



UNIVERSIDADE FEDERAL DO MARANHÃO  
Curso de Ciência da Computação

Anderson Silva Fonseca

**Estratégia de agrupamento baseada em  
algoritmo genético para a construção  
capacitada de unidades de leitura**

São Luís - MA  
2019

Anderson Silva Fonseca

**Estratégia de agrupamento baseada em algoritmo  
genético para a construção capacitada de unidades de  
leitura**

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. Anselmo Cardoso de Paiva

São Luís - MA

2019

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

Silva Fonseca, Anderson.

Estratégia de agrupamento baseada em algoritmo genético para a construção capacitada de unidades de leitura / Anderson Silva Fonseca. - 2019.

58 f.

Coorientador(a): Italo Francyles Santos da Silva.

Orientador(a): Anselmo Cardoso de Paiva.

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

1. Agrupamento Capacitado. 2. Algoritmo Genético. 3. K-Means. 4. Problema de Distritamento. 5. RCMean. I. Cardoso de Paiva, Anselmo. II. Francyles Santos da Silva, Italo. III. Título.

Anderson Silva Fonseca

# **Estratégia de agrupamento baseada em algoritmo genético para a construção capacitada de unidades de leitura**

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 19/07/2019 - São Luís - MA:

---

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

---

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

---

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

---

Prof. Dr. Daniel Lima Gomes Junior  
Examinador  
Instituto Federal do Maranhão

São Luís - MA

2019

À minha família e meus amigos.

# Agradecimentos

Primeiramente aos meus pais, Josimario e Silvia, que sempre foram minha maior razão para continuar nos estudos.

Aos meus familiares, em especial Ivana, Inaldo e Vitor, por terem me acompanhado e ajudado na maioria dos meus obstáculos.

A Paulo Renato, por ter enfrentando o mesmo caminho que escolhi e se tornou umas das maiores inspirações pra continuar no curso.

A Italo Francyles, por ter toda a paciência de me instruir para a conclusão deste trabalho e acima de tudo, por se tornar umas das melhores companhias que tive.

A Ninguém (Pedro Almeida), por ter perdido um tempo valioso comigo estudando, jogando e conversando.

A Erik, Eduardo, Luciano, Mayara, Nelia e Wesley, por sempre estarem de prontidão para maioria dos Roles Gourmet.

A Carlos, Gabriel Monteles, Francisco, Rodrigo Garcês, Vinícius, Zaynab e a todos os outros membros do Codebuilders por terem me acompanhado em toda a minha jornada no curso.

A todos os membros do PET Computação que trabalharam, viajaram, comeram, roncaram, xingaram comigo. Obrigado pelos ensinamentos e pelo horário de intervalo entre as 12 e 14 horas.

A Great (Diego), Sky (Celielma), Lucas, Rod (Rodrigo Nascimento) e Victor, por falar mal de todos, por falar mal de nós mesmos, por dar conselhos duvidosos e por sempre terem despojado algum tempo para ajudar no trabalho do Sr. França.

A equipe maravilhosa do SAUL, formada por Pedro Henrique, Italo, Mourão, Robert, Mayara, Polyana e Victor, que tornaram o meu trabalho um simbolo de união, de amor, de carinho e de raiva. Obrigado por não soltaram a mão de ninguém.

A todos os integrantes do NCA, por terem me ajudado e me acolhido durante toda a minha vivência no laboratório.

A Alexsandra, Arthur, Dayane, David e Nunes, por terem me dado dicas de sobrevivência, conselhos bons e ruins e por terem ouvidos todos os meus desafios. Mais uma vez, obrigado.

A Geraldo, Anselmo e Daniel, por não terem só interpretado os papéis de professores, mas também de amigos, de conselheiros e de como se deve comer em um rodízio.

*“My thirst for knowledge  
cannot be quenched.”  
(Rubick, Defense of the Ancients 2)*

# Resumo

No Brasil, as empresas responsáveis pela distribuição de energia elétrica devem também se responsabilizar pela medição e do faturamento de seus clientes. Para facilitar a organização das medições em um dia de trabalho, essas empresas criam planejamentos de leitura. Esses planejamentos consistem em organizar seus clientes em grupos, os quais recebem a visita de um funcionário leiturista responsável por fazer a medição do consumo do cliente. Porém, a criação destes planejamentos corresponde a um problema de agrupamento capacitado com uma série de restrições de capacidade e homogeneidade. Para resolver esse problema é proposto uma estratégia de agrupamento capacitado baseado em algoritmo genético. Os planejamentos resultantes deste método foram criados com a intenção de obter o melhor agrupamento possível de uma região com a minimização dos custos de implementação. Estes resultados são comparados com a situação atual da empresa, mostrando as melhorias da aplicação deste método.

**Palavras-chaves:** Agrupamento Capacitado, K-Means, Algoritmo Genético, RCMMeans, Problema de Distritamento



# Abstract

In Brazil, electrical energy companies are responsible for the distribution, measuring, billing of their customers. To make it easier to organize a workday's measurements, these companies create a reading plan. These plans organize groups their clients, which are visited by several employees who measure the customer's consumption. However, the creation of these plans is a capacitated clustering problem with numerous capacity and homogeneity restrictions. To solve this problem a capacitated clustering strategy based on genetic algorithm is proposed. The resulting plans are created with the intention of obtaining the best possible clustering of a region, which minimizes implementation costs. This results are compared with the current situation of the company, showing the improvements of this method.

**Keywords:** Capacitated Clustering, K-Means, Genetic Algorithm, RCMean, Districting Problems

# Lista de ilustrações

Figura 1 – Estrutura de funcionamento de um Algoritmo Genético tradicional. . . . .	25
Figura 2 – Fluxo de Trabalho do <i>RCMeans</i> . . . . .	26
Figura 3 – Etapa de pós-processamento do <i>RCMeans</i> . . . . .	27
Figura 4 – Visão Geral do Método. . . . .	28
Figura 5 – Fluxo de Entrada do Método . . . . .	30
Figura 6 – Exemplo de agrupamento com destaque para a região densa . . . . .	32
Figura 7 – Amostra das instalações distribuídas no espaço bidimensional. . . . .	32
Figura 8 – Esquema da execução do processo de localização das regiões densas. . . . .	33
Figura 9 – Constituição do indivíduo. . . . .	34
Figura 10 – Esquema de cruzamento dos indivíduos. . . . .	35
Figura 11 – Fluxo de Execução do <i>RCMeans</i> . . . . .	36
Figura 12 – Visualização do Planejamento Vigente de Acailândia . . . . .	40
Figura 13 – Visualização do Planejamento Vigente de Imperatriz . . . . .	41
Figura 14 – Visualização do Planejamento Vigente de Marabá . . . . .	41
Figura 15 – Visualização do Planejamento Vigente de Santarém . . . . .	42
Figura 16 – Comparativos dos Planejamentos Urbanos Vigente com o Inédito de Acailândia . . . . .	43
Figura 17 – Comparativos dos Planejamentos Rurais Vigente com o Inédito de Acailândia . . . . .	43
Figura 18 – Comparativos dos Planejamentos Urbanos Vigente com o Inédito de Imperatriz . . . . .	44
Figura 19 – Comparativos dos Planejamentos Rurais Vigente com o Inédito de Imperatriz . . . . .	45
Figura 20 – Comparativos dos Planejamentos Urbanos Vigente com o Inédito de Marabá . . . . .	45
Figura 21 – Comparativos dos Planejamentos Rurais Vigente com o Inédito de Marabá	46
Figura 22 – Comparativos dos Planejamentos Urbanos Vigente com o Inédito de Santarém . . . . .	46
Figura 23 – Comparativos dos Planejamentos Rurais Vigente com o Inédito de Santarém . . . . .	47
Figura 24 – Comparativos dos Planejamentos Urbanos Vigente com o Planejamento Baseado de Acailândia . . . . .	48
Figura 25 – Comparativos dos Planejamentos Rurais Vigente com o Planejamento Baseado de Acailândia . . . . .	49
Figura 26 – Comparativos dos Planejamentos Urbanos Vigente com o Planejamento Baseado de Imperatriz . . . . .	50

Figura 27 – Comparativos dos Planejamentos Rurais Vigente com o Planejamento Baseado de Imperatriz . . . . .	50
Figura 28 – Comparativos dos Planejamentos Urbanos Vigente com o Planejamento Baseado de Marabá . . . . .	51
Figura 29 – Comparativos dos Planejamentos Rurais Vigente com o Planejamento Baseado de Marabá . . . . .	51
Figura 30 – Comparativos dos Planejamentos Urbanos Vigente com o Planejamento Baseado de Santarém . . . . .	52
Figura 31 – Comparativos dos Planejamentos Rurais Vigente com o Planejamento Baseado de Santarém . . . . .	52

# Lista de tabelas

Tabela 1 – Amostra dos dados de instalação fornecidos . . . . .	29
Tabela 2 – Dados resumidos da situação atual de Açailândia . . . . .	39
Tabela 3 – Dados resumidos da situação atual de Imperatriz . . . . .	40
Tabela 4 – Dados resumidos da situação atual de Marabá . . . . .	40
Tabela 5 – Dados resumidos da situação atual de Santarém . . . . .	41
Tabela 6 – Dados comparativos do planejamento vigente com o planejamento inédito do município de Açailândia . . . . .	42
Tabela 7 – Dados comparativos do planejamento vigente com o planejamento inédito do município de Imperatriz . . . . .	44
Tabela 8 – Dados comparativos do planejamento vigente com o planejamento inédito do município de Marabá . . . . .	44
Tabela 9 – Dados comparativos do planejamento vigente com o planejamento inédito do município de Santarém . . . . .	46
Tabela 10 – Dados comparativos do planejamento vigente com o planejamento baseado do município de Açailândia . . . . .	48
Tabela 11 – Dados comparativos do planejamento vigente com o planejamento baseado do município de Imperatriz . . . . .	49
Tabela 12 – Dados comparativos do planejamento vigente com o planejamento baseado do município de Marabá . . . . .	49
Tabela 13 – Dados comparativos do planejamento vigente com o planejamento baseado do município de Santarém . . . . .	51

# Lista de abreviaturas e siglas

AG	Algoritmo Genético
ANEEL	Agência Nacional de Energia Elétrica
BFS	Breadth-First Search
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 Capacitado com Centro Geométrico
UL	Unidade de Leitura

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
<b>1.1</b>	<b>Objetivo Geral</b>	<b>16</b>
1.1.1	Objetivos Específicos	16
<b>1.2</b>	<b>Justificativa</b>	<b>16</b>
<b>1.3</b>	<b>Organização do Trabalho</b>	<b>16</b>
<b>2</b>	<b>TRABALHOS RELACIONADOS</b>	<b>18</b>
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>20</b>
<b>3.1</b>	<b>Agrupamento Capacitado</b>	<b>20</b>
<b>3.2</b>	<i>K-Means</i>	21
<b>3.3</b>	<b>Algoritmo Genético</b>	<b>23</b>
<b>3.4</b>	<i>RCMeans</i>	24
<b>4</b>	<b>METODOLOGIA</b>	<b>28</b>
<b>4.1</b>	<b>Visão Geral</b>	<b>28</b>
<b>4.2</b>	<b>Aquisição dos dados</b>	<b>29</b>
<b>4.3</b>	<b>Geração de Etapas</b>	<b>30</b>
4.3.1	Reconhecimento da Situação	30
4.3.2	Localização das Regiões Densas	30
4.3.3	Execução do Algoritmo Genético	33
4.3.4	Geração de Unidades de Leitura	35
<b>4.4</b>	<b>Métricas de Avaliação</b>	<b>37</b>
4.4.1	Minimização do Desvio do Padrão	37
4.4.2	Coeficiente de Silhueta	37
<b>5</b>	<b>RESULTADOS</b>	<b>39</b>
<b>5.1</b>	<b>Análise da Situação Vigente</b>	<b>39</b>
<b>5.2</b>	<b>Agrupamento a partir de uma Situação Inédita</b>	<b>42</b>
<b>5.3</b>	<b>Agrupamento a partir da Situação Vigente</b>	<b>48</b>
<b>6</b>	<b>CONCLUSÕES</b>	<b>54</b>
	<b>REFERÊNCIAS</b>	<b>55</b>

# 1 Introdução

O processo de medição, revisão e supervisão do consumo de eletricidade no Brasil é realizado pelas empresas de distribuição de energia elétrica. Essas distribuidoras também são responsáveis por realizar a leitura do consumo e do faturamento de cada cliente. A Agência Nacional de Energia Elétrica (ANEEL) regulamenta como as distribuidoras de energia devem realizar o fornecimento da eletricidade, com a finalidade de aperfeiçoar e definir condições gerais para o relacionamento entre a distribuidora e o cliente.

A ANEEL estabelece, entre outras regulamentações, que uma concessionária deve efetuar o faturamento de uma unidade consumidora no ciclo de 30 dias, tendo variações entre 27 e 33 dias. Caso seja necessário realizar um remanejamento ou alteração do período de leitura, essas datas podem ser alteradas entre o intervalo de no mínimo 15 dias e no máximo 47 dias, avisando antecipadamente o consumidor por escrito e com atenciedade mínima de um ciclo de faturamento ([ANEEL, 2010](#)).

Dada a regulamentação estabelecida, as distribuidoras de energia elétrica organizam a medição e a entrega das faturas de cada unidade consumidora através de grupos chamados de Unidades de Leitura (UL). A leitura dos medidores de uma UL é feita por um único funcionário leiturista, que mede e emite a fatura durante a visita de cada unidade consumidora. Com a finalidade de facilitar a organização da medição e respeitar o critério de dias exigidos, as ULs próximas são reunidas em grupos maiores chamados de Etapas, que por sua vez correspondem ao dia no qual será feita a leitura das unidades consumidoras.

Para cada ciclo de faturamento, é criado um planejamento das medições para um município ou grupo de municípios, definido como plano de leitura, no qual são criadas diversas ULs contendo as unidades consumidoras da região-alvo que posteriormente serão agrupadas em Etapas. Durante um dia do ciclo, uma etapa receberá a vistoria dos leituristas, e assim sucessivamente até o final do ciclo de faturamento.

Atualmente, a criação dos planos de leitura é um trabalho manual realizado por diversos especialistas que conhecem os critérios exigidos para a formação dos agrupamentos. Por sua vez, esse trabalho é feito usando apenas a experiência do funcionário, ou seja, não há maneiras garantidas de encontrar a melhor otimização para os grupos. Cada planejamento, por se tratar de uma tarefa que atinge diretamente o faturamento do cliente e os lucros da companhia, é submetido a diversas revisões e refatorações até que se encontre a solução mais agradável para a empresa.

Durante a criação dos agrupamento são respeitados dois critérios: a homogeneidade, que impacta diretamente no tempo de trabalho dos leituristas e, portanto, na sua carga

de trabalho; e a compacidade, que se refere a proximidade geográfica de cada unidade consumidora, a qual a forma geográfica impacta em como uma área é explorada de maneira eficiente pelo funcionário.

O procedimento de planejamento de um plano de leitura é um exemplo prático de um problema abordado na literatura como distritamento. [Assis, França e Usberti \(2010\)](#) definem o distritamento como problema de otimização combinatória, que possui o objetivo de agrupar  $n$  unidades territoriais contíguas em  $p$  agrupamentos ( $p < n$ ), buscando encontrar a melhor solução de acordo com um critério de otimização.

Existem diversas soluções usadas para criar agrupamentos de indivíduos semelhantes. Estes algoritmos, por sua vez, requerem um ou mais critérios para que estes grupos (*clusters*) sejam criados ou classificados. Um método muito popular conhecido como *K-Means* é bastante usado em aplicações científicas e industriais para agrupamentos. O *K-Means* usa como parâmetro um determinado número de *clusters*, sendo que cada *cluster* é caracterizado por um centroide que representa, na maioria das vezes, a média do grupo ([BERKHIN, 2006](#)).

Devido à natureza do problema, encontrar um resultado ótimo para a organização do plano de leitura remete a um agrupamento baseado na situação geográfica da região da leitura, dos critérios de criação e no esforço que será exercido pelo funcionário leiturista. Esta situação pode ser aplicada como problema de otimização multiobjetivo (POM), em que busca-se alcançar a solução mais viável dentre as restrições estabelecidas pelos parâmetros do problema ([ÁVILA et al., 2006](#)).

Encontrar uma organização objetiva que respeite os aspectos geográficos de uma região específica, as regras definidas pela ANEEL e os critérios específicos para os grupos de leitura é um trabalho árduo para os especialistas, pois, devido à dinamicidade das variáveis, não existe uma solução que se adapte a todos os casos de distritamento. Essa dificuldade de encontrar resoluções ótimas usando recursos convencionais torna o uso de algoritmos genéticos um método eficiente para otimização ([BARBOSA; RIBEIRO; ARANTES, 2010](#)). Algoritmos Genéticos são métodos inspirados em evolução. Esses algoritmos codificam diversas potenciais soluções para um problema específico que sofrem diversas recombinações, tentando sempre manter a informação crítica, no caso, a função objetivo ([WHITLEY, 1994](#)).

Este trabalho propõe usar os conhecimentos de agrupamento capacitado e algoritmo genético para elaborar um método de criação de grupos capacitados que melhorem o planejamento de leitura criado pelas distribuidoras de energia elétrica. Serão usados como base a organização de planejamento oferecidos pela Equatorial Energia. Para mostrar os resultados, serão realizadas comparações com a saída do método proposto e a situação da companhia elétrica.



## 1.1 Objetivo Geral

Este trabalho tem como objetivo o desenvolvimento de uma estratégia de agrupamento para as ULs, utilizando a técnica *K-Means* com centroides definidos por algoritmo genético. O método proposto almeja obter Etapas melhores e mais otimizadas do que os planejamentos criados manualmente pelos especialistas, que respeitem os critérios de homogeneidade e compacidade, as regulamentações da ANEEL e o tempo de trabalho do leiturista.

### 1.1.1 Objetivos Específicos

- Construir um método para realizar o balanceamento da carga de trabalho de um leiturista durante um dia de trabalho
- Contruir uma metodologia para a criação de agrupamentos que respeitem as normas estabelecidas pela ANEEL
- Análise comparativa entre o planejamento do cenário real e o obtido com o método proposto

## 1.2 Justificativa

O uso de métodos computacionais capazes de gerenciar a logística de criação de grupos de leitura de maneira eficiente possibilitam a geração de melhorias no processo de faturamento por otimização de recurso.

A automatização do processo de criação de plano de leitura irá permitir uma significativa redução do esforço exercido pelos funcionários, que por sua vez efetuam o trabalho manualmente para cada município ou conjunto de municípios. Além disso, melhorar o grupo de trabalho do leiturista e a entrega das faturas proporcionam uma diminuição do esforço exercido pelos especialistas para a criação de planos de leitura e maior qualidade de serviço ao cliente-final das empresas de energia elétrica.

## 1.3 Organização do Trabalho

O presente trabalho está organizado da seguinte maneira: o Capítulo 2 apresenta alguns trabalhos relacionados ao problema de agrupamento de unidades de leitura; o Capítulo 3 apresenta a fundamentação teórica das tecnologias necessárias para o entendimento do trabalho; no Capítulo 4 é descrito o funcionamento do método proposto, incluindo as métricas usadas para avaliação; 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 alguns trabalhos relacionados que exploram o problema de agrupamento de unidades de leituras das concessionárias de energia elétrica.

Costa, Costa e Góes (2007) 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 capacitado <sup>1</sup> (MLADENOVIC et al., 2007) e algoritmos genéticos. Foram comparados os resultados entre as duas abordagens propostas, e verificou-se que o método que utiliza algoritmos genéticos gera melhores soluções para o problema abordado. Verificou-se também que ambas as propostas obtiveram resultados melhores em comparação à criação manual dos agrupamentos, prática até então adotada pela concessionária.

França et al. (2007), Assis, França e Usberti (2009) e Assis, Franca e Usberti (2014) propuseram métodos heurísticos aplicados para 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, 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 propõem uma nova redefinição das áreas de faturamento com as instalações devidamente agrupadas obedecendo as restrições impostas.

Guerine et al. (2017) propuseram uma heurística híbrida para a solução do Agrupamento Capacitado com Centro Geométrico (PACC). Neste trabalho, usa-se uma combinação entre *Data Mining* (HAN; PEI; KAMBER, 2011), *Clustering Search* (CHAVES; LORENA, 2010) e *Simulated Annealing* (KIRKPATRICK; GELATT; VECCHI, 1983) para encontrar as melhores soluções no espaço de busca, sendo que a primeira é aplicada para detectar padrões e, a partir disso, guiar a busca por novas e melhores soluções. As principais aplicações reais para o PACC são aplicadas em projetos de zonas de coleta de

---

<sup>1</sup> A técnica usa os conceitos de agrupamento capacitado

lixo, definições de áreas de cobertura de vendedores, dentre outras.

Smiderle, Steiner e Carnieri (2003) propuseram um método baseado em técnicas de pesquisa operacional, para obter o caminho otimizado percorrido por leiturista para visitar os pontos de leitura, visando minimizar o tempo de deslocamento realizado. Esse trabalho é focado em dois processos principais: a primeira parte é focada na criação de grupos de atendimento, a qual utilizou-se algoritmo genético em conjunto com algoritmos de heurística clássica de Teitz e Bart (TEITZ; BART, 1968) aplicados ao problema das P-Medianas (MLADENOVIC et al., 2007) e a segunda parte, focada no roteamento dos agrupamentos gerados. Seus experimentos sobre os trajetos originais implementados obtiveram uma melhoria de 6,15%, reduzindo em média 7.200 metros no trajeto total em um grupo de 774 pontos.

Já Steiner et al. (2009) elaboraram um método para otimizar a logística e dimensionamento 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.

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.

Diferente das abordagens mencionadas, o presente trabalho apresenta uma metodologia para a construção capacitada de unidades de leitura utilizando algoritmo genético e a técnica de agrupamento K-Means.

## 3 Fundamentação Teórica

Neste Capítulo são descritos os conceitos e técnicas importantes para o entedimento do método proposto. Inicialmente é necessário explicar o que é Agrupamento Capacitado, no qual seus conceitos são utilizados durante a criação dos grupos. Logo em seguida o algoritmo de agrupamento *K-Means*, usado na geração de grupos de etapas. Em seguida o algoritmo genético, para encontrar a melhor distribuição para o planejamento de leitura. E por fim, a técnica de agrupamento *RCMeans* para a criação de grupos capacitados.

### 3.1 Agrupamento Capacitado

Podemos definir o agrupamento capacitado como um problema que consiste em agrupar um determinado número de indivíduos em diferentes grupos com uma capacidade limitada. Esse problema pode ser visualizado da seguinte maneira: dado um conjunto de  $n$  indivíduos com seus respectivos pesos, é preciso agrupar todos os  $n$  indivíduos em  $p$  agrupamentos, onde a soma total dos pesos do agrupamento não seja maior que a capacidade total dos grupos, isto é, cada  $p$  deve respeitar um critério da capacidade total (RECK, 2010).

O problema do agrupamento capacitado é definido por Garey e Johnson (2002) como um problema combinatorial complexo *NP-Completo*, isto é, um problema que apresenta característica não-determinística e pode ser provado em tempo polinomial. É esperado que não existam métodos ou soluções exatas que lidem com este tipo de problema e a busca de uma solução eficiente se vê em métodos fundamentados na heurística.

Podemos definir o problema do agrupamento capacitado a partir do modelo de Ahmadi e Osman (2004):

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

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  $I$  representa um conjunto de indivíduos,  $I = \{1, 2, \dots, n\}$ , com uma matriz  $n \times m$  simétrica  $d_{ij}$ , que representa a distância entre um par de indivíduos, a qual se assume que  $d_{ij} > 0$  e que  $d_{ii} = 0, \forall i, j \in I$ . Atribui-se também um valor positivo para os pesos de cada indivíduos  $i$ , chamado de  $w_i$ . Sendo assim, define-se como  $J = \{1, 2, \dots, m\}$  o conjunto de grupos e  $W_j$  suas capacidades.

A mediana de um grupo  $j$  é representada por um indivíduo onde a soma das distâncias de todos os outros indivíduos do mesmo grupo é minimizada. Neste modelo, a capacidade de cada grupo do conjunto é assumida como igual.

Por fim, a variável  $y_j$  assume um valor binário, sendo 1 caso o indivíduo  $j$  seja atribuído como mediana do grupo  $j \in J$ , e 0 caso contrário. A variável  $x_{ij}$  assume um valor binário, onde 1 indica se o indivíduo pertence ao grupo  $j$  e 0 caso contrário.

A Equação 3.2 garante que apenas os  $p$  grupos serão selecionados. A Equação 3.3 garante que todos os indivíduos do conjunto  $I$  serão atribuídos a um único grupo. A Equação 3.4 garante que indivíduos que não foram selecionados como medianas sejam atribuídos. A Equação 3.5 garante que a soma dos pesos de cada indivíduo de um grupo não ultrapasse a capacidade máxima. Por fim, a Equação 3.6 garante a binaridade de  $x_{ij}$  e  $y_{i,j}$ .

## 3.2 *K-Means*

O método de agrupamento *K-Means* (HARTIGAN; WONG, 1979) é um algoritmo popular bastante usado para mineração de dados. Consiste em agrupar  $n$  amostra de dados em  $k$  grupos semelhantes, isto é, cada amostra observada será associada a grupo que possui características similares, diferente dos outros  $k$  grupos.

A similaridade entre os elementos pode ser definida como uma medida de distância. Quanto maior a proximidade maior será a similaridade, e quanto maior a distância, maior a diferença. Chama-se de distância **intra-grupo** a distância entre as amostras de um mesmo grupo e a distância **inter-grupo** a distância entre as amostras de um grupo com a de outros grupos.

Na maioria das aplicações do *K-Means* é comum o uso da distância euclidiana como medida de similaridade. Em um espaço bidimensional, é o comum o uso do teorema de Pitágoras, mostrada na Equação 3.7:

$$d((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2} \quad (3.7)$$

onde  $(x, y)$  e  $(a, b)$  representa dois pontos que assumem os valores das amostras. Essa Equação é usada para fornecer a distância de um segmento de reta entre os dois pontos (BORTOLOSSI, 2002).

Dentro de um espaço  $n$ -dimensional a distância euclidiana é dado a partir da Equação 3.8 :

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.8)$$

onde  $x, x = x_1, x_2, \dots, x_n$  e  $y, y = y_1, y_2, \dots, y_n$  (BORTOLOSSI, 2002).

Definidos os conceitos do *K-Means*, podemos descrever o seu algoritmo em 4 passos:

1. Seleciona-se  $k$  amostras que serão usadas como centroides iniciais de um grupo
2. Para cada amostra será selecionado um grupo com distância mais próxima
3. Ao termino das atribuições, será recalculado os centroides
4. Os passos 2 e 3 são repetidos até que se atinja um ponto de convergência, isto é, quando os centroides não mudaram desde a última alteração, ou qualquer outra condição de parada.

Podemos definir a complexidade do *K-Means* como *NP-Difícil*, porém metodologias heurísticas aplicadas ao método convergem rapidamente a um ótimo local. A complexidade do tempo por iteração varia de acordo com a quantidade de amostras da base, as dimensões das amostras e ao número de  $k$  grupos (BEKKERMAN; BILENKO; LANGFORD, 2011).

O *K-Means* apresenta uma limitação de ser sensível às amostras iniciais selecionadas como centroides. A escolha de centroides diferentes resultam em agrupamentos completamente distintos. Geralmente esses centroides são escolhidos aleatoriamente.

É necessário também que exista uma normalização dos dados amostrais antes que inicie o processo, garantindo que os valores sejam expressos em uma unidade de medida equivalente

Por fim, estimar a quantidade de  $k$  grupos é uma tarefa árdua quando não se possui conhecimento prévio sobre a base. Na literatura, se encontra diversas técnicas para encontrar a quantidade ideal de grupos sem o conhecimento do domínio das amostras. Estas técnicas tentam encontrar possíveis centroides que melhor representaram um agrupamento semelhante.

### 3.3 Algoritmo Genético

Os algoritmos genéticos (AG) são métodos que usam os princípios da **seleção natural** para a resolução de problemas, em especial para otimização. Um algoritmo genético se caracteriza por ser robusto, genérico e adaptável, bastante usada em diversas áreas.

É comum ao procurar uma solução otimizada para um problema defrontar-se com situações em que é preciso decidir quais restrições devem ser aplicadas, com a finalidade de extrair o maior número possível de benefícios. Situações reais podem ser exemplos práticos da busca de uma solução, como escolher um trajeto de ônibus que maximize o número de passageiro e minimize a distância percorrida

O processo de otimização pode se classificar em dois tipos:

- **Otimização Numérica:** que é solucionada a partir de funções matemáticas bem definidas.
- **Otimização Combinatória:** que tem como objetivo encontrar possíveis soluções encontrando a melhor combinação de recursos e suas características.

Seguindo os princípios da otimização combinatória e a seleção natural, um AG consiste em criar uma população de possíveis respostas para um determinado problema e a partir de diversas épocas, submete-las aos processo de evolução e mutação.

Para aplicar os conceitos de AG é importante conhecer as unidades fundamentais para o funcionamento: os **indivíduos**. Os indivíduos codificam possíveis soluções para um problema a ser resolvido e é através de sua ‘evolução’ que estas respostas são encontradas.

Um indivíduo, também conhecido como **genoma** ou **cromossomo** se caracterizam por possuir um **genótipo**, que corresponde aos **genes** do indivíduo, 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** que representa o quão bem é a resposta contida no indivíduo, determinada por uma avaliação (LUCAS, 2002).

Para a execução do algoritmo genético realiza-se as seguintes etapas:

- **Inicialização:** A inicialização é o passo em que se gera uma população de indivíduos, na maioria das vezes, aleatória. Esta população será submetida aos passos posteriores.
- **Avaliação:** Para cada solução existente, isto é, indivíduos da população serão avaliados o grau de aptidão para o problema em questão.



- **Seleção:** A seleção constitui em escolher os indivíduos aptos para a reprodução. Existem diversas técnicas de seleção, como *ranking*, roleta, torneio, uniforme, etc comumente relacionados ao seu grau de aptidão.
- **Cruzamento:** Corresponde à etapa de recombinar os indivíduos escolhidos com a finalidade de criar novos indivíduos
- **Mutação:** Para cada novo indivíduo resultante do cruzamento existe a possibilidade de ocorrerem alterações no seu genótipo, criando variedade a população.
- **Atualização:** A atualização se dá a medida que os novos indivíduos são inseridos na população criando-se assim uma nova geração. Nessa etapa também existem técnicas de atualização como o **elitismo**, no qual os melhores indivíduos de uma geração anterior são enviados para a nova geração.
- **Finalização:** A finalização corresponde a etapa para verificar as condições de parada da evolução. Pode ser determinado por um número de gerações, execuções ou qualquer outro critério definido. Caso a condição de parada não seja atingida a população é submetida novamente ao processo de evolução, caso contrário se encerra a execução do algoritmo e retorna-se a população que melhor atendeu à função de avaliação.

A Figura 1 representa a estrutura de funcionamento de um algoritmo genético tradicional.

### 3.4 *RCMeans*

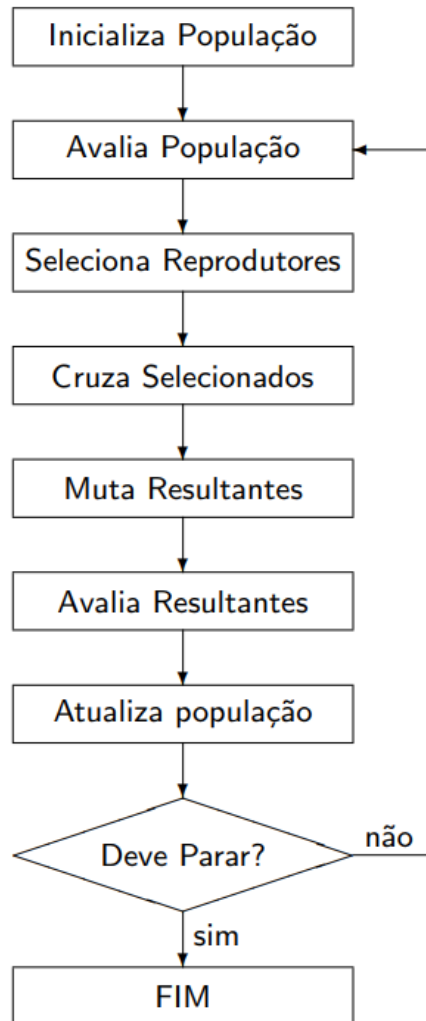
O *RCMeans* é um método proposto por Costa et al. (2018) para resolver o problema de agrupamento capacitado de Unidades de Leitura. A criação de agrupamentos é baseado no algoritmo *K-Means*, onde dados do conjunto são divididos em grupos de acordo com um critério de similaridade.

A Figura 2 exhibe como o algoritmo *RCMeans* funciona. O número inicial de grupos  $K$  é definido pelo total de pontos e na capacidade estabelecida. Para cada  $K$  grupo avalia-se a capacidade máxima e, se a capacidade for excedida junto a um valor de tolerância, os pontos de um grupo serão separados para um processo de agrupamento recursivo. Os novos grupos gerados também serão avaliados. As etapas de classificação, avaliação e divisão são realizadas recursivamente até que os limites de capacidades de todos os grupos estejam satisfeitos.

Este algoritmo garante duas características nos grupos criados:

1. Todos os pontos pertencem a um único grupo

Figura 1 – Estrutura de funcionamento de um Algoritmo Genético tradicional.

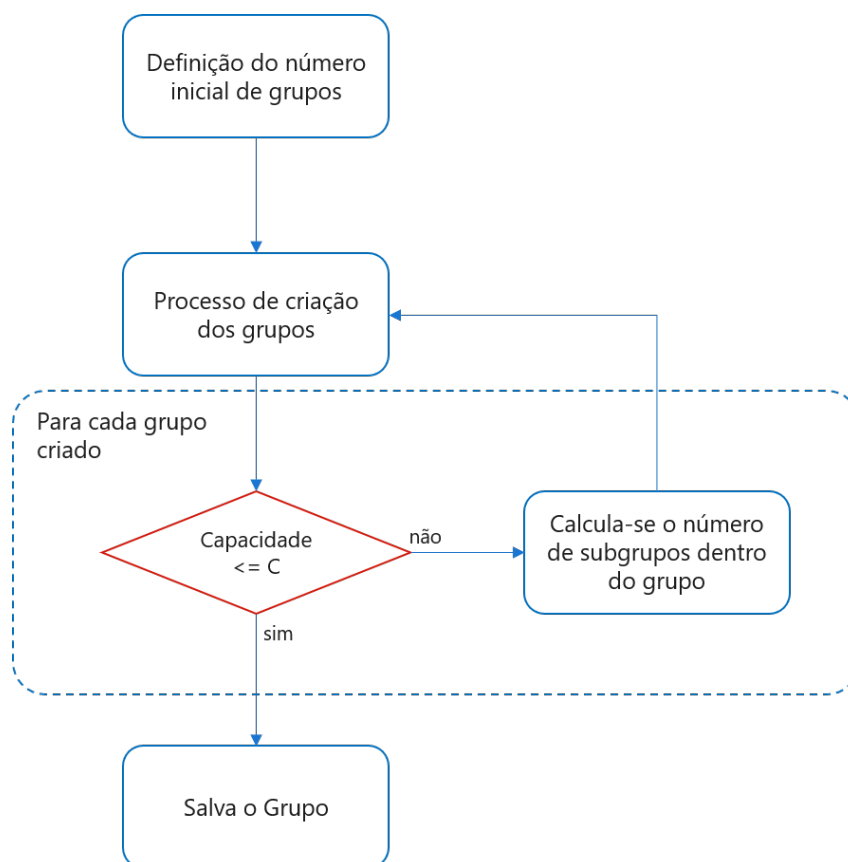


Fonte: Lucas (2002)

2. Todos os grupos satisfazem a capacidade desejada.

Para evitar a criação de grupos muito menores do que a capacidade total, o algoritmo realiza uma fusão de grupos com a finalidade de garantir mais homogeneidade entre os agrupamentos. Consiste em uma etapa de pós-processamento necessária para mesclar pequenos grupos para produzir outros grupos que ainda respeite os critérios de capacidade. Além da capacidade, este grupo deve ser mesclado a um outro agrupamento próximo para minimizar as alterações na capacidade.

Para mesclar os agrupamentos de acordo com a proximidade e capacidade usa-se um grafo conectado, no qual cada centroide do grupo é associado a um vértice e cada conexão com um grupo próximo é representado como uma aresta. O peso das arestas é dada pela distância euclidiana entre os centroides. O uso desta abordagem é baseada na

Figura 2 – Fluxo de Trabalho do *RCMeans*.

Adaptado de: [Costa et al. \(2018\)](#)

### Triangulação de Delaunay ([LEE](#); [SCHACHTER](#), 1980).

Após a construção do grafo, executa-se um algoritmo de busca *Breadth-First* (BFS). O BFS possui uma característica de descobrir os vértices vizinhos antes de avançar para um novo vértice buscado anteriormente. Devido a isso, durante o processo de busca, os vizinhos mais próximos de um vértice serão visitados primeiro e o método *RCMeans* tentará fundir os grupos adjacentes. Se a capacidade de um vértice não estiver satisfeita, o algoritmo irá executar a mesclagem com outros vértices até que soma dos pesos dos vizinhos ao peso do vértice não ultrapasse a capacidade. Esse pós-processamento resulta em menos grupos porém respeitando o limite de capacidade. A [Figura 3](#) mostra o fluxo de pós-processamento executado pelo *RCMeans*

O *RCMeans* é um algoritmo proposto para um problema de agrupamento capacitado que considera um tempo máximo de trabalho. Ele garante que grupos criados respeitem o critério de compacidade, isto é, grupos compactos e minimize a quantidade de grupos necessários dado uma quantidade de pontos, respeitando uma capacidade definida.

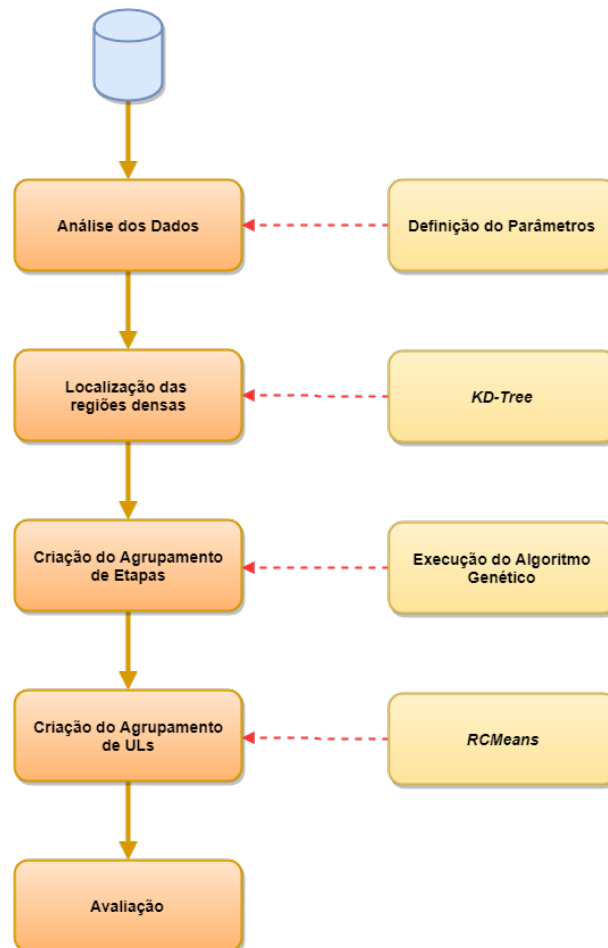
Figura 3 – Etapa de pós-processamento do *RCMeans*.Adaptado de: [Costa et al. \(2018\)](#)

## 4 Metodologia

Este capítulo descreve o método proposto para a criação das Etapas de leitura usando *K-means* baseado em algoritmo genético que concordem os critérios de homogeneidade e compacidade. Também será descrito no presente trabalho o tratamento dos dados e as métricas para avaliação.

### 4.1 Visão Geral

Figura 4 – Visão Geral do Método.



Fonte: Próprio Autor.

A Figura 4 apresenta as fases executadas para a execução do método. O primeiro passo corresponde a análise dos dados adquiridos e como eles são interpretados pelo método. Em seguida, é descrito como o algoritmo encontra as regiões mais densas populacionalmente, usando uma estrutura de dados chamada *KD-Tree*. Logo após, é apresentado o processo de execução do algoritmo genético junto ao *K-Means* que irá criar agrupamentos de Etapas.

Executa-se por fim o *RCMeans* sobre cada agrupamento de etapa para a criação de ULs. Ao final da execução é retornado um planejamento de leitura resultante de região que será usada para comparação e avaliação.

## 4.2 Aquisição dos dados

A base de dados usado no presente trabalho foi fornecida pela Equatorial Energia para o projeto P&D PD-00371-0031/2017. Os dados disponibilizados continham informações de unidades consumidoras, também denominada de instalação, que pertencem aos municípios de Açailândia e Imperatriz, do Maranhão e Santárem e Marabá, do Pará.

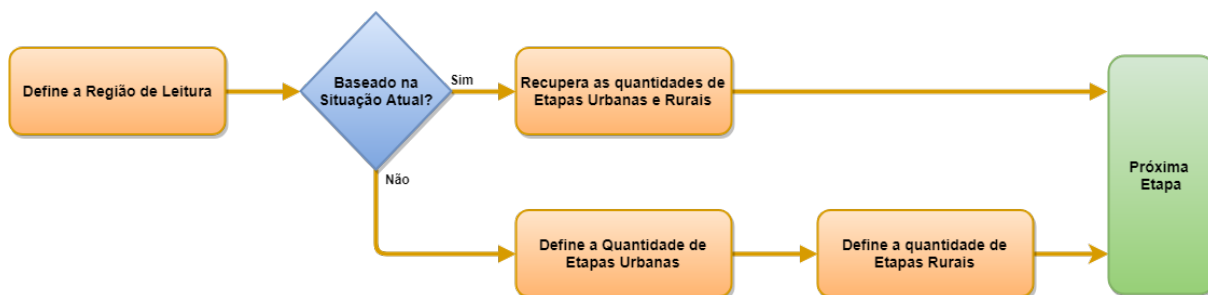
A Tabela 1 exibe uma amostra dos dados fornecidos. Cada instalação possui informações da sua região; latitude e longitude; a numeração da Etapa, representando o dia de leitura no planejamento vigente; o tempo de leitura da instalação, que informa em minutos o tempo médio que um funcionário leiturista consegue ler o registro de consumo; e se a unidade consumidora se encontra na área urbana ou rural da região, definidas pelas siglas U e R, respectivamente.

Região	Latitude(°)	Longitude(°)	Etapa	Tempo(min)	Tipo
MARABA	-5,3504077	-49,1375917	1	0,05416666666	U
MARABA	-5,6747134333	-49,80594031	6	4,2050720164	R
MARABA	-5,3418919	-49,0897745	8	2,20277777777	U
SANTAREM	-2,4338185	-54,7410037	2	0,07222222222	U
SANTAREM	-2,4336681978	-54,7405349581	2	0,03611111111	U
SANTAREM	-2,4825577966	-54,8196462633	9	2,23888888888	R
ACAILANDIA	-5,0600501	-47,5581488	27	0,01805555555	R
ACAILANDIA	-4,9506962666	-47,4888713266	10	0,03611111111	U
ACAILANDIA	-4,9494776	-47,4892488	10	0,05416666666	U
IMPERATRIZ	-5,4937074	-47,481681	10	0,03611111111	U
IMPERATRIZ	-5,4944809	-47,485428	10	0,07222222222	U
IMPERATRIZ	-5,4943815	-47,4830942	10	0,03611111111	U

Tabela 1 – Amostra dos dados de instalação fornecidos

A base fornecida contém os planejamentos de leitura vigente de uma região, isto é, qual agrupamento de Etapa cada unidade consumidora está atualmente inclusa, viabilizando a comparação dos resultados e sugestões de melhoria a partir da situação atual.

Figura 5 – Fluxo de Entrada do Método



Fonte: Próprio Autor.

## 4.3 Geração de Etapas

### 4.3.1 Reconhecimento da Situação

Para dar início ao método, se fez necessário uma análise dos dados recebidos pelas distribuidoras de energia elétrica para saber quais deveriam ser os parâmetros de entrada da estratégia de agrupamento proposta.

Foi observado que uma região-alvo possui dois planejamentos internos, um urbano e um rural, que são agrupados de maneira independente, ou seja, existe um agrupamento diferente tanto para instalações rurais quanto para urbanas.

Dependendo da situação organizacional da região, realizar uma otimização que alcance um resultado ótimo do planejamento de leitura pode resultar em um alto custo de implementação, devido o plano vigente anterior ser divergente ao plano gerado. Logo, o método proposto deve ser capaz de considerar tanto a situação atual quanto uma situação inédita.

A Figura 5 apresenta uma visão geral da entrada do método. Esses parâmetros de entrada do método são estabelecidos pela região de leitura alvo, que contém as instalações que serão usadas no planejamento. São eles: a quantidade de Etapas urbanas, que indica quantos grupos serão criados com as instalações urbanas; e a quantidade de Etapas rurais, que indica quantos grupos serão criados com as instalações rurais. Estes parâmetros foram estabelecidos com o propósito de permitir que planos de leitura sejam criadas a partir de um estado completamente novo ou a partir de um planejamento vigente pela companhia. Quando um plano vigente é considerado, a quantidade de Etapas urbanas e rurais correspondem a mesma quantidade da situação atual.

### 4.3.2 Localização das Regiões Densas

A partir dos parâmetros enviados, o próximo passo é definir os centroides para inicialização do método *K-Means*. Um centroide é definido como o melhor ponto que

representa o agrupamento ao redor e, na presente metodologia, correspondem à instalação que melhor representam a Etapa. O *K-Means* cria grupos usando pontos em torno de um centroide usando uma função de distância euclidiana e, devido à natureza desta abordagem, é comum que os agrupamentos gerados possuam forma semelhante à esférica. Porém uma má inicialização do método pode levar o agrupamento final a convergir a um mínimo local ou agrupamentos não proporcionais (JAIN, 2010).

Tendo em vista a organização do planejamento, quando uma instalação muda de Etapa é necessário avaliar uma série de restrições estabelecidas pela ANEEL. Entre estas, destaca-se o remanejamento ou alteração do período de leitura. Alterar as instalações inclusas em agrupamento de Etapa pode resultar em elevados custos de implementação, tornando inviável agrupamentos que diferem repentinamente da situação vigente.

Para minimizar os custos dos ajustes, é apropriado que instalações em regiões muito densas demograficamente sofram menos alterações do que as instalações que se encontram distantes de um centro urbano. Analisando este comportamento, as instalações que se encontrarem próximas as bordas do seu agrupamento de Etapa devem ser mais suscetíveis ao remanejamento para uma Etapa próxima. Essa característica deve ser mantida durante a criação do planejamento a fim de reduzir os gastos de remanejamento. A Figura 6 exemplifica um agrupamento com região densa: os pontos com cores mais frias indicam que os remanejamentos sobre estas instalações devem ser mínimos, diferente das cores mais quentes, por estarem próximas a borda, são mais remanejadas entre agrupamentos próximos.

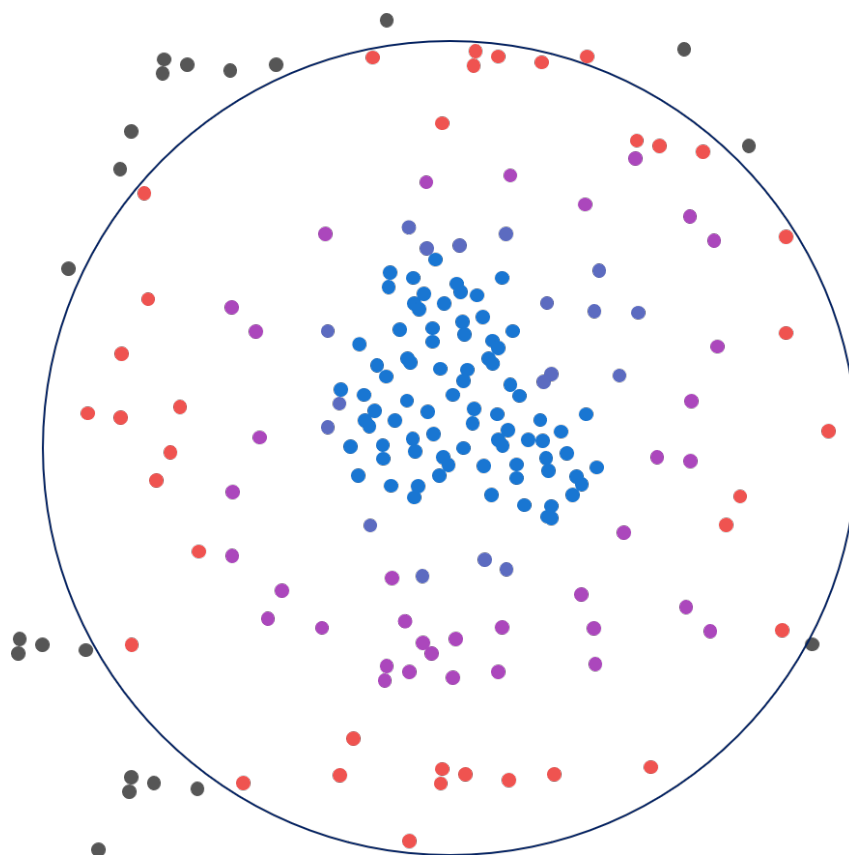
Logo, para a inicialização do *K-means* é viável enviar como entrada centroides aos quais as instalações próximas se encontrem nas áreas populacionais mais densas da região. Para esse fim, usaremos a estrutura de dados chamada de Árvore *k-d* (BENTLEY, 1975) que possui métodos otimizados para busca e exploração de porções do espaço.

Cada instalação é representada como um nó da árvore, cujas dimensões são representadas pela longitude e latitude. Neste caso, os pontos são retratados em um espaço bidimensional. A Figura 7 demonstra a distribuição das instalações no espaço.

Com a finalidade de encontrar as regiões mais densas, foi aplicada a seguinte estratégia: as instalações são separadas em urbanas e rurais e inseridas arbitrariamente como um novo nó de uma árvore *k-d* específica para cada tipo; pesquisa-se posteriormente todos os nós em que a distância euclidiana entre outros nós seja de 1000; para cada arranjo encontrado é conferido a quantidade de nós vizinhos; e por fim, os arranjos que tiverem os maiores valores representarão a área mais densa da região. Tais arranjos serão selecionados como possíveis centroides para inicialização do *K-Means*, sendo usada a quantidade de agrupamentos urbanos e rurais definidos no passo anterior. A Figura 8 exemplifica o processo de localização das regiões densas.

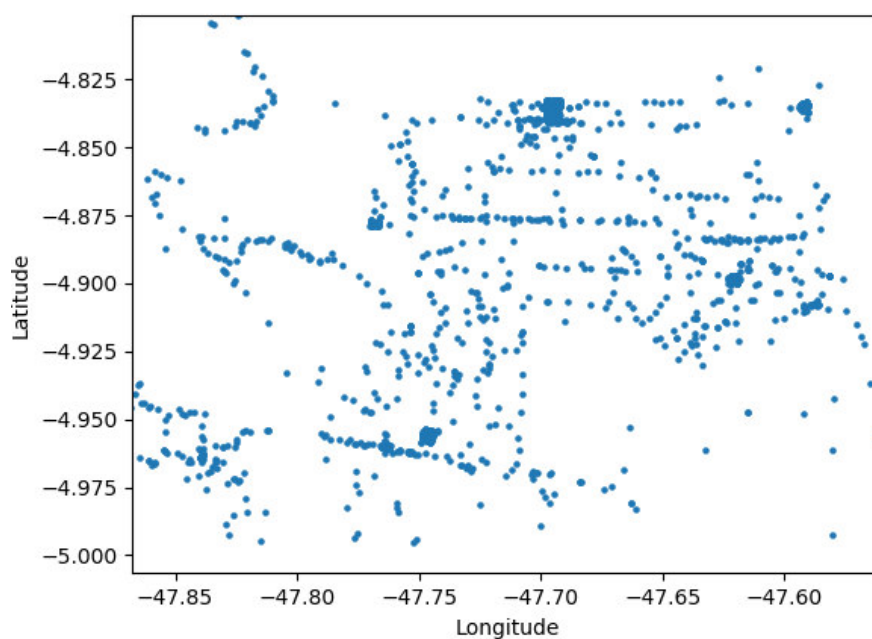


Figura 6 – Exemplo de agrupamento com destaque para a região densa



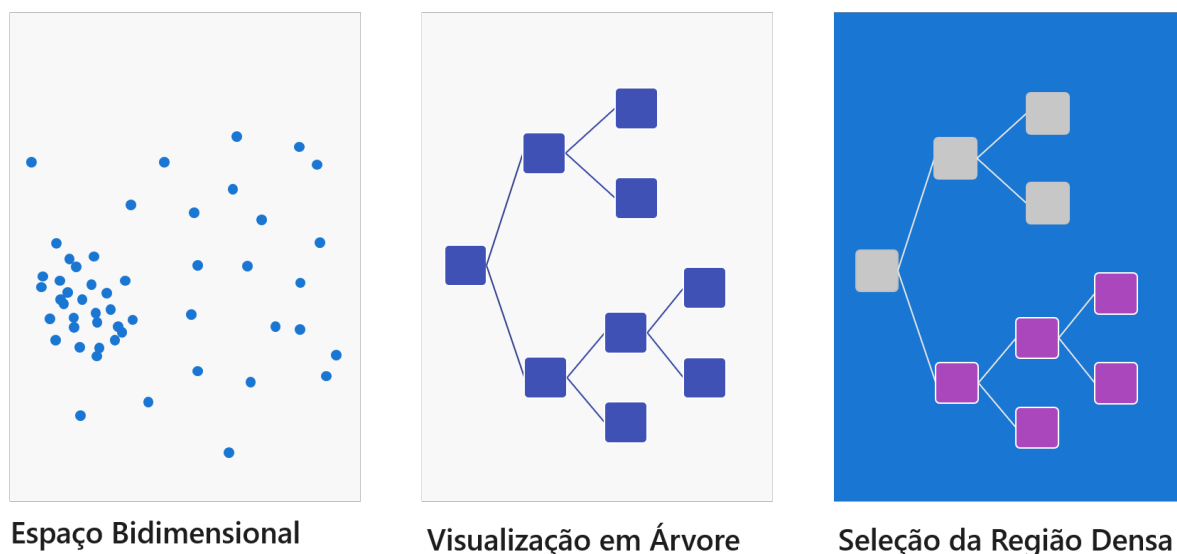
Fonte: Próprio Autor.

Figura 7 – Amostra das instalações distribuídas no espaço bidimensional.



Fonte: Próprio Autor.

Figura 8 – Esquema da execução do processo de localização das regiões densas.



Fonte: Próprio Autor.

### 4.3.3 Execução do Algoritmo Genético

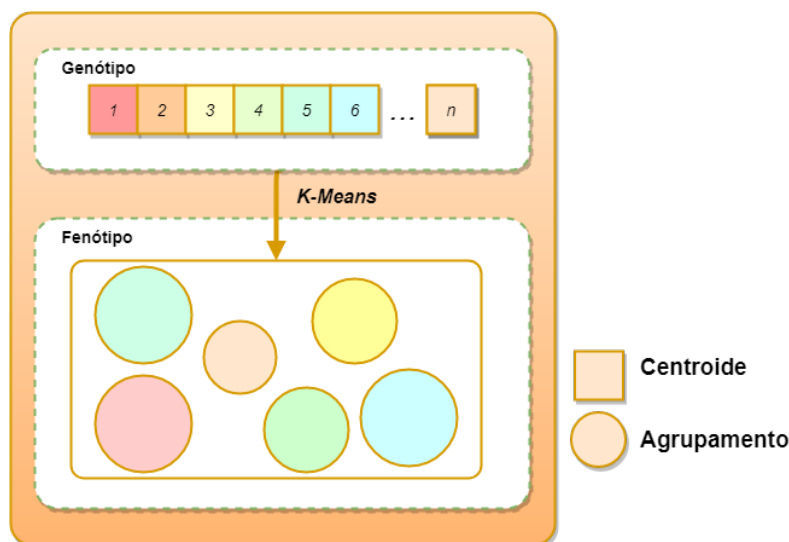
Tendo em mãos os parâmetros para inicialização do *K-Means*, ou seja, a quantidade de agrupamentos de Etapas que devem ser criadas para cada tipo de instalação, urbana e rural, junto com os arranjos de instalações das regiões densas que serão usadas como centroides viáveis, inicia-se preparação do algoritmo genético. Essa abordagem é executada duas vezes, uma execução focada apenas na região urbana, e a outra é voltada apenas a região rural.

Para o início do processo genético precisa-se, primeiramente, definir os indivíduos, a população, as gerações, a avaliação, a seleção, o cruzamento e a mutação aplicados ao contexto atual (LUCAS, 2002).

Cada indivíduo é formado por um planejamento de leitura que possui uma sequência diferente de centróides representando uma possível solução para o problema atual. Podemos definir como *genótipo* uma sequência de pontos usados como centroides para inicialização de um *K-Means* e o seu *fenótipo* como o agrupamento de Etapa resultante da clusterização. A Figura 9 exibe como um indivíduo é formado.

A população é definida como uma sequência de indivíduos que serão usados no processo de avaliação, seleção e cruzamento. Inicialmente a população é gerada a partir de um agrupamento inicial, no qual cada grupo incluso receberá um novo centroide viável, isto é, uma nova instalação de uma Etapa e pertencente a uma região densa será eleita como centroide de sua própria Etapa, que resultará em um novo agrupamento ao usar o algoritmo *K-means*. Se o planejamento esperado é baseado situação vigente, este plano é

Figura 9 – Constituição do indivíduo.



Fonte: Próprio Autor.

usado como agrupamento inicial. Porém, em uma situação inédita o agrupamento inicial enviado é uma clusterização simples do *K-Means* sobre as instalações enviadas, tendo como parâmetro a quantidade de Etapas definidas na parametrização.

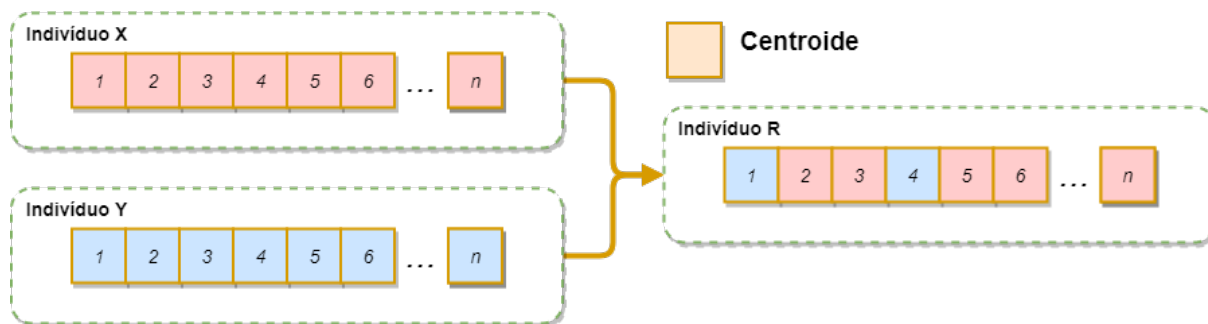
Para a inicialização do algoritmo genético será usada como geração inicial uma população de 300 indivíduos com diferentes características. Cada geração será submetida aos passos de seleção do algoritmo genético e para cada época, definida pela quantidade de execuções, será gerada uma nova população.

Em uma geração, cada indivíduo da população é submetido a uma avaliação para que seja estipulado seu grau de aptidão para a situação ótima. A função avaliadora, ou função objetiva, é determinada pela minimização do desvio padrão dos tempos de leitura entre os agrupamentos de Etapa. Usar o desvio padrão permite observar a variabilidade do tempo total em cada Etapa e a situação do plano gerado está próximo do resultado ideal. A minimização irá privilegiar planos homogêneos, ou seja, planos que possuam Etapas com tempos mais homogêneos entre si do que planos heterogêneos, isto é, planos com tempos mais dispersos (LUNET; SEVERO; BARROS, 2006).

Após a avaliação dos indivíduos da geração é realizada a seleção para o cruzamento. Para escolher os indivíduos mais aptos para cruzamento, usa-se uma seleção por *ranking*, em que cada indivíduo é ordenado de acordo com seu grau de aptidão e suas chances de seleção variam conforme a posição ocupada na sua classificação.

Visto que os indivíduos foram classificados com uma chance de cruzamento na seleção posterior, inicia-se o processo de reprodução. Durante o cruzamento, são selecionados os *genótipos* de dois indivíduos e realiza-se um sorteio da quantidade de genes que serão transmitidos de cada indivíduo. O novo indivíduo gerado possuirá outros genes herdados

Figura 10 – Esquema de cruzamento dos indivíduos.



Fonte: Próprio Autor.

do primeiro e segundo indivíduo, o que resultará em novo *fenótipo*, isto é, um novo agrupamento de Etapas. Esse passo é exibido na Figura 10

Após o cruzamento, existe uma probabilidade de mutação. A mutação consiste em alterar os genes dos indivíduos gerados no passo de reprodução. Um gene, representado por um ponto de uma região densa, é trocado aleatoriamente por outro ponto do arranjo de regiões densas, permitindo que outras instalações encontradas participem do conjunto solução para o melhor centróide de um agrupamento. Assim, define-se que a probabilidade de mutação usada no algoritmo é de 10%.

Ao término da mutação, os indivíduos resultantes são inseridos na população, dando início a uma nova geração. Para prevenir a perda dos melhores indivíduos da população anterior, é usada a técnica de elitismo. Essa técnica consiste em enviar os melhores indivíduos da população anterior para a nova população prevenindo a perda das melhores solução de agrupamento encontradas. Com este intuito, é usado o elitismo para enviar apenas o melhor indivíduo da população anterior para a nova geração.

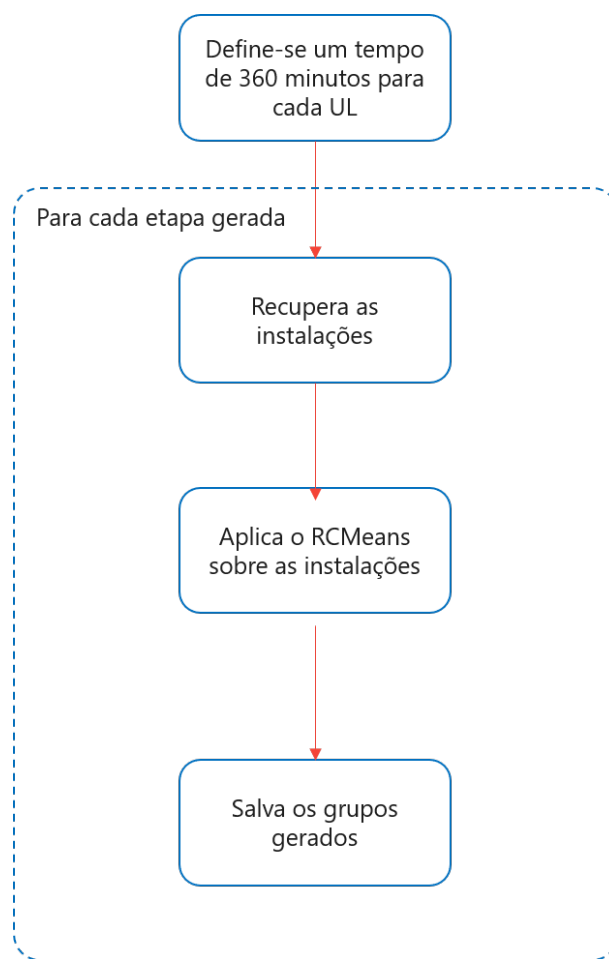
A condição de parada para o algoritmo genético é um total de 100 gerações. Ao usar um número baixo de gerações, aumentavam-se as chances de resultar em indivíduos com agrupamentos heterogêneos, enquanto um alto número de gerações resultava em uma grande demora de execução do processo completo.

Ao final o melhor agrupamento de Etapas será o melhor indivíduo da última população gerada. Este planejamento proverá Etapas com tempos de trabalho mais homogêneos, isto é, com uma menor variedade de tempos em cada Etapa.

#### 4.3.4 Geração de Unidades de Leitura

As distribuidoras de energia elétrica organizam dentro das Etapas os agrupamentos de Unidades de Leitura, nos quais um funcionário leiturista fica responsável pela medição e entrega das faturas.

Figura 11 – Fluxo de Execução do RCMMeans.



Fonte: Próprio Autor.

Para respeitar a carga de trabalho de um leiturista é importante que sejam criados grupos capacitados com um tempo de trabalho aceitável ao funcionário.

Para definir o agrupamento de Unidades de Leituras em uma Etapa do planejamento é usado o método chamado *RCMeans* (COSTA et al., 2018). Este método foi utilizada para a construção de ULs capacitadas respeitando os critérios de homogeneidade e compacidade. A Figura 11 mostra o fluxo de execução utilizado pela metodologia.

O algoritmo *RCMeans* pode ser categorizado como um método de aprendizagem não-supervisionada baseada no *K-Means*. São enviados como parâmetros de inicialização as instalações pertencentes a uma Etapa feita pelos passos anteriores. Estas instalações são divididas em grupos de acordo com um critério de similaridade.

O *RCMeans* cria grupos que respeitem a capacidade máxima de 360 minutos de trabalho por agrupamento de Unidades de Leituras. Uma outra característica é que esse algoritmo minimize a heterogeneidade dos grupos criados, isto é, não cria agrupamentos com uma grande diferença de tempos entre si.

## 4.4 Métricas de Avaliação

Com o propósito de avaliar o método proposto são usados dois tipos de métrica para justificar os benefícios do algoritmo diante da situação real das distribuidoras. São elas: a minimização do desvio padrão, usada para justificar o critério de homogeneidade entre os tempos de leitura; e o coeficiente de Silhueta, usada para justificar a compacidade do agrupamentos.

### 4.4.1 Minimização do Desvio do Padrão

Para avaliar o critério de homogeneidade dos tempos dos agrupamentos gerados é usado a minimização do desvio padrão. Ao verificarmos o desvio padrão é possível analisar se as Etapas estão balanceadas, isto é, quanto maior o desvio padrão, maior é a variabilidade dos tempos entre as Etapas, e quanto menor o desvio, menor é a variabilidade e por consequência, mais homogeneidade.

Para obter os valores necessários para o cálculo do desvio padrão são realizados os seguintes passos: para cada instalação de uma Etapa é recuperado seu tempo de trabalho; cada tempo recuperado é somado aos tempos de instalações da mesma Etapa; o somatório total do tempo é associado a cada Etapa gerada.

Com o tempo total de cada Etapa é possível aplicar a fórmula do desvio padrão sobre os tempos das Etapas, representada pela Equação 4.1

$$Desvio = \sqrt{\frac{\sum_{i=1}^n (Tempos_i - M_a)^2}{n}} \quad (4.1)$$

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

Ao aplicar a fórmula sobre um agrupamento, é possível verificar a variabilidade dos tempos entre as Etapas. Um valor próximo de 0 corresponderá a um desvio menor entre os tempos, e portanto, Etapas mais homogêneas.

### 4.4.2 Coeficiente de Silhueta

Para avaliar o critério de compacidade dos agrupamentos é usado o Coeficiente de Silhueta. Esta medida é usada para avaliar cada agrupamento individualmente.

Para calcular o coeficiente de silhueta de um grupo é selecionado uma instalação aleatória  $i$ , calcula-se a distância média desta 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 podemos expressar a Equação 4.2 como o coeficiente de silhueta

$$C_S = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (4.2)$$

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

A fórmula descrita retorna valores que variam de -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, todos os grupos que tenha um coeficiente de silhueta positivo.

## 5 Resultados

Para a validação da metodologia foram realizados testes nas bases de quatro municípios enviados pela Equatorial Energia: Açailândia e Imperatriz, do Maranhão; e Santárem e Marabá, do Pará. Para a visualização das grupos de instalações foi utilizada a biblioteca *OpenLayers*.

Foram construídos dois tipos de testes: o primeiro tipo não leva em consideração a situação vigente da distribuidora, isto é, cria agrupamentos a partir de uma condição inédita; o segundo tipo teste, por sua vez, considera a situação vigente da companhia, que é levada em consideração durante a criação dos grupos

### 5.1 Análise da Situação Vigente

Para comparar os resultados obtidos, é necessário tirar as métricas de cada agrupamento vigente na base. As métricas obtidas serão usadas para mostrar a eficiência do uso da metodologia aplicada.

Para o agrupamento vigente do município de Açailândia foram extraídas todas as instalações inclusas na região, totalizando um total de 44.371 instalações, com o tempo total de 20.055,77082 minutos (aprox. 334 horas), sendo distintas 17 etapas urbanas e 6 etapas rurais. A Tabela 2 representa os dados obtidos da base.

	Área Urbana	Área Rural
Quantidade de Etapas	17	6
Quantidade de Instalações	39.196	5.175
Média do Tempo por Etapa (min)	642,6301813	1.521,8429559
Desvio Padrão por Etapa (min)	196,9575918	764,5610848
Tempo Mínimo de Etapa (min)	380,7263929	646,2291666
Tempo Máximo de Etapa (min)	1.049,580581	2.855,996327
Coefficiente de Silhueta	0,0905402	0,1297999

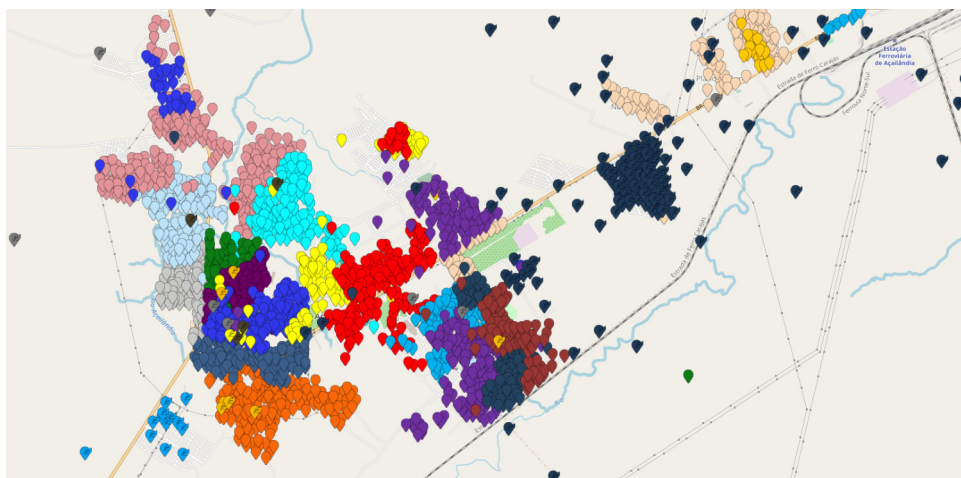
Tabela 2 – Dados resumidos da situação atual de Açailândia

A Figura 12 mostra a divisão dos agrupamentos vigentes da Região de Açailândia, onde cada cor é representada por um grupo de etapa. Os agrupamentos rurais são exibidos com um símbolo de folha.

Para a região de Imperatriz foram verificadas 117.592 instalações, com o tempo total de 46.758,1433767 minutos (aprox. 779 horas), contendo 19 etapas urbanas e 5 etapas rurais. A Tabela 3 representa os dados do planejamento de Imperatriz.



Figura 12 – Visualização do Planejamento Vigente de Acailândia



Fonte: Próprio Autor.

	Área Urbana	Área Rural
Quantidade de Etapas	19	5
Quantidade de Instalações	114.180	3.412
Média do Tempo por Etapa (min)	2.242,110893	831,6072825
Desvio Padrão por Etapa (min)	923,1528967	7.799,0397408
Tempo Mínimo de Etapa (min)	0,0541667	8,7027778
Tempo Máximo de Etapa (min)	3.687,210453	1.744,079769
Coefficiente de Silhueta	0,2145923	0,3654557

Tabela 3 – Dados resumidos da situação atual de Imperatriz

A Figura 13 mostra a divisão dos agrupamentos da Região de Imperatriz.

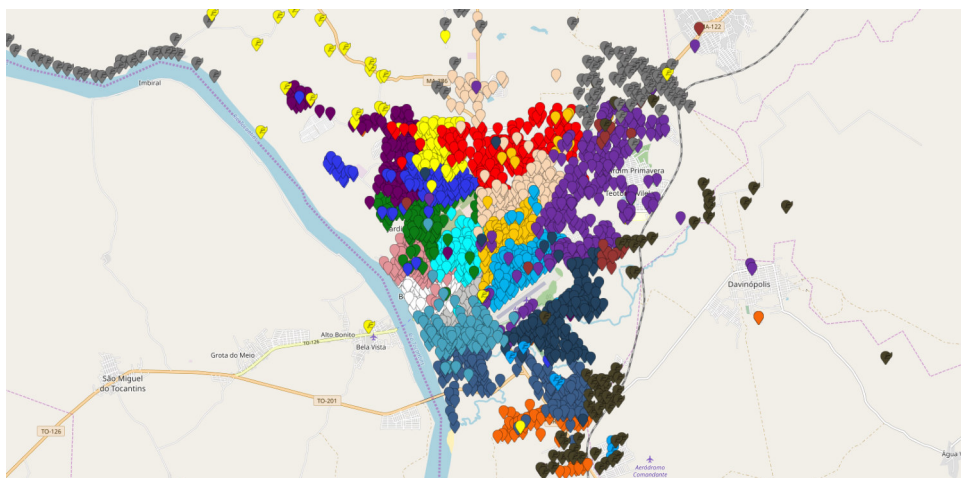
Em Marabá, foram analisadas 98.978 instalações distribuídas em 18 etapas urbanas e 18 etapas rurais, com o tempo total de 120.909,412206 minutos (aprox. 2015 horas). A Tabela 4 exhibe os dados resumidos do planejamento da região.

	Área Urbana	Área Rural
Quantidade de Etapas	18	18
Quantidade de Instalações	86.540	12.438
Média do Tempo por Etapa (min)	5.020,1603342	1.697,0292328
Desvio Padrão por Etapa (min)	1.775,7785675	681,1453408
Tempo Mínimo de Etapa (min)	2.729,7233749	552,9888889
Tempo Máximo de Etapa (min)	9.836,1901899	2.901,1216476
Coefficiente de Silhueta	-0,1254141	-0,2862414

Tabela 4 – Dados resumidos da situação atual de Marabá

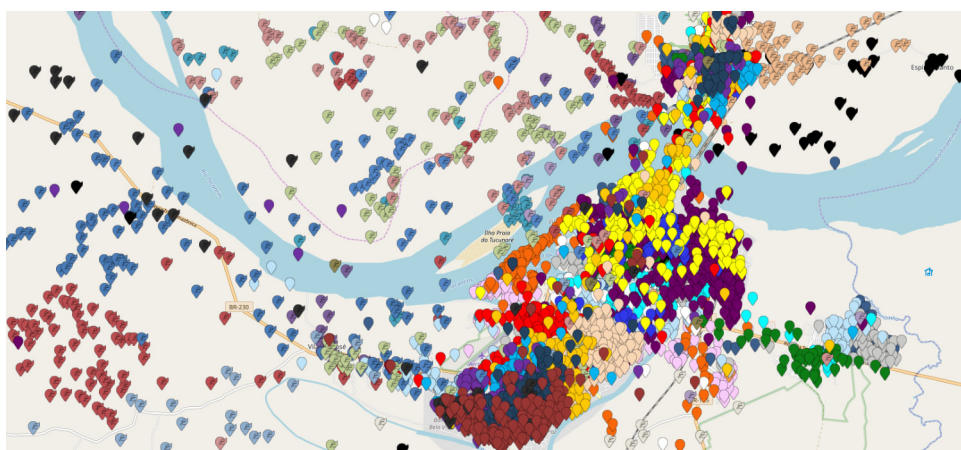
A Figura 14 mostra a divisão dos agrupamentos da Região de Marabá.

Figura 13 – Visualização do Planejamento Vigente de Imperatriz



Fonte: Próprio Autor.

Figura 14 – Visualização do Planejamento Vigente de Marabá



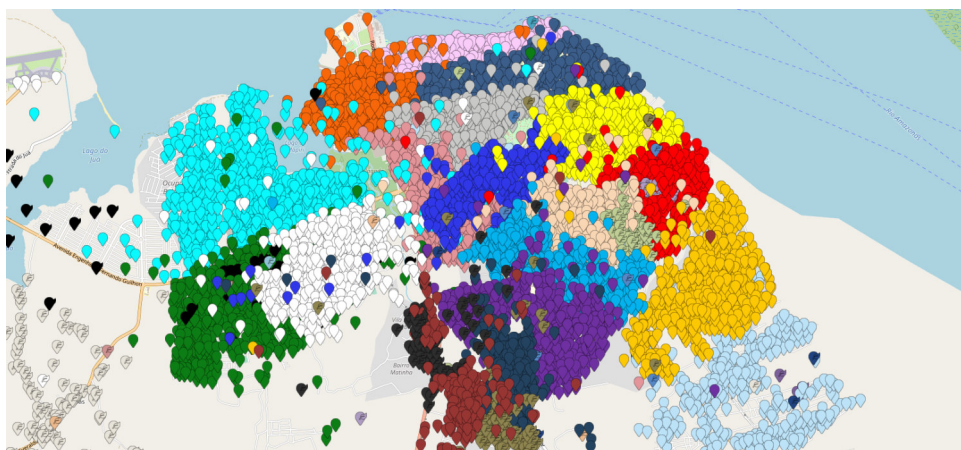
Fonte: Próprio Autor.

Por último, foi analisada a região de Santarém, contendo 10.6771 instalações distribuídas em 18 etapas urbanas e 18 etapas rurais, com o tempo total de 108.976,2143 minutos (aprox. 1816 horas). A Tabela 5 exibe os dados resumidos da região.

	Área Urbana	Área Rural
Quantidade de Etapas	18	18
Quantidade de Instalações	84.867	21.904
Média do Tempo por Etapa (min)	4.455,5537492	1.598,6803785
Desvio Padrão por Etapa (min)	931,9544986	388,8194972
Tempo Mínimo de Etapa (min)	2.777,8802095	967,3836039
Tempo Máximo de Etapa (min)	6.072,3362624	2.202,8608073
Coefficiente de Silhueta	-0,0709605	-0,1483548

Tabela 5 – Dados resumidos da situação atual de Santarém

Figura 15 – Visualização do Planejamento Vigente de Santarém



Fonte: Próprio Autor.

A Figura 15 mostra a divisão dos agrupamentos da Região de Santarém.

Com os dados obtidos, é possível comparar os resultados obtidos pelo método com a situação atual de cada região. Nas próximas seções serão realizados comparativos dos planejamentos gerados.

## 5.2 Agrupamento a partir de uma Situação Inédita

Seguindo a ordem em que os municípios foram apresentados serão exibidos os dados comparativos de cada região. Cada agrupamento gerado não considerou a situação vigente de seu município, isto é, incializado a partir de uma situação inédita. Para manter o quesito de comparação, foram mantidas as mesmas quantidades de etapas urbana e rurais do planejamento vigente.

	Planejamento Vigente		Planejamento Inédito	
	Área Urbana	Área Rural	Área Urbana	Área Rural
Desvio Padrão	196,9575918	764,5610848	324,1709961	115,1066925
Tempo Mínimo (min)	380,7263929	646,2291666	3,25	1.373,8236368
Tempo Máximo (min)	1.049,580581	2.855,996327	981,4977053	1.657,5995585
Coefficiente de Silhueta	0,0905402	0,1297999	0,3700692	0,3457767

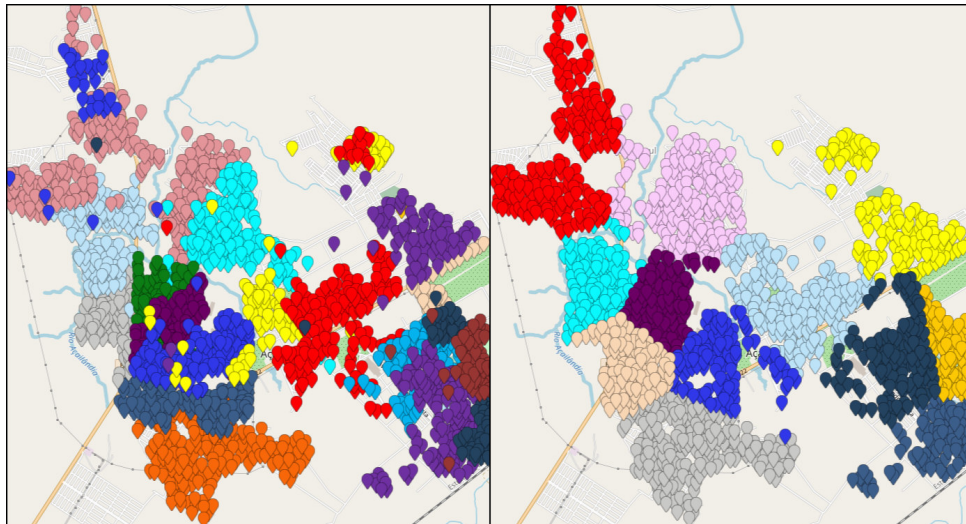
Tabela 6 – Dados comparativos do planejamento vigente com o planejamento inédito do município de Açailândia

A Tabela 6 apresenta os dados da situação vigente de Açailândia com o retorno da metodologia aplicada sobre a mesma região. Ao comparar os resultados, nota-se que os agrupamentos urbanos do planejamento gerado tiveram um aumento do desvio padrão. Analisando o menor tempo do agrupamento de etapa encontra-se um grupo com 3,25 minutos, o que ocasionou um aumento do desvio. Por sua vez nos agrupamentos rurais



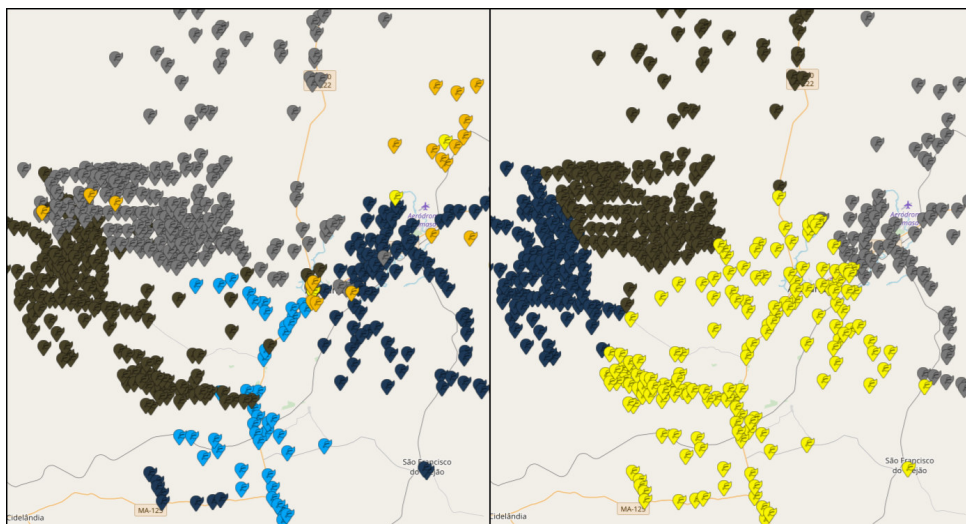
ocorreu uma grande minimização do desvio padrão comparada ao planejamento vigente. Nota-se também que o maior e o menor tempo de entre agrupamentos estão próximos, isto é, balanceados. Além disso, ocorreu um aumento do coeficiente de silhueta comparado a situação atual. As Figuras 16 e 17 exibem uma comparação do planejamento vigente com o planejamento resultante.

Figura 16 – Comparativos dos Planejamentos Urbanos Vigente com o Inédito de Acailândia



Fonte: Próprio Autor.

Figura 17 – Comparativos dos Planejamentos Rurais Vigente com o Inédito de Acailândia



Fonte: Próprio Autor.

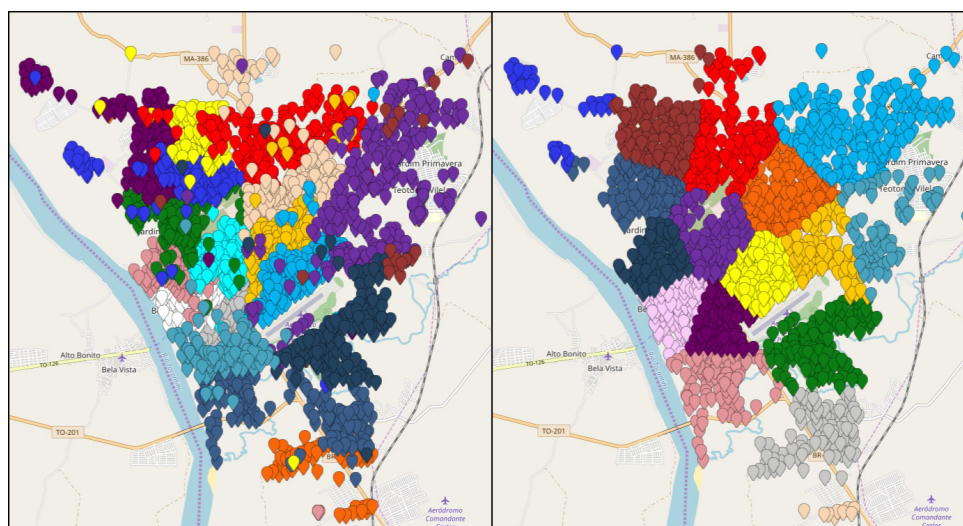
A Tabela 7 apresenta os dados obtidos de um planejamento inédito de Imperatriz. Nota-se que os agrupamentos urbanos gerados tiveram um aumento do desvio padrão, com uma leve redução do tempo máximo entre etapas. Nos agrupamentos rurais, obteve-se uma significativa melhora na minimização do desvio padrão, tornando os tempos entre as etapas

mais homogêneas. Em ambos os casos o coeficiente silhueta aumentou, demonstrando a criação de grupos compactos. As Figuras 18 e 19 exibem uma comparação entre os planejamentos.

	Planejamento Vigente		Planejamento Inédito	
	Área Urbana	Área Rural	Área Urbana	Área Rural
Desvio Padrão	923,1528967	799,0397408	1.060,3184939	19,1710431
Tempo Mínimo (min)	0,0541667	8,7027778	0,0180556	1.024,930168
Tempo Máximo (min)	3.687,210453	1.744,0797692	3.597,6383276	1.067,6314733
Coeficiente de Silhueta	0,2145923	0,3654557	0,3967004	0,5421482

Tabela 7 – Dados comparativos do planejamento vigente com o planejamento inédito do município de Imperatriz

Figura 18 – Comparativos dos Planejamentos Urbanos Vigente com o Inédito de Imperatriz



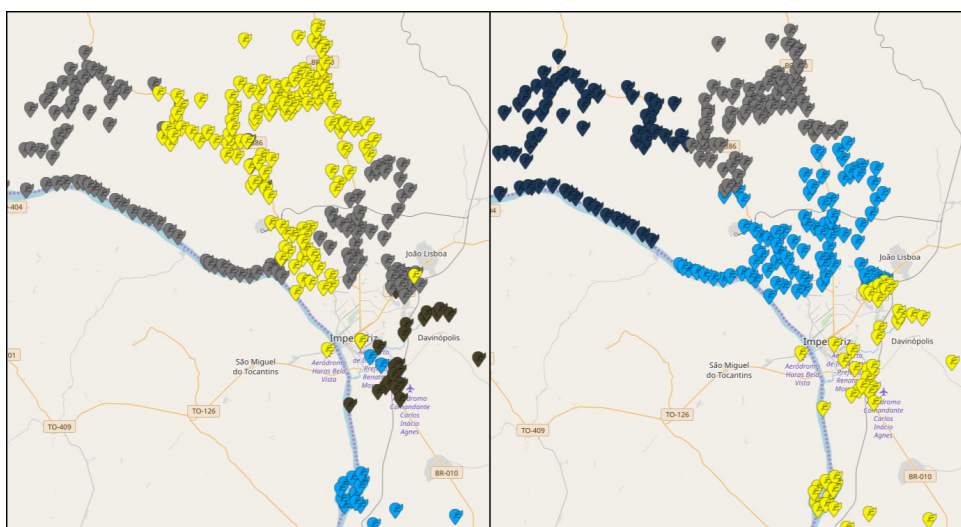
Fonte: Próprio Autor.

	Planejamento Vigente		Planejamento Inédito	
	Área Urbana	Área Rural	Área Urbana	Área Rural
Desvio Padrão	1.775,7785675	681,1453408	5.652,9355723	784,2826255
Tempo Mínimo (min)	2.729,7233749	552,9888889	0,0541667	0,9819444
Tempo Máximo (min)	9.836,1901899	2.901,1216476	15.063,4882053	446,9260096
Coeficiente de Silhueta	-0,1254141	-0,2862414	0,4855453	0,3664483

Tabela 8 – Dados comparativos do planejamento vigente com o planejamento inédito do município de Marabá

Para o município de Marabá foram obtidos os seguintes resultados exibidos na Tabela 8. Em ambos os agrupamentos, urbanos e rurais, ocorreu um aumento do desvio padrão. Ocorreu aumento do tempo de leitura máximo urbano, enquanto no agrupamento rural houve um grande decréscimo na comparação. Nesta análise, destaca-se o coeficiente

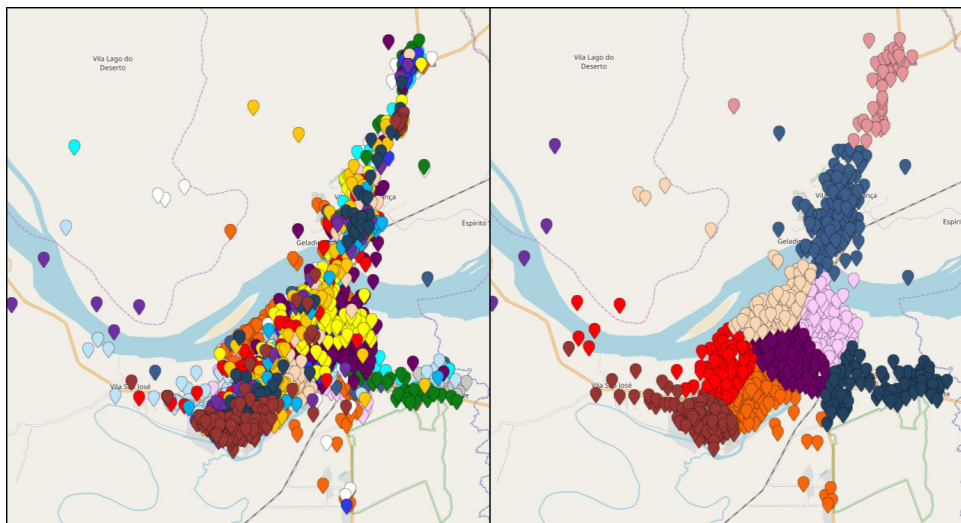
Figura 19 – Comparativos dos Planejamentos Rurais Vigente com o Inédito de Imperatriz



Fonte: Próprio Autor.

de silhueta, que teve um aumento considerável comparado ao planejamento vigente. As Figuras 20 e 21 exibem uma comparação entre os planejamentos.

Figura 20 – Comparativos dos Planejamentos Urbanos Vigente com o Inédito de Marabá



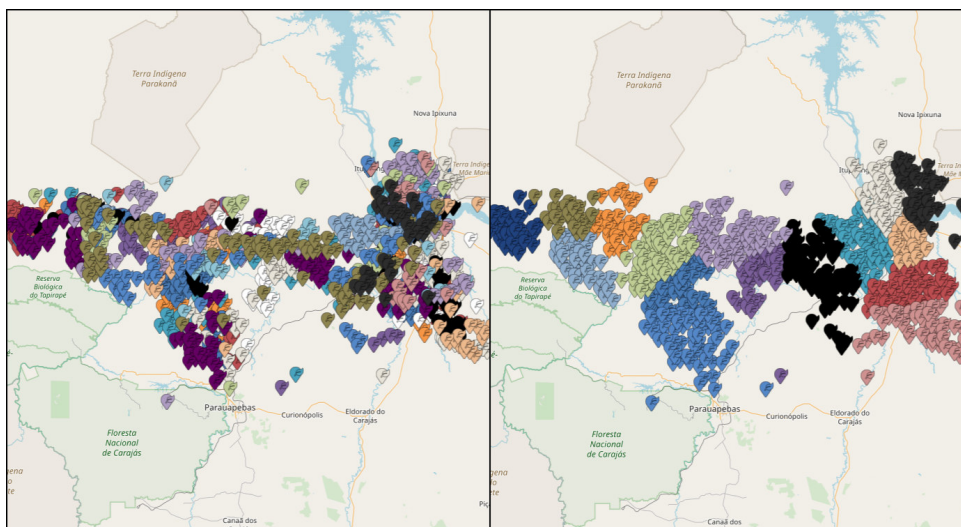
Fonte: Próprio Autor.

Por fim, na região de Santarém, os seguintes resultados são constatados na Tabela 9. No planejamento gerado, o desvio padrão aumentou de maneira significativa dado o aumento nos tempos mínimos dos agrupamentos urbanos e rurais. Também ocorreu um aumento também no tempo máximo, o que aumentou o valor do desvio padrão. Assim como os planos anteriores, o coeficiente de silhueta aumentou bastante, indicando grupos compactos. As Figuras 22 e 23 mostra os comparativos dos agrupamentos.

A comparação entre os agrupamentos gerados com os vigentes da companhia não é



Figura 21 – Comparativos dos Planejamentos Rurais Vigente com o Inédito de Marabá

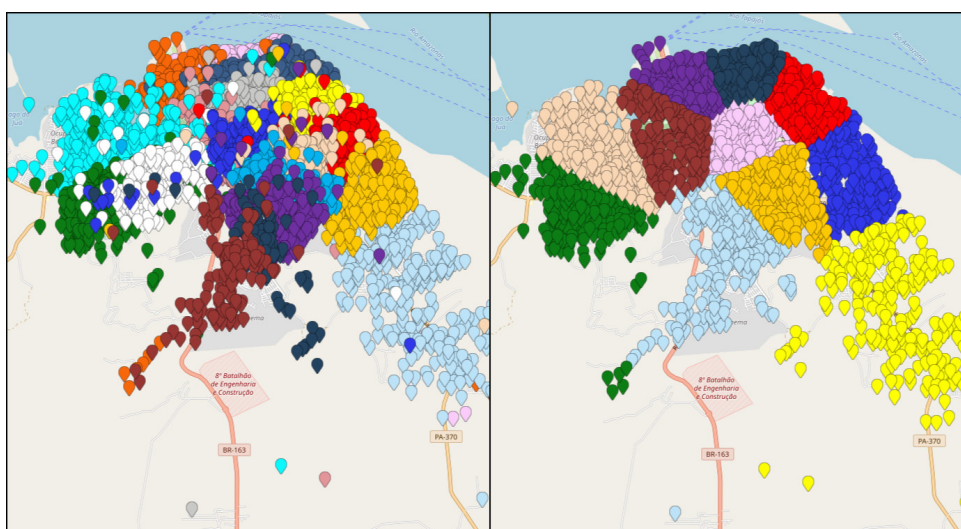


Fonte: Próprio Autor.

	Planejamento Vigente		Planejamento Inédito	
	Área Urbana	Área Rural	Área Urbana	Área Rural
Desvio Padrão	931,9544986	388,8194972	3.404,3552646	1.759,4789166
Tempo Mínimo (min)	2.777,8802095	967,3836039	146,046819	12,0430556
Tempo Máximo (min)	6.072,3362624	2.202,8608073	8.422,1719782	5.024,1928772
Coefficiente de Silhueta	-0,0709605	-0,1483548	0,4844796	0,4233167

Tabela 9 – Dados comparativos do planejamento vigente com o planejamento inédito do município de Santarém

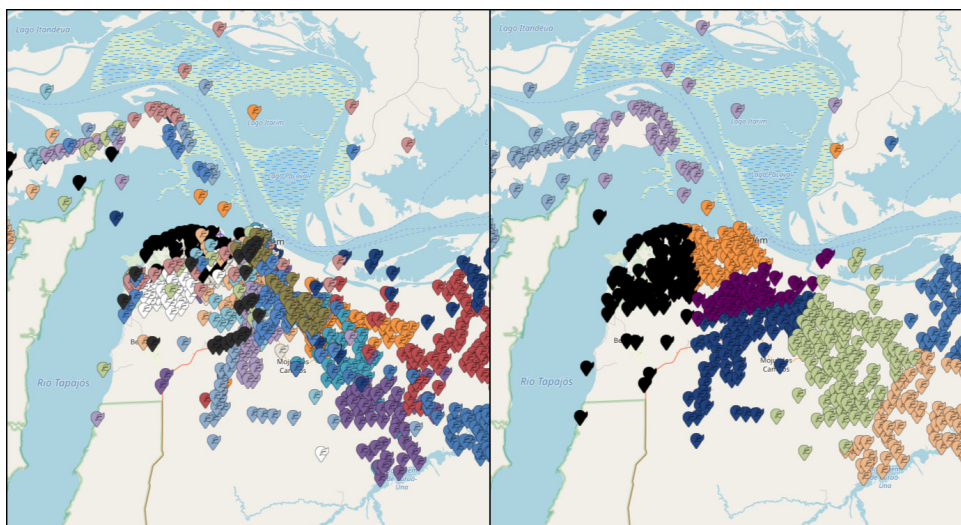
Figura 22 – Comparativos dos Planejamentos Urbanos Vigente com o Inédito de Santarém



Fonte: Próprio Autor.

totalmente justa, pois um planejamento vigente é um agrupamento que obteve uma série de modificações ao longo do tempo, feita por especialistas que possuem compreensão geográfica

Figura 23 – Comparativos dos Planejamentos Rurais Vigente com o Inédito de Santarém



Fonte: Próprio Autor.

da área, enquanto o planejamento gerado não leva em consideração o conhecimento da região, desfavorecendo o uso do método proposto. Outra limitação é a realização do agrupamento inicial pelo *K-Means*, que por sua vez, é sensível a *outliers*, isto é, agrupa instalações muito distantes de outros grupos, o que explica agrupamentos de etapas com tempos pequenos, pois estas instalações estão distantes das regiões mais populosas. A comparação não leva em consideração o tempo necessário para um leitorista chegar do centro urbano até a uma etapa distante.

Por outro lado, ao focar no coeficiente de silhueta, é perceptível que os grupos gerados pela abordagem são superiores ao planejamento vigente no quesito compacidade. Isto se deve também à inicialização baseado no *K-Means*, que pela natureza do algoritmo tende a criar grupos esféricos e próximos.

A vantagem de usar a abordagem inédita é permitir que regiões que não possuem um planejamento de leitura possam ser criadas já com o foco na compacidade dos agrupamentos. Além disso, estes planos inéditos podem ser usados como inicialização para os próximos planejamentos, o que faz com que os agrupamentos se tornem mais otimizados do que os gerados inicialmente.



### 5.3 Agrupamento a partir da Situação Vigente

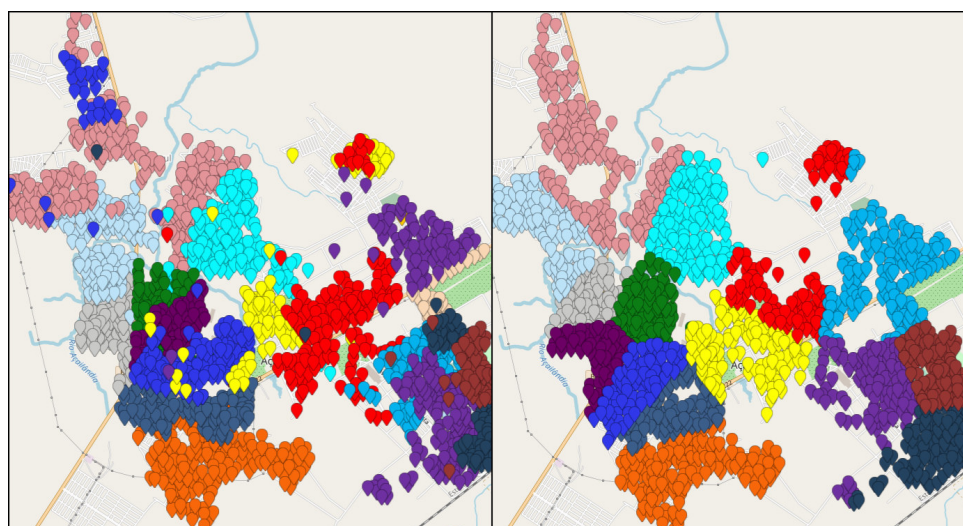
Serão analisados nesta Seção os agrupamentos gerados a partir de uma situação vigente. Para facilitar o discernimento, será chamado de planejamento baseado todo agrupamento que teve como base uma situação anterior.

A Tabela 10 mostra a comparação do planejamento vigente com o planejamento baseado da região de Açailândia. Observa-se uma grande redução do desvio padrão tanto para os os agrupamentos urbanos quanto para os rurais. Os tempos mínimos e máximos da área urbana e rural se tornarem mais próximos, fazendo jus ao desvio padrão. Observa-se também um aumento no coeficiente de silhueta, o que promove a compacidade do agrupamentos gerados. As Figuras 24 e 25 exibem comparativos entre o agrupamento urbano e rural.

	Planejamento Vigente		Planejamento Baseado	
	Área Urbana	Área Rural	Área Urbana	Área Rural
Desvio Padrão	196,9575918	764,5610848	61,9885289	115,1066925
Tempo Mínimo (min)	380,7263929	646,2291666	518,5597262	1.074,9559028
Tempo Máximo (min)	1.049,580581	2.855,996327	712,1148593	1.696,8846213
Coeficiente de Silhueta	0,0905402	0,1297999	0,3191258	0,5892947

Tabela 10 – Dados comparativos do planejamento vigente com o planejamento baseado do município de Açailândia

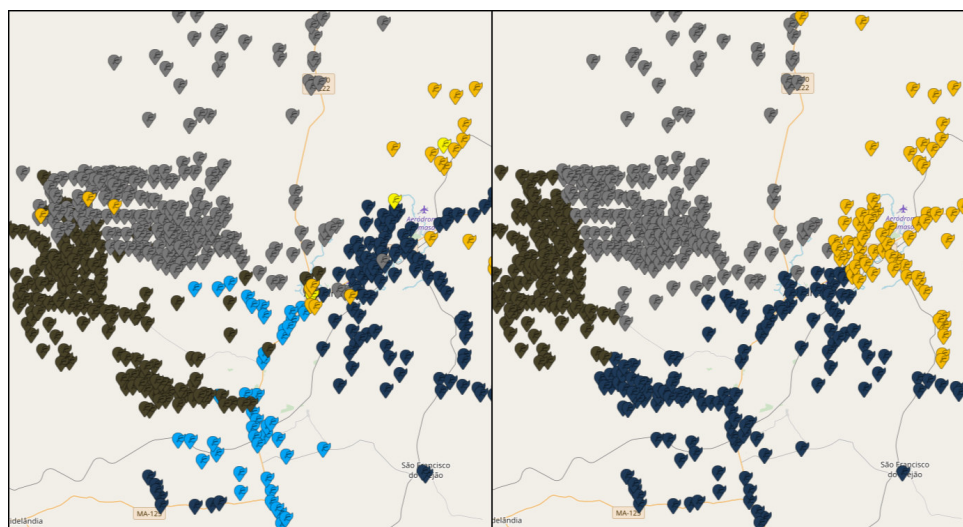
Figura 24 – Comparativos dos Planejamentos Urbanos Vigente com o Planejamento Baseado de Açailândia



Fonte: Próprio Autor.

A Tabela 11 mostra a comparação dos dados do município de Imperatriz. Pode-se constatar uma redução do desvio padrão da área urbana devido a descentralização dos tempos de etapas muito altas para outros agrupamentos e um leve aumento no coeficiente

Figura 25 – Comparativos dos Planejamentos Rurais Vigente com o Planejamento Baseado de Acailândia



Fonte: Próprio Autor.

	Planejamento Vigente		Planejamento Baseado	
	Área Urbana	Área Rural	Área Urbana	Área Rural
Desvio Padrão	923,1528967	799,0397408	565,4348177	54,6129913
Tempo Mínimo (min)	0,0541667	8,7027778	0,0541667	764,0137096
Tempo Máximo (min)	3.687,210453	1.744,0797692	2.576,3023357	902,169964
Coefficiente de Silhueta	0,2145923	0,3654557	0,2861341	0,20518

Tabela 11 – Dados comparativos do planejamento vigente com o planejamento baseado do município de Imperatriz

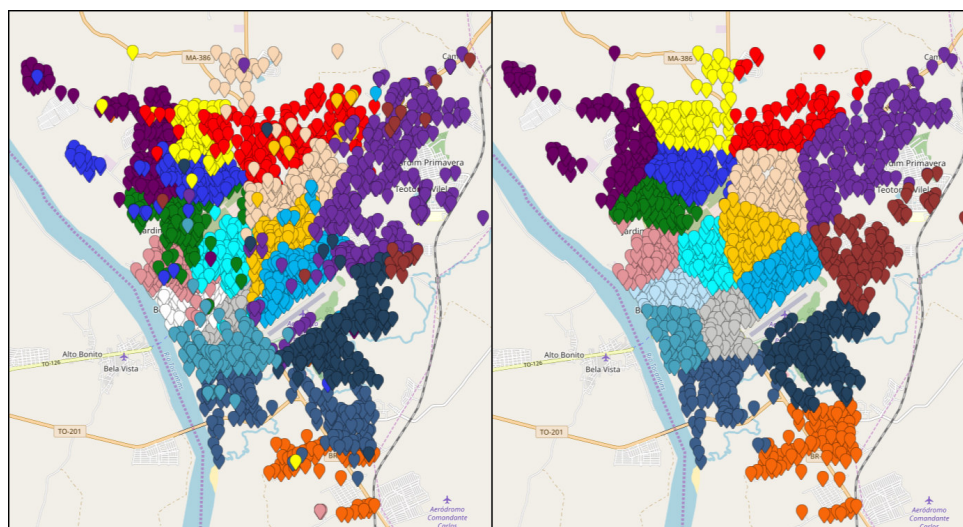
de silhueta. Um detalhe singular durante a geração deste planejamento foi a conservação de um agrupamento de leitura, o que ocasionou um tempo mínimo igual ao vigente. Na área rural, os tempos entre as etapas ficaram próximos, ocasionando uma diminuição significativa do desvio padrão, porém ocorreu uma leve degradação no coeficiente silhueta. As Figuras 26 e 27 exibem comparativos entre os dois agrupamentos.

	Planejamento Vigente		Planejamento Baseado	
	Área Urbana	Área Rural	Área Urbana	Área Rural
Desvio Padrão	1775,7785675	681,1453408	804,29258	319,1041035
Tempo Mínimo (min)	2729,7233749	552,9888889	3840,0509954	1195,342875
Tempo Máximo (min)	9836,1901899	2901,1216476	7408,5982321	2215,6071703
Coefficiente de Silhueta	-0,1254141	-0,2862414	0,2600981	0,3342394

Tabela 12 – Dados comparativos do planejamento vigente com o planejamento baseado do município de Marabá

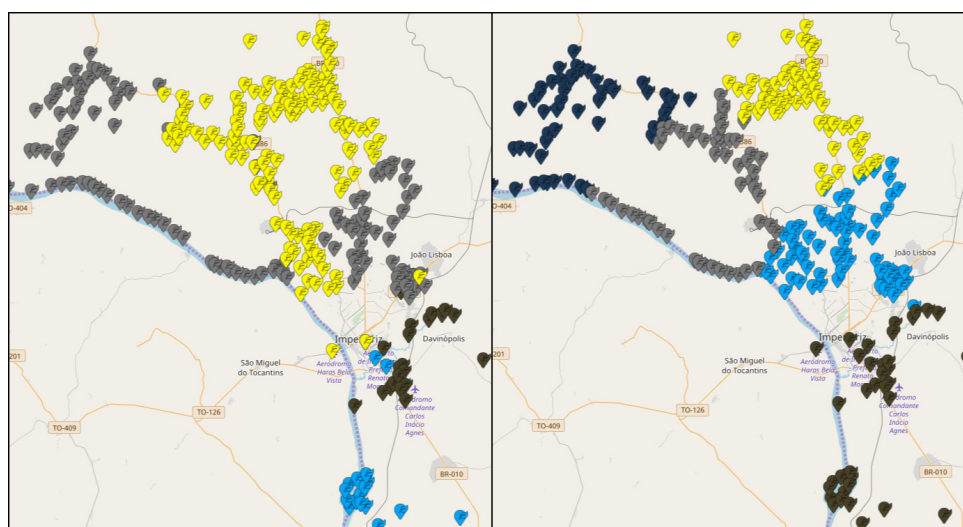
A seguir, compara-se na Tabela 12 os resultados obtidos sobre a região de Marabá.

Figura 26 – Comparativos dos Planejamentos Urbanos Vigente com o Planejamento Baseado de Imperatriz



Fonte: Próprio Autor.

Figura 27 – Comparativos dos Planejamentos Rurais Vigente com o Planejamento Baseado de Imperatriz



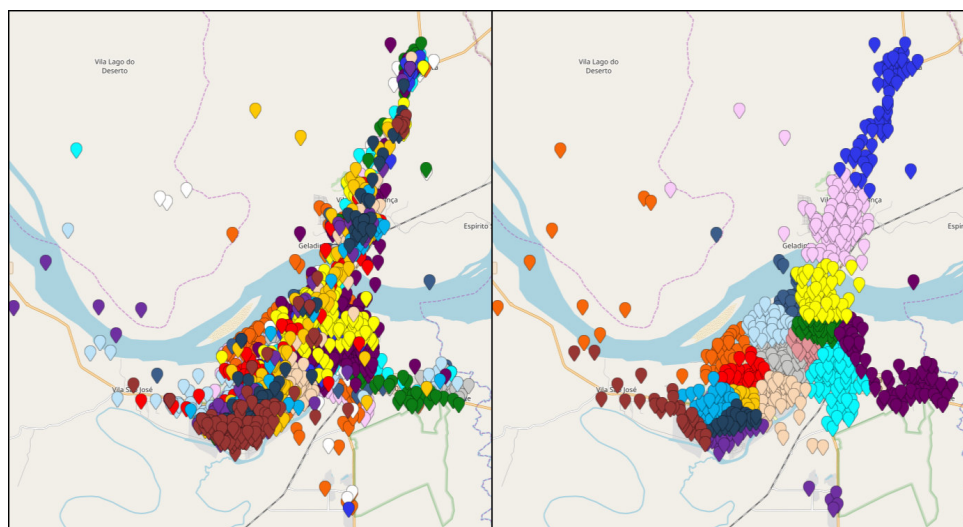
Fonte: Próprio Autor.

Assim como nos planos anteriores, ocorreram uma significativa redução dos desvios padrões da área urbana e área rural, o que balanceou os tempos mínimos e máximos de cada agrupamento. O coeficiente de silhueta do planejamento gerado também aumentou comparado ao vigente. As Figura 28 e 29 exibem os comparativos dos planejamentos.

Por fim, compara-se na Tabela 13 os resultados sobre a região de Santarém. Na área urbana, visualiza-se uma melhora no desvio padrão, enquanto na área rural este valor sobe, o que se deve a um agrupamento com o tempo muito menor do que o apresentado no vigente. Contrapondo o desvio, observa-se um aumento no coeficiente de silhueta para

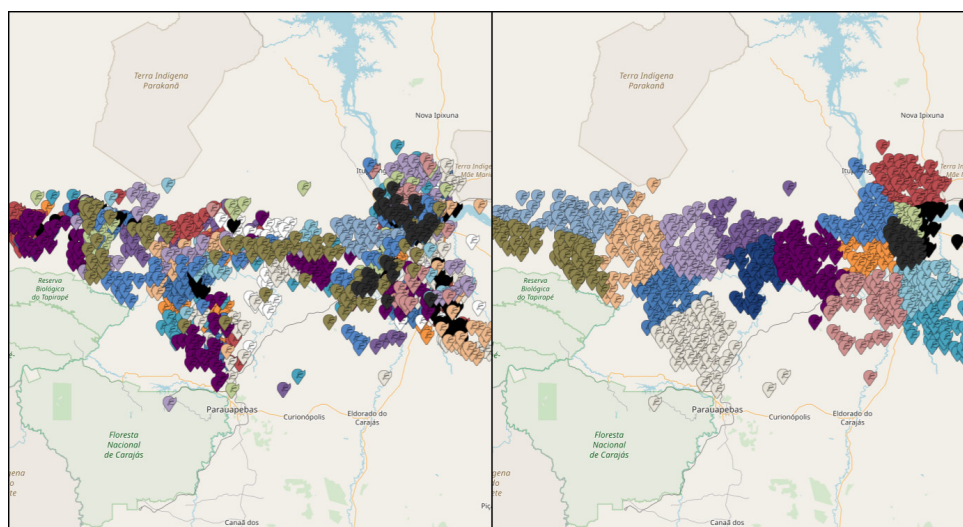


Figura 28 – Comparativos dos Planejamentos Urbanos Vigente com o Planejamento Baseado de Marabá



Fonte: Próprio Autor.

Figura 29 – Comparativos dos Planejamentos Rurais Vigente com o Planejamento Baseado de Marabá



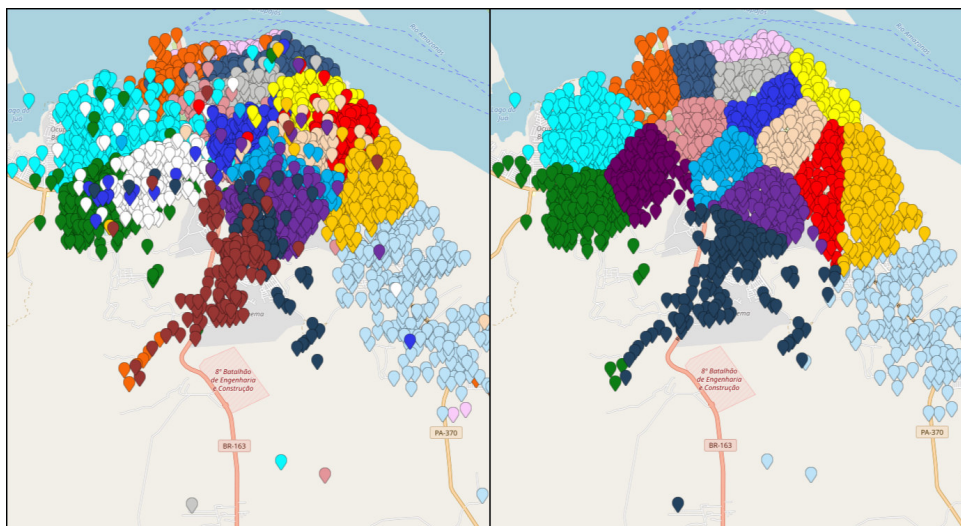
Fonte: Próprio Autor.

	Planejamento Vigente		Planejamento Baseado	
	Área Urbana	Área Rural	Área Urbana	Área Rural
Desvio Padrão	931,9544986	388,8194972	827,2172853	408,3152814
Tempo Mínimo (min)	2777,8802095	967,3836039	1401,6503395	212,7847222
Tempo Máximo (min)	6072,3362624	2202,8608073	5071,5933963	2078,7078802
Coefficiente de Silhueta	-0,0709605	-0,1483548	0,1576152	0,28515

Tabela 13 – Dados comparativos do planejamento vigente com o planejamento baseado do município de Santarém

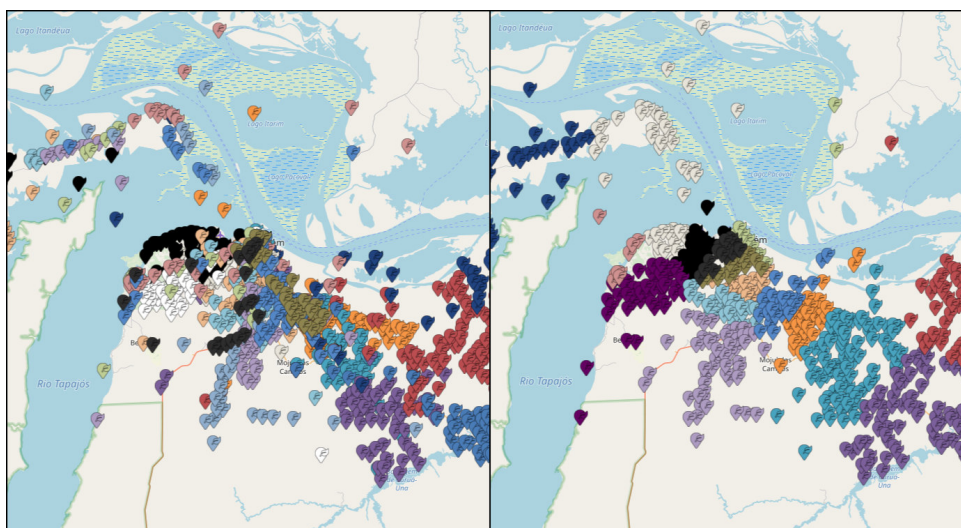
ambas as áreas. As Figuras 30 e 31 exibem os comparativos do planejamento vigente e resultante.

Figura 30 – Comparativos dos Planejamentos Urbanos Vigente com o Planejamento Baseado de Santarém



Fonte: Próprio Autor.

Figura 31 – Comparativos dos Planejamentos Rurais Vigente com o Planejamento Baseado de Santarém



Fonte: Próprio Autor.

Ao comparar estes resultados é possível verificar uma melhora significativa no balanceamento dos tempos. Isto se deve ao desvio padrão encontrado no planejamento baseado que são menores que o do planejamento vigente. Ao usar um agrupamento inicial como base é possível recuperar o planejamento prévio e otimiza-lo a partir do ponto em que a região se encontra. Usar este tipo de abordagem permite que os planejamentos não se distanciem da situação original, levando em consideração os conhecimentos dos especialistas.

Além disso, usar este método corresponde a menos remanejamento, viabilizando a implementação da estratégia.

Outra vantagem ao utilizar esta abordagem, é o aumento do coeficiente de silhueta. Como são realizadas diversas distribuições de grupos criadas pelo *K-means*, estes grupos gerados também tendem a ser mais compactos, com o diferencial de ter um planejamento prévio como base. A usabilidade deste método se torna bastante eficaz quando aplicado a um planejamento existente, respeitando os critérios exigidos e melhorando a situação atual.

## 6 Conclusões

O processo de criação de um planejamento para as operações de leitura e cobrança de unidades de consumidoras apresentam um problema complexo relacionado à otimização logística, principalmente por se tratar de uma tarefa com diversas alterações ao longo do tempo que foi estudada, revisada e avaliada antes de entrar em vigor.

Este trabalho apresentou um método de agrupamento baseado em algoritmo genético feito para resolver o problema do agrupamento capacitado dado os critérios de homogeneidade e compacidade. Os resultados obtidos mostram a viabilidade do uso da metodologia ao aplicada a um planejamento existente, mostrando uma redução do desvio padrão das etapas, balanceando o tempo entre as etapas. Outro fator importante é a criação de agrupamento com um bom coeficiente de silhueta comparadas a situação atual, o que demonstra a compacidade dos grupos.

Esta metodologia oferece também uma maneira rápida de criação de planos de leitura automáticos, visto que este trabalho é realizado manualmente pelos especialistas na área. Os resultados podem ser usados como uma implementação ou como uma segunda opinião para os responsáveis pela criação do planejamento.

O método proposto será testado por especialistas na área através do projeto P&D PD-00371-0031/2017 em parceria com a Equatorial Energia. Será avaliado junto aos funcionários a viabilidade e implementação dos benefícios apresentados pela abordagem.

Para trabalhos futuros, planeja-se mudar o critério de avaliação do algoritmo genético para trabalhar também com a maximização do coeficiente de silhueta. É almejado também um melhor tratamento para os *outliers*, considerando o tempo de distância de uma região populada até as instalações mais distantes. Também é válido investigar como os *outliers* afetam na criação de agrupamentos. E, por fim, considerar a geografia da região de leitura a partir do arruamento da cidade, visando o melhorar os grupos criados.

## Referências

- AHMADI, S.; OSMAN, I. H. Density based problem space search for the capacitated clustering p-median problem. *Annals of Operations Research*, Springer, v. 131, n. 1-4, p. 21–43, 2004. Citado na página 20.
- ANEEL. *Resolução Normativa Nº 414*. 2010. Disponível em: <[www.aneel.gov.br/cedoc/bren2010414.pdf](http://www.aneel.gov.br/cedoc/bren2010414.pdf)>. Citado na página 14.
- ASSIS, L. S. D.; FRANCA, P. M.; USBERTI, F. L. A redistricting problem applied to meter reading in power distribution networks. *Computers & Operations Research*, Elsevier, v. 41, p. 65–75, 2014. Citado na página 18.
- ASSIS, L. S. de; FRANÇA, P. M.; USBERTI, F. L. Reagrupamento capacitado: problema de redistribuição de lotes de faturamento. 2009. Citado na página 18.
- ASSIS, L. S. de; FRANÇA, P. M.; USBERTI, F. L. Reagrupamento capacitado multicritério: Problema de redistribuição de lotes de faturamento. *XLII Simpósio Brasileiro de Pesquisa Operacional*, p. 219–230, 2010. Citado na página 15.
- ÁVILA, S. L. et al. Otimização multiobjetivo e análise de sensibilidade para concepção de dispositivos: aplicação: síntese de antenas refletoras para comunicação via satélite. Florianópolis, SC, 2006. Citado na página 15.
- BARBOSA, A. M.; RIBEIRO, L. d. C.; ARANTES, J. d. O. Algoritmo genético multiobjetivo: Sistema adaptativo com elitismo. *Citado*, v. 2, p. 16, 2010. Citado na página 15.
- BEKKERMAN, R.; BILENKO, M.; LANGFORD, J. *Scaling up machine learning: Parallel and distributed approaches*. [S.l.]: Cambridge University Press, 2011. Citado na página 22.
- BENTLEY, J. L. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, ACM, v. 18, n. 9, p. 509–517, 1975. Citado na página 31.
- BERKHIN, P. A survey of clustering data mining techniques. In: *Grouping multidimensional data*. [S.l.]: Springer, 2006. p. 25–71. Citado na página 15.
- BORTOLOSSI, H. J. *Cálculo Diferencial à várias variáveis*. [S.l.]: Edicoes Loyola, 2002. Citado na página 22.
- BRENT, R. P. An improved monte carlo factorization algorithm. *BIT Numerical Mathematics*, Springer, v. 20, n. 2, p. 176–184, 1980. Citado na página 19.
- 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 18.
- COSTA, C.; COSTA, D.; GÓES, A. Determinação de setores de atendimento em uma concessionária de energia. *Trends in Applied and Computational Mathematics*, v. 8, n. 3, p. 381–390, 2007. Citado na página 18.



- COSTA, P. B.; SILVA, I. F. S. da; VIEIRA, P. H. C.; SANTOS, R. D. A.; SILVA, M. G.; LIMA, C. S. C.; JÚNIOR, D. L. G.; GARROS, E. M.; SILVA, I. F. S. da; PINHEIRO, L. P. 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. Citado 4 vezes nas páginas 24, 26, 27 e 36.
- 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 18.
- FRANÇA, P. M.; GARCIA, V. J.; MORELATO, A.; USBERTI, F. L. Enfoque multicritério para o problema de redistribuição capacitado. *XXXIX Simpósio Brasileiro de Pesquisa Operacional-SBPO*, p. 1–12, 2007. Citado na página 18.
- GAREY, M. R.; JOHNSON, D. S. *Computers and intractability*. [S.l.]: wh freeman New York, 2002. v. 29. Citado na página 20.
- GUERINE, M.; STOCKINGER, M. B.; ROSSETI, I.; PLASTINO, A. Heurística híbrida com mineração de dados para o problema de agrupamento capacitado com centro geométrico. 2017. Citado na página 18.
- HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011. Citado na página 18.
- HARTIGAN, J. A.; WONG, M. A. A k-means clustering algorithm. *JSTOR: Applied Statistics*, v. 28, n. 1, p. 100–108, 1979. Citado na página 21.
- JAIN, A. K. Data clustering: 50 years beyond k-means. *Pattern recognition letters*, Elsevier, v. 31, n. 8, p. 651–666, 2010. Citado na página 31.
- 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 18.
- LEE, D.-T.; SCHACHTER, B. J. Two algorithms for constructing a delaunay triangulation. *International Journal of Computer & Information Sciences*, Springer, v. 9, n. 3, p. 219–242, 1980. Citado na página 26.
- LUCAS, D. C. Algoritmos genéticos: uma introdução. *Universidade Federal do Rio Grande do Sul-Porto Alegre*, 2002. Citado 3 vezes nas páginas 23, 25 e 33.
- LUNET, N.; SEVERO, M.; BARROS, H. Desvio padrão ou erro padrão. *Arquivos de Medicina*, ArquiMed-Departamento de Edições Científicas da AEFMUP, v. 20, n. 1-2, p. 55–59, 2006. Citado na página 34.
- MLADENOVIĆ, N.; BRIMBERG, J.; HANSEN, P.; MORENO-PÉREZ, J. A. The p-median problem: A survey of metaheuristic approaches. *European Journal of Operational Research*, Elsevier, v. 179, n. 3, p. 927–939, 2007. Citado 2 vezes nas páginas 18 e 19.
- MURTY, K. G. Linear and combinatorial programming. *ROBERT E. KRIEGER PUBLISHING COMPANY, MELBOURNE, FL(USA), 1985, 592*, 1985. Citado na página 19.

- RECK, W. d. M. Algoritmos de agrupamento capacitado aplicado ao problema de despacho de ordens de serviço. Universidade Federal do Pampa, 2010. Citado na página 20.
- SILVA, A. A. Implementação capacitada do growing neural gas aplicada à logística de leitura em companhias de distribuição de energia elétrica. Universidade Federal do Maranhão, 2019. Citado na página 19.
- SMIDERLE, A.; STEINER, M. T. A.; CARNIERI, C. Problema de cobertura de arcos – um estudo de caso. *science*, XXIII Encontro Nac. de Eng. de Produção, 2003. Citado na página 19.
- 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 19.
- TEITZ, M. B.; BART, P. Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations research*, INFORMS, v. 16, n. 5, p. 955–961, 1968. Citado na página 19.
- WHITLEY, D. A genetic algorithm tutorial. *Statistics and computing*, Springer, v. 4, n. 2, p. 65–85, 1994. Citado na página 15.