



UNIVERSIDADE FEDERAL DO MARANHÃO

Curso de Ciência da Computação

Robert Douglas de Araujo Santos

**Agrupamento capacitado de unidades de leitura  
utilizando particionamento binário do espaço e  
algoritmo genético**

São Luís - MA

2019

Robert Douglas de Araujo Santos

**Agrupamento capacitado de unidades de leitura  
utilizando particionamento binário do espaço e algoritmo  
genético**

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão como parte dos requisitos necessários para obtenção do grau de bacharel em Ciência da Computação.

Orientador: Prof. Dr. João Dallyson Sousa de Almeida

São Luís - MA

2019

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).  
Núcleo Integrado de Bibliotecas/UFMA

Araujo Santos, Robert Douglas de.

Agrupamento capacitado de unidades de leitura  
utilizando particionamento binário do espaço e algoritmo  
genético / Robert Douglas de Araujo Santos. - 2019.  
59 f.

Coorientador(a): Italo Francyles Santos da Silva.

Orientador(a): João Dallyson Sousa de Almeida.

Monografia (Graduação) - Curso de Ciência da  
Computação, Universidade Federal do Maranhão, São Luís,  
2019.

1. Agrupamento Capacitado. 2. Algoritmo Genético. 3.  
Logística de Leitura. 4. Particionamento Binário do  
Espaço. I. Almeida, João Dallyson Sousa de. II. Silva,  
Italo Francyles Santos da. III. Título.

Robert Douglas de Araujo Santos

## **Agrupamento capacitado de unidades de leitura utilizando particionamento binário do espaço e algoritmo genético**

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão como parte dos requisitos necessários para obtenção do grau de bacharel em Ciência da Computação.

Trabalho aprovado em 19/12/2019 - São Luís - MA:

---

Prof. Dr. João Dallyson Sousa de Almeida  
Orientador  
Universidade Federal do Maranhão

---

Me. Italo Francyles Santos da Silva  
Coorientador  
Universidade Federal do Maranhão

---

Prof. Dr. Anselmo Cardoso de Paiva  
Examinador  
Universidade Federal do Maranhão

---

Prof. Dr. Geraldo Braz Júnior  
Examinador  
Universidade Federal do Maranhão

São Luís - MA

2019

À minha família e meus amigos.

# Agradecimentos

Primeiramente aos meus pais, Irismar e Nonato por sempre me incentivarem em meus estudos, pelos ensinamentos de vida, que independente das dificuldades sempre me proporcionaram o melhor que eu poderia ter e por sempre apoiarem em minhas decisões sejam elas pessoais ou acadêmicas. A eles minha eterna gratidão.

A minha namorada Letícia, por estar ao meu lado do início ao fim desta caminhada, pela paciência, carinho e mais importante, por sempre acreditar no melhor de mim.

A todos os sobreviventes da turma 2015.1 e integrantes do grupo "Codebuilders", os quais eram desconhecidos no primeiro de dia de aula e hoje se tornaram amigos para a vida toda. Por proporcionarem momentos de descontração, felicidade e sempre estarem dispostos a ajudar independente problema.

A todos os colegas que conheci durante minha passagem no PET de Computação, que mesmo sendo breve, me acolheram como um velho amigo e propiciaram momentos inesquecíveis e por permitirem integrar essa grande família.

A todos os integrantes do NCA, por me acolherem e transformarem o que deveria ser um "simples" laboratório em um segundo lar. Por serem pessoas dotadas de grande sabedoria e humildade, as quais estão sempre dispostas a compartilhar seus conhecimentos.

A todos que compuseram a equipe maravilhosa do projeto SAUL em que tive a honra de trabalhar, a qual me ensinou a raiva e o valor do trabalho em equipe e mais importante, por nunca soltarem as mãos. Em especial Italo, Denes, Polyana, Pedro Henrique, Morão, Mr. Anderson, Mayara, Vitu (Victor). Por fim, um obrigado especial a nossa querida mãe, Daniel, por nos defender em momentos tensos nas reuniões de "sprint". A Weldson e Alexandre que mesmo que não compoem a equipe, agora fazem parte do "grupo bom".

A meu coorientador Italo, pela paciência de ensinar um novato em um laboratório que achava saber alguma coisa. Compartilhando conhecimentos nas mais diversas áreas, os quais permitiram o despertar do meu interesse pela pesquisa e pelos assuntos abordados nesse trabalho. Obrigado pela inestimável orientação no desenvolvimento desse trabalho e em como beber uma mag gelada.

Como prometido, um paragrafo de agradecimento para meu amigo Anderson, por sempre assustar os compiladores e permitir que meu códigos voltem a funcionar.

A meu amigo Alexandre que mesmo conhecendo de verdade a pouco tempo, salvou minha pele em momentos cruciais.

A meu orientador prof<sup>o</sup> João Dallyson, pelas oportunidades que me concedeu, pela orientação e liberdade durante a realização desse trabalho, me permitindo sempre levar minhas ideias adiante.

A todos os professores que tive o prazer de ter durante a jornada da graduação, os quais ensinaram muito além de conteúdos didáticos, conhecimentos que sempre levarei comigo. Um agradecimento especial a querida coordenadora do curso de Ciência da Computação, a prof<sup>a</sup> Simara Vieira, por se prontificar de imediato a defender os meus direitos como aluno no momento em que mais precisei e por ser uma verdadeira mãe para todos os alunos do curso. Aos professores Gerado e Anselmo, agradeço imensamente pelas oportunidades de integrar o PET de Computação e o NCA. Essas oportunidades, mudaram completamente a minha percepção do curso, expandiram o meu horizonte de forma imensurável e permitiram o despertar do meu interesse pela pesquisa.

Por fim, a todos aqueles que me ajudaram de forma direta ou indireta durante a minha jornada até aqui.

*“I saw that ending from miles away.”*  
*(Templar, Defense of the Ancients 2)*



# Resumo

O processo de faturamento das companhias de energia no Brasil está ligado à logística de leitura do consumo elétrico de seus clientes. Um processo eficiente e equilibrado traz benefícios, como a redução de custos operacionais e a melhora da qualidade do serviço prestado. A criação de grupos de medição do consumo energia elétrica, que em geral é realizado manualmente por analistas especializados, compõe o processo de leitura. Nesse cenário, este trabalho propõe uma abordagem computacional para o problema de agrupamento capacitado, aplicado à logística de leitura em companhias de distribuição de energia elétrica atuantes nos estados do Maranhão e Pará. No agrupamento capacitado, os elementos são associados a pesos para a construção de grupos com capacidades limitadas. O objetivo para solucionar essa modalidade de problema, é a criação de partições que minimizem a dispersão interna dos grupos. Neste trabalho, é apresentado um método baseado em algoritmo genético e particionamento binário do espaço, com a inclusão da restrição de capacidade para a definição dos grupos. Os resultados obtidos pelo método proposto são comparados com a situação atual sob a análise do coeficiente de silhueta e minimização do desvio padrão dos tempos de leitura. A partir dos resultados obtidos tem-se que, o método proposto apresenta uma redução expressiva no desvio padrão da capacidade do agrupamento, redução da quantidade de grupos e um aumento do coeficiente de silhueta.

**Palavras-chaves:** Logística de Leitura, Agrupamento Capacitado, Algoritmo Genético, Particionamento Binário do Espaço.

# Abstract

In Brazil, the billing process in power companies is tied to electric consumption readings. A balanced and efficient process brings benefits such as operational costs reduction and improves the quality of services. Creating a measurement group, which a specialized analyst do manually, compose the reading process. A capacitated clustering applied to the reading planning of Maranhão's and Pará's power companies using a genetic approach and capacity restrained binary space partitioning is proposed. In capacitated clustering, all elements have a weight assigned to them and it aims to minimize intra-cluster dispersion. Silhouette score and minimization of standard deviation used to compare the resulting clusters with actual plannings used in the companies. Considering the results, the method proposed presents a expressive reduction in the standard deviation of the grouping load, a reduction in the number of groups and an silhouette coefficient increase.

**Keywords:** Readings process logistics, Capacitated clustering, genetic algorithm, Binary space partitioning

# Lista de ilustrações

Figura 1 – Divisão do espaço 3D. . . . .	24
Figura 2 – BSP após duas execuções . . . . .	24
Figura 3 – BSP gerada utilizando os eixos. . . . .	25
Figura 4 – Funcionamento de um Algoritmo Genético . . . . .	26
Figura 5 – Processo de cruzamento . . . . .	28
Figura 6 – Representação gráfica de uma KD-Tree: (a) representação no plano (b) representação em árvore . . . . .	29
Figura 7 – Visão geral do método proposto . . . . .	31
Figura 8 – Comparação entre as áreas urbana e rural da cidade de Açailândia: (a) área urbana (b) área rural . . . . .	33
Figura 9 – Instalações distantes do centro urbano . . . . .	33
Figura 10 – Instalações no espaço 2D . . . . .	35
Figura 11 – Execução do processo de localização da região densa. . . . .	36
Figura 12 – Antes e depois da aplicação do pré-processamento na cidade de Açailândia-MA . . . . .	36
Figura 13 – Representações gráficas de um gene e um cromossomo . . . . .	37
Figura 14 – Variedade de aresta da população . . . . .	37
Figura 15 – Processo de mutação . . . . .	39
Figura 16 – Particionamento de uma etapa . . . . .	40
Figura 17 – Planejamentos vigentes da cidade de Açailândia: (a) agrupamento urbano (b) agrupamento rural . . . . .	44
Figura 18 – Planejamentos propostos para a cidade de Açailândia: (a) agrupamento urbano (b) agrupamento rural . . . . .	45
Figura 19 – Comparação entre ULS urbanas dos planejamentos de Açailândia-MA: (a) ULs vigentes (b) ULs propostas . . . . .	45
Figura 20 – Planejamentos vigentes da cidade de Imperatriz-MA: (a) agrupamento urbano (b) agrupamento rural . . . . .	46
Figura 21 – Planejamentos proposto para a cidade de Imperatriz-MA: (a) agrupamento urbano (b) agrupamento rural . . . . .	47
Figura 22 – Comparação entre ULS urbanas dos planejamentos de Imperatriz-MA: (a) ULs vigentes (b) ULs propostas . . . . .	47
Figura 23 – Planejamentos vigentes da cidade de Abaetetuba-PA: (a) agrupamento urbano (b) agrupamento rural . . . . .	48
Figura 24 – Planejamentos proposto para a cidade de Abaetetuba-PA: (a) agrupamento urbano (b) agrupamento rural . . . . .	49

Figura 25 – Comparação entre ULS urbanas dos planejamentos de Abaetetuba-PA: (a) ULS vigentes (b) ULS propostas . . . . .	49
Figura 26 – Planejamentos vigentes da cidade de Santarém-PA: (a) agrupamento urbano (b) agrupamento rural . . . . .	50
Figura 27 – Planejamentos propostos para a cidade de Santarém-PA: (a) agrupamento urbano (b) agrupamento rural . . . . .	51
Figura 28 – Comparação entre ULS urbanas dos planejamentos de Santarém-PA: (a) ULS vigentes (b) ULS propostas . . . . .	52
Figura 29 – <i>BoxPlot</i> da Carga de trabalho de ULs: (a) Área urbana Açailândia; (b) Área rural Açailândia; (c) Área urbana Abaetetuba; (d) Área rural Abaetetuba; (e) Área urbana Imperatriz; (f) Área rural Imperatriz; (g) Área urbana Santarém; (h) Área rural Santarém. . . . .	54

# Lista de tabelas

Tabela 1 – Comparação de métricas para cidade de Açailândia-MA . . . . .	44
Tabela 2 – Comparação de métricas para cidade de Imperatriz-MA . . . . .	46
Tabela 3 – Comparação de métricas para cidade de Abaetetuba-PA . . . . .	48
Tabela 4 – Comparação de métricas para cidade de Santarém-PA . . . . .	50

# Lista de abreviaturas e siglas

2D	Segunda Dimensão
3D	Terceira Dimensão
ACCG	Agrupamento Capacitado com Centro Geométrico
AG	Algoritmo Genético
ANEEL	Agência Nacional de Energia Elétrica
BSP	Binary Space Partition
C-GNG	Capacitated Growing Neural Gas
GNG	Growing Neural Gas
GRASP	Greedy Randomized Adaptive Search Procedure
NP	Tempo polinomial não determinístico
PAC	Problema de Agrupamento Capacitado
PACC	Problema de Agrupamento Centrado Capacitado
UL	Unidade de Leitura

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>16</b>
1.1	Justificativa	17
1.2	Objetivo Geral	18
1.2.1	Objetivos Específicos	18
1.3	Organização do Trabalho	18
<b>2</b>	<b>TRABALHOS RELACIONADOS</b>	<b>20</b>
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>22</b>
3.1	Agrupamento Capacitado	22
3.2	Particionamento Binário do Espaço	23
3.3	Algoritmo Genético	25
3.4	KD-Tree	28
3.5	Z-Score	29
<b>4</b>	<b>METODOLOGIA</b>	<b>30</b>
4.1	Visão Geral	30
4.2	Aquisição dos dados	32
4.3	Caracterização do cenário de aplicação	32
4.4	Pré-Processamento	34
4.5	Agrupamento	35
4.5.1	Execução do Algoritmo Genético	36
4.5.2	Particionamento do espaço	39
4.6	Pós-processamento e Avaliação	41
4.6.1	Minimização do desvio padrão	41
4.6.2	Coeficiente de Silhueta	42
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>43</b>
5.0.1	Planejamento para a cidade de Açailândia	44
5.0.2	Planejamento para a cidade de Imperatriz	46
5.0.3	Planejamento para a cidade de Abaetetuba	48
5.0.4	Planejamento para a cidade de Santarém	50
<b>5.1</b>	<b>Discussão</b>	<b>53</b>
<b>6</b>	<b>CONCLUSÕES</b>	<b>55</b>

<b>REFERÊNCIAS</b>	<b>56</b>
--------------------	-----------



# 1 Introdução

As companhias que atuam no setor energético brasileiro têm, como algumas de suas principais atribuições a gestão da leitura do consumo de energia e o processo de faturamento. No Brasil, essas companhias são regulamentadas pela Agência Nacional de Energia Elétrica (ANEEL), a qual visa melhorar o relacionamento entre as empresas fornecedoras de energia e os consumidores. Também é estipulado pela ANEEL, através da resolução normativa N° 414, um intervalo de dias para que as concessionárias finalizem o processo de medição do consumo de energia elétrica (ANEEL, 2010).

A coleta de dados, referente ao consumo dos clientes, e a entrega das faturas, têm um grande impacto no lucro das empresa. Nos estados do Maranhão e Pará, esse processo tende a ser realizado de forma manual, ou seja, faz-se necessária a visita de um funcionário ao endereço do cliente para a realização da leitura do consumo de energia elétrica registrado nos medidores de cada instalação ativa. As instalações são a associação de um medidor a uma localização física atrelada a uma cliente.

No planejamento, as instalações de um município são organizadas em grupos, denominados etapas, as quais representam um determinado dia do mês e também devem respeitar a resolução N° 414 da ANEEL. Cada etapa é dividida em grupos menores de instalações denominados de Unidades de Leitura (UL). Para cada UL existe um funcionário responsável pela leitura e entrega das faturas de cada cliente deste subgrupo de instalações.

Ressaltam-se alguns fatores, que podem influenciar na execução do planejamento mensal, como por exemplo, instalações que são ativadas ou desativadas e a quantidade de municípios que na prática possuem organizações distintas.

O processo de criação do planejamento mensal é abordado na literatura como distritamento. Esse problema consiste em agrupar espaços de território em distritos contínuos, não sobrepostos, dada uma função objetivo e um conjunto de restrições.

Para a execução do processo de leitura, precisa-se levar em consideração a carga de trabalho do funcionário, que se deslocará até o medidor e realizará a leitura do consumo de energia elétrica, portanto caracteriza-se como a especificação do problema de distritamento, conhecido como agrupamento capacitado.

Existem diversas aplicações do problema do distritamento como, por exemplo, a alocação de distritos eleitorais (BUSSAMRA, 1995) e (GARFINKEL; NEMHAUSER, 1970). Outra aplicação realizada é o distritamento domiciliar para clínica de saúde comunitária, observado em (BLAIS; LAPIERRE; LAPORTE, 2003). Também existem trabalhos desenvolvidos na área de transporte público, em (TAVARES-PEREIRA et al.,

2007) e (NOVAES et al., 2009). Nos cenários de vendas e marketing também existem aplicações deste problema, como foi abordado por (RÍOS-MERCADO; FERNÁNDEZ, 2009), (ZOLTNERS; SINHA, 1983), (EASINGWOOD, 1973) e (LODISH, 1975).

O problema de agrupamento é relativamente antigo na literatura e diversos métodos já foram propostos com o objetivo de solucioná-lo. Entre os mais clássicos estão o K-Means, apresentado pela primeira vez em (HARTIGAN; WONG, 1979), DBscan (ESTER et al., 1996) e sua variação GDBscan (SANDER et al., 1998) e P-Cluster (Judd; McKinley; Jain, 1998).

O problema de agrupamento capacitado (PAC) é uma especificação do problema de distritamento, em que cada grupo tem uma capacidade limitada, seja ela de quantidade de indivíduos ou área de extensão. Para o problema de agrupamento capacitado existem algumas variações como, P-Medianas apresentado em (STEFANELLO; MÜLLER, 2009) e Agrupamento centrado capacitado visto em (CARVALHO; SALDANHA; MARTINS, 2015).

Com o objetivo de solucionar o problema de agrupamento capacitado, vários trabalhos foram desenvolvidos, dentre eles: enfoque multicritério (FRANÇA et al., 2007), Recursive Capacitated Means (COSTA et al., 2018), o tabu search (NEGREIROS; PALHANO, 2006), métodos baseados em algoritmos evolutivos (MÜHLENBEIN; GORGES-SCHLEUTER; KRÄMER, 1988) e simulated annealing (SELIM; ALSULTAN, 1991).

Nesse contexto, este trabalho propõe um método, utilizando os conhecimentos de agrupamento capacitado, algoritmo genético e particionamento binário do espaço, para gerar grupos de forma capacitada e que melhorem os planejamentos de leitura criados pelas companhias de distribuição elétrica. Serão utilizados como base de dados, os planejamentos de leitura fornecidos pela Equatorial Energia, no decorrer do projeto P&D PD-00371-0031/2017. Os resultados serão expostos através da realização de comparações entre o agrupamento gerado pelo método proposto e a situação original dos planejamentos de leitura da companhia.

## 1.1 Justificativa

No Maranhão e Pará o processo de planejamento de leitura de consumo de energia elétrica é realizado mensalmente pelas concessionárias de forma manual. Essa atividade tem influência diretamente nos funcionários que realizam o planejamento e nos que executam as leituras, além de impactar nos clientes das companhias.

Os responsáveis pela organização dos planos de leitura se deparam com algumas dificuldades: o manuseio de quantidades massivas de dados e a construção de forma manual e empírica das unidades de leitura, pois contam apenas com as suas experiências prévias.

Os leiturista, funcionários responsáveis pela leitura do medidor, sofrem principalmente com problemas relacionados ao tamanho e disformidades dos grupos, que podem ser grandes, ultrapassando à carga de trabalho diária, ou pequenos, deixando o funcionário ocioso. Ambos prejudicam sua produtividade e sua relação com o empregador. Os clientes das companhias também são afetados no que diz respeito as suas faturas, pois podem recebê-las em datas erradas ou seu período de consumo pode ser afetado de alguma forma.

Por fim, a própria companhia pode sofrer perdas, tais como, fragilização da relação com seus funcionários, atrasos no processo de faturamento, necessidade de contratação de mais leituristas – na tentativa de realizar o processo de leitura e faturamento no tempo correto – e reclamações vindas dos clientes.

Neste cenário, faz-se necessário o uso de um método computacional que automatize boa parte deste processo, principalmente na geração das unidades de leitura. Com esse método, deve ser possível organizar instalações em grupos compactos, uniformes e que não ultrapassem a carga de trabalho máxima dos leituristas. Além de permitir a aceleração do processo de planejamento, tirando dos funcionários a necessidade de manipulação de uma grande quantidade de dados e por fim, minimizar custos com a realização de todo o processo de leitura do consumo.

## 1.2 Objetivo Geral

Este trabalho tem por objetivo propor um método capaz de agrupar instalações de energia de forma automática, utilizando as técnicas de particionamento binário de espaço e algoritmo genético. O método proposto almeja a obtenção de grupos melhores do que os gerados atualmente de forma manual.

### 1.2.1 Objetivos Específicos

- Construir um método para a geração de unidades de leitura de forma capacitada;
- Utilizar métricas da literatura para avaliar os grupos gerados;
- Analisar e comparar a solução proposta com o estado atual das companhias.

## 1.3 Organização do Trabalho

Este trabalho está organizado da seguinte forma: o Capítulo 2 apresenta trabalhos relacionados ao problema de agrupamento de unidades de leitura e planejamento de leitura; o Capítulo 3 apresenta conceitos e técnicas fundamentais para o entendimento da metodologia proposta; no Capítulo 4 é descrito o funcionamento do método proposto,

neste capítulo também são mostradas as métricas de avaliação utilizadas para avaliar o método; no Capítulo 5 são apresentados os resultados, sendo comparados com a situação vigente das distribuidoras de energia; e, por fim, o Capítulo 6 apresenta as conclusões e trabalhos futuros.

## 2 Trabalhos Relacionados

Este Capítulo apresenta trabalhos relacionados que exploram o problema de agrupamento de unidades de leituras em companhias de energia elétrica.

França et al. (2007) propuseram métodos heurísticos aplicados para a resolução de forma genérica dos problemas de otimização do agrupamento capacitado. Em ambas as propostas cada unidade consumidora é associada a um grafo conexo não-orientado do arruamento da cidade, no qual cada nó é representado pelo cruzamento de uma rua e o peso das arestas é baseado na quantidade de medidores próximos entre dois cruzamentos. Usando esse modelo de grafo, foram propostas soluções que utilizam o método Greedy Randomized Adaptive Search Procedure (GRASP) (FEO; RESENDE, 1995) tendo como objetivo a criação de grupos compactos, respeitando-se também, além da restrição de capacidade, fatores geográficos. Os resultados obtidos pela aplicação do algoritmo sugerem uma nova redefinição das áreas de faturamento com as instalações devidamente agrupadas obedecendo as restrições impostas.

Stefanello e Müller (2009) propuseram um método para agrupamento dos clientes de uma distribuidora de energia elétrica, com o objetivo de diminuir o tempo de execução dos serviços solicitados pelos clientes através de uma melhor distribuição de tarefas entre as equipes disponíveis. Cada cliente é associado a uma área de atendimento e a um setor específico, sendo que cada setor é acompanhado por uma equipe de eletricitas. Porém, por se tratar de uma distribuição manual, é comum que determinados setores possuam maior carga de trabalho enquanto outros tenham a carga menor. A metodologia aplicada envolve o uso do algoritmo das P-medianas de forma capacitada, que obteve resultados melhores em comparação à criação manual dos agrupamentos, prática até então adotada pela concessionária.

Já Steiner et al. (2009) elaboraram um método para otimizar o dimensionamento e logística das equipes de trabalho das concessionárias de energia elétrica para realizarem atendimento de uma determinada ocorrência. Propuseram uma solução em duas etapas, são elas: o desenvolvimento de um modelo matemático de Programação Linear Inteira, desenvolvido por Murty (1985), para otimização do número de equipes necessárias para o dia de trabalho; e a utilização do Algoritmo de Floyd (BRENT, 1980), para determinar a equipe mais adequada para fazer o atendimento.

Guerine et al. (2017) expuseram uma heurística híbrida para a solução do Agrupamento Capacitado com Centro Geométrico (ACCG). A técnica proposta utilizava uma combinação entre *Data Mining* (HAN; KAMBER; PEI, 2012), *Clustering Search* (CHAVES; LORENA, 2010) e *Simulated Annealing* (KIRKPATRICK; GELATT; VECCHI, 1983)

para buscar as melhores soluções no espaço de busca. A primeira é aplicada para detectar padrões e, a partir disso, guiar a busca por melhores soluções no espaço de busca. Aplicações reais para o ACCG são feitas, em projetos de zonas de coleta de lixo, definições de áreas de cobertura de vendedores, dentre outras.

Em 2018, [Costa et al. \(2018\)](#) apresentou o método RCMeans, baseado em uma abordagem recursiva e no algoritmo clássico de agrupamento K-Means, junto com a aplicação de uma restrição de capacidade para a definição do grupos. Obteve resultados superiores ao agrupamento realizado de forma manual pelos funcionários da companhia elétrica. Os resultados após a aplicação da técnica levaram em consideração a coesão, separação, número de grupos, índice de silhueta e tempo médio de leitura dos grupos gerados.

Por fim, [Silva \(2019\)](#) propôs um método para resolver o problema do agrupamento capacitado de uma distribuidora de energia elétrica do Maranhão. O trabalho apresenta o *Capacitated Growing Neural Gas* (C-GNG), um algoritmo baseado no *Growing Neural Gas* (GNG) que utiliza os conceitos de agrupamento capacitado. Seus experimentos foram comparados com a situação atual da distribuidora e demonstraram que o método apresentado cria agrupamentos mais eficientes, isto é, com tempos bem distribuídos e com grupos compactos.

Este trabalho se diferencia dos demais, pois apresenta um método para a criação capacitada de unidades de leitura utilizando o particionamento binário do espaço, comumente empregado na computação gráfica durante as etapas de renderização de cenas na terceira dimensão(3D), e algoritmo genético. Com a utilização do particionamento binário do espaço, garante-se que a restrição de não sobreposição de grupos do agrupamento capacitado seja respeitada. Sem a necessidade de implementação direta dessa restrição, se permite a redução da dimensionalidade do problema e dessa forma a redução do esforço computacional.

## 3 Fundamentação Teórica

O seguinte capítulo descreve técnicas e conceitos importantes para compreensão do método proposto. Primeiro, é necessário explicar sobre Agrupamento Capacitado, do qual tem-se conceitos e regras utilizados na criação de grupos. Logo após, será explanado sobre o particionamento binário do espaço usado na geração de grupos capacitados, algoritmo genético, o qual define a melhor configuração de particionamento do espaço. Por fim, é realizada uma breve descrição da árvore k-d e a z-score utilizadas na etapa de pré-processamento.

### 3.1 Agrupamento Capacitado

O problema de agrupamento capacitado (PAC) consiste em uma especificação do problema clássico de agrupamento, no qual dado um conjunto  $n$  de indivíduos, que possuem um peso ou demanda, devem ser distribuídos em  $p$  grupos com capacidades limitadas e sem sobreposição. Cada elemento do conjunto deve fazer parte de apenas um grupo. Por fim, ao somar-se os pesos dos componentes que constituem um grupo, a soma não deve ultrapassar a capacidade estipulada de maneira prévia. Logo, o objetivo é encontrar a melhor configuração, em que a distância dos indivíduos para o centro de seu grupo seja minimizada e a sua distância para os grupos vizinhos seja maximizada (FRANÇA; SOSA; PUREZA, 1999).

O agrupamento capacitado é um problema combinatorial complexo e foi provado ser NP-completo (GAREY; JOHNSON, 1979). Pela complexidade do problema, várias soluções são propostas baseadas em heurísticas e meta-heurísticas (DESALE et al., 2015). Por fim, dadas as aplicações reais, que diversas vezes são cenários dinâmicos, não se espera que algoritmos exatos solucionem este problema.

O PAC pode ser definido formalmente por um conjunto de equações que formam o modelo a seguir:

$$F(P) = \text{Min} \sum_{i \in I} \sum_{j \in J} d_{ij} x_{ij} \quad (3.1)$$

sujeito a:

$$\sum_{j \in J} y_j = p \quad (3.2)$$

$$\sum_{j \in J} x_{ij} = 1 \quad i \in I \quad (3.3)$$

$$x_{ij} \leq y_j \quad i \in I, j \in J \quad (3.4)$$

$$\sum_{i \in I} w_i x_{ij} \leq W \quad j \in J \quad (3.5)$$

$$x_{ij}, y_j \in \{0, 1\} \quad i \in I, j \in J \quad (3.6)$$

onde o conjunto de indivíduos  $I = \{1, 2, \dots, n\}$ , em que a matriz  $n \times n$  simétrica  $d_{ij}$ , que representa a distância entre um par de indivíduos; assumindo que a distância  $d_{ij} > 0$  e que  $d_{ii} = 0, \forall i, j \in I$ ; por sua vez, um peso positivo  $w_i$  é atribuído a cada indivíduo  $i$  do conjunto; sendo  $J = \{1, 2, \dots, m\}$  o conjunto de grupos e  $W_j$  suas capacidades respectivas. O *centroid* do grupo  $j$  é definido pelo indivíduo no qual a soma das distâncias a todos os outros indivíduos do grupo  $j$  é minimizada. Para fins de simplificação da notação, as capacidades  $W$  dos grupos, serão todas consideradas iguais a o valor de capacidade pré-estabelecida.

A variável  $y_j$  é binária, assumindo o valor 1 se o indivíduo  $j$  é atribuído como mediana do grupo  $j \in J$  e assume o valor 0 caso contrário. A variável binária  $x_{ij}$  indica se o indivíduo  $i$  pertence ao grupo  $j$  ou não, no caso afirmativo seu valor é 1 do contrário será 0.

A Equação 3.2 garante que somente  $p$  grupos serão selecionados. A restrição 3.3 garante que todos os indivíduos serão atribuídos a um grupo, já a restrição 3.4 previne a atribuição de indivíduos que não foram selecionados como medianas. A restrição 3.5 garante que a soma dos pesos atribuídos ao grupo não extrapole a capacidade máxima.

## 3.2 Particionamento Binário do Espaço

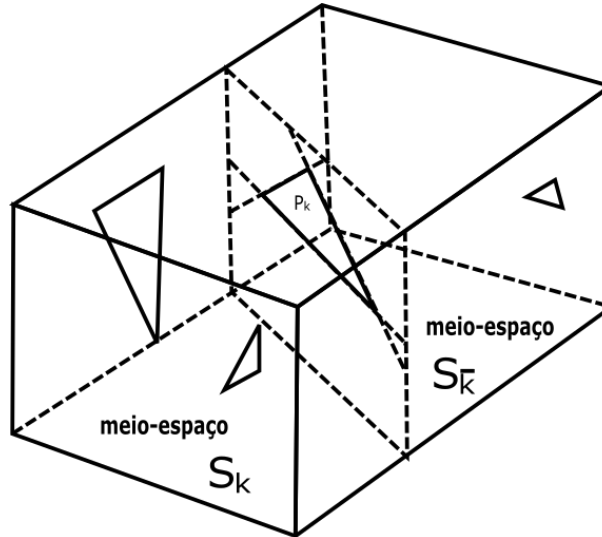
Particionamento binário do espaço (BSP), em inglês *binary space partitioning* é um processo de divisão do espaço proposto por Fuchs, Kedem e Naylor (1980) para resolver o problema de superfície oculta, além de gerar mais rapidamente imagens realistas de cenas em 3D compostas por polígonos. Esse processo constituía uma etapa de pré-processamento que tinha como objetivo principal a diminuição do tempo de execução e cálculos necessários para a geração de uma cena 3D, a qual comumente é construída por diversos polígonos. A etapa de pré-processamento é baseada na geração de uma estrutura de dados do tipo árvore chamada de *BSP-Tree*. Considerando que a cena é estática e ocorre apenas a mudança do ponto de observação, após a construção da árvore, a visitação produzirá uma ordem linear dos polígonos a serem renderizados.

No processo de construção da árvore BSP, considera-se um conjunto de polígonos  $P = \{P_1, P_2, \dots, P_n\}$  que definem o ambiente 3D; um polígono arbitrário  $P_k$  pertencente ao conjunto é escolhido; pode-se notar que o plano que contém o polígono  $P_k$  divide o espaço 3D em dois "meio-espacos", os quais são nomeados de  $S_k$  e  $\bar{S}_k$ . Os dois "meio-



espaços", são identificados como o lado positivo e negativo do polígono  $P_k$ . Considerando a prioridade de visibilidade e que o ponto de observação esteja em  $S_k$ , nenhum polígono pertencente a  $S_{\bar{k}}$  irá obstruir a visão de qualquer polígono  $P_k$  ou qualquer polígono em  $S_k$ , como observado na Figura 1.

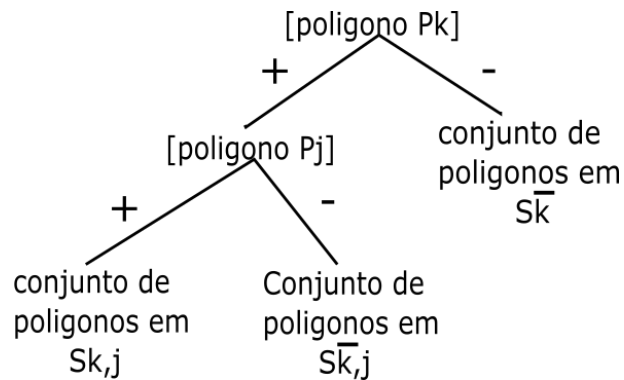
Figura 1 – Divisão do espaço 3D.



Fonte: Adaptado de Fuchs, Kedem e Naylor (1980)

O Processo de construção continua dividindo cada um dos polígonos em  $P - P_k$  ao longo do plano  $P_k$  dando origem a uma hierarquia de espaços que pode ser representada por uma estrutura de dados do tipo árvore, em que a raiz contém o polígono  $P_k$  inicial e cada subárvore  $s$  contém os polígonos associados aos "meio-espaços", como mostrado na Figura 2.

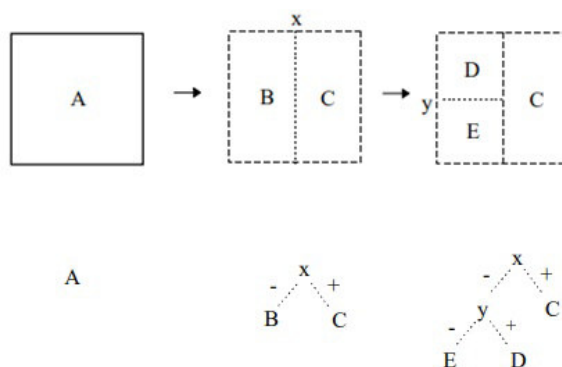
Figura 2 – BSP após duas execuções



Fonte: Adaptado de Fuchs, Kedem e Naylor (1980)

Conforme exposto anteriormente, a BSP apresenta uma subdivisão hierárquica e recursiva do espaço, cujo resultado são dois subespaços convexos. Mesmo que inicialmente tenha sido proposta para o ambiente 3D, em que o espaço é particionado por um hiperplano, esse conceito de particionamento pode ser facilmente aplicado em espaços com d-dimensões. Portanto, é possível aplicá-la em um espaço 2D onde o hiperplano se torna uma reta (STOWELL et al., 2011). Caso a reta de particionamento seja considerada como um dos dois eixos do plano cartesiano, um polígono pode ser particionado conforme a Figura 3.

Figura 3 – BSP gerada utilizando os eixos.



Fonte: Stowell et al. (2011)

O particionamento binário do espaço compõe a etapa de agrupamento do método proposto. Neste trabalho, utiliza-se a estrutura da BSP na criação e organização das unidades de leitura, tirando proveito da característica de convexidade desse tipo de particionamento. A BSP foi escolhida para compor a etapa de agrupamento do método proposto, por permitir flexibilidade no particionamento do espaço e por intrinsecamente satisfazer restrições básicas do problema de agrupamento capacitado.

### 3.3 Algoritmo Genético

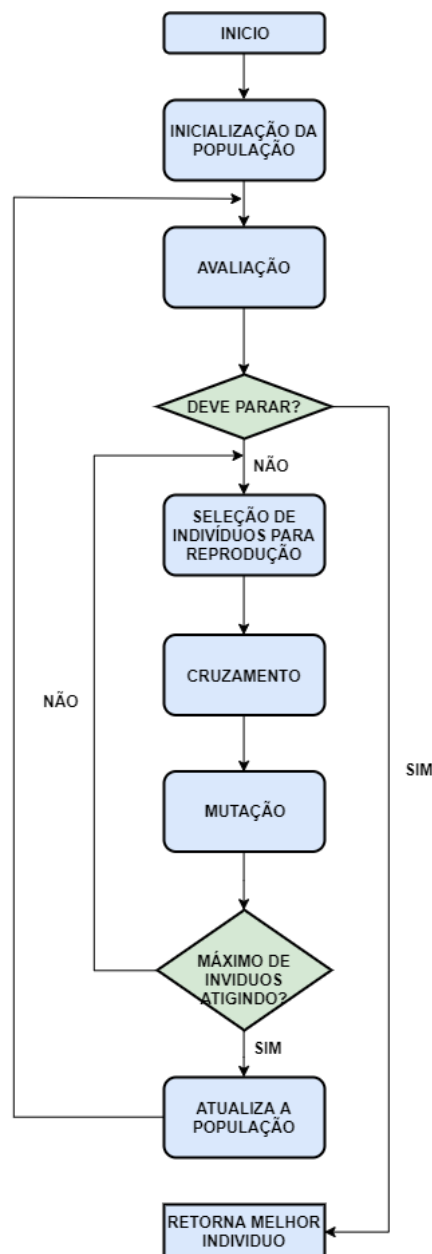
Algoritmos genéticos compõem uma gama de algoritmos, os quais são chamados de bio-inspirados, pois, baseiam-se em conceitos da biologia. Os algoritmos genéticos (AG) são métodos que usam os princípios da seleção natural para a resolução de problemas, em especial problemas de otimização combinatória (HOLLAND, 1992). Um algoritmo genético se caracteriza por ser robusto, genérico e adaptável, características que permitem sua aplicação nas mais diversas áreas. É comum que, ao procurar uma solução otimizada para um determinado problema, seja necessário decidir quantas e quais restrições aplicar para obter a maior quantidade de benefícios. Diversas vezes essas soluções estão ligadas à maximização de ganhos ou minimização de perdas.

O processo de otimização pode ser dividido em duas classes distintas:

- **Otimização Numérica:** encontra a solução a partir de funções matemáticas claras e bem definidas;
- **Otimização Combinatória:** encontra possíveis soluções através da combinação de recursos e suas características.

Optando-se pelo caminho da otimização combinatória, um AG irá consistir na criação de uma população em que cada indivíduo representa uma possível resposta, encontrada após a sua execução. As etapas de funcionamento de um AG podem ser observadas na Figura 4.

Figura 4 – Funcionamento de um Algoritmo Genético



Fonte: Adaptado de Fulcher (2008)

Os **indivíduos** que integram a população codificam possíveis soluções para um problema a ser resolvido e através do processo de "evolução", em que a resposta mais próxima da solução ideal do problema será encontrada. Um indivíduo, também conhecido como **cromossomo** se caracteriza por possuir um **genótipo**, isto é, as informações de uma possível solução, e um **fenótipo**, que é o resultado do processo de decodificação do genoma. Além disso, todo indivíduo possui um **grau de adaptação** comumente chamado de *fitness*, o qual mensura a qualidade da resposta que o indivíduo representa. Este valor é determinado por uma função de avaliação (HOLLAND, 1992).

A execução de um AG segue algumas etapas, as quais podem ser divididas em:

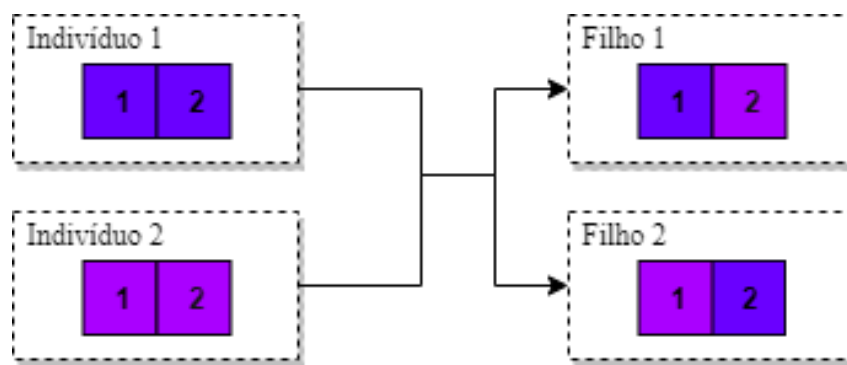
- **Inicialização:** etapa a qual define-se os parâmetros de execução do AG, dentre elas: número de gerações, quantidade de indivíduos, probabilidade de cruzamento, condição de parada, probabilidade de mutação e etc. Ainda nesta etapa é gerada a população inicial de cromossomos, comumente de forma aleatória, porém também é possível inicializar a população com indivíduos conhecidos;
- **Avaliação:** responsável por mensurar a qualidade da resposta que cada cromossomo da população contém e definir o grau de aptidão do indivíduo para a solução do problema em questão;
- **Seleção:** processo de escolha dos indivíduos aptos para a reprodução. Existem diversas técnicas para executar a seleção dentre elas, roleta, torneio, uniforme, etc. A seleção leva em consideração diretamente a aptidão do candidato;
- **Cruzamento:** processo que dará origem a população da próxima geração. A reprodução ocasiona a recombinação de dois indivíduos com a finalidade de originar novos;
- **Mutação:** o indivíduo tem a possibilidade de sofrer uma alteração em seu genótipo, criando assim a possibilidade de variação na população;
- **Atualização:** gera novos indivíduos criando-se assim uma nova geração. Nesta etapa é possível existir o **elitismo**, na qual indivíduos com aptidão muito boa sejam passados para a geração posterior, para que não sejam perdidos e influenciem positivamente no processo de evolução;
- **Finalização:** se dá quando alguma condição de parada pré-estabelecida é alcançada, podendo ser uma quantidade de gerações, execuções, um limiar de aptidão atingido por um indivíduo dentre outras. Enquanto nenhuma condição de parada seja satisfeita, a população é submetida ao processo de evolução caso contrário, o algoritmo retorna o cromossomo que apresenta a melhor solução encontrada;

A técnica de seleção utilizada foi a seleção por "roleta", na qual os integrantes da população são ordenados de acordo com seu *fitness* e lhe são atribuídas probabilidades

decrecentes, considerando à razão entre o *fitness* individual e a soma dos *fitness* da população. Portanto, de forma aleatória a escolha é feita de acordo com as probabilidades atribuídas. Dessa maneira, consegue-se escolher como pais indivíduos bem adaptados, porém sem desprezar características de diversidade dos indivíduos menos aptos, os quais também possuem chance de serem escolhidos.

Com os indivíduos selecionados, é iniciado o processo de cruzamento. Estes indivíduos e têm seus genes intercalados para originar dois novos indivíduos, conforme a Figura 5, nos quais também terão novos fenótipos.

Figura 5 – Processo de cruzamento



Fonte: Próprio Autor

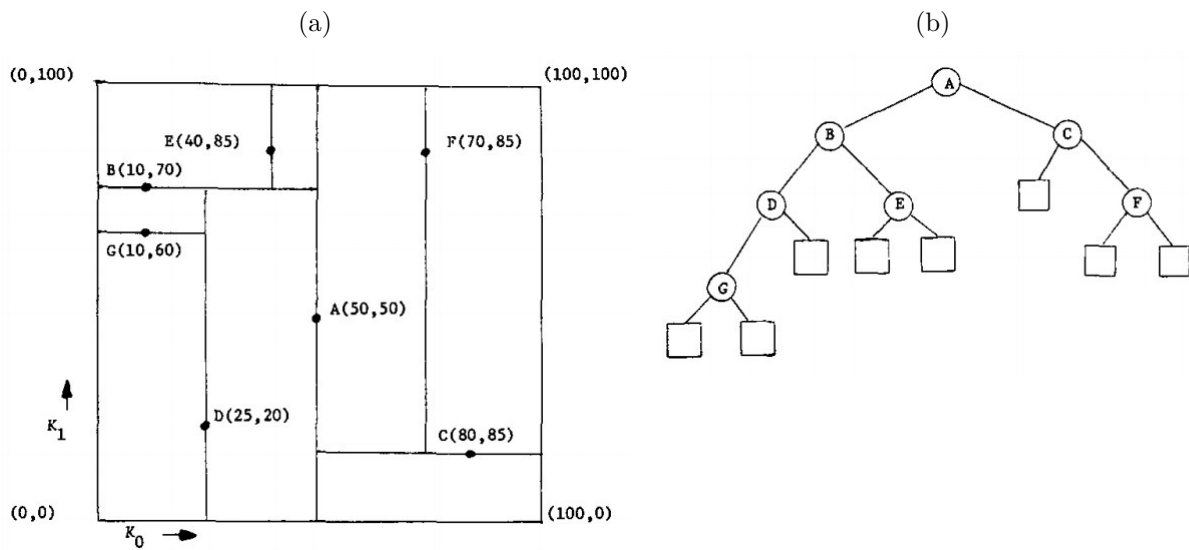
O algoritmo genético foi escolhido por ser flexível, permitindo que a modelagem do problema seja mais clara e simples, além de ser uma técnica robusta e amplamente explorada na literatura. O AG é utilizado em conjunto com a BSP (Seção 3.2) na etapa de agrupamento do método proposto (Seção 4.5). Neste trabalho, é responsável por encontrar a melhor configuração de corte dos polígonos que irão gerar a BSP.

### 3.4 KD-Tree

A *Kd-tree* é uma árvore de pesquisa binária multidimensional especializada em pesquisa associativa, proposta por (BENTLEY, 1975) como solução para o problema da recuperação associativa. Essa estrutura fornece um conjunto de algoritmos eficientes para as operações de inserção, remoção e busca.

Para diversos tipos de aplicações, dada uma função de distância  $D$ , uma coleção de pontos  $B$ , um espaço com  $K$  dimensões e um ponto  $P$  também neste espaço, é desejado encontrar o vizinho mais próximo de  $P$  em  $B$ . Essa consulta é um caso clássico de busca associativa, chamada de vizinho mais próximo, em que a árvore  $k$ - $d$  resolve em tempo logarítmico. Um exemplo *KD-Tree* é exposto na Figura 6. A esquerda, é mostrado os pontos no plano junto com a parcela do espaço que cada ponto representa e a direita é exibida a representação da árvore em forma de um grafo.

Figura 6 – Representação gráfica de uma KD-Tree: (a) representação no plano (b) representação em árvore



Fonte: (BENTLEY, 1975)

Por ser uma estrutura de dados que permite encontrar os vizinhos mais próximos de maneira eficiente, a árvore k-d é utilizada na fase de pré-processamento do método proposto, para localizar a região mais densa de uma cidade e posteriormente permitir a remoção de ruídos, caso seja necessário.

### 3.5 Z-Score

O *z-score* ou *score* padrão, pode ser definido como a quantidade de desvios-padrão entre um elemento do conjunto e a média do conjunto de dados (ABDI; WILLIAMS, 2010). Na prática, o *score* padrão pode ser positivo ou negativo, podendo ser colocado em uma curva de distribuição normal. Caso o *score* z seja negativo, estará localizado a extrema esquerda da curva, se for positivo estará localizado a extrema direita da curva de distribuição. Desta forma, se um ponto do conjunto de dados extrapola esses limites, o mesmo pode ser considerado como um *outlier*.

No método proposto, utiliza-se o *z-score* para verificar a distância de uma instalação para a região densa de uma cidade e avaliar se a instalação pode ser considerada como *outlier*, permitindo dessa forma a identificação e remoção de tais instalações.

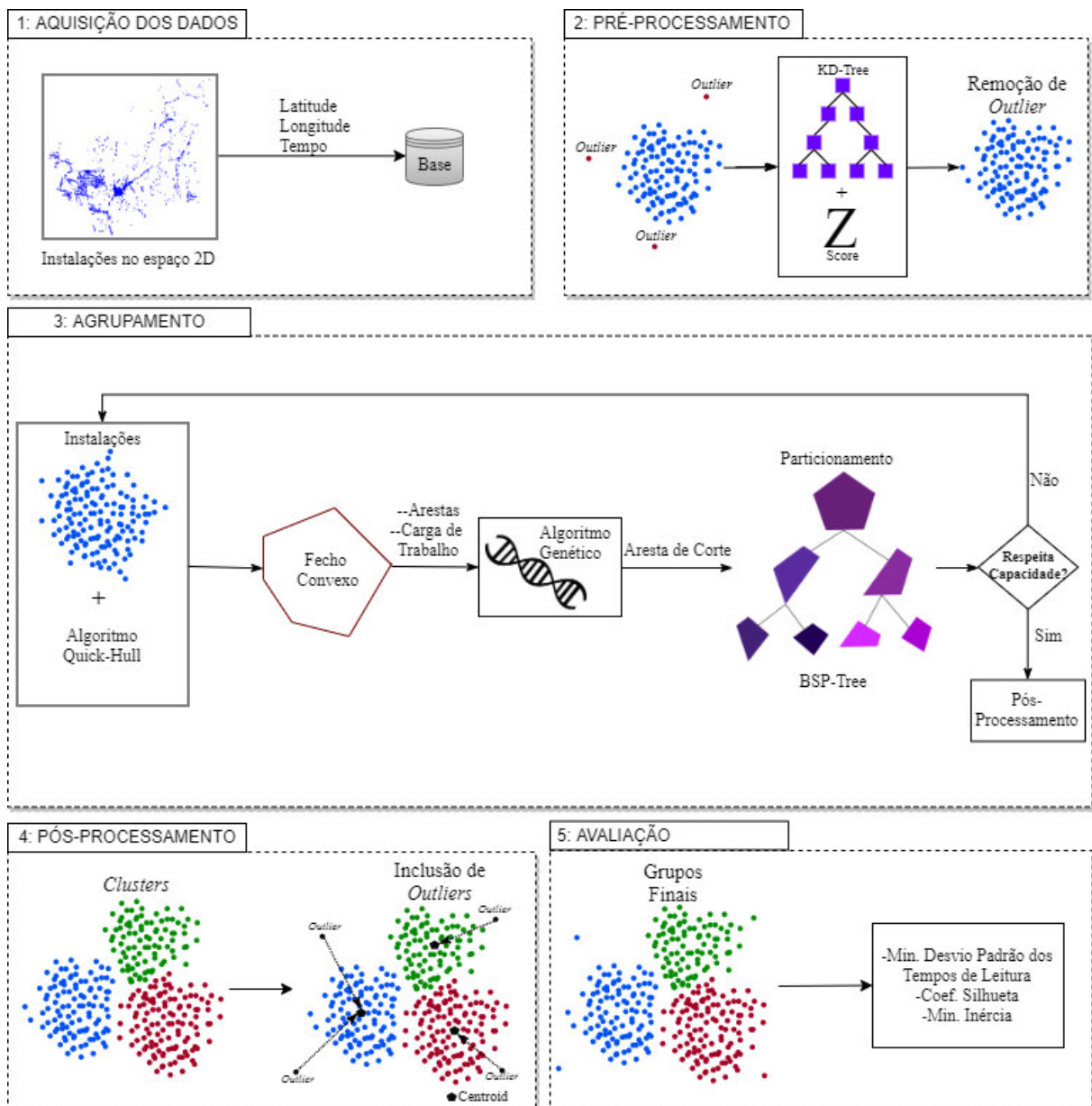
## 4 Metodologia

Este Capítulo descreve o método proposto para a criação das unidades de leitura utilizando particionamento binário do espaço, baseado em algoritmo genético, que atendam aos critérios de homogeneidade e compacidade. Ainda neste capítulo, será descrito a aquisição da base de dados e o tratamento realizado sobre os dados.

### 4.1 Visão Geral

O método proposto está organizado em 5 estágios. A Figura 7 apresenta os estágios de execução do método proposto. O primeiro estágio refere-se à aquisição, análise e a forma de interpretação dos dados pelo método. No segundo, realiza-se o pré-processamento dos dados, utilizando-se as técnicas *KD-Tree* e *Z-Score*. O terceiro é a etapa central, responsável pelo agrupamento das instalações e geração das ULs através do algoritmo genético e particionamento do espaço. No quarto estágio é realizado o pós-processamento, incluindo os *outliers* nos grupos formados. Por último, é realizada a avaliação dos grupos na etapa de pós-processamento.

Figura 7 – Visão geral do método proposto



Fonte: Próprio Autor



## 4.2 Aquisição dos dados

A base de dados usada neste trabalho foi fornecida pela Equatorial Energia para o projeto de pesquisa e desenvolvimento, P&D PD-00371-0031/2017. Os dados disponibilizados contam com informações referentes as unidades de consumo, que no âmbito deste trabalho, também são denominadas de instalações. Essas instalações pertencem as cidades de Açaílândia e Imperatriz, do estado do Maranhão, e Abaetetuba e Santarém, do estado do Pará. As cidades de Açaílândia e Abaetetuba foram escolhidas por serem cidades que devido a suas quantidades de unidades consumidoras podem ser classificadas como pequeno porte. Imperatriz e Santarém, foram escolhidas por serem duas das maiores cidades presentes na base de dados.

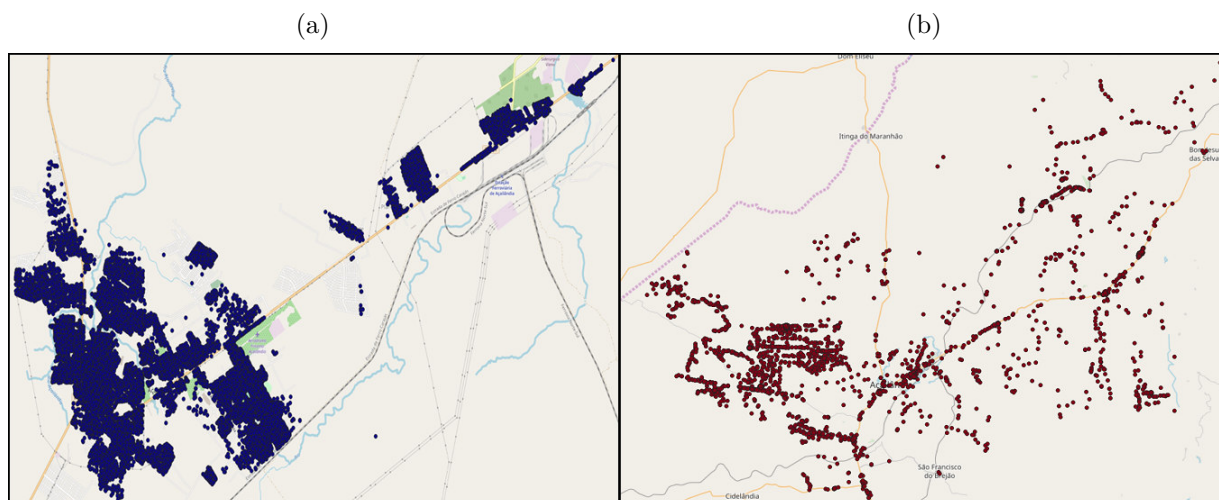
Para cada instalação, são fornecidos os campos referentes a: cidade de origem; latitude; longitude; etapa de Leitura, que representa o dia em que a instalação deve ser lida, dado o planejamento de leitura; o tempo de leitura da instalação, o qual representa o tempo médio gasto pelo funcionário leiturista para aferir o consumo registrado; e o tipo da unidade consumidora, U caso seja urbana e R caso seja Rural.

## 4.3 Caracterização do cenário de aplicação

Para dar início ao método, primeiramente é necessário analisar os dados fornecidos. De início, foi observado que para cada cidade existe uma divisão no conjunto de instalações em duas áreas, uma urbana e outra rural, em que cada área é agrupada de forma independente. Através da visualização das instalações em um mapa, observa-se que a região urbana costuma ser bastante densa, ou seja, possui mais instalações por km<sup>2</sup>. Por sua vez, as instalações classificadas como rurais são mais distantes uma das outras, conforme pode ser observado na Figura 8.

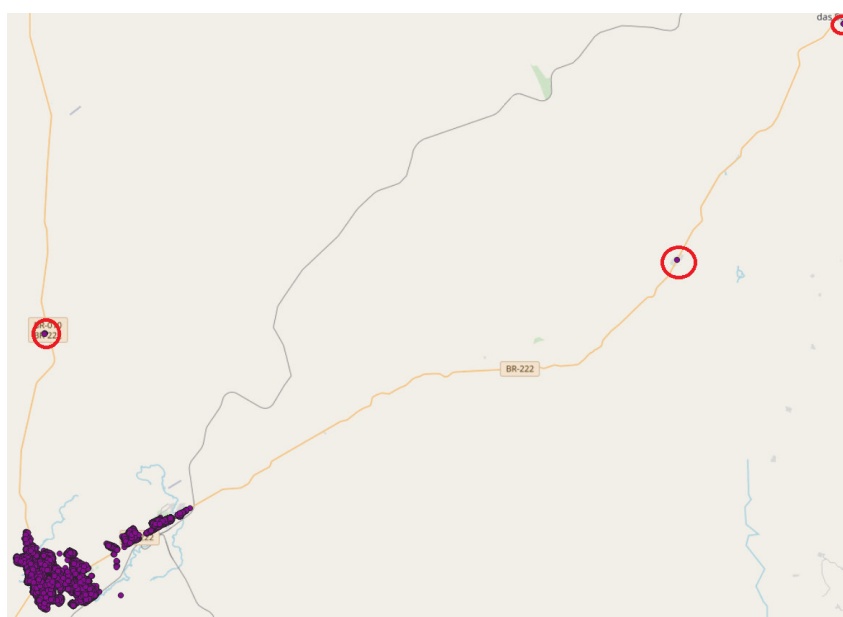
Através da visualização dos dados como na Figura 9, também é possível observar instalações muito distantes das demais, algumas vezes estão localizadas fora dos limites da cidade. Essa discrepância se torna mais perceptível ao se observar a área urbana, que como mostrado anteriormente, é mais densa em comparação com a área rural.

Figura 8 – Comparação entre as áreas urbana e rural da cidade de Açailândia: (a) área urbana (b) área rural



Fonte: Próprio Autor

Figura 9 – Instalações distantes do centro urbano



Fonte: Próprio Autor

As instalações distantes e isoladas podem ser consideradas como *outliers*, ou seja, ruídos na base de dados. Pois, podem ser amostras de dados com algum tipo de erro, podendo tratar-se de instalações classificadas com tipos errados ou até mesmo instalações de cidades vizinhas que foram atribuídas a um planejamento de leitura de forma errônea. Para a identificação de remoção dessas instalações, foi desenvolvida uma etapa de pré-processamento, que será explicada na próxima seção.

## 4.4 Pré-Processamento

Após a análise dos dados fornecidos, realizada na Seção 4.3, foi constatada a presença de ruídos na base de dados. Como a correção de dados não faz parte do escopo deste trabalho, fez-se necessário a construção de uma etapa que realize um pré-processamento sobre o conjunto de dados para identificar e separar as instalações ruidosas. Para a execução da etapa de agrupamento do método proposto, que será discutida mais adiante, deseja-se como entrada um conjunto de instalações sem a presença de *outliers*, ou seja, um conjunto de dados mais compacto.

Portanto, a etapa de pré-processamento utilizada pode ser dividida em duas etapas: na primeira, encontra-se o conjunto de instalações que compõem a região mais densa da cidade e a segunda será responsável pela identificação e remoção das instalações ruidosas.

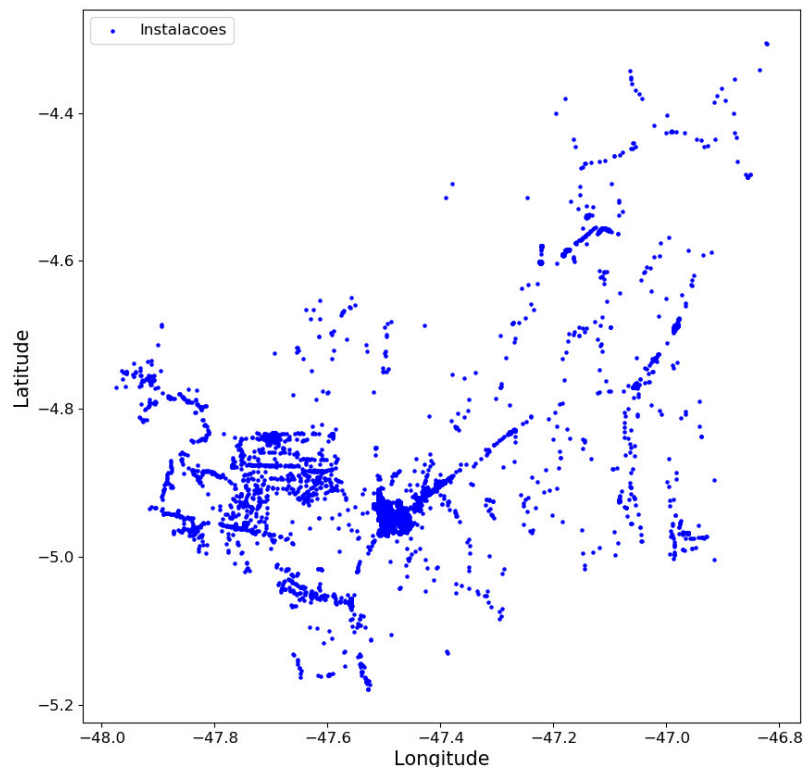
A região densa é encontrada utilizando uma *KD-Tree*, a qual possui mecanismos de busca otimizados e especializados para exploração do espaço. O conjunto de instalações é representado em um espaço bidimensional, no qual as posições são definidas pelas coordenadas de longitude e latitude como exemplificado na Figura 10. Por fim, para encontrar a região densa são executados os seguintes passos: as instalações são divididas em urbanas e rurais e inseridas na árvore *k-d* referente a seu tipo, logo cada instalação constitui um nó desta árvore; logo após, são pesquisados os nós em que a distância euclidiana é menor ou igual a 100 metros; para cada grupo retornado, confere-se a quantidade de membros vizinhos; e por fim, o grupo com mais indivíduos é considerado a região mais densa, conforme ilustrado na Figura 11.

A segunda etapa do pré-processamento identifica e remove os *outliers* através da aplicação da técnica de normalização *z-score* (ABDI; WILLIAMS, 2010). Dado o conjunto de instalações que compõem a região densa encontrada na etapa anterior, o processamento segue os seguintes passos: calcula-se o centróide da região densa; calcula-se a distância euclidiana de todos os pontos para o centróide; computa-se a média e o desvio padrão das distâncias; cada ponto é submetido a Equação 4.1, em que  $D(i)$  representa a distância do ponto  $i$  para o centróide e  $ft$  é o desvio padrão máximo permitido. Caso a Equação seja satisfeita, a instalação é considerada um ponto de ruído.

$$\frac{D(i) - Média}{DesvioPadrão} > ft \quad (4.1)$$

Por fim, as instalações ruidosas são identificadas e removidas de forma automática. Um exemplo da aplicação desta etapa pode ser observada na Figura 12, em as marcações em vermelho são as instalações consideradas como *outliers* e as marcações em verde mostram a ausência dessas instalações após a execução da etapa de pré-processamento.

Figura 10 – Instalações no espaço 2D



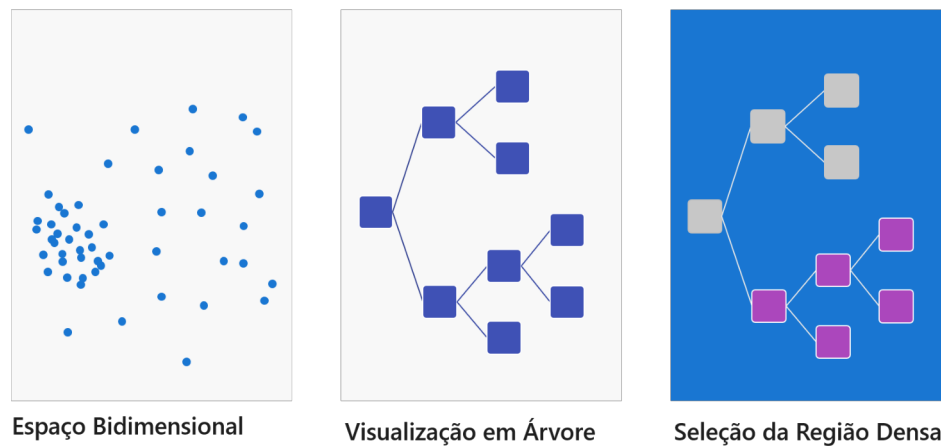
Fonte: Próprio Autor

## 4.5 Agrupamento

Esta é responsável pela geração capacitada das unidades de leitura. O processo de geração das ULs é iniciado recebendo como entrada o conjunto de instalações após a remoção dos *outliers*, realizada na etapa anterior (Seção 4.4). Neste ponto as instalações são divididas em grupos, cada grupo representa o dia de leitura das instalações informados pela companhia, denominados etapas de leitura. Para cada etapa, são geradas as ULs necessárias para a execução do processo de leitura do consumo, referente a este dia de trabalho.

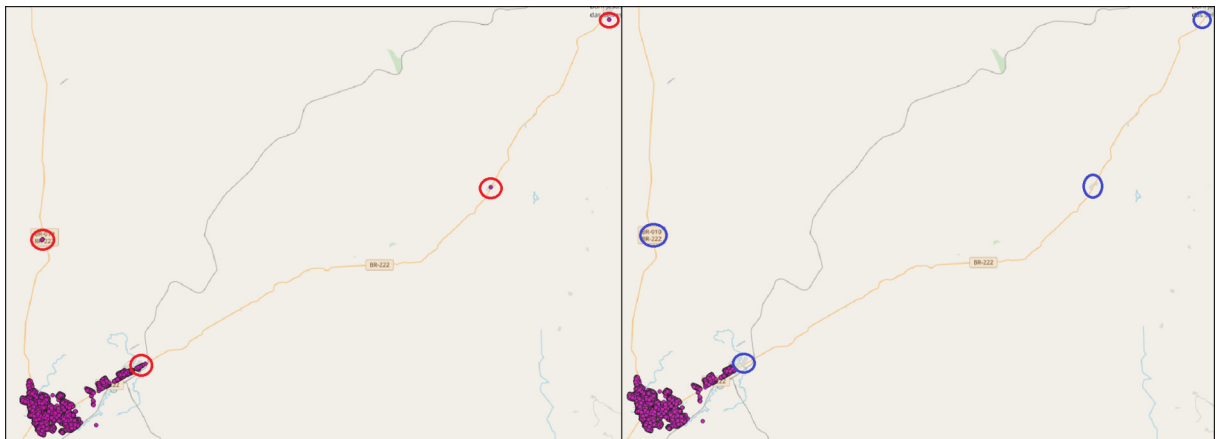
Contudo, uma das principais fases do processo de agrupamento é o particionamento do espaço 2D e a construção da árvore BSP, como definido anteriormente. A árvore é construída a partir do particionamento de polígonos convexos. Portanto, define-se o fecho convexo das instalações para cada etapa de leitura, utilizando o algoritmo *quick-hull* (BARBER et al., 1996). O polígono resultante, é o menor polígono convexo que contém todas as instalações da etapa e é utilizado na fase de execução do algoritmo genético, que será discutida mais adiante.

Figura 11 – Execução do processo de localização da região densa.



Fonte: Fonseca (2019)

Figura 12 – Antes e depois da aplicação do pré-processamento na cidade de Açailândia-MA



Fonte: Próprio Autor

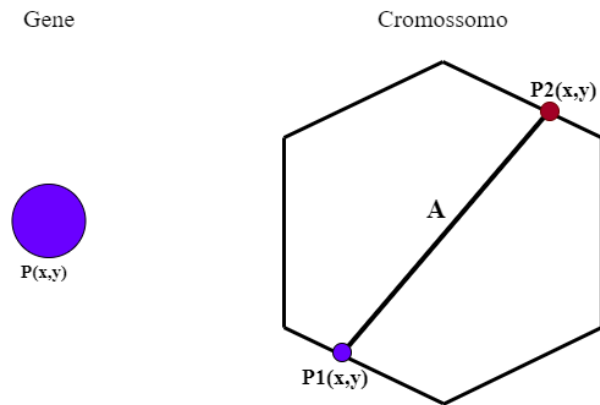
#### 4.5.1 Execução do Algoritmo Genético

Nesta etapa do método, as instalações estão separadas em etapas de leitura e com informações referentes as coordenadas geográficas, tempo de leitura para cada instalação e um polígono convexo que representa cada etapa. Com essas definições, é possível iniciar a preparação necessária para executar o algoritmo genético. O último estágio da preparação do AG corresponde à definição da estrutura de seus indivíduos, a população, quantidade de gerações, e suas funções de avaliação, seleção e mutação para o contexto do problema.

Para definir a estrutura de cada indivíduo (cromossomo) da população do AG, é necessário iniciar pela descrição do seus elementos básicos, os genes. Cada gene do cromossomo representa um ponto  $P(x, y)$  no espaço e está situado sobre uma aresta do

fecho convexo da etapa de leitura. Cada cromossomo é constituído por dois genes, os quais devem ser de arestas distintas, e a relação entre o gene 1 e o gene 2 constitui uma aresta  $A(p1,p2)$  conforme a Figura 13.

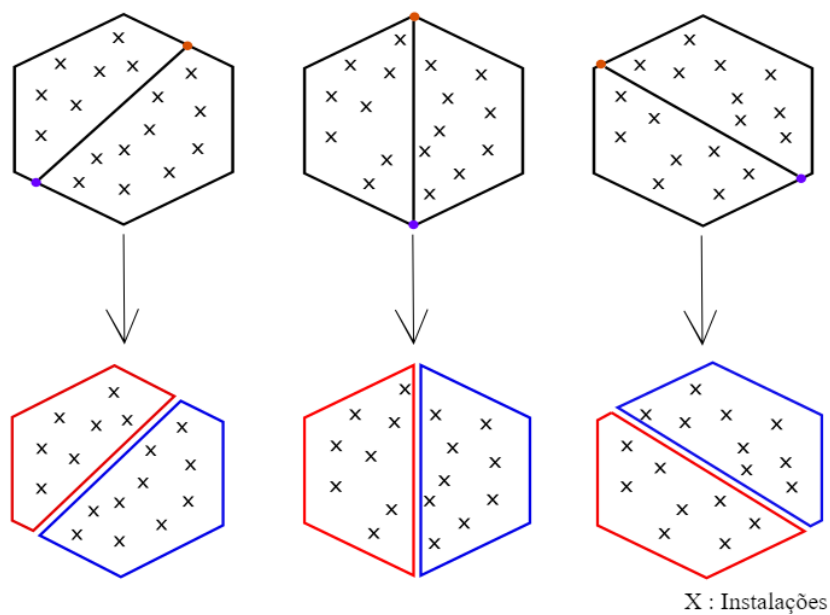
Figura 13 – Representações gráficas de um gene e um cromossomo



Fonte: Próprio Autor

A população é formada por um conjunto de 100 indivíduos, os quais representam arestas distintas. Cada aresta da população cria uma separação entre as instalações que compõem a etapa de leitura. Esse processo fica mais claro ao observar a Figura 14. A população será submetida aos processos de avaliação, e as operações genéticas de cruzamento e mutação durante as gerações.

Figura 14 – Variedade de aresta da população



Fonte: Próprio Autor

Durante cada geração da execução do AG, os indivíduos da população são submetidos a uma função de avaliação, ou função objetivo, para medir seu *fitness* como solução. A função objetivo, no contexto desse trabalho, é determinada pela combinação de dois fatores. O primeiro fator considera o critério da soma dos quadrados dentro do grupo, também conhecido como inércia (OLIVEIRA, 2011). Quando o polígono do fecho convexo das instalações é particionado pela aresta que está codificada em um indivíduo, conforme a Figura 14, esse polígono dá origem a dois novos, os quais são chamados de A e B. Cada polígono contém uma parcela das instalações do polígono original. A inércia de um agrupamento é calculada conforme a Equação 4.2.

$$\sum_{i=1}^n \text{Min}_{\mu_j \in C} (||x_i - \mu_j||^2) \quad (4.2)$$

onde  $X_i$  representa cada ponto que foi agrupado e  $\mu_j$  é um centroide pertencente a  $C = \{1, 2, \dots, n\}$ , em que  $n$  é o número total de centroides do agrupamento obtido.

O segundo fator é o módulo da diferença das cargas de trabalho dos polígonos A e B, resultantes do particionamento. Esta carga de trabalho é obtida quando se realiza a soma dos tempos das instalações contidas nos polígonos, podendo ser calculada como descrito na Equação 4.3, em que  $C_a$  é a carga de trabalho total do polígono A e  $C_b$  é a carga de trabalho do polígono B.

$$||Dif(C_a - C_b)|| \quad (4.3)$$

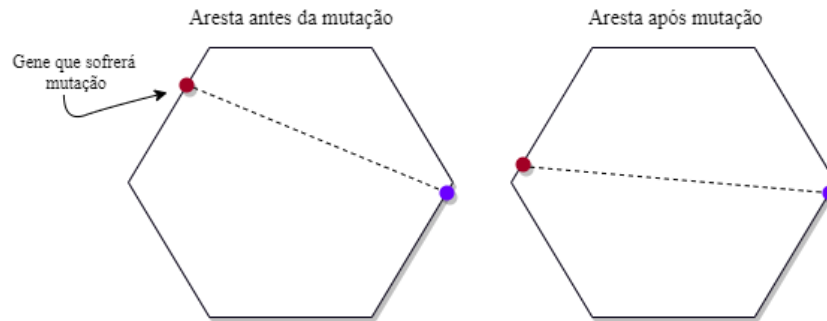
Espera-se obter a partir da aplicação da Equação 4.2 grupos coesos, ao mesmo tempo em que se minimize o desvio padrão dos tempos de leitura dos grupos, através da Equação 4.3. Esses dois fatores são ponderados através dos coeficientes  $\alpha$  e  $\beta$ . Logo, a função objetivo é descrita pela Equação 4.4.

$$Fitness = \sum_{i=1}^n \text{Min}_{\mu_j \in C} (||x_i - \mu_j||^2) * \alpha + ||Dif(C_a - C_b)|| * \beta \quad (4.4)$$

Após à avaliação da população, realiza-se a seleção dos indivíduos para o cruzamento – conforme explicado na Seção 4.5.1 – utilizando a seleção por roleta e intercalação dos genes do cromossomo. Em seguida, existe uma probabilidade de mutação para cada indivíduo. A mutação é a alteração dos genes de um elemento. Para o método proposto foi utilizado uma probabilidade de 10% de ocorrer mutação. Caso um indivíduo sofra mutação, é sorteado um de seus dois genes e o gene sorteado será alterado. O gene, como explicado no começo desta seção, é um ponto pertencente a uma aresta do fecho convexo do conjunto de instalações. A mutação consiste em alterar o posicionamento deste ponto sobre a aresta,

como apresentado na Figura 15. Os indivíduos mutantes são inseridos de volta na nova população e constituirão, portanto, uma nova geração.

Figura 15 – Processo de mutação



Fonte: Próprio Autor

Outra técnica empregada é o elitismo, na qual o indivíduo mais apto da geração anterior é passado diretamente para a nova população, para que a melhor solução até o momento não seja perdida. Neste trabalho, a condição de parada do algoritmo genético foi definida pela quantidade de gerações, configurada em 100. A melhor aresta, isto é, o melhor indivíduo da população da última geração é retornado como solução. Esta aresta proporcionará um particionamento das instalações de forma a gerar grupos coesos e com desvio padrão dos tempos minimizado.

#### 4.5.2 Particionamento do espaço

Conforme explicado na Seção 3.2, o particionamento binário do espaço é realizado a partir da construção de uma estrutura de dados do tipo árvore, conhecida como BSP. Também foi mostrado naquela seção que o espaço 2D pode ser particionado através dos eixos das abscissas ou das ordenadas, considerando-se o plano cartesiano. Neste trabalho, o espaço 2D é limitado pela área do fecho convexo da etapa de leitura e a reta de particionamento é obtida através da execução do AG descrito na Seção 4.5.1, com a finalidade de permitir maior flexibilidade no particionamento.

Na construção da árvore sempre que se particiona um polígono convexo, o resultado são dois polígonos também convexos e não sobrepostos. Se for considerado que um polígono representa um conjunto de dados, e ao particioná-lo originam-se dois novos grupos, os grupos resultantes não apresentarão sobreposição entre si, esta característica vai ao encontro de uma das principais restrições de um agrupamento capacitado, conforme explicado na Seção 3.1.

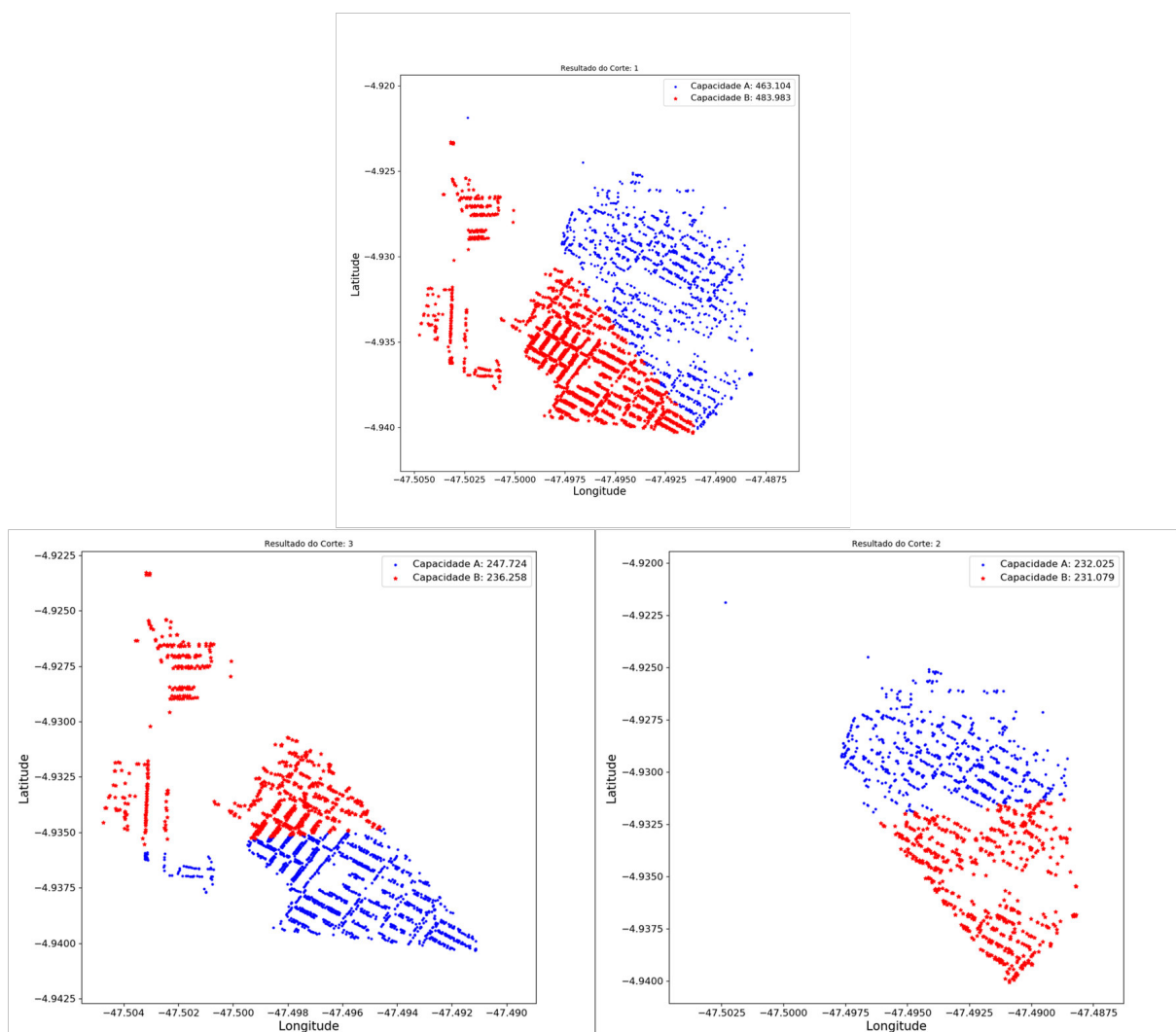
Ao se iniciar a fase de agrupamento do método proposto, o fecho convexo da etapa de leitura será a raiz da árvore BSP. Após a execução do AG, a melhor aresta de corte encontrada durante a execução é recebida e as instalações da etapa de leitura são divididas



em dois grupos, pelo particionamento do polígono. Para cada grupo é calculada sua nova carga de trabalho, caso a carga de trabalho ultrapasse o limite estipulado, este conjunto de instalações é reintroduzido no fluxo de agrupamento e dará origem a um processo recursivo, como ilustrado na Figura 7, caso contrário o grupo é mantido.

No momento em que todas as recursões do processo de agrupamento terminarem, a BSP estará completa. Sua raiz principal representa a etapa de leitura e suas folhas representam as unidades de leitura, as quais não ultrapassam o limite estipulado e não apresentam sobreposição entre si. Um exemplo pode ser observado na Figura 16, em que a etapa 1 da cidade de Açailândia foi particionada inicialmente em dois grupos, os quais possuem capacidades superiores ao limite estipulado, portanto, cada grupo é dividido novamente. Logo abaixo, é mostrado que os dois grupos iniciais deram origem a 4 novos grupos, que por sua vez não ultrapassam a capacidade limite, portanto, são mantidos.

Figura 16 – Particionamento de uma etapa



Fonte: Próprio Autor

## 4.6 Pós-processamento e Avaliação

Após o término do processo de agrupamento, todas as ULs são submetidas a etapa de pós-processamento, juntamente com as instalações que foram classificadas como *outliers*. As instalações que são consideradas como ruídos, por apresentarem algum tipo de erro e consequentemente poderem comprometer o método, devem ser lidas e faturadas, sendo assim, não podem ser excluídas do planejamento de leitura gerado. Portanto, o processo de atribuição das instalações ruidosas em uma UL é realizado da seguinte forma: calcula-se o centro de cada UL; em seguida, calcula-se a distância euclidiana de cada instalação para o centro da UL, por último, aloca-se a instalação a UL de centro mais próxima.

Quando não existem instalações desalocadas, os grupos são enviados para etapa final do método proposto, responsável por avaliar a qualidade dos grupos formados. A avaliação é feita através da minimização do desvio padrão (Seção 4.6.1) e do coeficiente de silhueta (Seção 4.6.2).

### 4.6.1 Minimização do desvio padrão

A minimização do desvio padrão dos tempos de leitura é utilizado para avaliar o critério de homogeneidade dos grupos formados. Ao se analisar o desvio padrão dos tempos de leitura, é possível verificar se as ULs estão balanceadas, pois, quanto maior o desvio padrão, maior é a diferença dos tempos de leitura entre os grupos e quanto menor o desvio padrão, mais homogêneas serão as ULs. Indicando na prática, que a carga de trabalho da etapa está sendo dividida de forma uniforme entre os funcionários responsáveis pela realização das leituras.

Para cada UL gerada, é realizado o somatório do tempo de leitura de cada instalação que a compõem. Dessa forma é calculada sua carga de trabalho total. Com o tempo total de todas as ULs é possível calcular o desvio padrão, utilizando a Equação 4.5.

$$Desvio = \sqrt{\frac{\sum_{i=1}^n (Tempos_i - M_a)^2}{n}} \quad (4.5)$$

onde  $n$  é o número total de ULs do agrupamento,  $Tempos_i$  é o  $i$ -ésimo tempo de uma UL e  $M_a$  é a média aritmética dos tempos.

Ao aplicar a Equação 4.5 sobre um agrupamento, é possível verificar a variabilidade dos tempos entre as ULs. Um valor próximo de 0 corresponderá a um desvio menor entre os tempos, portanto, significa que foram geradas ULs mais homogêneas.

### 4.6.2 Coeficiente de Silhueta

Para avaliar o critério de compacidade dos agrupamentos é usado o Coeficiente de Silhueta. Essa medida é utilizada para avaliar cada grupo individualmente.

Para calcular o coeficiente de silhueta de um grupo, grupo, no contexto desse trabalho, seleciona-se uma instalação aleatória  $i$ . Calcula-se a distância média dessa instalação para todas as outras instalações do grupo, definida como  $a_i$ ; calcula-se também a distância média das instalações para as instalações do grupo mais próximo, definida como  $b_i$ . Assim o coeficiente de silhueta é obtido pela Equação 4.6.

$$C_S = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (4.6)$$

onde a função  $Max(a, b)$  retorna o maior valor entre  $a$  e  $b$ .

Essa métrica retorna valores que variam entre -1 e 1. Se o coeficiente tiver um valor negativo, significa que o agrupamento é esparso. Quanto mais próximo de 1 for o coeficiente, mais compacto é um grupo. É considerado como um bom agrupamento caso todos os grupos tenham um coeficiente de silhueta positivo.

## 5 Resultados e Discussão

Com o objetivo de validar o método proposto, foram realizados testes utilizando base de dados e planejamentos de leitura – fornecidos pela Equatorial energia. Esses planejamentos são referentes a quatro cidades: Açailândia e Imperatriz, do Maranhão; e Santarém e Abaetetuba, do Pará. Para a visualização dos grupos de instalações, foi utilizado o *software Qgis* (QGIS Development Team, 2019), que é especializado em manipulação e visualização de dados geográficos.

Durante a execução dos testes, a capacidade máxima permitida para um grupo foi determinada em 360 minutos, aproximadamente 6 horas de trabalho, e uma tolerância de 10%, portanto, a capacidade total permitida para uma UL é de 396 minutos. Para o algoritmo genético, empregado na fase de agrupamento, foram utilizados os seguintes parâmetros:

- População composta por 100 indivíduos;
- Probabilidade de *crossover* estabelecida em 80%;
- Probabilidade de mutação estabelecida em 10%;
- Numero de gerações definido em 100 gerações;
- Utilizado 1 indivíduo elite.

Ainda no algoritmo genético, em sua função objetivo, foram utilizados os valores de 0,8 e 0,2 para os coeficientes  $\alpha$  e  $\beta$  respectivamente. Esses coeficientes ponderam minimização da inércia e a diferença de carga de trabalho dos grupos.

Os experimentos foram realizados seguindo algumas etapas. A primeira etapa, consistiu em separar o planejamento de cada cidade de acordo com seu tipo - planejamento urbano e planejamento rural. Na segunda, os planejamentos vigentes da cidade, foram processados para a obtenção das métricas, as quais são: a quantidade de ULs; coeficiente de silhueta; e desvio padrão dos tempos de leitura. Na terceira etapa, os parâmetros de capacidade máxima dos grupos e os parâmetros referentes ao algoritmo genético foram definidos. Chegando à quarta etapa, as instalações foram divididas em urbanas e rurais e submetidas ao método proposto. Na quinta, os agrupamentos gerados foram avaliados para a obtenção das mesmas métricas utilizadas nos planejamentos vigentes. Por fim, a última etapa consistiu em realizar uma análise comparativa entre o método proposto e a solução vigente adotada pela companhia.

### 5.0.1 Planejamento para a cidade de Açailândia

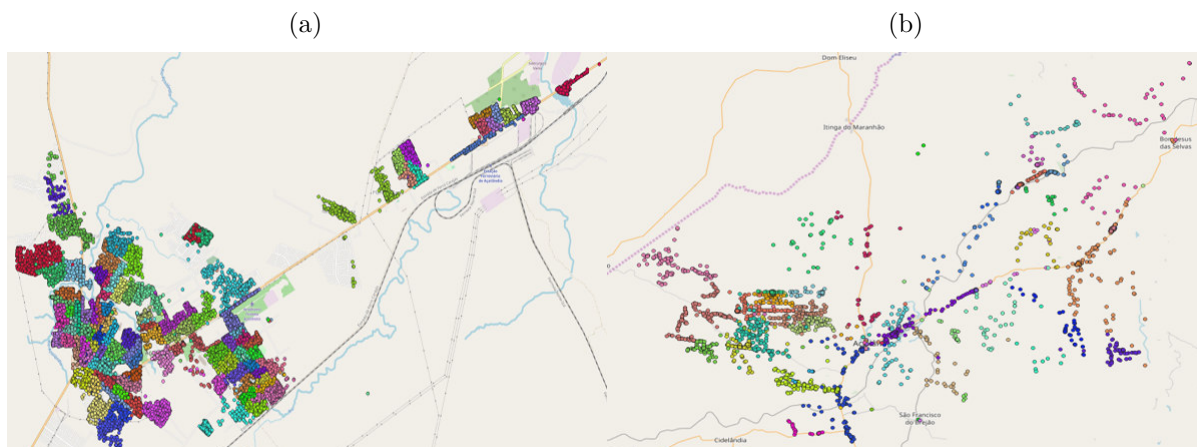
O planejamento vigente da cidade de Açailândia, que pode ser observado na Figura 17, foi analisado conforme as métricas de avaliação propostas: o desvio padrão dos tempos de leitura e o coeficiente de silhueta. Além desses dois valores também foi obtida a quantidade de unidades de leitura.

Planejamento da área urbana possui 118 ULs, um desvio padrão de 75,76 minutos e um coeficiente de silhueta de 0,2462. Se comparado com o planejamento urbano gerado pelo método proposto, pode-se observar que a quantidade de ULs diminuiu para 41, o desvio padrão dos tempos de leitura foi reduzido para 53,21 minutos, por fim, o coeficiente de silhueta foi melhorado, alcançando 0,3888 pontos conforme exposto na Tabela 1. Ainda nessa tabela, é possível observar o melhoramento do planejamento rural da cidade, quando se compara o planejamento vigente com o gerado pelo método. A quantidade de ULs rurais diminuiu de 55 para 35. Também é importante visualizar, que o desvio padrão dos tempos diminuiu para 82,33 minutos, aproximadamente a metade do original, indicando um melhor balanceamento da carga de trabalho. Por último, o coeficiente de silhueta das ULs rurais do planejamento vigente, que originalmente era de 0,2659, foi melhorado, adquirindo o valor de 0,4901, indicando ULs mais compactas e não sobrepostas.

	Planejamento Vigente		Planejamento Proposto	
	Urbano	Rural	Urbano	Rural
Quantidade de ULs	118	55	<b>41</b>	<b>35</b>
Desvio Padrão	75,76	161,48	<b>53,21</b>	<b>82,33</b>
Coef. Silhueta	0,2462	0,3888	<b>0,2659</b>	<b>0,4901</b>

Tabela 1 – Comparação de métricas para cidade de Açailândia-MA

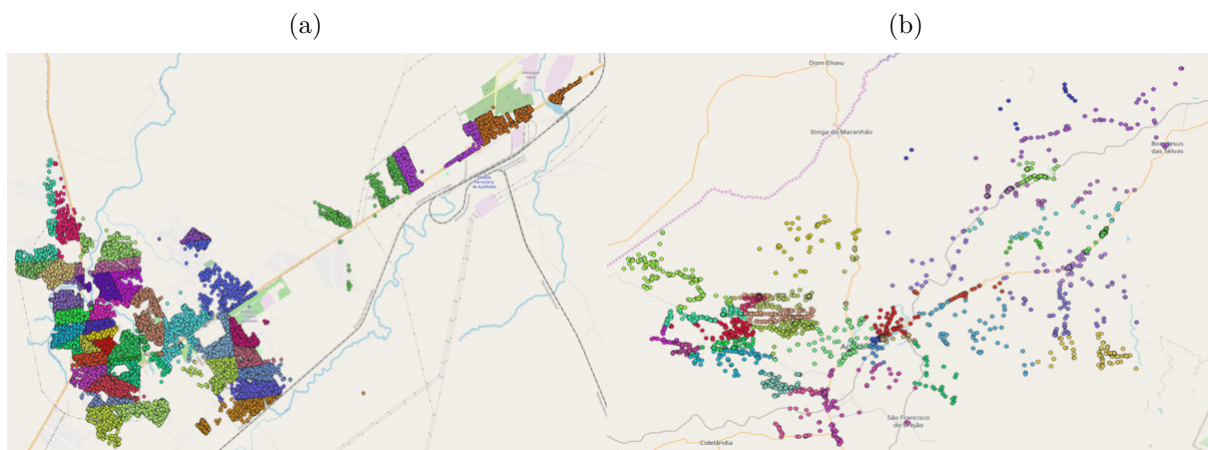
Figura 17 – Planejamentos vigentes da cidade de Açailândia: (a) agrupamento urbano (b) agrupamento rural



Fonte: Próprio Autor

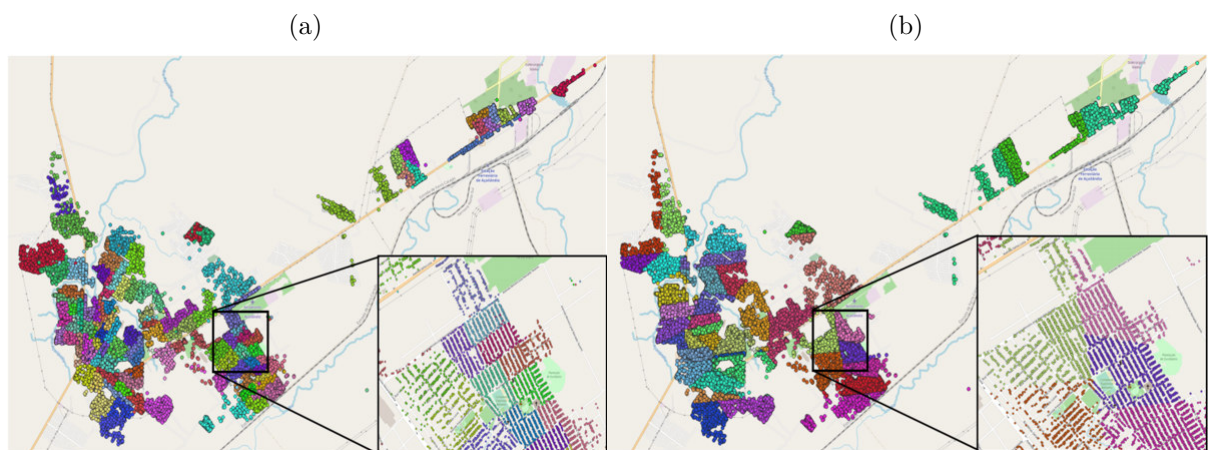
O planejamento obtido pelo método proposto é apresentado na Figura 18. Com a visualização dos grupos gerados em um mapa, observa-se que as ULs não se sobrepõem, situação diferente do agrupamento vigente. Um exemplo em que a sobreposição das ULs foi corrigida é ilustrada na Figura 19, na qual uma mesma região geográfica é colocada em foco. Ao comparar a situação das ULs nesse mesmo local, para ambos os planejamentos. Essa característica é garantida pelo particionamento dos polígonos convexos e uso da BSP.

Figura 18 – Planejamentos propostos para a cidade de Açailândia: (a) agrupamento urbano (b) agrupamento rural



Fonte: Próprio Autor

Figura 19 – Comparação entre ULS urbanas dos planejamentos de Açailândia-MA: (a) ULS vigentes (b) ULS propostas



Fonte: Próprio Autor

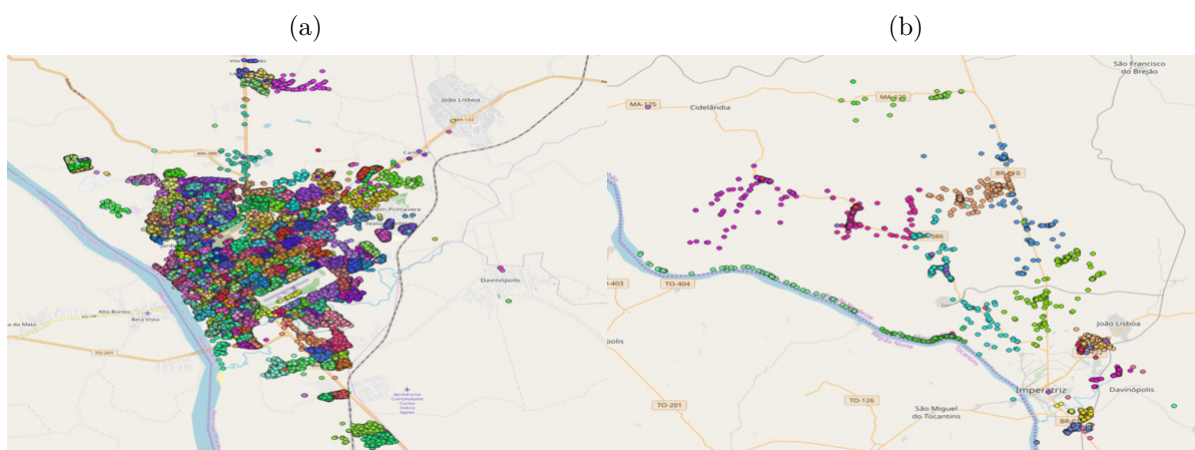
### 5.0.2 Planejamento para a cidade de Imperatriz

A segunda cidade na qual o método proposto foi executado, é a cidade de Imperatriz, situada no estado do Maranhão. Os dados do planejamento vigente e o resultados do método são comparados na Tabela 2. A região urbana da cidade é composta por 347 ULs, com suas cargas de trabalho apresentando um desvio padrão igual a 77,23 minutos e um coeficiente de silhueta de 0,2590. Em Imperatriz também existe um planejamento rural, constituído por 34 ULs, as quais possuem um desvio padrão de 159,25 minutos e o coeficiente de silhueta equivale a -0,0094. Os dois planejamentos da cidade são exibidos na Figura 20. Nessa figura é possível observar a esquerda o agrupamento vigente da área urbana e ao lado o agrupamento da área rural.

	Planejamento Vigente		Planejamento Proposto	
	Urbano	Rural	Urbano	Rural
<b>Quantidade de ULs</b>	347	34	<b>132</b>	<b>16</b>
<b>Desvio Padrão</b>	77,23	159,25	<b>69,48</b>	<b>26,51</b>
<b>Coef. Silhueta</b>	0,2590	-0,0094	<b>0,2602</b>	<b>0,5141</b>

Tabela 2 – Comparação de métricas para cidade de Imperatriz-MA

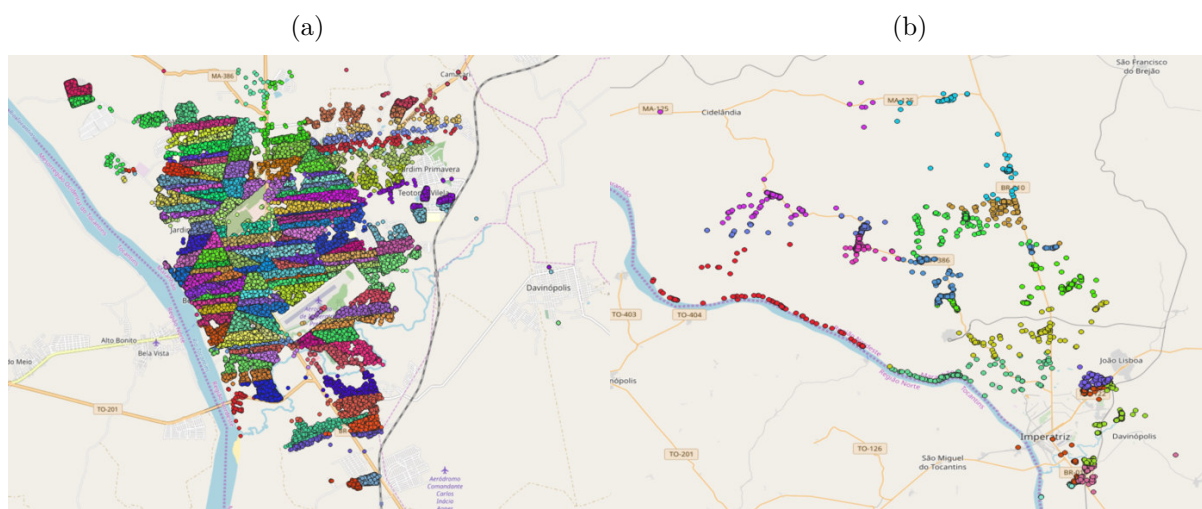
Figura 20 – Planejamentos vigentes da cidade de Imperatriz-MA: (a) agrupamento urbano (b) agrupamento rural



Fonte: Próprio Autor

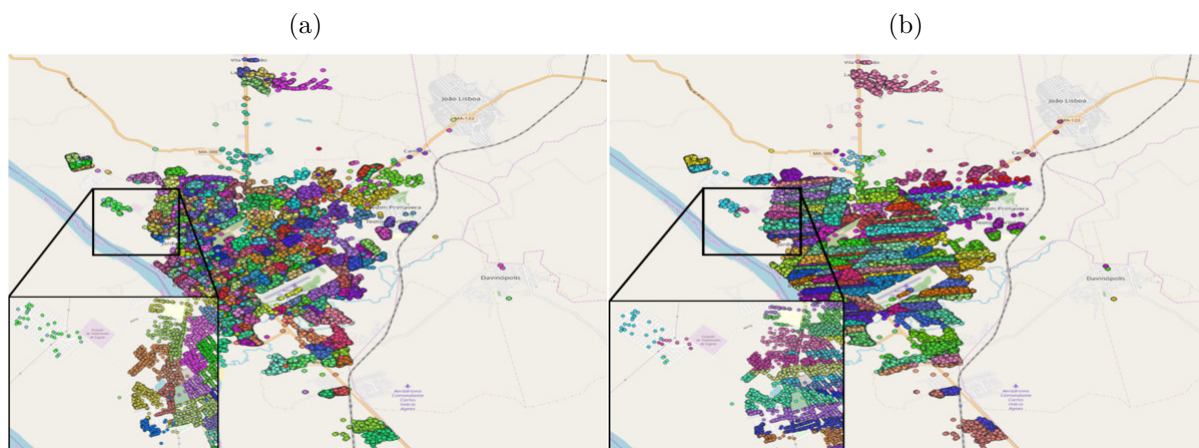
A Figura 21 revela o planejamento proposto. Nesse planejamento a quantidade de ULs urbanas obtidas foi menor que a quantidade do planejamento vigente. Ainda na área rural, houve uma redução no desvio padrão dos tempos de leitura das ULs, o qual originalmente era 77,23 minutos e agora equivale a 69,48 minutos. Por fim, o coeficiente de silhueta foi elevado para 0,2602 sendo considerado um ganho modesto, garantindo que não existam ULs sobrepostas. Um exemplo de correção do problema de sobreposição é ilustrado na Figura 22

Figura 21 – Planejamentos proposto para a cidade de Imperatriz-MA: (a) agrupamento urbano (b) agrupamento rural



Fonte: Próprio Autor

Figura 22 – Comparação entre ULS urbanas dos planejamentos de Imperatriz-MA: (a) ULS vigentes (b) ULS propostas



Fonte: Próprio Autor



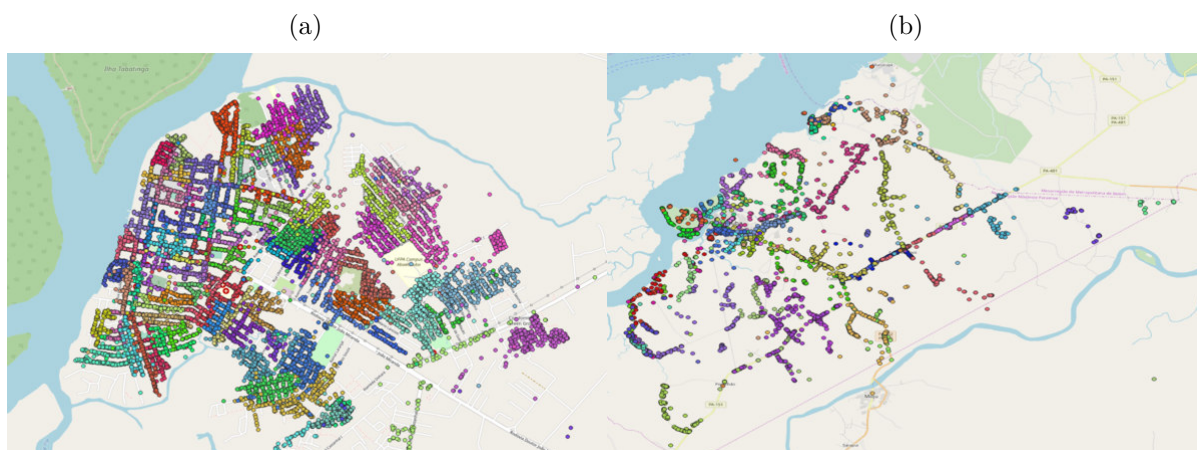
### 5.0.3 Planejamento para a cidade de Abaetetuba

O município de Abaetetuba, situado no estado do Pará, também possui duas organizações para o planejamento de leitura vigente. Esses planos são mostrados na Figura 23. Após a aplicação das métricas obteve-se os dados apresentados na Figura 23.

	Planejamento Vigente		Planejamento Proposto	
	Urbano	Rural	Urbano	Rural
Quantidade de ULs	91	60	<b>80</b>	<b>51</b>
Desvio Padrão	90,70	129,38	<b>71,21</b>	<b>60,42</b>
Coef. Silhueta	0,1764	0,1431	<b>0,2142</b>	<b>0,2789</b>

Tabela 3 – Comparação de métricas para cidade de Abaetetuba-PA

Figura 23 – Planejamentos vigentes da cidade de Abaetetuba-PA: (a) agrupamento urbano (b) agrupamento rural

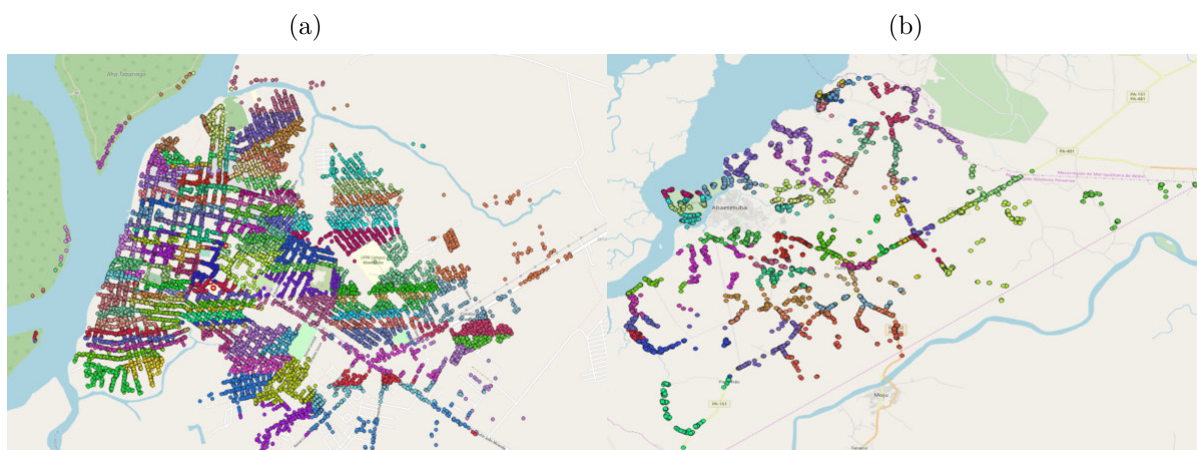


Fonte: Próprio Autor

A área urbana, é composta por 91 unidades de leitura, as quais possuem um desvio padrão da carga de trabalho equivalente a 90,70 minutos e um coeficiente de silhueta de 0,1764. Por sua vez, a área rural engloba 60 ULs, com um desvio padrão igual a 129,38 minutos e 0,1431 para o coeficiente de silhueta. Como mostrado na Tabela 3, o planejamento gerado pelo método proposto para essa cidade também foi superior à situação vigente.

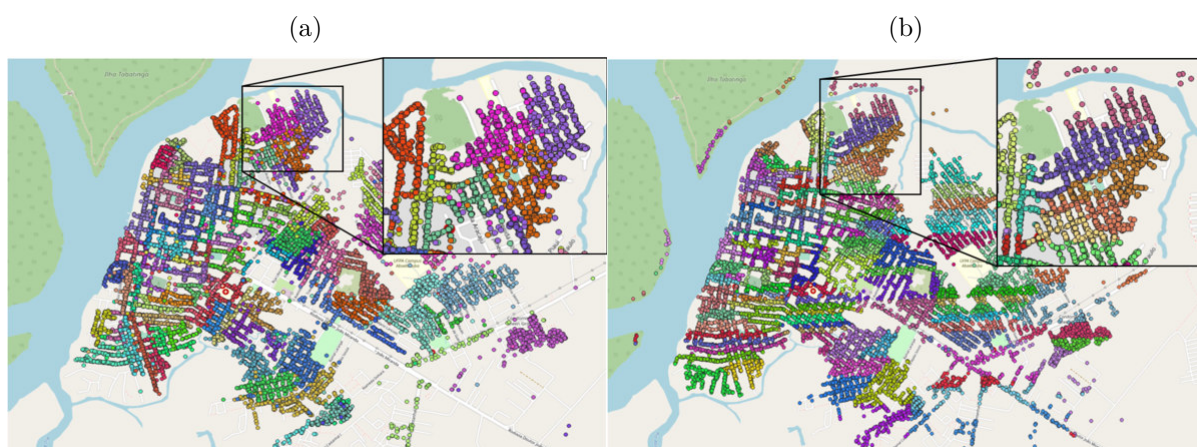
No planejamento proposto, a quantidade de ULs foi reduzida e o desvio padrão foi minimizado, indicando uma melhor distribuição de instalações entre os grupos; e o coeficiente de silhueta foi melhorado, tanto para a área urbana quanto para rural, indicando grupos mais coesos e sem sobreposição, conforme exibido na Figura 24. Novamente, é importante observar, que problemas como a sobreposição de grupos não ocorre no planejamento proposto. Para evidenciar esse fato, algumas ULs na área urbana do planejamento vigente foram colocadas em foco na Figura 25 junto com ULs obtidas pelo método proposto.

Figura 24 – Planejamentos proposto para a cidade de Abaetetuba-PA: (a) agrupamento urbano (b) agrupamento rural



Fonte: Próprio Autor

Figura 25 – Comparação entre ULS urbanas dos planejamentos de Abaetetuba-PA: (a) ULS vigentes (b) ULS propostas



Fonte: Próprio Autor

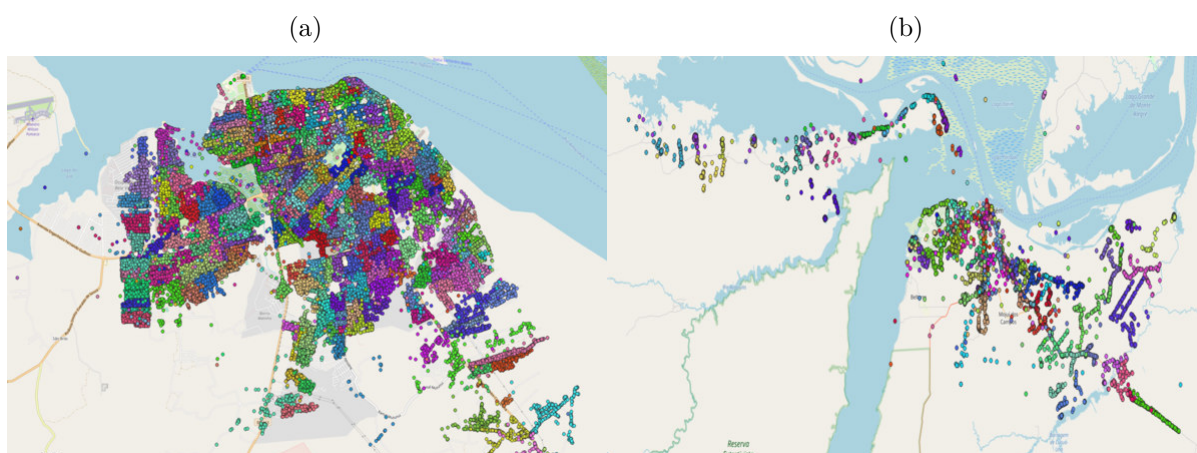
#### 5.0.4 Planejamento para a cidade de Santarém

O planejamento vigente da cidade de Santarém foi analisado e pode ser observado na Figura 26. O planejamento da área urbana é composto por 314 ULs, um desvio padrão de 134,69 minutos e um coeficiente de silhueta equivalente a 0,1521. Se comparado ao planejamento urbano gerado pelo método proposto, é observa-se um leve crescimento da quantidade de ULs. Esse aumento na cidade de Santarém, se originou devido ULs originais sofrerem com uma sobrecarga acentuada, até mesmo quando comparadas com as ULs da situação vigente das outras três cidades apresentadas anteriormente. Essa situação é corroborada ao se analisar a diminuição do desvio padrão dos tempos de leitura para 71,30 minutos, o qual originalmente alcançava 134,69 minutos. É possível perceber que o desvio padrão foi reduzido quase que pela metade. Finalmente, ao se comparar os coeficientes de silhueta, o obtido pelo método alcançou o valor de 0,2037 se mostrando superior ao original, conforme exposto na Tabela 4.

	Planejamento Vigente		Planejamento Proposto	
	Urbano	Rural	Urbano	Rural
<b>Quantidade de ULs</b>	<b>314</b>	<b>113</b>	354	121
<b>Desvio Padrão</b>	134,69	158,35	<b>71,30</b>	<b>62,48</b>
<b>Coef. Silhueta</b>	0,1521	0,2326	<b>0,2037</b>	<b>0,2816</b>

Tabela 4 – Comparação de métricas para cidade de Santarém-PA

Figura 26 – Planejamentos vigentes da cidade de Santarém-PA: (a) agrupamento urbano (b) agrupamento rural

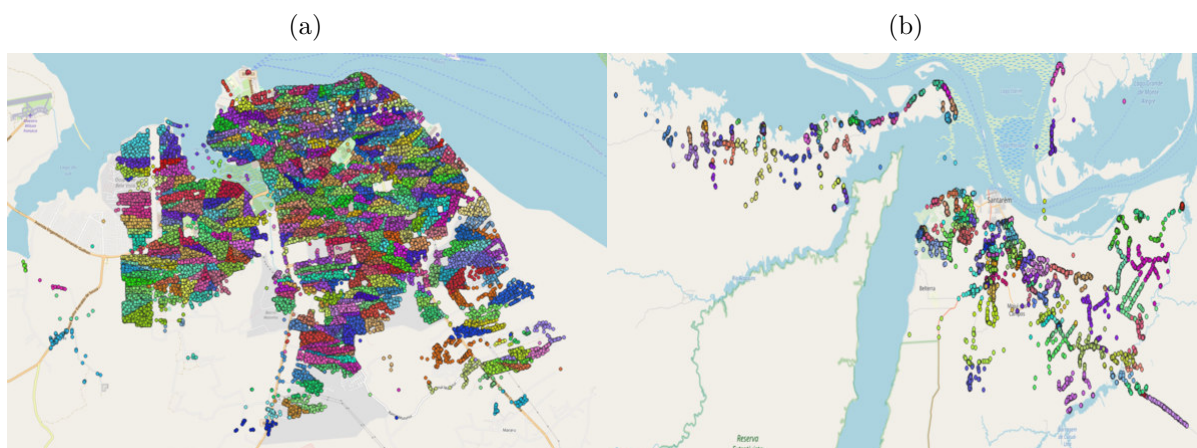


Fonte: Próprio Autor

Na Tabela 4, é possível observar que no planejamento rural obtido ocorre uma situação semelhante ao ocorrido na área urbana. O desvio padrão foi reduzido de 158,35 para 62,48 minutos. No entanto, a quantidade de de ULs também sofreu um leve aumento, uma vez que foram obtidas 121 ULs e no planejamento rural vigente existiam 113, esse

aumento conforme explicado na área urbana, se deu a sobrecarga acentuada das unidades de leitura originais. Ainda na área rural, o coeficiente de silhueta foi aprimorado, antes equivalente a 0,2326 e após a execução do método alcançando 0,2816. Os planejamentos propostos, são apresentados na Figura 27.

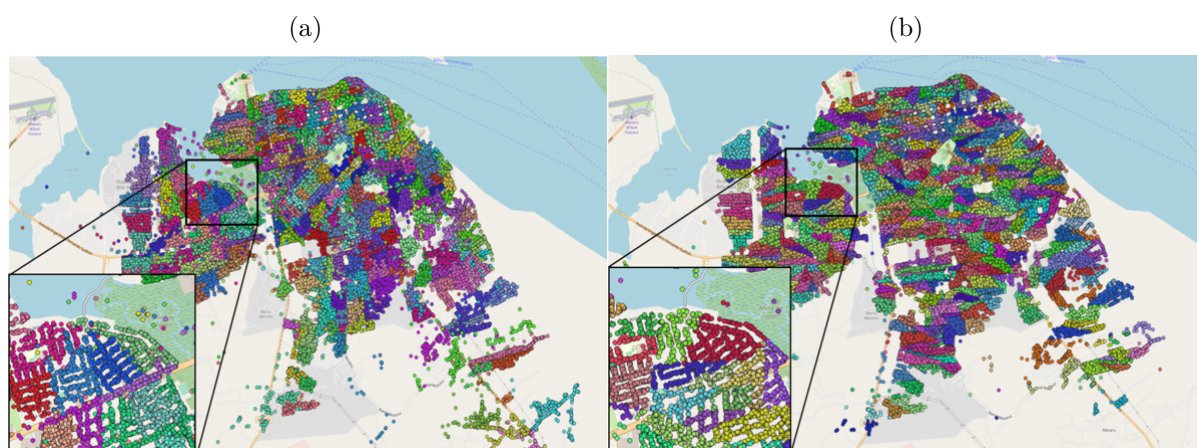
Figura 27 – Planejamentos propostos para a cidade de Santarém-PA: (a) agrupamento urbano (b) agrupamento rural



Fonte: Próprio Autor

No novo planejamento para a cidade de Santarém, obtido com o método proposto, também ocorre a mesma situação apresentada para as cidades anteriores, em que o problema de sobreposição encontrado no planejamento vigente é corrigido. A correção é exibida, através do exemplo da Figura 28.

Figura 28 – Comparação entre ULS urbanas dos planejamentos de Santarém-PA: (a) ULS vigentes (b) ULS propostas



Fonte: Próprio Autor

## 5.1 Discussão

A partir dos resultados apresentados, é possível avaliar a eficácia da utilização do método proposto para a geração de planejamentos de leitura. Ao se comparar as unidades de leitura vigentes com as obtidas pelo método, percebe-se que houve melhoras significativas nos desvios padrão das ULs, indicando grupos mais homogêneos. Com isso, os leituristas – responsáveis por executar a leitura das instalações contidas no planejamento – terão uma carga de trabalho mais igualitária, contribuindo para reduzir a ociosidade e a sobrecarga desses funcionários.

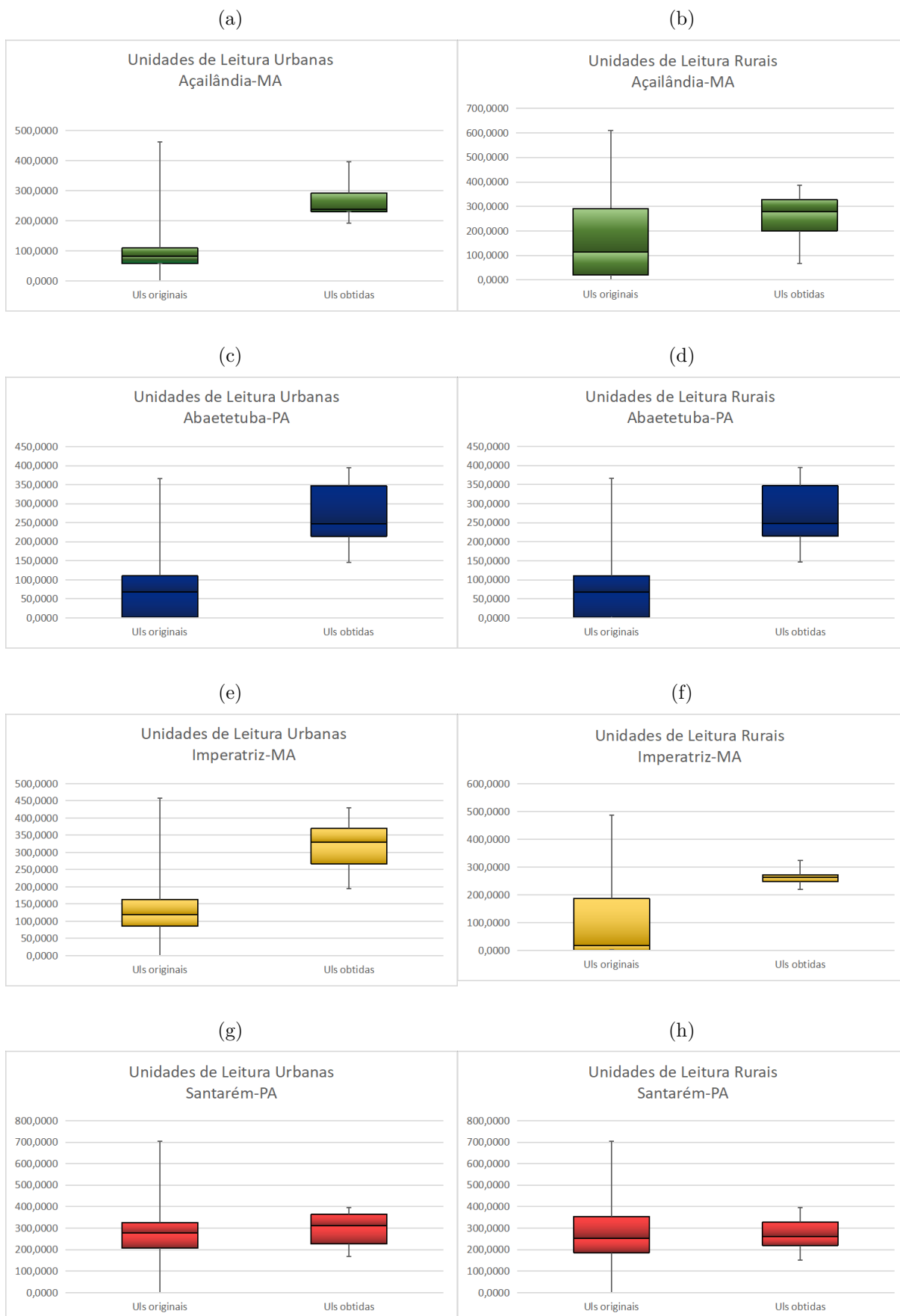
Essa situação é corroborada pelos *boxplots* apresentados na Figura 29. Ao analisar os gráficos, é possível observar que a média de instalações aumentou nas ULs obtidas. Esse aumento significa que, a capacidade de cada UL foi melhor aproveitada. Dessa maneira, ULs na situação vigente que tinham uma carga de trabalho muito abaixo da capacidade, tiveram suas instalações redistribuídas e dessa forma não se faziam mais necessárias. Também é possível observar que a variação de carga de trabalho entre as ULs diminuiu, pois a amplitude dos tempos de leitura, em todos os casos, foram menores em comparação à situação vigente.

Um caso peculiar acontece na cidade de Santarém no estado do Pará, em que o desvio padrão diminuiu mesmo com o modesto aumento da quantidade de unidades de leitura. Esse acontecimento justifica-se pelo valor máximo de carga de trabalho ser muito alto na no agrupamento atual da companhia, indicando que existem ULs que sofrem com grave sobrecarga. No agrupamento obtido após a aplicação do método proposto, essa sobrecarga foi corrigida ao realizar a divisão de carga com novas ULs.

O coeficiente de silhueta, responsável por avaliar a coesão e a separação dos grupos, foi outra métrica em que as novas ULs se mostraram melhores que as atuais. Todas as cidades apresentadas tiveram seus coeficientes urbanos e rurais melhorados. Um caso interessante para analisar são os agrupamentos gerados para a cidade de Imperatriz-MA. Nessa cidade, o coeficiente de silhueta das ULs rurais da situação vigente deixou de ser um valor negativo, o que indica ULs com muita sobreposição, e alcançou um valor de 0,5141 indicando ULs coesas e não sobrepostas.

É importante ressaltar que o emprego da BSP, como explicado na Seção 4.5.2, garantiu a geração de grupos contíguos e não sobrepostos, restrição fundamental do problema de agrupamento capacitado explicado na Seção 3.1. A utilização do método proposto, também assegura a soma dos pesos atribuídos ao grupo não extrapole a capacidade máxima estipulada. Contudo, existem limitações no método proposto como, por exemplo, a geração de grupos lineares mesmo que em poucos casos, que geograficamente podem não ser a melhor resposta esperada.

Figura 29 – *BoxPlot* da Carga de trabalho de ULs: (a) Área urbana Açailândia; (b) Área rural Açailândia; (c) Área urbana Abaetetuba; (d) Área rural Abaetetuba; (e) Área urbana Imperatriz; (f) Área rural Imperatriz; (g) Área urbana Santarém; (h) Área rural Santarém.



## 6 Conclusões

Este trabalho apresentou um método baseado em algoritmo genético e particionamento binário do espaço que permite a construção de agrupamentos de forma automática, para solucionar o problema de agrupamento capacitado aplicado à logística de leitura em companhias de distribuição de energia elétrica, respeitando os critérios de compacidade e homogeneidade. O método pode ser uma alternativa eficiente para a geração de planejamentos de leitura, em virtude, da elaboração dos planejamentos vigentes serem construídos de forma manual pelos funcionários das companhias.

O método proposto, foi descrito detalhadamente no Capítulo 4. Nesse Capítulo, foi mostrado a aquisição dos dados. Devido a presença de *outliers* nos dados adquiridos, foi desenvolvida uma etapa de pré-processamento, em que se utiliza a *KD-tree* e a normalização por *z-score* para identificar e separar os *outliers*. Na fase seguinte do método, responsável pelo agrupamento das instalações e definição das ULs, foi explicada a modelagem, a execução do algoritmo genético e o particionamento realizado utilizando-se a BSP. Por fim, a etapa de pós-processamento foi definida, alocando os *outliers* nas UL mais próxima. Após essa etapa os grupos são avaliados conforme o desvio padrão dos tempos de leitura e o coeficiente de silhueta.

A partir dos resultados apresentados na Seção 5, observa-se a eficácia do método proposto, apresentando significativa diminuição do desvio padrão da carga de trabalho das unidades de leitura, redução da quantidade de ULs na maioria dos casos e um aumento do coeficiente de silhueta. Isso indica a geração de ULs coesas, não sobrepostas, homogêneas e que respeitam a capacidade estipulada.

Para trabalhos futuros, sugere-se utilizar, durante o processo de agrupamento, informações geográficas além das coordenadas de latitude, longitude e a distância euclidiana como, por exemplo, informações do arruamento da região, com o objetivo de facilitar a rota seguida pelo do funcionário responsável pela leitura do grupo. Propõe-se, ainda, investigar as situações, em que o método tende a gerar grupos com uma forma linear, com a finalidade de corrigir essa limitação e que seja possível a geração de grupos mais circulares, os quais geograficamente são uma resposta mais adequada. Por fim, pretende-se estudar heurísticas como o *simulated annealing* (LAARHOVEN; AARTS, 1987) e aplicá-las a otimização dos fatores  $\alpha$  e  $\beta$  utilizados na função objetivo do algoritmo genético, para permitir um aprimoramento do particionamento do espaço.



## Referências

- ABDI, H.; WILLIAMS, L. J. Normalizing data. *Encyclopedia of research design*, Sage Thousand Oaks CA, v. 1, 2010. Citado 2 vezes nas páginas 29 e 34.
- ANEEL. *Resolução Normativa Nº 414, de 9 de Setembro de 2010*. 2010. Disponível em: <<http://www.aneel.gov.br/documents/656877/14486448/bren2010414.pdf/3bd33297-26f9-4ddf-94c3-f01d76d6f14a?version=1.0>>. Acesso em: 10/12/2019. Citado na página 16.
- BARBER, C. B.; DOBKIN, D. P.; DOBKIN, D. P.; HUHDANPAA, H. The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.*, ACM, New York, NY, USA, v. 22, n. 4, p. 469–483, dez. 1996. ISSN 0098-3500. Disponível em: <<http://doi.acm.org/10.1145/235815.235821>>. Citado na página 35.
- BENTLEY, J. L. Multidimensional binary search trees used for associative searching. *Commun. ACM*, ACM, New York, NY, USA, v. 18, n. 9, p. 509–517, set. 1975. ISSN 0001-0782. Disponível em: <<http://doi.acm.org/10.1145/361002.361007>>. Citado 2 vezes nas páginas 28 e 29.
- BLAIS, M.; LAPIERRE, S. D.; LAPORTE, G. Solving a home-care districting problem in an urban setting. *Journal of the Operational Research Society*, Taylor Francis, v. 54, n. 11, p. 1141–1147, 2003. Disponível em: <<https://doi.org/10.1057/palgrave.jors.2601625>>. Citado na página 16.
- BRENT, R. P. An improved monte carlo factorization algorithm. *BIT Numerical Mathematics*, Springer, v. 20, n. 2, p. 176–184, 1980. Citado na página 20.
- BUSSAMRA, N. M. *Distritamento eleitoral : uma metodologia para definir o recorte dos distritos*. Dissertação (Mestrado) — Unicamp, <http://repositorio.unicamp.br/jspui/handle/REPOSIP/261569>, 1995. Citado na página 16.
- CARVALHO, R. de; SALDANHA, R. R.; MARTINS, A. X. Heurística aplicada ao problema de agrupamento centrado capacitado. p. 2691–2702, 09 2015. Disponível em: <<http://www.din.uem.br/sbpo/sbpo2015/pdf/142902.pdf>>. Citado na página 17.
- CHAVES, A. A.; LORENA, L. A. N. Clustering search algorithm for the capacitated centered clustering problem. *Computers & Operations Research*, Elsevier, v. 37, n. 3, p. 552–558, 2010. Citado na página 20.
- COSTA, P.; SILVA, I. S. da; VIEIRA, P. C.; SANTOS, R. A.; SILVA, M.; LIMA, C. C.; JÚNIOR, D. G.; GARROS, E.; SILVA, I. S. da; PINHEIRO, L. A. Rmeans: A recursive capacitated means for districting problem. *Journal on Advances in Theoretical and Applied Informatics*, v. 4, n. 1, p. 28–35, 2018. ISSN 2447-5033. Disponível em: <<https://revista.univem.edu.br/jadi/article/view/2745>>. Citado 2 vezes nas páginas 17 e 21.
- DESALE, S.; RASOOL, A.; ANDHALE, S.; RANE, P. Heuristic and meta-heuristic algorithms and their relevance to the real world: A survey. *INTERNATIONAL JOURNAL*

OF COMPUTER ENGINEERING IN RESEARCH TRENDS, v. 351, p. 2349–7084, 01 2015. Citado na página 22.

EASINGWOOD, C. A heuristic approach to selecting sales regions and territories. *Journal of the Operational Research Society*, v. 24, n. 4, p. 527–534, Dec 1973. ISSN 1476-9360. Disponível em: <<https://doi.org/10.1057/jors.1973.102>>. Citado na página 17.

ESTER, M.; KRIEGEL, H.-P.; SANDER, J.; XU, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: . ACM, 1996. v. 96, p. 226–231. Disponível em: <<https://dl.acm.org/citation.cfm?id=3001507>>. Citado na página 17.

FEO, T. A.; RESENDE, M. G. Greedy randomized adaptive search procedures. *Journal of global optimization*, Springer, v. 6, n. 2, p. 109–133, 1995. Citado na página 20.

FONSECA, A. S. *Estratégia de agrupamento baseada em algoritmo genético para a construção capacitada de unidades de leitura*. 2019. Monografia (Bacharel em ciência da Computação), UFMA (Universidade Federal do Maranhão), Maranhão, Brasil. Citado na página 36.

FRANÇA, P. M.; GARCIA, V. J.; MORELATO, A.; USBERTI, F. L. Enfoque multicritério para o problema de redistribuição capacitada. *XXXIX Simpósio Brasileiro de Pesquisa Operacional-SBPO*, p. 1–12, 2007. Citado 2 vezes nas páginas 17 e 20.

FRANÇA, P.; SOSA, N. M.; PUREZA, V. An adaptive tabu search algorithm for the capacitated clustering problem. *International Transactions in Operational Research*, v. 6, n. 6, p. 665–678, 1999. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-3995.1999.tb00180.x>>. Citado na página 22.

FUCHS, H.; KEDEM, Z. M.; NAYLOR, B. F. On visible surface generation by a priori tree structures. *SIGGRAPH Comput. Graph.*, ACM, New York, NY, USA, v. 14, n. 3, p. 124–133, jul. 1980. ISSN 0097-8930. Disponível em: <<http://doi.acm.org/10.1145/965105.807481>>. Citado 2 vezes nas páginas 23 e 24.

FULCHER, J. Computational intelligence: An introduction. In: \_\_\_\_\_. *Computational Intelligence: A Compendium*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p. 3–78. ISBN 978-3-540-78293-3. Disponível em: <[https://doi.org/10.1007/978-3-540-78293-3\\_1](https://doi.org/10.1007/978-3-540-78293-3_1)>. Citado na página 26.

GAREY, M.; JOHNSON, D. Computers and intractability: A guide to the theory of np-completeness. In: \_\_\_\_\_. ACM, 1979. Disponível em: <<https://dl.acm.org/citation.cfm?id=574848>>. Citado na página 22.

GARFINKEL, R. S.; NEMHAUSER, G. L. Optimal political districting by implicit enumeration techniques. *Manage. Sci.*, INFORMS, Institute for Operations Research and the Management Sciences (INFORMS), Linthicum, Maryland, USA, v. 16, n. 8, p. B-495–B-508, abr. 1970. ISSN 0025-1909. Disponível em: <<http://dx.doi.org/10.1287/mnsc.16.8.B495>>. Citado na página 16.

GUERINE, M.; STOCKINGER, M. B.; ROSSETI, I.; PLASTINO, A. Heurística híbrida com mineração de dados para o problema de agrupamento capacitado com centro geométrico. 09 2017. Disponível em: <<http://www.sbp2017.iltc.br/pdf/169441.pdf>>. Citado na página 20.

- HAN, J.; KAMBER, M.; PEI, J. 3 - data preprocessing. In: HAN, J.; KAMBER, M.; PEI, J. (Ed.). *Data Mining (Third Edition)*. Third edition. Boston: Morgan Kaufmann, 2012, (The Morgan Kaufmann Series in Data Management Systems). p. 83 – 124. ISBN 978-0-12-381479-1. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9780123814791000034>>. Citado na página 20.
- HARTIGAN, J. A.; WONG, M. A. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, JSTOR, v. 28, n. 1, p. 100–108, 1979. Citado na página 17.
- HOLLAND, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1992. ISBN 0262082136. Citado 2 vezes nas páginas 25 e 27.
- Judd, D.; McKinley, P. K.; Jain, A. K. Large-scale parallel data clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 20, n. 8, p. 871–876, Aug 1998. ISSN 0162-8828. Citado na página 17.
- KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by simulated annealing. *science*, American Association for the Advancement of Science, v. 220, n. 4598, p. 671–680, 1983. Citado na página 20.
- LAARHOVEN, P. J. M. van; AARTS, E. H. L. *Simulated annealing*. Dordrecht: Springer Netherlands, 1987. 7–15 p. ISBN 978-94-015-7744-1. Disponível em: <[https://doi.org/10.1007/978-94-015-7744-1\\_2](https://doi.org/10.1007/978-94-015-7744-1_2)>. Citado na página 55.
- LODISH, L. M. Sales territory alignment to maximize profit. *Journal of Marketing Research*, v. 12, n. 1, p. 30–36, 1975. Disponível em: <<https://doi.org/10.1177/002224377501200105>>. Citado na página 17.
- MURTY, K. G. Linear and combinatorial programming. *ROBERT E. KRIEGER PUBLISHING COMPANY, MELBOURNE, FL(USA), 1985, 592*, 1985. Citado na página 20.
- MÜHLENBEIN, H.; GORGES-SCHLEUTER, M.; KRÄMER, O. Evolution algorithms in combinatorial optimization. *Parallel Computing*, v. 7, n. 1, p. 65 – 85, 1988. ISSN 0167-8191. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0167819188900981>>. Citado na página 17.
- NEGREIROS, M.; PALHANO, A. The capacitated centred clustering problem. *Computers Operations Research*, v. 33, p. 1639–1663, 06 2006. Citado na página 17.
- NOVAES, A. G.; CURSI, J. S. de; SILVA, A. C. da; SOUZA, J. C. Solving continuous location–districting problems with voronoi diagrams. *Computers Operations Research*, v. 36, n. 1, p. 40 – 59, 2009. ISSN 0305-0548. Part Special Issue: Operations Research Approaches for Disaster Recovery Planning. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0305054807001323>>. Citado na página 17.
- OLIVEIRA, M. D. B. d. Monitorizando a evolução de clusters. Faculdade de Economia da Universidade do Porto, 2011. Tese de Mestrado em Análise de Dados e Sistemas de Apoio à Decisão. Citado na página 38.

QGIS Development Team. *QGIS Geographic Information System*. [S.l.], 2019. Disponível em: <<http://qgis.osgeo.org>>. Citado na página 43.

RÍOS-MERCADO, R. Z.; FERNÁNDEZ, E. A reactive grasp for a commercial territory design problem with multiple balancing requirements. *Computers Operations Research*, v. 36, n. 3, p. 755 – 776, 2009. ISSN 0305-0548. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0305054807002249>>. Citado na página 17.

SANDER, J.; ESTER, M.; KRIEGEL, H.-P.; XU, X. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data Mining and Knowledge Discovery*, v. 2, n. 2, p. 169–194, Jun 1998. ISSN 1573-756X. Disponível em: <<https://doi.org/10.1023/A:1009745219419>>. Citado na página 17.

SELIM, S. Z.; ALSULTAN, K. A simulated annealing algorithm for the clustering problem. *Pattern Recognition*, v. 24, n. 10, p. 1003 – 1008, 1991. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/003132039190097O>>. Citado na página 17.

SILVA, A. A. *Implementação capacitada do Growing Neural Gas aplicada à logística de leitura em companhias de distribuição de energia elétrica*. 2019. Monografia (Bacharel em ciência da Computação), UFMA (Universidade Federal do Maranhão), Maranhão, Brasil. Citado na página 21.

STEFANELLO, F.; MÜLLER, F. M. Um estudo sobre problemas de agrupamento capacitado. *Anais do XLI SBPO, Simpósio Brasileiro de Pesquisa Operacional*, p. 2819–2828, 2009. Citado 2 vezes nas páginas 17 e 20.

STEINER, M. T. A.; COSTA, C. E. da S.; COSTA, D. M. B.; ZAMBENEDETTI, V. C. et al. Técnicas da pesquisa operacional aplicadas à logística de atendimento aos usuários de uma rede de distribuição de energia elétrica. *Sistemas & Gestão*, v. 1, n. 3, p. 229–243, 2009. Citado na página 20.

STOWELL, T.; SCORESBY, J.; COATS, K. C.; CAPELL, M. R.; SHELTON, B. E. Leveraging open source technology in 3d game engine development. In: *Discoveries in Gaming and Computer-Mediated Simulations: New Interdisciplinary Applications*. [S.l.]: IGI Global, 2011. p. 78–109. Citado na página 25.

TAVARES-PEREIRA, F.; FIGUEIRA, J. R.; MOUSSEAU, V.; ROY, B. Multiple criteria districting problems. *Annals of Operations Research*, v. 154, n. 1, p. 69–92, Oct 2007. ISSN 1572-9338. Disponível em: <<https://doi.org/10.1007/s10479-007-0181-5>>. Citado na página 17.

ZOLTNERS, A. A.; SINHA, P. Sales territory alignment: A review and model. *Management Science*, v. 29, n. 11, p. 1237–1256, 1983. Disponível em: <<https://doi.org/10.1287/mnsc.29.11.1237>>. Citado na página 17.