatorial problem complexity. BSSP is a subproblem with a few approaches in the literature (PARK; KIM, 2010), because usually the points are already set in the roads. However, it is usually an urban region characteristic, where the density of users is higher and the routes are better defined.

Figure 1 represents a real instance of a rural environment. It is important to highlight the differences between urban and rural enviromnents, since those characteristics may require special approaches in the problem.

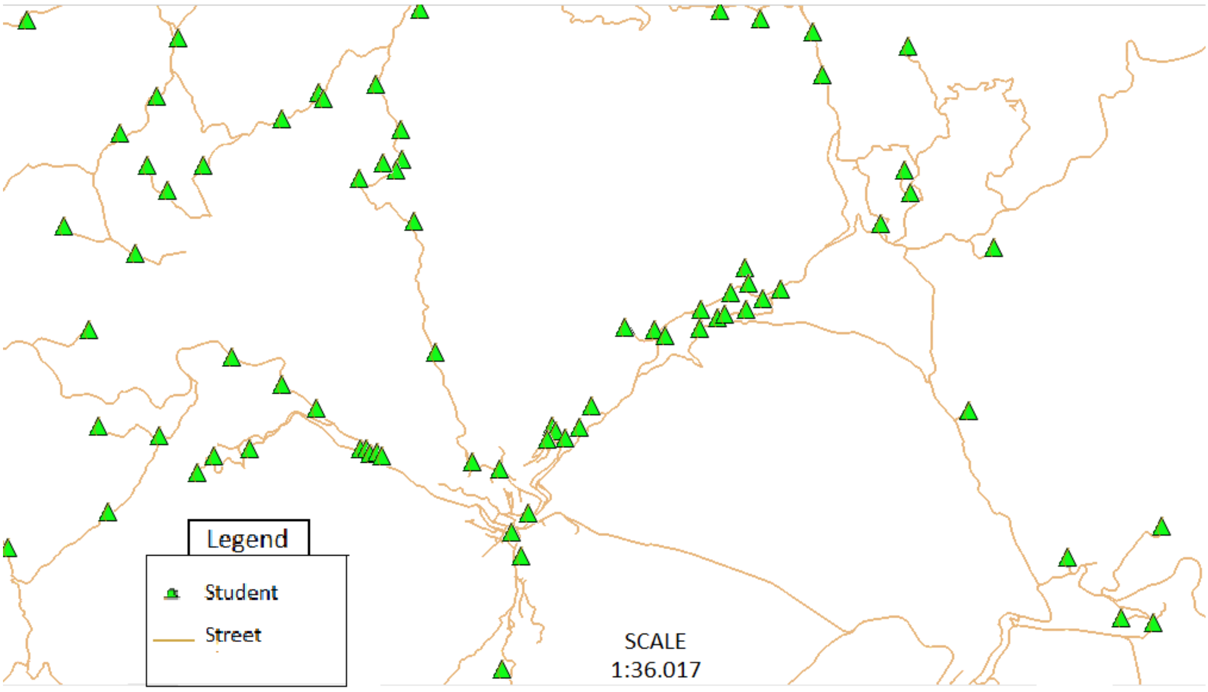

Figure 1 – Rural environment. Students in green triangles and roads in brown.

Urban environment in turn, as shown in Figure 2, has more intersections, alternative routes and a more fluid traffic. Students density is higher, and preset bus stops are already in use, disposed in suitable locations.

Students' residences are represented by green triangles and the roads' network is in brown. One of the factors to be noticed is the network roads distribution. In rural environments, the roads tend to be more sinuous, with less intersections and sometimes dead-end streets (GALDI; THEBPANYA, 2016). Students' disposition is sparser, with a considerable number of individuals at a significant distance from their neighbors. These characteristics may make it difficult to map certain routes, especially when picking a student at home - considering that some types of vehicles cannot perform certain maneuvers - and may increase the students' riding time, which is considered as a problem restriction (PARK; KIM, 2010).

The BSS problem is usually executed seeking to minimize at least one of the three classic Objectives: number of bus stops, shared bus stops or total walking distance (ELLEGOOD et al., 2019). The number of bus stops brings different solutions, since the number of bus stops reduce the vehicles transport cost and the total walking distance tends to increase the students' life quality (BRONSHTEIN; VAGAPOVA; NAZMUTDINOVA, 2014).

A few works in the recent literature deal with the BSSP. Sarubbi et al. (2016) algorithm has been chosen as a baseline comparison because, as this work, it also aims to

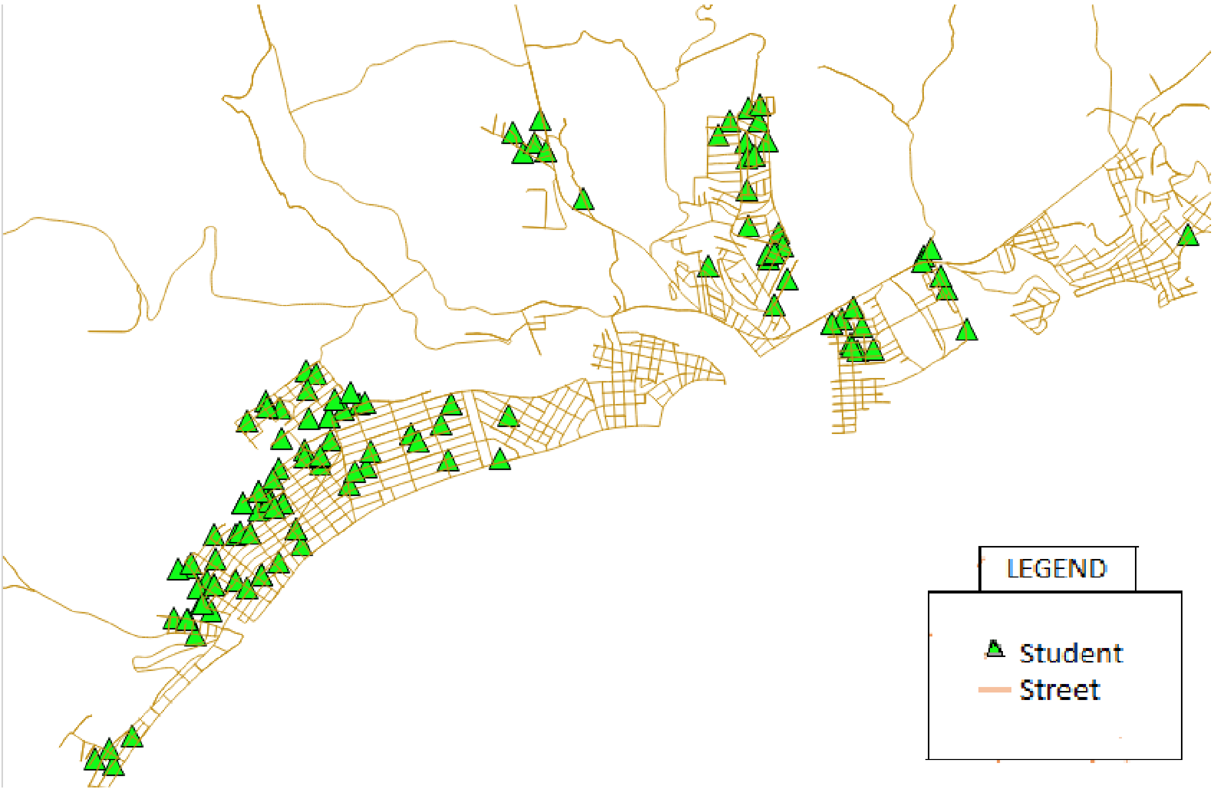

Figure 2 – Urban environment. Students in green triangles and roads in brown.

minimize the bus stops number. Though Sarubbi's work used a different georeferenced data, in this work it is implemented and executed over the same conditions as the presented approaches. Sarubbi et al. (2016) algorithm is based in a LAR approach, where the first step is to determine which points will be available to the process of selection. They consider, in a Location-Allocation-Routing (LAR) approach, that the bus stop can be allocated all along the street. Based on a set of equidistant points, generated at a distance λ, where the number of points is given by $numPoints \longleftarrow (\lfloor street\_size/\lambda \rfloor)$ and must be calculated for every street in the roads network. The lower the λ, the higher the number of different points created, consequently more points can be analyzed and probably, better solutions can be found. However, it also increases the complexity of the problem as more points need to be compared. It is necessary to find an adequate value to λ, that is higher enough to consider possible key points, but not too low to overload the algorithm's complexity.

In this work, it is presented two approaches to select the candidate points: the Two Students and the *Isochrone* approaches. Whereas the comparison between the students and each possible point is a very expensive operation, the approaches presented look for - each with its own characteristic - work around that issue.

2.3 Solving the BSS problem

The BSS problem can be executed in two stages (RIERA-LEDESMA; SALAZAR-GONZÁLEZ, 2013)(SARUBBI et al., 2016)(LIMA et al., 2016):

- Bus Stops Selection: In order to generate a bus stop set, this process aims to define which points are going to be candidates to become a bus stop. Some studies (FARAJ et al., 2014) considers that only the network natural points, such as intersections and end points are elegible as candidate points.

- Allocation Process:

  Since the points set is defined, it is necessary to allocate the students to a bus stop (PARK; KIM, 2010). The allocation process is a combinatorial problem, usually addressed as a GRASP approach, but is deeply tied to constraints such as maximum walked distance (ELLEGOOD et al., 2019), (RIERA-LEDESMA; SALAZAR-GONZÁLEZ, 2013) and the students' mobility or special needs (RUSSELL; MORREL, 1986).

Ellegood et al. (2019), Chapleau et al. (1985) mention that the BSS problem usually deals with three different objectives: Stops Number, where the routing efficience aims to reduce the transportation cost, grouping as much students as possible on a single bus stop; Minimizing the total students walking distance, where the students' quality of life is considered at first. However, this objective function leads to the situation where every student must be taken at home; Therefore, the third objective is, according to Solomon et al. (2019) and Caceres et al. (2016), the bi-objective model that considers both stops number and the students total walking distance.

3 Problem Definition

The formulation that describes this work is quite similar to Sarubbi et al (2016), since in both works the goal is to find the minimum set of bus stops, respecting all the restrictions and attending all the students. One contribution over Sarubbi et al (2016) in this work is the individual maximum walking distance per student. Such flexibility is important due to the real problems' sensibility. Assuming a young child at age of 5 could be walking as far as a teenager at age of 15 is inconsistent. And as this work deals with real data, each student has its own characteristics, that have to be considered and respected. Same is applied to reduced mobility students, that not just need to be collected at home, but may also need special companion through the bus riding.

In this study, the roads representation is based on a real georeferenced data, stored in PostgreSQL database system. Elements that compose the roads network are polygonal lines which represent the streets, points that represent the beginning, the end and also the streets intersections. Such structure can be formally classified as a graph (DIESTEL, 2000), where analogously the streets represent the edges and the points are the vertices that bound the edges. Thus it is possible to rely on graph theories to contribute with the development.

Considering multiple schools, the problem addressed also consists of a set of students S, where each S_i has his own characteristics such as reduced mobility, age and study shift; a set of visitation points V where initially, each $S_i \in$ V; a graph G representing the roads network; a maximum riding distance and a maximum walking distance for each S_i. Since the vehicles are only needful later in the Routing Problem, they are abstracted. This problem can be represented as the following Integer Linear Programming (IPL) formulation:

Table 1 – Sets, parameters and variables

S =	Students' Set
P =	Candidate Points' Set
λ =	Max walking distance
d_{ij} =	Distance between student i and point j
y_j = (0,1),	1 = if point j is chosen to receive a bus stop 0 = otherwise
x_{ij} = (0,1),	1 = if student i is assigned to bus stop j 0 = otherwise

Describing the Θ set to prevent the persistence of distances between students i and bus stops j that measure more than λ:

$$\Theta = \{(i, j) \in S \times P | d_{i,j} \leqslant \lambda\}$$

The Bus Stop modeling:

$$min \sum_{j=1}^{|P|} y_j \tag{1}$$

subject to:

$$\sum_{(i,j)\in\Theta} x_{(i,j)} = 1 \qquad \forall i \in S \tag{2}$$

$$x_{(i,j)} \leq y_i \qquad \forall_i \in S, j \in P | (i, j) \in \Theta \tag{3}$$

$$y_j \in \{0, 1\} \qquad \forall_j \in P \tag{4}$$

$$x_{(i,j)} \in \{0, 1\} \qquad \forall_{(i,j)} \in \Theta \tag{5}$$

The objective function (1) aims to minimize the selected bus stops' number. Constraint (2) ensures that a student will be associated to exactly one bus stop, Contraint (3) ensures that no one bus stop will exist without students associated. Constraints (4) and (5) defines that the variables are binary.

4 Development

4.1 Bus Stops Selection Approaches

Firt step in Bus Stop Selection problem is to define the initial points to be considered in the problem. As usual, those points are the native vertices presented by the georeferenced topology: considering that each street is an Graph edge, the two vertices (initial and final) that compose the edge are candidate points. It is also considered intersection points from different edges. A specific feature of this problem is that students' residences are also considered as candidate points. Therefore, if a student have reduced mobility, his house will be a bus stop and other students can be allocated to that point.

4.1.1 Isochrone Approach

Given a d_i parameter, where d represents the maximum walking distance of a i student, the Isochrone Approach returns from a PostGre Isochrone Function all the points in the map that exists inside d_i topology distance. The function parameters are the start point, in this case, the student i, and its respective maximum walking distance d_i.

Isochrone Approach does not require the massive comparison between every pair of students, neither the comparison between all vertices and student. It means every known vertex inside the distance range will be returned. Therefore, it is just necessary to create the adjacency relationship between the student and the vertices and add the points in the set of selected points.

Figure 3 represents how occurs the students-vertices association. Given a reference student (orange triangle), all candidate points considered in the problem (initial, final, intersection points and students' residences; represented by green circles) that are inside the referenced student's maximum walk distance (blue lines) are selected in the student's adjacency list.

Algorithm 1 represents the Isochrone Strategy. The inputs are the students set (S), the instance points set (P) and the students' distances set (d_i). The expected output is the points set P', where $P' \in P$.

For each student considered in the problem, the Isochrone function colects all points inside P that are inside the distance d_i. The function returns a set of points that are associated to the student s.

The required parameters for the function are:

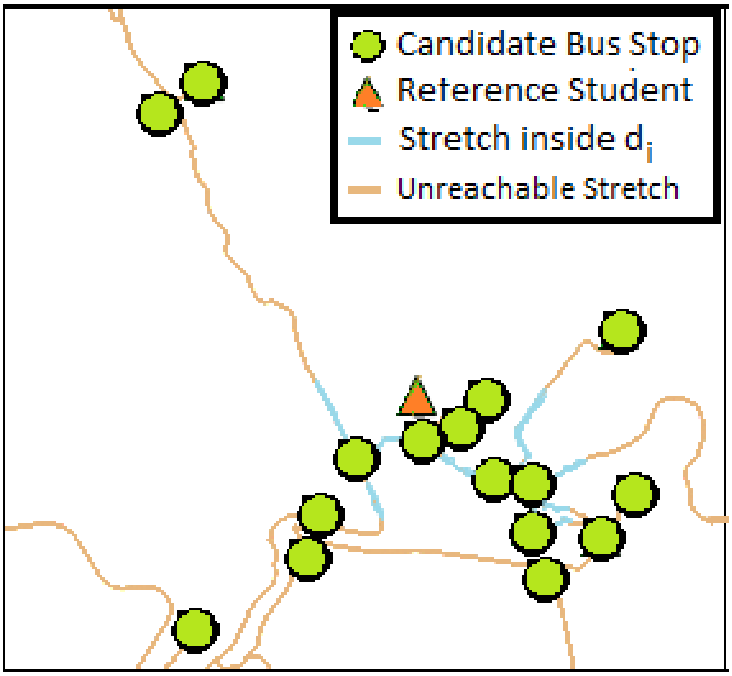

Figure 3 – Isochrone Approach. The reference Student (orange triangle) and the reachable vertices inside the distance constraint (blue lines)

- A students set S;
- A points set P, addressed in the problem;
- A distance vector D_s, that indicates for each student s, its maximum walking distance;

Data: S, P, D
Result: P'
for $s \in S$ **do**
  $P' \leftarrow P' \cup Isochrone(s, P, D_s)$;
end

Algoritmo 1: Isochrone

4.1.2 Two Students Approach

Two Students Approach makes a comparison between every pair of students ($_{i,j}$) in *S*. For each comparison, it creates a path whose initial and final points are (p, p') $\in S$, utilizing *Dijikstra's Minimum Cost Path* function in *Postgre*. Considering that the path contains all the intermediate vertices, the next step is to check if both students are able to walk to some of these points on the path. For that, the students' individual distances constraints d_i and d_j must be: $d_i + d_j \geq d_{i,j}$, where $d_{i,j}$ is the total distance of the path between them. in affirmative case, an adjacency relationship is created between the points inside that intersection and the students and the point are added to the set of selected points, as shown in Algorithm 2.

The Two Students algorithm (Algorithm 2) uses as parameters:

- A students set S;
- A points set P, addressed in the problem;
- A distance vector D_s, that indicates for each student s, its maximum walking distance;
- A distance matrix $N_{i,j}$ that indicates the distance between point i to j

For each students pair, if the distance $d_i + d_{i'}$ is equal or higher than the distance $N_{s,s'}$ between then, the points in the path are considered reacheable and the students are associated to these points adjacency list.

Data: S, P, D, N
Result: P'
for $s \in S_n$ **do**
    for $s' \in S_n$ **do**
        if($d_i + d_j \geq N_{i,j}$)
            $P' \leftarrow P' \bigcup reachable\_points\_in\_path(s, s', P)$
    end
end

Algoritmo 2: Two Students Algorithm

In this work as well as (SARUBBI et al., 2016), the point allocation is not limited to be in a student position or intersections, but can be all along the streets disposed in the graph. This freedom in the allocation may increase the chances of finding a more interesting solution, especially with regard to equal distance distribution of walking distance between students. (SARUBBI et al., 2016) approach to select the candidates involves the discretization of equidistant points all along the road, given a λ parameter that represents the distance between each new point.

Reproducing the (SARUBBI et al., 2016) selection approach, the implementation followed the Two Students' idea, by comparing a pair of students, all the mutual reachable vertices are set as candidates. However, instead of creating the path from native vertices – that just includes start, ending and intersection vertices -, it uses the function ST_Line_Interpolate_Point from Postgre, that discretize a points sequence – with the addition of generate_series function – in a line/road represented as a geometry object - and make the same validations as Two Students Approach, verifying which vertices are reachable for both students.

When applying the *Routing Problem* along with the *Bus Stops Selection Problem*, two heuristics approaches have been used in literature (PARK; KIM, 2010), the Location-Allocation-Routing (LAR) approach and the Allocation-Routing-Location (ARL). Location process is the task to search and define which points will be selected as possible Bus Stops. Allocation process involves directing the students in the set to the possible bus-stops, respecting the problem restrictions such as maximum distance walked. Routing problem is the process of generating routes between the points considered in the problem. In this project, it is used an approach based in LAR, where at first it is selected a set of points that can cluster students together. After that, it uses a approach based on a GRASP construction (SARUBBI et al., 2016) to associate the students to the bus stops, in order to minimize the number of bus stops required without infringing the restrictions. Routing problem is under development and will not be addressed in this work.

As this work deals with real geo-referenced data, it uses besides PostgreSQL as mentioned before, QGis as a graphic tool to visualize the topology and check the results' consistence. It is also used some of the geographical extensions from PostgreSQL, like PostGis. It provides functions such as *Dijkstra's Minimum Path, Interpolate Points and ST_LineInterpolate*, all fundamental to the implementation of the presented approaches and Sarubbi's.

4.2 Bus Stops Allocation

Since the candidate points have already been selected, the allocation process determines which one will remain as point and which student will be associated to it. Our Allocation Approach, named "Approach One", uses a GRASP approach to select the candidate bus stops with more students associated. Always the bus stop that is accessible to the larger students' number is selected as a bus stop. (SARUBBI et al., 2016) approach adds a randomical factor to the selection, with a α parameter, that defines a subset with α bus stops. The buses in the subset are the ones with most students associated to be elected as a bus stop.

Algorithm 3 represents the general algorithm and the event sequence to the BSSP execution. After the slection of candidate points in function *points_selection*, the allocation process is executed in *Students_allocation* function, returning the bus stops set.

The algorithm parameters are:

- A students set S;
- A points set P, addressed in the problem;
- A distance vector D_s, that indicates for each student s, its maximum walking distance;
- A distance matrix $N_{i,j}$ that indicates the distance between point i to j

Data: S, P, N, D
Result: P'
P'← $Points\_Selection(S, P)$;
B ← $Students\_Allocation(S, P')$;
Algoritmo 3: General Algorithm

In the proposed allocation approach (Algorithm 4, considering that P possess an adjacency list of students and the students possess a reference to its vertex, the function parameters are:

- Students set S;
- Points set P, returned by Selection Approach;
- Adjacency Matrix M, with the students adjacency also modified in Selection Approach;

The Allocation Approach One will return a bus stops set B, with the students allocated respecting the integrality constraints listed in Chapter 4. In the initialization, the B set is empty and the points set is sorted by number of students associated. A single point is selected, executing as a greedy approach (GRASP).

```
Data: S, P, M
Result: B
B ← ∅
P ← sort_by_associated_students(P)
p ← P_0
for s ∈ M_{s,p} do
    if(is_not_allocated(s))
        s.allocate_to(p)
        p.allocate(s)
end
B ← B ∪ p
P ← P − p refresh(M)
if(exist_pendant_students(S))
    Allocation_Approach_One(S,P,M)
```

Algoritmo 4: Allocation_Approach_One

For each student s associated to p, it is verified if s is already allocated using a boolean function *is_not_allocated(student)*. If the student is not allocated, it is associated to p.

Once the iterations are executed to p, the points list is refreshed and p is allocated to B set. At this moment, p is officially a bus stop.

The next step is recursive, only executed if exists allocation pendent students.

5 Results

5.1 Methodology

The approaches presented in this work have been implemented with conditional flexibility, allowing the settings parameterization for different comparison metrics. The whole project is based on Java language. The versatility and dynamism present in the language scope led us to its use. Efficience was strongly considered, so the objects are usually referenced and related to each other. For database application, PostGre SQL was used, along with an institution server that hosted the real georeferenced data. Geographical application was based on PostGIS and its functions.

As introduced in Section 3, the problem's objective function considered in this work aims to minimize the bus stops number. However, different metrics can be considered in the execution:

- students distribution per bus stops: It may be limited by the city topology, associated to the walking distance constraint. It can also be restrictive for a maximum students number per bus stop, in a way to not overload the point nor to force more buses to the same route.
- maximum walked distance: Frequently used in literature as a Objective Function, the students' maximum walked distance is a relevant metric, specially to measure the students' life quality in a execution. It is a trade-off to the bus stops number metric, since to minimize the walked distance, the buses should travel more to attend individually, consquently, more bus stops are generated.

The comparison objective of this work is to present the results generated by the proposed approaches in comparison with the related work (SARUBBI et al., 2016), utilizing the minimal bus stops number metric. It is also presented different metrics that altought not the objective, can sharpen different discussions for future works.

To generate results, the approaches presented in chapter 4 were applied in different real georeferenced data sets, all related to Brazilian cities. The cities were selected arbitrarily, however with peculiar characteristics, such as streets topology and students number, to diversify the approaches behaviors.

Per city, one execution has been tabulated for each of the following combinations:

As introduced in Chapter 4, the bus stop selection approaches are responsible for generating a candidate bus stops set. For this work, the two proposed approaches were

Table 2 – Combination of Selection and Allocation Approaches used on each instance

Bus Stop Selection Approach	**Allocation Approach**
Isochrone	Approach One
Two Students	Approach One
Sarubbi	Approach One
Sarubbi	Sarubbi Approach

tested, along with (SARUBBI et al., 2016) selection approach. The allocation process involves the allocation approach proposed in section 4.2 along with the proposed bus stop selection approaches. For a imersive comparison, the (SARUBBI et al., 2016) selection strategy was executed along with the proposed and their own allocation strategy.

5.2 Parameters

The proposed approaches are based on a parameter d_i. That parameter represents the limit distance a student i can walk. Instead of using a standard parameter value for all instance students, in this work the d_i value is individual and depends on each students' characteristics, highlighting its mobility.

Students with reduced mobility, such as wheelchair users, are identified by their special needs using the d_i parameter. If $d_i = 0$, the student can be identified as a student with lower mobility capacity.

Isochrone Approach uses d_i as a function parameter to limit the reach range, given a start point. For each student i, all the reachable points inside the distance d_i are associated to the student.

Two students Approach uses d_i for each pair of student. Considering the students i and j, will have intersections if the topology distance x_{ij} between them is $x_{ij} \leq d_i + d_j$. All reachable points inside the intersection will be associated to the students.

Considering the presented approaches' deterministic nature, the study of the distinct parameters behavior was not included in this work analysis. The only parameters that might require some adjustments are the Sarubbi et al (SARUBBI et al., 2016), since their approach characteristics is based in a random nature. However, it is assumed the best cases presented in their work as $\lambda = 0.05$ and $\alpha = 1$. α parameter presses the algorithm to a greedy approach, since its value determines the range of best solutions that might be selected.

5.3 Results and Comparison

5.3.1 Results Overview

The tests were applied in a total of 63 executions, in the scope of 16 different Brazilian cities. Grouping the executions by cities, each execution per group disposes of different approaches setups. The most interesting comparison is between the approach Isochrone, combined with the presented allocation method, the baseline algorithm approach and baseline allocation method. Analyzing the number of selected bus stops between the main approaches Isochrone and the baseline work (Table 3), it is noticeable in most of the cases, the Isochrone approach got slightly better results (Graphic 4). Cities 7 and 12 are special cases, where respectively, exists a difference of 45% and 31% between the number of bus stops achieved by the main approaches. In Figure 4, the proportion between the approaches results is shown for all cities. The graphic is based on the bus stops number comparison between the cities. In a percentage range (0 to 1), the worst approach result is taken as maximum reachable value (100%) by city. The other values per city are proportional to the worse execution, so the lower the value, the better the solution. According to the Objective Funcion, the target solution is the one with the lower number of bus stops. It means that by the goal of reducing the number of bus stops, the Isochrone obtained better results. However for instance, in city 17, Table 3, the (SARUBBI et al., 2016) got better results (10% better than Isochrone and Two Students). It is important to hisghlight that the analysis is not limited to the bus stops' number but also includes the number of students considered in the problem, the number of students handicapped, the number of students per bus stop and also the avarage distances considered by the executions, as show in Table 3. For this purpose, this work treated the students' walked distances individually. Some works in literature (PARK; KIM, 2010), (MIRANDA, 2018), (SARUBBI et al., 2016) consider a standard value for all students, however there are discrepancies in the locomotion capacities of different students. A handicapped student has special needs and cannot be processed as the other kids. The same is valid to kids with different age ranges. Children in elementary school have not the same walking capacity as high school students. It is noticeable in Table 3 the column that indicates the number of students handicapped considered in the execution. In addition, the walking distances are defined by range of age in a parameters table. The values were collected from a national database, that standardized the recommended walking distance.

Despite optimization works do not focus on execution time, it can be a relevant metric to analyze in Table 3 the computational difference between *Isochrone Approach* and *Two Students Approach*. Since Two Students Approach involves a students comparison pair by pair, it is necessary a large ammount of processing to execute the approach. Execution time is more expressive in executions with a larger number of students, such as city 6 and city 4 (931 and 705 students, respectively). City 4 had Two Students Approach execution

Table 3 – Results Table - For each city, following the Objective Function presented, the bus stop number is highlighted in blue.

	BSS Approach	Allocation Approach	Execution Time (s)	Bus Stops Number	Handicapped Students	Students Number	Average Students/ Bus Stop	Max. Number of Students/ BS	Standard Deviation Students/ BS
1	Isochrone	S. One	191	41	0	93.0	2.268	17.0	2.71131
	TwoStudents	S. One	167	44	0	93.0	2.114	15.0	2.37444
	Sarubbi	S. One	142	44	0	93.0	2.114	15.0	2.34487
	Sarubbi	S. Sarubbi	132	47	0	93.0	1.979	15.0	2.31713
2	Isochrone	S. One	426	47	0	135.0	2.872	12.0	2.39217
	TwoStudents	S. One	468	51	0	135.0	2.647	8.0	2.02804
	Sarubbi	S. One	424	51	0	135.0	2.647	8.0	1.87428
	Sarubbi	S. Sarubbi	360	53	0	135.0	2.547	17.0	2.87941
3	Isochrone	S. One	1016	42	0	343.0	8.167	37.0	8.88934
	TwoStudents	S. One	5097	51	0	343.0	6.725	36.0	7.49954
	Sarubbi	S. One	4333	48	0	343.0	7.146	48.0	9.06945
	Sarubbi	S. Sarubbi	4002	53	0	343.0	6.472	65.0	10.65292
4	Isochrone	S. One	1785	70	0	705.0	10.071	91.0	14.06540
	TwoStudents	S. One	12055	89	0	705.0	7.921	53.0	10.12785
	Sarubbi	S. One	9686	94	0	705.0	7.500	67.0	10.64152
	Sarubbi	S. Sarubbi	9511	107	0	705.0	6.589	86.0	11.68288
5	Isochrone	S. One	1119	62	20	346.0	5.581	31.0	6.05047
	TwoStudents	S. One	4436	67	20	346.0	5.164	31.0	5.94825
	Sarubbi	S. One	3768	57	20	346.0	6.070	40.0	7.00729
	Sarubbi	S. Sarubbi	3500	64	20	346.0	5.406	42.0	7.88955
6	Isochrone	S. One	2984	94	9	931.0	9.904	47.0	8.26193
	TwoStudents	S. One	17251	109	9	931.0	8.541	36.0	7.09138
	Sarubbi	S. One	15486	105	9	931.0	8.867	36.0	8.11109
	Sarubbi	S. Sarubbi	14004	109	9	931.0	8.541	57.0	11.78082
7	Isochrone	S. One	1873	74	2	568.0	7.676	23.0	5.28919
	TwoStudents	S. One	7728	82	2	568.0	6.927	20.0	5.04983
	Sarubbi	S. One	6844	82	2	568.0	6.927	30.0	6.10764
	Sarubbi	S. Sarubbi	6017	89	2	568.0	6.382	35.0	8.23588
8	Isochrone	S. One	1069	37	2	310.0	8.378	47.0	10.65716
	TwoStudents	S. One	6133	47	2	310.0	6.596	35.0	8.35792
	Sarubbi	S. One	5025	45	2	310.0	6.889	46.0	9.69354
	Sarubbi	S. Sarubbi	4765	53	2	310.0	5.849	52.0	9.46328
9	Isochrone	S. One	1098	60	0	300.0	5.000	16.0	3.62197
	TwoStudents	S. One	2730	66	0	300.0	4.545	15.0	3.47378
	Sarubbi	S. One	2569	62	0	300.0	4.839	18.0	4.14173
	Sarubbi	S. Sarubbi	2225	67	0	300.0	4.478	26.0	5.68986
10	Isochrone	S. One	1232	69	27	360.0	5.217	24.0	4.50121
	TwoStudents	S. One	3955	77	27	360.0	4.675	31.0	4.70845
	Sarubbi	S. One	3387	68	27	360.0	5.294	32.0	5.86174
	Sarubbi	S. Sarubbi	3118	80	27	360.0	4.500	30.0	6.20127
11	Isochrone	S. One	1671	126	0	531.0	4.214	113.0	10.41200
	TwoStudents	S. One	10087	140	0	531.0	3.793	111.0	9.71648
	Sarubbi	S. One	8384	133	0	531.0	3.992	110.0	9.72851
	Sarubbi	S. Sarubbi	8280	139	0	531.0	3.820	111.0	9.84463

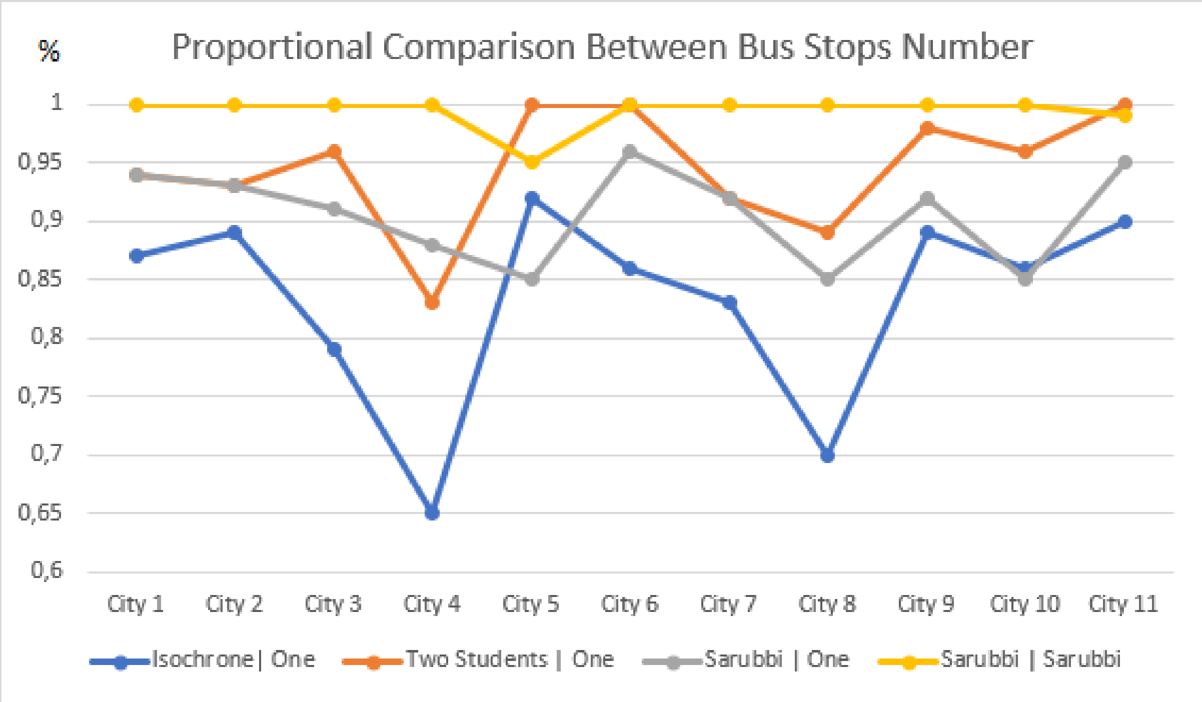

Figure 4 – Percentual proportion (Y axis) between approaches per city (X axis). The further from 1 (Y axis), the greater the difference between approaches.

time 6.7 times longer than Isochrone Approach execution time. City 6 had Two Students Approach execution time 5.8 times longer than Isochrone Approach. However, executions with fewer students may be more efficient with Two Students Approach, such as cities 1, 2 and 3.

5.3.2 Detailed Comparison

Figure 5 represents one of the problem executions. The green triangles represent the students, which all of them belongs to the same shift hour, however may belong to different schools and time schedules. Each student may also have different characteristics such as reduced mobility or a different maximum walking distance. The environment represented in Figure 5 is predominantly rural, with sinuous roads and few alternative routes. The lower density of students can also be identified.

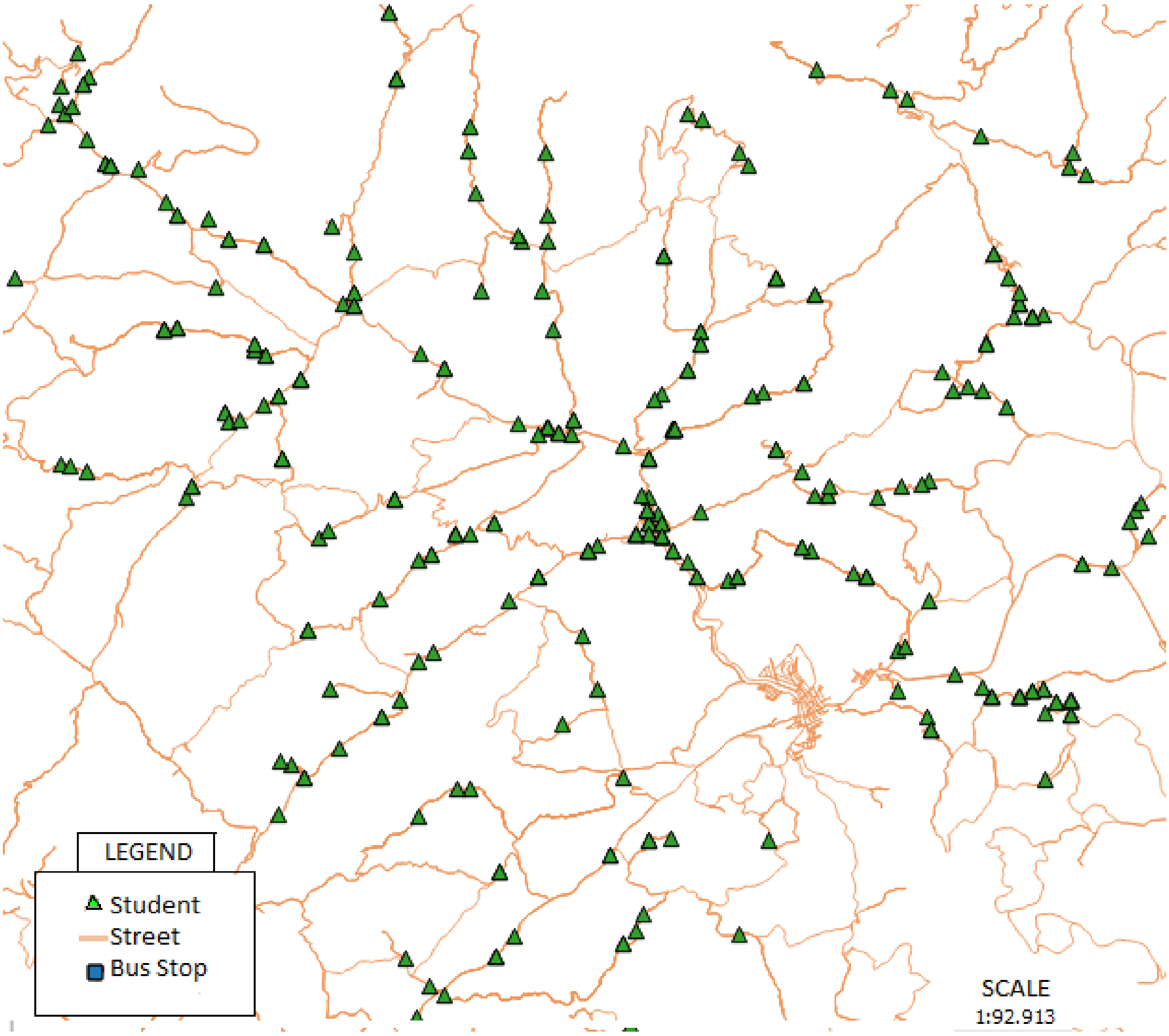

Figure 5 – One of the executions before the bus stops selection. Rural environment characteristics.

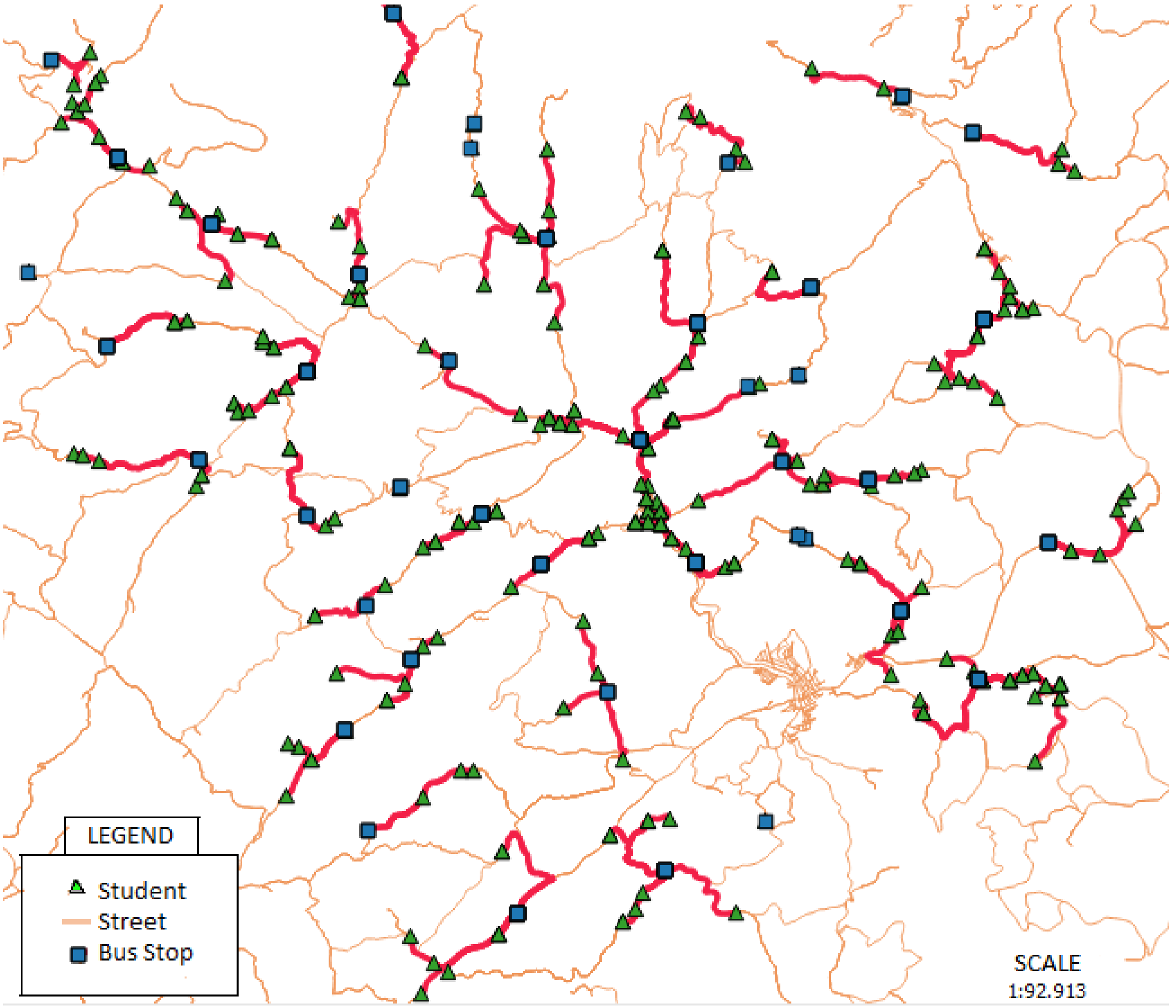

Figure 6 – The execution after the bus stop selection. Students are associated to bus stops. Lines on red represents the students walked path between their homes and the bus stops.

6 Conclusion

6.1 Final Considerations

In this work, the SBRP is described and contextualized in real-world applications, highlighting the relevance of the project, especially in Brazil's context, due to the political effort to reduce the distances between the students and a suitable teaching. It also presents the developed strategies and the comparison with another recent related work in literature. The BSSP can be a powerful tool in the Routing process, reducing the initial complexity of the problem by reducing the number of points to be visited, clustering the students in Bus Stops, respecting restrictions as the maximum distance the students can individually walk. The presented aproach in the BSSP showed to be efficient, both in rural and urban environments. However it needs improvements and more tests to be performed.

6.2 Future work

Some improvement can be done in the strategies, especially with regard to the disparity between the students' walk distance, avoiding unfair/fostering situations. There should also be a way to prioritize a bus stop according to its proximity to main roads which occasionally can be used in the main route to the schools. It is also important to check availability to implement non-deterministic approaches, since the project allows this kind of integration, so better results may be found. In order to deeply compare the approaches performance, it would be interesting to use the same instances than Sarubbi et al (2016), since the topology is relevant for the analysis.

Bibliography

BRONSHTEIN, E. M.; VAGAPOVA, D. M.; NAZMUTDINOVA, A. V. On constructing a family of student delivery routes in minimal time. *Automation and Remote Control*, v. 75, n. 7, p. 1195–1202, Jul 2014. ISSN 1608-3032. Disponível em: <https://doi.org/10.1134/S0005117914070029>. Citado na página 15.

CARVALHO, W. L. et al. Rural school transportation in emerging countries: The brazilian case. *Research in Transportation Economics*, v. 29, n. 1, p. 401–409, 2010. Citado na página 12.

DIESTEL, R. *Graph Theory*. second. [S.l.]: Springer, 2000. v. 173 of Graduate Texts in Mathematics. Citado na página 18.

ELLEGOOD, W. et al. School bus routing problem: Contemporary trends and research directions. *Omega*, 03 2019. Citado 3 vezes nas páginas 14, 15, and 17.

FARAJ, M. F. et al. A real geographical application for the school bus routing problem. *Intelligent Transportation Systems (ITSC)*, v. 2014 IEEE, p. 2762 – 2767, 2014. Citado na página 17.

GALDI, M.; THEBPANYA, P. Optimizing school bus stop placement in howard county, maryland: A gis-based heuristic approach. *IJAGR*, v. 7, p. 30–44, 2016. Citado na página 15.

INEP. *National Institute of Study and Research Anisio Teixeira - INEP. School Census*. 2017. Disponível em: <http://portal.inep.gov.br/censo-escolar>. Citado na página 12.

JOZEFOWIEZ, N.; SEMET, F.; TALBI, E. Multi-objective vehicle routing problems. *European Journal of Operational Research*, v. 189, n. 2, p. 293–309, 2008. Citado na página 12.

KRISMER, N. et al. Computing isochrones in multimodal spatial networks using tile regions. In: *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*. New York, NY, USA: ACM, 2017. (SSDBM '17), p. 33:1–33:6. ISBN 978-1-4503-5282-6. Disponível em: <http://doi.acm.org/10.1145/3085504.3085538>. Citado na página 13.

LAPORTE, G.; NOBERT, Y.; TAILLEFER, S. Solving a family of multi-depot vehicle routing and location-routing problems. *Transportation Science*, v. 22, n. 3, p. 161–172, 1988. Disponível em: <https://doi.org/10.1287/trsc.22.3.161>. Citado na página 14.

LI, L.; FU, Z. The school bus routing problem: a case study. *Journal of the Operational Research Society*, Springer, v. 53, n. 5, p. 552–558, 2002. Citado na página 12.

LIMA, F. M. S. et al. A mixed load capacitated rural school bus routing problem with heterogeneous fleet: Algorithms for the brazilian context. *Expert Systems with Applications*, v. 56, p. 320–334, 2016. Citado 2 vezes nas páginas 12 and 17.

MIRANDA, D. M. A multi-loading school bus routing problem. *Expert Systems with Applications*, v. 101, 02 2018. Citado 2 vezes nas páginas 12 and 27.

MUROTA, K.; SHIOURA, A. Dijkstra's algorithm and l-concave function maximization. *Mathematical Programming*, Springer, v. 145, n. 1-2, p. 163–177, 2014. Citado na página 13.

PARK, J.; KIM, B.-I. The school bus routing problem: A review. *European Journal of Operational Research*, v. 202, n. 2, p. 311–319, 2010. Disponível em: <https://EconPapers.repec.org/RePEc:eee:ejores:v:202:y:2010:i:2:p:311-319>. Citado 6 vezes nas páginas 12, 14, 15, 17, 22, and 27.

RIERA-LEDESMA, J.; SALAZAR-GONZÁLEZ, J. J. A column generation approach for a school bus routing problem with resource constraints. *Computers & Operations Research*, Elsevier, v. 40, n. 2, p. 566–583, 2013. Citado na página 17.

RIERA-LEDESMA, J.; SALAZAR-GONZáLEZ, J.-J. Solving school bus routing using the multiple vehicle traveling purchaser problem: A branch-and-cut approach. *Computers Operations Research*, v. 39, n. 2, p. 391 – 404, 2012. ISSN 0305-0548. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0305054811001225>. Citado na página 12.

ROA, M. A.; SUÁREZ, R. Grasp quality measures: review and performance. *Autonomous robots*, Springer, v. 38, n. 1, p. 65–88, 2015. Citado na página 13.

RUSSELL, R. A.; MORREL, R. B. Routing special-education school buses. *INFORMS Journal on Applied Analytics*, v. 16, n. 5, p. 56–64, 1986. Disponível em: <https://doi.org/10.1287/inte.16.5.56>. Citado na página 17.

SARUBBI, J. F. M. et al. A strategy to clustering students minimizing the number of bus stops for solving the school bus routing problem. *European Journal of Operational Research*, v. 202, n. 2, p. 311–319, 2016. Citado 7 vezes nas páginas 12, 17, 22, 23, 25, 26, and 27.

SCHITTEKAT, P. et al. A metaheuristic for the school bus routing problem with bus stop selection. *European Journal of Operational Research*, v. 229, n. 2, p. 518 – 528, 2013. ISSN 0377-2217. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0377221713001586>. Citado na página 12.

TOTH, P.; VIGO, D. The vehicle routing problem. *SIAM, Philadelphia, PA*, 2002. Citado na página 12.

TERMO DE RESPONSABILIDADE

Eu, **Henrique Queiroz do Amaral** declaro que o texto do trabalho de conclusão de curso intitulado "***Students clustering approaches aiming to minimize the bus stop number: a comparative study in real problems***" é de minha inteira responsabilidade e que não há utilização de texto, material fotográfico, código fonte de programa ou qualquer outro material pertencente a terceiros sem as devidas referências ou consentimento dos respectivos autores.

João Monlevade, 12 de Julho 2019

Henrique Queiroz do Amaral