XXVI IJCIEOM 2020
International Joint Conference on Industrial Engineering and Operations Management
ABEPRO-ADINGOR-IISE-AIM-ASEM (IJCIEOM 2020)

Solving a University Admission Exam Location Problem: An Application in Brazil

Thiago Edmar de Oliveira[1,2] and Rafael Martinelli[1]

[1] Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio),
Departamento de Engenharia Industrial,
22451-900, Rio de Janeiro, RJ, Brazil.
martinelli@puc-rio.br
[2] Universidade Federal de Juiz de Fora (UFJF),
36036-330, Juiz de Fora, MG, Brazil.
thiago.edmar@engenharia.ufjf.br

Abstract. This work presents a study on a University Admission Exam Location problem that solves a real case scenario from the admission exam of a university in Brazil. The objective is to assign candidates to exam locations happening the same day and time. Since the candidates must take the exam every year during secondary school, different types of exams must be offered. Each location can only offer one type of exam to reduce the complexity of the logistics. The objective is to decide which locations to open and the candidates' allocation to locations, minimizing the overall distance and opening costs. We propose a methodology that uses a binary programming formulation based on the Single Source Capacitated Facility Location Problem to solve the real case. Computational experiments using real information obtained solutions that reduce the travel distances in about 10%, for cases with more than 3,000 candidates and seven exam locations. For candidates who live close to the exam locations, this reduction can reach up to 24%.

Keywords: exam location, binary programming, facility location, application

1 Introduction

University access is scarce in many countries. Upon finishing secondary school, students compete for the few spots available for the different available courses. In Brazil, many kinds of selection processes can be found, but regardless of the institution funding (private or public) and equity actions, most of them still rely on written exams to fully or partially select the students among a large number of candidates.

When a university is responsible for organizing its selection process, many complex tasks are needed to be performed. The candidates have to be registered in a specific system, several different written exams have to be elaborated and graded, external locations must be found to apply the exams if the number of candidates exceeds the university's capacity, among others.

We are interested in the problem which arises in the case where the university is unable to apply the exam only in its campus, requiring the selection of external locations to host most of the candidates. This limitation is a common issue not only for the university selection process but for any other selection process with a large number of candidates, such as for the selection of public jobs.

The problem is then defined as the selection of exam locations from a set of available locations, each one with a given capacity and incurring in a different cost. This selection is influenced by the candidates who are not supposed to travel long distances to take their exams. Moreover, different candidates may take different tests due to field-specific requirements or different levels of knowledge. The objective is then to minimize a balance between the total location opening costs and the total travel distance of all candidates. We call this problem the University Admission Exam Location Problem (UAELP).

The improvement in the UAELP planning may result in several significant impacts. On the social side, the reduction in candidates' average journey size, also reducing public traffic before and after the exam. On the economic side, for the university, by optimally choosing the exam locations, and for the students, again by reducing their journeys, thus reducing the expenses with fuel, taxis, apps rides. Finally, we can also mention the environmental impact, like any optimization involving transportation, which indirectly reduces fuel consumption.

We present a methodology to solve the UAELP. It uses an integer programming formulation that extends a facility location formulation [18], by considering different types of exams and imposing a given location to host only one type of exam. This formulation is then used to solve a real-case arising from a Brazilian university. The results are compared against the real planning created and executed by the university at the time.

In the next section, we present a brief literature review. In Section 3, we present the methodology for solving the problem. Section 4 explains the real-life application, shows the computational experiments, and analyze the results with real planning. Finally, Section 5 concludes the work and points to the avenues of future research.

2 Literature Review

Across the literature, it is possible to find works related to the planning of exam times, such as the University Exam Timetabling Problem [21], which, besides the exam times, also schedules rooms, people to watch the tests, among others. Another type of university-related topic is the College Admission Problem [7], which asks a matching between schools and candidates, based on their preferences, part of what is commonly called the College-Choice Process [5]. Some works apply facility location models in the subject of Education [2, 3], but with different objectives from the one discussed in this paper.

To the best of our knowledge, no work deals directly with the UAELP. Some studies with similar content can be found, such as [1], which conducted research that addresses our study's subject, but with some limitations. Although it has the same objective – to reduce the displacement of the candidates – the approach is significantly different. To achieve the objective, they propose choosing the cities with a higher concentration of candidates in the vicinity, not involving the distances performed within those cities. Still, their approach does not assign the candidates to a location or city and, most importantly, it does not use the current exam data, but uses the data from the last selection process to plan the next one. This decision can present a significant limitation to the model, since there could be notable changes from one exam to another, mainly in the number of candidates, making the optimization ineffective. This difference can be seen in the article itself, in which only one in five simulations has shown significant improvement; in two of these simulations, there were even negative results or no gains.

The study developed by [11] shows strong similarities with this paper, even presenting a resembling mathematical model. However, both works have entirely different purposes, with their approach aimed at reducing the displacement performed by students within the area of a university and also not assigning individuals to a location, assigning subjects to classrooms instead.

Despite the lack of direct correlation, as the UAELP is an extension of the Single Source Capacitated Facility Location Problem [18], a vast literature can be found for its base problem. There is a large production of works related to the facility location subject with different approaches, presenting a wide range of mathematical models, heuristic algorithms, and theoretical analyses [10, 16]. It is a very relevant topic when it comes to supply chain management, as shown by the in-depth analysis of [14], but also especially in areas such as health care [4, 8, 20] and humanitarian logistics [6, 12, 15]. Moreover, like the approach presented in this work, it is widespread in facility location problems, the use of geographic information systems (GIS), as in [13, 17]. Finally, the reader is referred to [19] for a survey of discrete facility location problems.

3 Methodology

We propose a methodology to allocate candidates to the exam locations. We divided the task into four consecutive stages, as shown in Figure 1, the first one being the data collection, followed by the acquisition of distances, a pre-processing stage, the mathematical model application, and finally the solution analyses and implementation.

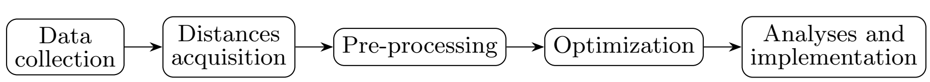

Fig. 1. Stages of the Proposed Methodology

3.1 Data Collection and Distance Acquisition

The input data needed are the addresses of the candidates and the addresses of the exam locations. For the cases on which there is more than one type of exam – tests with different contents to specific candidates, the information on which kind of exam each candidate has to take is also required. The addresses collection must be done with the collaboration of the institution that owns the exam, and consequently, its database. Thus, the university must provide the candidates' addresses, as well as the exam type that each candidate has registered in case of different contents. The same goes for acquiring the addresses of the test sites.

One of the main steps of the proposed method, and possibly the most difficult, as it requires careful treatment and analysis of the data, is the phase of obtaining the distances between candidates and examination sites. For this, it is necessary to use a tool that provides these distances. Currently, many tools can perform this task, most of which charge for their use (commonly, they do not charge a limited amount of data at the beginning of the use). Among them, we can highlight Google Maps[3] and Bing Maps[4]. There are also free platforms, with Open Street Maps[5] and QGIS[6] being the most notorious. As a large amount of distance queries is expected, there is the need to develop an automated script for the task of querying these distances on the chosen platform, which usually offers an API (application programming interface).

3.2 Pre-processing

After the acquisition of the distances, a pre-processing stage must be performed before using the mathematical formulation to obtain a solution. This stage is essential to identify data inconsistencies and to define which information will be passed for the formulation. A paramount concern is how to deal with candidates for whom the map tool was unable to locate using the given address or to obtain the distance to at least one exam location. In such a situation, we decided to disregard any distance from the candidate to exam locations, but still considering his allocation, as he is going to take the exam anyway. If the percentage of candidates with this inconsistency is high, the whole result may be compromised.

Another concern is about the candidates living far from all exam locations. Those candidates will not leave their homes on the exam day. Usually, they travel beforehand to a place closer to the exam location and travel from there on the day. Since it is not possible to know the exact travel origin for them, a distance cutoff must be considered. Candidates exceeding this cutoff will have the same treatment as the ones with data inconsistency. All distances between their addresses and exam locations will be set to zero, but they will be allocated to an exam location anyway.

[3] https://www.google.com/maps
[4] https://www.bing.com/maps
[5] https://www.openstreetmap.org
[6] https://www.qgis.org

3.3 Mathematical Formulation

As mentioned previously, the proposed mathematical formulation is an extension of the original formulation of the Single Source Capacitated Facility Location Problem [18]. Let C, L, and T be the set of candidates, locations, and exams types, respectively, obtained in the previous stages. The set of candidates eligible to take the exam of a given type is defined as C_t. The distance between the candidates and the locations is given by d_{ij}, where $i \in C$ and $j \in L$. The cost of using a location to offer any exam is c_j, and, if it is opened, it provides a capacity of Q_j candidates. We introduce two binary variables, x_{ij} and y_j^t, representing whether candidate $i \in C$ is assigned to location $j \in L$, and whether location $j \in L$ offers the exam of type $t \in T$, respectively. Note that variables y_j^t are also responsible for controlling whether a location $j \in L$ is opened at all. The integer programming formulation is presented as follows.

$$\min \sum_{i \in C} \sum_{j \in L} d_{ij} x_{ij} + \sum_{j \in L} \sum_{t \in T} c_j y_j^t \tag{1}$$

subject to

$$\sum_{j \in L} x_{ij} = 1 \qquad \forall i \in C \tag{2}$$

$$\sum_{t \in T} y_j^t \leq 1 \qquad \forall j \in L \tag{3}$$

$$\sum_{i \in C_t} x_{ij} \leq Q_j y_j^t \qquad \forall j \in L, t \in T \tag{4}$$

$$x_{ij} \in \{0, 1\} \qquad \forall i \in C, j \in P \tag{5}$$

$$y_j^t \in \{0, 1\} \qquad \forall j \in P, t \in T \tag{6}$$

The objective function (1) minimizes the total distance between the candidates and their chosen exam locations. Constraints (2) force all candidates to be associated with exactly one location. Constraints (3) imposes a given place to offer only one type of exam. Constraints (4) bind both variables and also limits the number of candidates in a given place. Variables domains are given by (5) and (6).

4 Application

4.1 The Real Problem

In Brazil, most of the top-ranked universities are publicly funded. In the past, and for many years, these universities were responsible for organizing their selection processes, called *vestibular* in Portuguese. Nowadays, there are still many universities that continue to use the *vestibular* to evaluate the students interested in entering its ranks, despite the Brazilian government's efforts to offer a nationwide test. One of the institutions that have its selection process is the Federal University of Juiz de Fora (UFJF), one of the thirty largest federal colleges in the country in the number of students [9], having offered, for admission in the year 2020, a total of 4,592 positions in its undergraduate courses distributed in two campi, with about 50%, therefore 2,303 vacancies, for its selection process, called Mixed Selective Admission Program (in Portuguese, *Programa de Ingresso Seletivo Misto* – PISM).

This exam is divided into three stages – one for each of the years of high school in Brazil –, each stage having a different exam with the content of the specific series in which the candidate is. Thus, this exam is only for students attending regular secondary schools, and it has the goal of integration between higher education and secondary school since it assesses the student's knowledge year by year. At the end of the last stage – which occurs when the candidate completes the last year of secondary school – the grades of the candidate in each of the previous stages are summed so that students can be classified according to the intended course.

In its last edition, there were exams applied in five cities in two states, with thousands of candidates distributed in 78 test locations, requiring the participation of more than 4,500 people directly involved in the application of the exams, with still hundreds of people working with planning, logistics and subsequent correction and processing of the results. Due to its large scale, the process officially begins in July with the publication of the public call. The exams are applied on the first weekend of December (in two days), with the grading being started already in the week following the exam. The results are published starting in January, for candidates who finished secondary school and, therefore, competing for the available positions. Until March, all grades are published, including candidates from the first stages exams who still do not compete for positions.

In recent years, the PISM has seen a substantial increase in the number of candidates and, consequently, the logistic challenges into providing a safe, affordable, and efficient exam. For the 2018's admission, for instance, the number of candidates did not reach 30,000, with this amount increasing to 34,000 in 2019, and to more than 40,000 in 2020, with a total increase of 78% from 2015 to the current exam. Such growth has caused considerable impacts in the cities where the exam is applied. Traffic jams and other difficulties caused not only by a large number of people in transit in the exam days but also by the fact that many of these people have to move to exam locations far away of their residences – since their addresses are not taken into account for the allocation of candidates. Those are significant concerns of the university in its effort to provide a calm environment for the candidates.

4.2 Data Collection and Distance Acquisition

We applied the proposed methodology for the last edition of the exam (PISM 2020). We used real data from candidates in two of the five cities where the tests were applied, Petrópolis and Volta Redonda, both located in the state of Rio de Janeiro. For candidates of each city, the following procedure was executed.

The candidates' addresses were made available by the university in tables using a random identification code, containing their postal codes and the exam locations where the candidates were allocated in the previous edition of the exam. Thus, there was no access to personal data other than the candidates' addresses, with the impossibility to identify the individuals to which the addresses belong. It is worth highlighting, in this way, the compliance of this work with personal data protection and the right to information privacy for all candidates. The test locations were also made available by the institution, and their addresses were obtained after simple internet searches.

To acquire the distances between each candidate and each exam location, we developed an algorithm in Python that, using an access key created directly on the Google Maps platform, allows access to the company's routing service through the API Distance Matrix. This service provides distance queries between two registers, taking into account the actual shortest available path and road directions. We chose not to use the Euclidean distance between the points – or any tool that offers this type of distance – due to the obvious problems inherent in this type of calculation, as it completely disregards existing road sections. We decided to use the Google Maps platform in detriment of other alternatives, such as Open Street Maps or Bing Maps, as it is the most widely used geographic information system today, which offers high reliability of the available data.

For the city of Petrópolis, 25,767 distance queries were performed to the Google Distance Matrix API, corresponding to 3,681 candidates who took the tests in the city multiplied by seven possible exam locations. The query of all requests took an hour and three minutes to run. From the queries results, only 770 (3.0%), returned empty values. The distance queries between the candidates from Volta Redonda and the exam locations took 44 minutes and generated a table of 18,249 entries, considering the total of 2,607 candidates and seven locations. Three hundred fifty queries did not return values, which represent 1.9% of the total.

4.3 Pre-processing

From the possession of the distance matrices, we then started processing the data for later use in the mathematical formulation. First, as explained in the methodology, we chose not to discard the candidates with empty distances because even if we do not have their distances to the exam locations, they will be allocated. Thus, these candidates remain in the matrix, but their distances are considered as zero.

We noted vast distances between exam locations and many candidates, such as some who live more than 3,000 kilometers away from the city of exams, which is equivalent to almost two days of uninterrupted driving. Therefore, as also explained in the methodology section, these candidates have their distances set to zero, so they could be allocated, but the formulation would be indifferent about distance costs. The challenge here is to define what distance would be considered as the cutoff parameter. Considering that the average distance between the exam locations in each city does not exceed five kilometers, we decided to set to zero all distances for candidates that are ten times farther than this value, i.e., candidates who live more than 50 kilometers from the exam locations. Most of these candidates will reach the city of the exam previously, not being possible to check their real origin addresses.

Thus, after applying the proposed treatment, 1,634 entries of the distance matrix for Petrópolis were not set to zero, i.e., 44% of the candidates who take the test in this city have addresses that are less than 50km from all exam locations and, therefore, they will be considered in order to minimize their displacement. Applying the same procedure for data referring to the city of Volta Redonda, it was found that about 41% of the candidates attended the established cutoff parameter, which is equivalent to 1,071 candidates.

4.4 Results

From the distance matrices generated and treated previously, we solve the mathematical formulation proposed in Section 3.3 in an Intel Core i5 6400 2.7GHz with 16GB of RAM using the AIMMS software in its academic version. Besides the distance matrices, we also input the parameters of the exam site capacity, the parameter that identifies which candidate takes which type of test, and the parameter of logistical cost to open an exam site. It is defined as a large value that impacts the calculation of the objective function.

The running times for both cities were small, 58 seconds for Petrópolis, and 38 seconds for Volta Redonda. After subtracting the logistical cost of opening the exam locations from the objective function value, we obtained a total travel distance of 15,078 kilometers for the city of Petrópolis and 15,326 kilometers for the city of Volta Redonda.

To assess the quality of the solutions, we have calculated the total distance between the candidate addresses and the exam locations where they were allocated in the previous selection process. We used the same distance matrices previously obtained after the pre-processing stage, disregarding candidates with query issues or with any distance higher than 50 kilometers. Summing all values, the total displacement performed by the candidates in Petrópolis was 17,337 kilometers, and in the case of Volta Redonda, the total value was 17,089 kilometers. These totals mean an average reduction of 13% in the optimal total displacement for the first city, and a 10% for the second city. In absolute values, reductions of 1.4 and 1.6 kilometers, on average, were obtained for Petrópolis and Volta Redonda, respectively.

Table 1 summarizes the results. For each city, we present the total number of candidates, the number of candidates with distances different than zero, the original total distance and average distance, the optimized total distance and average distance, and the total percentage reduction on the distances.

It is also important to mention the results of the proposed methodology when analyzing the average displacement separated by distance ranges. We separate the candidates, based on their average distances to exam locations, in ranges of 12.5 km and obtain the performed average displacement. Subsequently, fol-

		Petrópolis	Volta Redonda
Candidates	Total	3,681	2,607
	With distances	1,634	1,071
Original	Total distance	17,337	17,089
	Average distance	10.6	16.0
Optimized	Total distance	15,078	15,326
	Average distance	9.2	14.3
	Reduction	13.0%	10.3%

Table 1. Summary results for Petrópolis and Volta Redonda

lowing the defined separation, we measure the original and optimized average displacement.

We present the results for different ranges in Table 2. For each range, we present the average original and optimized distances in meters, and the total distance reduction, for both cities. From this table, it is possible to notice that the most substantial reduction appears for candidates who live closer to the exam locations. As expected, as the distance range increases, the reduction decreases since the impact for a location change would not be high for a candidate who lives far from the exam locations.

	Petrópolis			Volta Redonda		
Range	Original	Optimized	Reduction	Original	Optimized	Reduction
[0.0, 12.5]	5,461	4,334	20.6%	6,196	4,682	24.4%
[12.5, 25.0]	18,209	16,702	8.3%	17,520	15,131	13.6%
[25.0, 37.5]	30,466	27,499	9.7%	34,943	33,481	4.2%
[37.5, 50.0]	41,491	38,677	6.8%	39,906	38,266	4.1%

Table 2. Results considering different distance ranges

To further illustrate the improvement obtained with the methodology, we present in Figures 2 and 3 the comparison between the original average distances and the optimized average distances, for each distance range. We can notice a definite improvement for all distance ranges.

5 Conclusion

This work proposed a methodology to solve a university admission exam location problem. We discussed the assumptions made during all stages, covering the data collection, distance acquisition, pre-processing, and the problem solution using a mathematical formulation. We further apply our methodology in a real-life application arising from a Brazilian university, using real data, and performed extensive experimental analyses.

The results are in line with an old demand from candidates who are, this way, allocated more efficiently, reducing the need to travel long distances to take

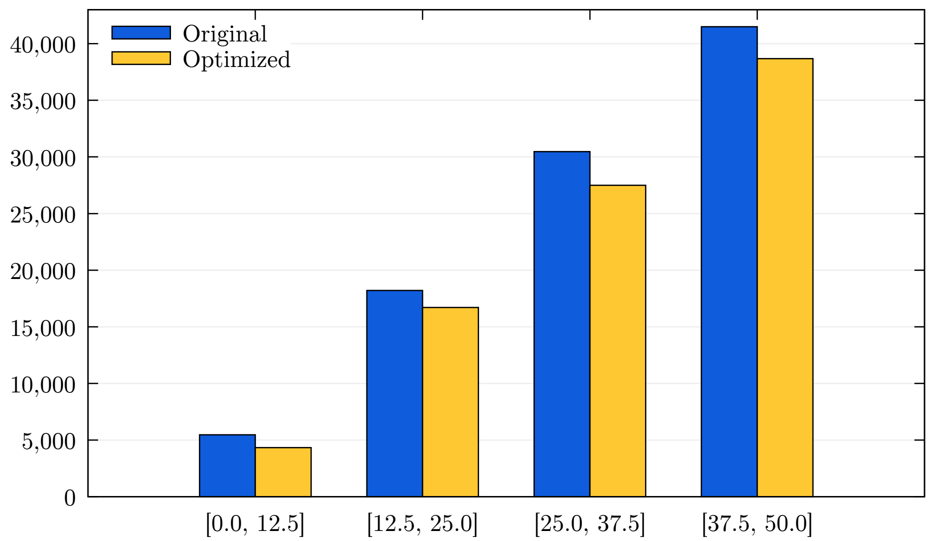

Fig. 2. Reduction on the average distance for each distance range - Petrópolis

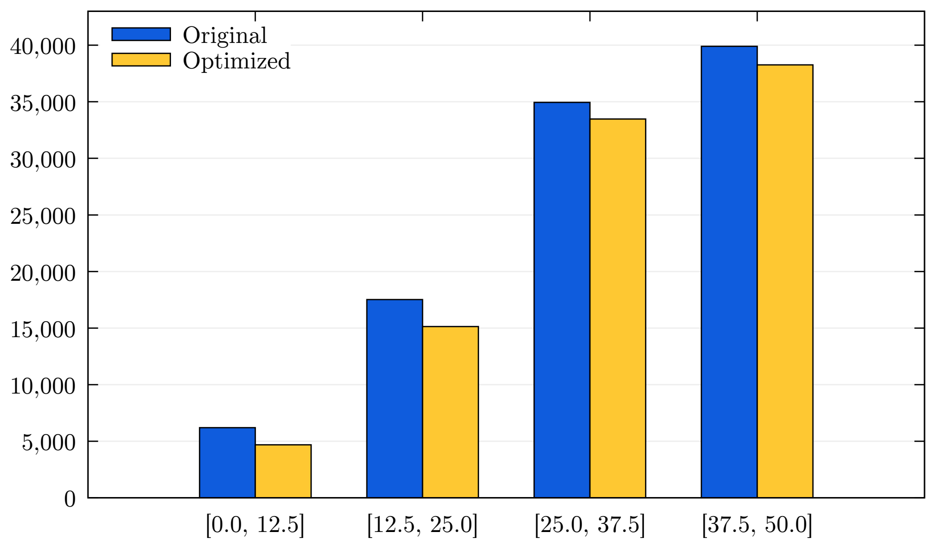

Fig. 3. Reduction on the average distance for each distance range - Volta Redonda

their exams, especially for those who live in the cities where they occur, thus reducing some of the stress arising on such an important date, in addition to relieving traffic in the city, especially on its central roads. Also noteworthy is the reduction on travel distances and even a possible reduction in the number of absent candidates (which requires future studies), considering that part of candidates missed the exam because they were unable to arrive in time to the locations where they were allocated due to difficulties in transfer caused by the long distances they had to travel.

It should be noted that the exam locations used in the optimization are the same as in the previous real process, with no possibility of including another exam location, maybe closer to the candidates' addresses. In this way, it is expected that if this formulation is used with many available exam locations, a number significantly higher than the current ones, mainly if these locations are distributed in different regions of the city, the reduction in displacement may be even more accentuated – making this a possibility for future research.

6 Acknowledgements

This research was partially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), grants 313521/2017-4 and 425962/2016-4, and by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), Financing Code 001.

We also would like to thank the Universidade Federal de Juiz de Fora (UFJF), and the Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), which supported the development of this work. All support is gratefully acknowledged.

References

1. de Andrade, R.Q., Silva, A.L.: Metodologia baseada no algoritmo de branch-and-bound para localização de sedes de vestibulares da UFOP. In: Simpósio Brasileiro de Pesquisa Operacional. pp. 2876–2885 (2011), in Portuguese
2. Antunes, A., Peeters, D.: A dynamic optimization model for school network planning. Socio-Economic Planning Sciences 34, 101–120 (2000)
3. Bruno, G., Esposito, E., Genovese, A., Piccolo, C.: Institutions and facility mergers in the italian education system: Models and case studies. Socio-Economic Planning Sciences 53, 23–32 (2016)
4. Burkey, M., Bhadury, J., Eiselt, H.: A location-based comparison of health care services in four u.s. states with efficiency and equity. Socio-Economic Planning Sciences 46, 157–163 (2012)
5. Cabrera, A.F., La Nasa, S.M.: Understanding the college-choice process. New directions for institutional research 2000(107), 5–22 (2000)
6. Cotes, N., Cantillo, V.: Including deprivation costs in facility location models for humanitarian relief logistics. Socio-Economic Planning Sciences 65, 89–100 (2019)
7. Gale, D., Shapley, L.S.: College admissions and the stability of marriage. The American Mathematical Monthly 69(1), 9–15 (1962)
8. Huang, R., Kim, S., Menezes, M.B.: Facility location for large-scale emergencies. Annals of Operations Research 181, 271–286 (2010)

9. Instituto Nacional de Estudos e Pesquisas Educacionais: Censo da Educação Superior (2018 – accessed February 2, 2020), http://portal.inep.gov.br/censo-da-educacao-superior, in Portuguese
10. Klose, A., Drexl, A.: Facility location models for distribution system design. European Journal of Operational Research 162(1), 4–29 (2005)
11. Kripka, R.M.L., Kripka, M., da Silva, M.C.: Formulação para o problema de alocação de salas de aula com minimização de deslocamentos. In: Simpósio Brasileiro de Pesquisa Operacional. pp. 1941–1951 (2011), in Portuguese
12. Lin, Y.H., Batta, R., Rogerson, P.A., Blatt, A., Flanigan, M.: Location of temporary depots to facilitate relief operations after an earthquake. Socio-Economic Planning Sciences 46, 112–123 (2012)
13. Matisziw, T.C., Murray, A.T.: Siting a facility in continuous space to maximize coverage of a region. Socio-Economic Planning Sciences 43, 131–139 (2009)
14. Melo, M., Nickel, S., da Gama, F.S.: Facility location and supply chain management – a review. European Journal of Operational Research 196, 401–412 (2009)
15. Murali, P., Ordóñez, F., Dessouky, M.M.: Facility location under demand uncertainty: Response to a large-scale bio-terror attack. Socio-Economic Planning Sciences 46, 78–87 (2012)
16. Owen, S.H., Daskin, M.S.: Strategic facility location: A review. European Journal of Operational Research 111, 423–447 (1998)
17. Shen, G.: Location of manufactured housing and its accessibility to community services: a gis-assisted spatial analysis. Socio-Economic Planning Sciences 39, 25–41 (2005)
18. Sridharan, R.: The capacitated plant location problem. European Journal of Operational Research 87(2), 203–213 (1995)
19. Ulukan, Z., Demircioglu, E.: A survey of discrete facility location problems. Int. J. Soc. Behav. Educ. Econ. Bus. Ind. Eng 9(7), 2487–2492 (2015)
20. Vieira, I.F.G., de Barros, M.F., Cormack, A.: Hierarchical facility location model for allocating cancer treatment units in interior of rio de janeiro. Production 29 (2019)
21. de Werra, D.: An introduction to timetabling. European journal of operational research 19(2), 151–162 (1985)