

UNIVERSIDADE FEDERAL DE GOIÁS  
INSTITUTO DE INFORMÁTICA

MAX GONTIJO DE OLIVEIRA

# **Sistema de Localização de Facilidades**

**Uma abordagem para mensuração de pontos de demanda  
e localização de facilidades**

Goiânia  
2012

UNIVERSIDADE FEDERAL DE GOIÁS  
INSTITUTO DE INFORMÁTICA

**AUTORIZAÇÃO PARA PUBLICAÇÃO DE DISSERTAÇÃO  
EM FORMATO ELETRÔNICO**

Na qualidade de titular dos direitos de autor, **AUTORIZO** o Instituto de Informática da Universidade Federal de Goiás – UFG a reproduzir, inclusive em outro formato ou mídia e através de armazenamento permanente ou temporário, bem como a publicar na rede mundial de computadores (*Internet*) e na biblioteca virtual da UFG, entendendo-se os termos “reproduzir” e “publicar” conforme definições dos incisos VI e I, respectivamente, do artigo 5º da Lei nº 9610/98 de 10/02/1998, a obra abaixo especificada, sem que me seja devido pagamento a título de direitos autorais, desde que a reprodução e/ou publicação tenham a finalidade exclusiva de uso por quem a consulta, e a título de divulgação da produção acadêmica gerada pela Universidade, a partir desta data.

**Título:** Sistema de Localização de Facilidades – Uma abordagem para mensuração de pontos de demanda e localização de facilidades

**Autor(a):** Max Gontijo de Oliveira

Goiânia, 08 de Outubro de 2012.

---

Max Gontijo de Oliveira – Autor

---

Cedric Luiz de Carvalho – Orientador

---

Telma Woerle de Lima Soares – Co-Orientadora

MAX GONTIJO DE OLIVEIRA

# **Sistema de Localização de Facilidades**

**Uma abordagem para mensuração de pontos de demanda  
e localização de facilidades**

Dissertação apresentada ao Programa de Pós-Graduação do Instituto de Informática da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

**Área de concentração:** Ciência da Computação.

**Orientador:** Prof. Cedric Luiz de Carvalho

**Co-Orientadora:** Profa. Telma Woerle de Lima Soares

Goiânia  
2012

MAX GONTIJO DE OLIVEIRA

# **Sistema de Localização de Facilidades**

## **Uma abordagem para mensuração de pontos de demanda e localização de facilidades**

Dissertação defendida no Programa de Pós-Graduação do Instituto de Informática da Universidade Federal de Goiás como requisito parcial para obtenção do título de Mestre em Ciência da Computação, aprovada em 08 de Outubro de 2012, pela Banca Examinadora constituída pelos professores:

---

**Prof. Cedric Luiz de Carvalho**

Instituto de Informática – UFG

Presidente da Banca

---

**Profa. Telma Woerle de Lima Soares**

Instituto de Informática – UFG

---

**Prof. Cássio Leonardo Rodrigues**

Instituto de Informática - UFG

---

**Prof. Antônio César Baleeiro Alves**

Escola de Engenharia Elétrica, Mecânica e de Computação - UFG

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador(a).

### **Max Gontijo de Oliveira**

Graduou-se em Ciência da Computação na UCG - Universidade Católica de Goiás. Durante sua graduação, foi monitor no Departamento de Computação da UCG. Durante o Mestrado, na UFG - Universidade Federal de Goiás, desenvolveu um trabalho teórico e prático na resolução de um problema de localização de viaturas, financiado pela empresa de distribuição de energia elétrica Celg. Atualmente, possui vínculo com a Segplan - Secretaria do Planejamento, atuando como analista e desenvolvedor de software.

Dedico esse trabalho à Deus, por possibilitar que tudo acontecesse.

Aos meus pais Joaquim e Maria de Fátima, que sempre me apoiaram e jamais desacreditaram de mim.

À minha irmã Lorena, por estar do meu lado e pela motivação, bem como pelas vezes que me surpreendia jogando e dizia "estudar assim deve ser bom".

À minha sobrinha Bianca, por todo o carinho e por todas as vezes que me obrigou a parar os trabalhos de pesquisa para poder brincar até ela (eu) cansar.

À minha namorada Gabriela, pelo companheirismo, pela compreensão nos meus momentos de estresse e por todo amor e carinho que renovavam minhas forças.

---

## Agradecimentos

---

Agradeço plenamente ao professor e meu orientador Dr. Cedric, por todos os ensinamentos, sugestões, conselhos, atenção, críticas, elogios, incentivos e, sobretudo, pela paciência que teve para comigo durante toda a realização do trabalho.

À Professora Dra. Telma Woerle de Lima Soares, pela avaliação do presente trabalho e pelas sugestões feitas, as quais muito contribuíram para a melhoria do texto original.

Ao amigo Francisco, que foi meu companheiro de batalhas durante todo o Mestrado, compartilhando comigo de todas as frustrações e vitórias diante dos desafios que enfrentamos juntos.

Aos amigos Leonardo e Gilmar pelo tempo de estudo que passamos juntos compartilhando nossas dúvidas e conhecimentos.

À CELG Distribuição S.A. por todo o apoio prestado no projeto desenvolvido.

Nossas virtudes e nossos defeitos são inseparáveis, como força e matéria.  
Quando eles se separam, o homem perde sua essência.

**Nikola Tesla,**

*De origem sérvia, foi um engenheiro e inventor que contribuiu com importantes invenções de grande interesse industrial, sobretudo no campo da engenharia elétrica..*

---

## Resumo

---

Gontijo de Oliveira, Max. **Sistema de Localização de Facilidades**. Goiânia, 2012. 89p. Dissertação de Mestrado. Instituto de Informática, Universidade Federal de Goiás.

Diversas organizações precisam lidar com o problema de localizar e alocar facilidades em uma região geográfica. Problemas de localização e alocação podem ser vistos, por exemplo, na distribuição de viaturas policiais, ambulâncias, viaturas de contenção de falhas em redes elétricas, taxistas, pontos de ônibus dentre outras inúmeras situações onde a localização de tais facilidades é um fator estratégico para a organização. Em problemas de localização/alocação de facilidades, geralmente aloca-se cada ponto de demanda à facilidade mais próxima e, localiza-se essa facilidade no centro dos pontos de demanda, considerando o valor da demanda como peso nessa distância. Entretanto, comumente, problemas reais de localização de facilidades possuem restrição de capacidade. Assim, cada facilidade possui uma certa capacidade em função do tipo de demanda. Problemas de localização de facilidades podem ser contínuos ou discretos. Em problemas contínuos (também chamados de problema de Weber com múltiplas fontes), qualquer ponto no plano é um potencial local para se instalar uma facilidade. Existem várias abordagens para trabalhar com modelos contínuos e outras tantas para trabalhar com modelos com restrição de capacidade, mas a maioria dessas abordagens realiza uma discretização do modelo. Assim, o objetivo desse trabalho é apresentar uma abordagem para gerar boas distribuições de facilidades para o problema de localização/alocação contínuo com restrição de capacidade. Um caso de estudo será apresentado com a finalidade de avaliar os resultados obtidos.

### Palavras-chave

localização de facilidades, algoritmos de cluster, k-means

---

## Abstract

---

Gontijo de Oliveira, Max. **Facility Location System**. Goiânia, 2012. 89p. MSc. Dissertation. Instituto de Informática, Universidade Federal de Goiás.

Several organizations need to solve the problem of locate and allocate facilities within a geographic area. There are location/allocation problems in various situations, like the distribution of police cars, ambulances, taxi drivers, bus stops among other numerous situations where the location of such facilities is strategic for organization. In location/allocation problems, usually is necessary allocate each demand point to the closest facility. So, each facility will be located in the center of demand points, considering the demand as weight. However, the majority of the real location problems have capacity constraint. Therefore, each facility has a certain capacity based on the type of demand. Facility location problems can be continuous or discrete. In continuous problems (also called Weber problem with multiple sources), any point in the plane is a potential site for the instalation of the facility. There are several approaches for working with continuous models. Furthermore, there are many others works approaches presenting models with capacity constraint. But most of these approaches turns the continous model to a discrete model. The objective of this work thesis is to present an approach to distribution of facilities in instances of the capacitated facility location problem. A case study will be presented with the purpose of evaluating the results.

### Keywords

facility location, clustering algorithms, k-means

---

# Sumário

---

Lista de Figuras	11
Lista de Algoritmos	14
<b>1</b> Introdução	<b>15</b>
1.1 Motivação	16
1.2 Objetivo	17
1.3 Metodologia	17
1.4 Organização da Dissertação	18
<b>2</b> Fundamentação Teórica	<b>19</b>
2.1 Problemas de localização de facilidades	19
2.1.1 Características de problemas de localização de facilidades	19
Modelos contínuos e o problema de Weber	21
2.2 <i>Clusters</i> e o método <i>k-means</i>	23
2.2.1 <i>Clusters</i> e o problema de localização	24
2.2.2 O método <i>k-means</i>	25
2.3 Processos de Poisson Modulados a Markov	28
2.3.1 Descrição do modelo	29
2.4 Discussão	33
<b>3</b> Trabalhos Relacionados	<b>34</b>
3.1 Melhorias no método <i>k-means</i>	34
3.1.1 Algoritmos genéticos: GKA e FGKA	34
3.1.2 Método <i>k-means</i> repetitivo	37
3.2 O problema das p-medianas capacitado	38
3.3 Desenho de grafos	41
3.3.1 O método <i>Springs</i> de Eades	41
3.3.2 Discussão	42
<b>4</b> O Problema de Localização de Facilidades com Restrição de Capacidade em um Modelo Contínuo	<b>44</b>
4.1 A relaxação do problema	45
4.2 A Modelagem	46
4.3 Histórico de demanda	48
4.3.1 Estatística Simples	49
4.4 A proposta de solução	50
4.4.1 Sistema de forças	52
Força de atração pelos pontos de demanda à facilidade mais próxima	53

	Força de atração pelos pontos de demanda atendidos parcialmente por outra(s) facilidade(s)	54
	Força de repulsão entre duas facilidades	57
4.5	Discussão	58
<b>5</b>	<b>Estudo de Caso e Resultados</b>	<b>59</b>
5.1	Determinação dos pontos de demanda	60
5.1.1	Estatística Simples com Média nas Janelas de Tempo	61
5.1.2	Usando os parâmetros de caracterização do MMPP no problema da CELG-D	63
5.1.3	Aplicação e resultados da mensuração dos pontos de demanda	64
5.2	Localização das Viaturas	68
5.2.1	O uso de temperatura e resfriamento no algoritmo	73
5.2.2	Aplicação e resultados da localização das viaturas	75
	Objetivos atingidos	76
	Validação da distribuição de pontos de demanda com a localização de viaturas	78
	Desempenho	78
5.3	Discussão	81
<b>6</b>	<b>Considerações Finais</b>	<b>84</b>
6.1	Conclusões	84
6.2	Trabalhos Futuros	85
	Referências Bibliográficas	<b>87</b>

---

## Lista de Figuras

---

2.1	Visualização de uma base de dados com destaque para conjuntos de dados aglomerados ( <i>clusters</i> ) que determinam as classes	24
2.2	Exemplo de execução do método <i>k-means</i>	27
(a)	Distribuição inicial - círculos vermelhos representam os pontos de demanda enquanto quadrados verdes representam os centroides	27
(b)	Alocação dos pontos de demanda ao centroide mais próximo	27
(c)	Centroides se movem para o centro dos <i>clusters</i>	27
(d)	Nova alocação dos pontos de demanda ao centroide mais próximo - alocação muda, logo, algoritmo continua a execução	27
(e)	Centroides se movem para o centro dos <i>clusters</i>	27
(f)	Algoritmo para, pois não existe mais mudança na alocação dos centroides aos pontos de demanda	27
2.3	Divisão estatística em uma função distribuição cumulativa das demandas ocorridas em janelas de tempo de uma célula da região sudoeste de Goiás no ano 2010	31
(a)	Função distribuição cumulativa dos valores medidos	31
(b)	Divisão do eixo das probabilidades em cinco intervalos de 0,2 (20%) cada - para os cinco intervalos, são necessários seis pontos no mesmo eixo, sendo um deles com valor 0 e o outro com valor 1	31
(c)	Mapeamento das divisões do eixo das probabilidades para pontos no eixo dos valores medidos	31
(d)	Divisões de tamanhos iguais dentro do intervalo das taxas	31
2.4	Limites dos processos de Poisson	33
3.1	Exemplo de mutação no GKA	36
(a)	Primeira solução utilizada na mutação	36
(b)	Segunda solução utilizada na mutação	36
(c)	Mutação entre os oito primeiros valores do cromossomo da primeira solução com os quatro últimos valores do cromossomo da segunda solução	36
(d)	Nova alocação (cromossomo mutante) gerada pela mutação e após os centroides serem devidamente reposicionados no centro dos pontos de dados a ele ligados	36
(e)	Solução final (novo cromossomo) após o método <i>k-means</i> ser executado a partir da nova solução gerada pela mutação	36
3.2	Identificação de nós candidatos à substituição na geração da vizinhança para $k' = 2$ - são selecionadas duas facilidades para serem removidas	40

(a)	Solução corrente. As estrelas representam as facilidades instaladas - duas delas serão selecionadas para serem removidas	40
(b)	Duas facilidades foram removidas (quadrados). A vizinhança passa a ser gerada a partir de todos os demais pontos (círculos)	40
(c)	Para $h = 3$ , são selecionadas os três pontos mais próximos para serem excluídos do espaço de solução (triângulos)	40
(d)	As linhas pontilhadas delimitam as margens periféricas, onde estarão os pontos excluídos	40
4.1	Passos para um possível mapeamento de pontos de demanda em um mapa geográfico plotado em um plano cartesiano para uma representação em grafo.	47
(a)	Configuração inicial do problema no plano com a representação dos pontos de demanda	47
(b)	Mapa é dividido em células de tamanhos iguais (nesse caso, optou-se por dividir em uma grade de $8 \times 8$ )	47
(c)	São identificadas as células que não possuem nenhum ponto de demanda e as mesmas são excluídas do espaço de solução	47
(d)	Cada célula que possui pelo menos um ponto de demanda recebe um vértice (ponto) e os vértices gerados são ligados por arestas conforme alguma regra de fronteira entre as células (nesse caso, optou-se como regra, considerar apenas fronteiras nos lados esquerdo, direito, de cima e de baixo)	47
(e)	O grafo gerado passa a ser o resultado da conversão, onde cada aresta tem como valor, a distância euclidiana do centro de cada célula ao centro da célula vizinha	47
4.2	Cenário com duas facilidades localizadas utilizando o método <i>k-means</i> que encontra os centroides de $k = 2$ clusters	50
4.3	Cenário com duas facilidades de capacidade 5ud localizadas utilizando o método <i>k-means</i> e alocadas a pontos de demanda com valor de demanda de 1ud cada	51
4.4	Cenário com duas facilidades de capacidade 5ud alocadas a pontos de demanda com valor de demanda de 1ud cada, mas localizadas de uma forma que minimize a distância entre a facilidade e os pontos de demanda a ela alocados	52
4.5	Força de atração exercida pelos pontos de demanda (pontos marcados com um raio) sobre a facilidade (pontos marcados com um carro e um número de identificação) mais próxima	53
4.6	Força de atração proveniente da demanda remanescente dos pontos (marcados com um raio) sobre facilidades (pontos marcados com um carro e um número de identificação) que ainda tenham capacidade remanescente do atendimento de demanda	55
(a)	Forças exercidas sobre as facilidades 1 e 2	55
(b)	Forças exercidas pelos pontos de demanda com foco apenas na facilidade 2	55
4.7	Conjunto das forças do sistema proposto	57
(a)	Forças de atração exercidas sobre as facilidades 1 e 2	57

(b)	Uma visualização mais aproximada da força de repulsão existente entre as facilidades 1 e 2	57
5.1	Comparação entre demandas que ocorrem em duas células distintas em um período de vinte dias utilizando dados hipotéticos	61
5.2	Demanda de uma célula representada em janelas de 5 dias com destaque para a média. Os valores são de uma célula da região sudoeste de Goiás com tamanho de 2.5 km em cada lado no período entre 01/01/2010 e 31/12/2010.	62
5.3	Demanda de uma célula representada em janelas de 5 dias com destaque para faixas definidas arbitrariamente e que possuem médias próprias.	63
5.4	Aplicação que determina os pontos de demanda a partir de um período e uma região	65
(a)	Parametrização do processo de determinação dos pontos de demanda	65
(b)	Mapa mostrando os pontos de demanda encontrados	65
5.5	Pontos de demanda da região sudoeste de Goiás no ano de 2010 encontrados utilizando o método simples que não considera o tempo	66
5.6	Comparação entre janela de tempo real medida com janela de tempo inferida, correspondente ao período entre os dias 01/01/2011 e 05/01/2011	66
(a)	Janela de tempo com demandas reais medidas	66
(b)	Janela de tempo com demandas inferidas a partir do método MMPP	66
5.7	Aplicação que determina a localização das viaturas da CELG-D em função dos pontos de demanda pré-definidos	75
(a)	Parametrização do algoritmo de determinação da localização das viaturas	75
(b)	Mapa mostrando os pontos de demanda e as viaturas em movimento	75
5.8	Visualização de viaturas localizadas no plano de acordo com os pontos de demanda existentes	77
(a)	Viaturas distribuídas segundo objetivos pré-estabelecidos	77
(b)	Destaque na viatura que está sendo atraída por outros pontos de demanda distantes dos pontos que atende por estar mais próxima	77
5.9	Comparação entre localizações de facilidades realizadas a partir de pontos de demanda reais medidos e a partir de pontos de demanda mensurados pelo modelo MMPP. A janela apresentada compreende ocorrências de falhas entre os dias 01/01/2011 e 05/01/2011 na região sudoeste do Estado de Goiás.	79
(a)	5 viaturas - demanda real	79
(b)	5 viaturas - demanda MMPP	79
(c)	9 viaturas - demanda real	79
(d)	9 viaturas - demanda MMPP	79
(e)	10 viaturas - demanda real	79
(f)	10 viaturas - demanda MMPP	79
5.10	Desempenho do algoritmo de localização de facilidades com 20 viaturas enquanto a quantidade de pontos de demanda é aumentada	80
5.11	Desempenho do algoritmo de localização de facilidades com 300 pontos de demanda enquanto a quantidade de viaturas é aumentada	81

---

## Lista de Algoritmos

---

2.1	<i>PWS(K)</i>	22
2.2	<i>k-means</i>	26
3.1	<i>FNS(k', maxiter, N)</i>	39
5.1	<i>Location(K, J)</i>	69
5.2	<i>CalcularCusto(K, J)</i>	71
5.3	<i>CalcularCustoNivel(K, J)</i>	72
5.4	<i>LocationChangeMoveSize(K, J)</i>	74

## Introdução

---

Decidir sobre como realizar a distribuição de facilidades<sup>1</sup> em uma região geográfica limitada (uma cidade, por exemplo), é uma atividade estratégica para muitas organizações. Quando a disponibilização de um recurso envolve um custo muito alto ou o número de recursos disponíveis é limitado, esse problema fica ainda mais preocupante, pois uma distribuição mal realizada pode acarretar em prejuízo certo para a organização.

Uma distribuição mal realizada de recursos disponíveis pode ser responsável por uma série de consequências indesejáveis por parte da organização. Tais consequências podem incluir perda de clientes para outras organizações concorrentes, dificuldades de atingir metas de qualidade de serviço (prazo para atendimento) e baixo uso do recurso (considerando sua capacidade). Somente para ilustrar, pode-se tomar como exemplo, redes de supermercados (que precisam decidir onde instalar uma nova filial), empresas que coordenam o serviço de divulgação por panfletagem (que precisam decidir como alocar as pessoas que irão distribuir panfletos), empresas distribuidoras de energia elétrica (que precisam decidir os melhores pontos em uma rede elétrica onde inserir chaves e transformadores), hospitais, postos de saúde, delegacias e semáforos em uma malha rodoviária (que precisam ser distribuídos de forma que haja a maior cobertura possível em uma cidade, considerando a demanda e necessidade de cada unidade em uma cada região).

Originalmente proposto por Weber [29], o problema de localização consistia em posicionar a facilidade no plano de modo que fosse minimizada a soma de todas as distâncias entre a mesma e os pontos de demanda. Quando o número de facilidades aumenta, a solução deixa de ser trivial. De fato, a partir do momento em que necessitasse de localizar mais de uma facilidade, o problema passa a ser de localização e alocação. Desse modo, além de buscar a melhor distribuição, é preciso realizar a alocação da melhor forma possível.

---

<sup>1</sup>Nesse contexto, uma facilidade é um recurso que provê algum tipo de serviço ou produto para atender uma demanda medida ou estimada que, geralmente, está próxima a esse recurso. De modo geral, facilidade e recurso terão o mesmo significado nesse trabalho.

Nos problemas de localização/alocação sem restrição de capacidade, a alocação é feita de forma natural, alocando cada ponto de demanda à facilidade mais próxima. Assim, de modo geral, localizar  $n$  facilidades sem restrição de capacidade em um plano consiste basicamente de encontrar  $n$  *clusters* entre os pontos de demanda. Entretanto, é muito mais comum em problemas reais que as facilidades tenham uma capacidade limitada. Portanto, distribuí-las sem considerar esse aspecto geraria alocações inviáveis, uma vez que facilidades poderiam estar alocadas a uma demanda superior à sua capacidade.

Existem diversos modelos de problemas de localização [15]. Frequentemente, cada problema real pode ser encaixado em algum dos diversos modelos já apresentados e estudados ao longo do tempo. Com a finalidade de simplificar o problema e utilizar alguma solução mais simples, existem diversos problemas que, originalmente se encaixariam em um modelo contínuo, mas que podem ser convertidos para um modelo discreto sem que haja um grande prejuízo dos resultados. Mas, nem sempre é desejável que seja feita essa conversão. Portanto, é interessante ter disponível uma solução eficiente, que retorne bons resultados para problemas de localização contínuos.

A dificuldade em encontrar soluções eficientes é que, segundo Benati e Laporte[1], problemas de localização de facilidades são NP-Difíceis, tornando inviável a solução ótima até mesmo de instâncias relativamente pequenas. Dessa forma, esse trabalho visa buscar uma abordagem inteligente embasada em propostas feitas ao longo do tempo na literatura, de modo que seja viável a geração de boas sugestões de distribuição de recursos de um determinado problema de localização de facilidades.

## 1.1 Motivação

Existem diversos tipos de organizações que enfrentam problemas com a distribuição de facilidades em uma região geográfica. A maior motivação da realização deste trabalho está justamente na necessidade crucial de se realizar essa distribuição de forma adequada, visando melhores resultados no atendimento da demanda. Em alguns casos, a localização das facilidades pode ser de importância crucial para a vida humana, como ocorre na localização de viaturas médicas e viaturas policiais.

Uma outra motivação é o fato de haver poucas abordagens para a localização de facilidades em modelos contínuos sem que haja uma discretização do plano. Essa discretização restringe o mesmo à uma pequena e limitada lista de pontos candidatos a receberem a facilidade. Essa abordagem de discretização é por vezes utilizada visando uma melhoria do desempenho. Entretanto, as facilidades não podem mais serem instaladas em qualquer lugar do plano. Dessa forma, criar uma abordagem realmente focada no modelo contínuo se tornou um desafio à parte, ainda que seja conhecido o fato de que muitos problemas contínuos podem ser discretizados.

## 1.2 Objetivo

Embora solucionar o problema de determinar pontos de demanda a partir de dados históricos seja de fundamental importância para diversos problemas de localização reais e uma solução para tal problema ter sido proposta, o foco deste trabalho está na localização das facilidades. Assim, o objetivo principal deste trabalho é desenvolver e propor uma nova abordagem para encontrar uma boa solução para o problema de localização de facilidades no plano contínuo considerando a restrição de capacidade.

Como parte integrante do trabalho, foi ainda desenvolvida uma aplicação para atender um problema de localização de uma empresa de distribuição de energia elétrica.

## 1.3 Metodologia

O trabalho realizado seguiu uma metodologia que pode ser descrita pelos seguintes passos:

- **Definição do escopo do problema e estudo da fundamentação teórica.** Nessa etapa foi definido o escopo do problema a ser estudado. À medida em que o escopo era melhor definido, foram estudados os fundamentos teóricos a respeito do problema. Assim, conceitos como problemas de localização, identificação de agrupamentos de dados (*cluster*) e processos estocásticos foram exaustivamente estudados com a finalidade de prover fundamentação para o desenvolvimento de uma solução.
- **Análise de trabalhos relacionados.** Com o escopo definido, foram buscados outros trabalhos desenvolvidos para resolver o mesmo problema ou problemas correlatos. Foi também nessa etapa que percebeu-se que a maioria das soluções já propostas para o problema de localização de facilidades no plano restringiam o plano a um subconjunto de pontos candidatos para a localização das facilidades. Assim, fez-se necessário o desenvolvimento de uma nova abordagem que não limitasse o plano. Alguns dos trabalhos relacionados foram de grande influência na solução proposta.
- **Desenvolvimento e proposta de uma nova solução.** Nessa etapa foi desenvolvida e proposta uma nova abordagem para solucionar o problema de localização contínuo com restrição de capacidade.
- **Aplicação da nova abordagem proposta.** A validação de todo o trabalho desenvolvido se dá nessa etapa, onde a abordagem desenvolvida ao longo da pesquisa foi implementada, segundo as necessidades em um estudo de caso real.

## **1.4 Organização da Dissertação**

Além deste capítulo introdutório, o restante do trabalho está dividido em cinco capítulos. O Capítulo 2 apresenta os fundamentos teóricos utilizados para a realização deste trabalho. No Capítulo 3 são descritos alguns trabalhos relacionados, incluindo alguns que foram utilizados neste trabalho. O problema de localização abordado neste trabalho, bem como a solução proposta para o mesmo são apresentados no Capítulo 4. O Capítulo 5 detalha o uso da solução proposta em um estudo de caso realizado na CELG Distribuição S.A., apresentando os resultados obtidos nesse estudo. Por fim, o Capítulo 6 apresenta conclusões obtidas e propõe alguns trabalhos futuros.

## Fundamentação Teórica

---

Problemas de localização de facilidades são amplamente estudados. Existem diversos modelos, métodos e abordagens para tratar os diferentes problemas enfrentados pelas organizações. Embora este trabalho ataque um tipo específico de problema de localização, é importante entender alguns conceitos básicos, que direta ou indiretamente, influenciaram na solução apresentada neste trabalho.

Este capítulo pretende exibir as características de problemas de localização, focando mais no tipo de problema que é atacado. Além disso, o conceito de *cluster* e um método para realizar a busca de *clusters* em um plano serão abordados.

### 2.1 Problemas de localização de facilidades

De modo resumido, os problemas de localização de facilidades se referem à decisão de onde posicionar recursos que uma organização possui de modo que atenda, da melhor forma possível, aos critérios estabelecidos pela organização referentes aos pontos de demanda. Soluções para resolver problemas dessa natureza são importantes para as organizações. A distribuição de recursos bem realizada pode garantir melhorias na qualidade do serviço e mesmo contribuir no orçamento da organização. Por razões assim, existe um grande interesse que já gerou inúmeros trabalhos abordando esse assunto, dos quais vários foram relacionados no trabalho de Klose et al. [15]. Entretanto, a busca de soluções cada vez melhores continua sendo um grande desafio, uma vez que problemas dessa natureza são NP-Difíceis, conforme provado por Benati et al. [1].

Certamente, não existem soluções genéricas para resolver todo tipo de problema de localização. O primeiro passo para a busca de uma solução para esse tipo de problema é analisar suas características.

#### 2.1.1 Características de problemas de localização de facilidades

Os problemas de localização de facilidades abrangem uma grande quantidade de características variáveis que podem determinar diversos tipos de modelos. Conseqüente-

mente, a forma como são abordadas as soluções para os mesmos são diretamente ligadas a essas características.

Segundo Klose e Drexler [15], os modelos desse tipo de problema podem ser classificados em diversas formas:

1. **Organização da região geográfica.** Alguns problemas de localização permitem que uma facilidade seja colocada em qualquer lugar de uma região planar. Quando não se deseja perder nenhuma possibilidade de localização, problemas assim podem ser modelados de forma planar. Em contrapartida, existem problemas que podem ser mapeados para modelos com um número restrito de possíveis pontos de localização de facilidades. Modelos de rede e modelos discretos possuem essa característica. Eventualmente, o próprio problema obriga que o mesmo seja mapeado em um modelo de rede. De qualquer forma, modelos assim, em geral, são mais abordados na literatura [15].
2. **Objetivos.** Os objetivos podem incluir a minimização de uma variável (como a soma do custo de instalação de facilidades ou a soma das distâncias entre pontos de demanda e a facilidade mais próxima) ou a maximização de outra (como a cobertura de uma facilidade dado um limite de distância para atendimento).
3. **Capacidade da facilidade.** Os modelos de localização podem ou não ter que considerar a restrição de capacidade. Essa restrição refere-se à capacidade de atendimento que uma facilidade tem. De fato, de acordo com Melkote e Daskin [22], modelos que não possuem a restrição de capacidade podem ser vistos como modelos que possuem tal restrição, desde que a capacidade de cada uma das facilidades seja maior ou igual à soma de todas as demandas do problema. Assim, uma facilidade teria capacidade para atender todos os pontos de demanda, podendo então, essa restrição, ser desconsiderada. Assim, restariam apenas as demais restrições do problema.
4. **Quantidade de estágios.** A logística de atendimento de demanda pode ter um único estágio ou pode ter vários estágios. Modelos com apenas um único estágio são aqueles onde todas as facilidades atendem diretamente os clientes da organização. Modelos multi-estágios são aqueles em que, além das facilidades que atendem os clientes, tem-se ainda facilidades que atendem outras facilidades, como por exemplo, pontos de abastecimento de estoque. A quantidade de estágios nessa estrutura é determinada pelo problema.
5. **Quantidade de produtos/serviços.** Os modelos podem contemplar facilidades que forneçam apenas um tipo de produto/serviço ou que forneçam mais de um tipo de produto/serviço. São ainda considerados modelos de um único produto/serviço aqueles modelos em que os diversos produtos/serviços podem ser condensados em um único produto/serviço.

6. **Influência da demanda.** A maioria dos problemas de localização consideram que a demanda existe independente da localização das facilidades. Entretanto, em alguns casos, a demanda pode ser influenciada pela existência ou não de uma facilidade.
7. **Dinamismo.** Modelos podem ser estáticos ou dinâmicos. Modelos estáticos buscam otimizações para períodos específicos. Modelos dinâmicos consideram o tempo como fator determinante a ser considerado na otimização.

Ainda analisando as classificações de modelos de localização, segundo Farahani, SteadieSeifi e Asgari [8], esses modelos ainda podem ser distinguidos quanto a quantidade de critérios a serem otimizados. Eles apresentam um estudo sobre o estado-da-arte de problemas de localização com múltiplos critérios.

### Modelos contínuos e o problema de Weber

Quando o problema de localização real se refere à distribuição de recursos em uma região geográfica planar, o meio mais natural de se tratar a solução é modelar o problema de forma contínua. Assim, cada ponto do plano passa a ser uma localização potencial onde pode-se instalar uma facilidade.

Partindo dessa premissa, em 1909, Weber [29] propôs o problema de encontrar o melhor ponto para instalar uma facilidade, de modo que a soma das distâncias euclidianas<sup>1</sup> entre todos os pontos de demanda e a facilidade instalada fosse minimizada.

No problema de Weber simples (PWS), existe apenas uma facilidade a ser localizada. Dessa forma, dado que  $d_k(x,y)$  é a distância do ponto de demanda  $k$  à facilidade localizada na posição  $(x,y)$  e que  $p_k$  é o peso de um ponto de demanda (podendo ser obtido por uma função que considere diversos aspectos como prioridade, custo, frequência de uso da facilidade, etc.), o problema pode ser descrito conforme (2-1).

$$PWS = \min_{(x,y)} \sum_{k \in K} p_k d_k(x,y) \quad (2-1)$$

Se forem consideradas coordenadas discretas ( $x \in \mathbb{N}$  e  $y \in \mathbb{N}$ ), a solução para o problema (2-1) poderia ser implementada com o uso de um algoritmo iterativo simples, como apresentado no algoritmo 2.1.

---

<sup>1</sup>Distância em linha reta entre dois pontos

**Algoritmo 2.1:**  $PWS(K)$ 

**Entrada:** conjunto  $K = \{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$  de coordenadas  $(k_x, k_y)$  dos pontos de demanda

**Saída:** coordenadas  $(x, y)$  do melhor posicionamento para a facilidade

```

1  $xMaior \leftarrow$  maior valor de coordenada x do conjunto  $K$ 
2  $xMenor \leftarrow$  menor valor de coordenada x do conjunto  $K$ 
3  $yMaior \leftarrow$  maior valor de coordenada y do conjunto  $K$ 
4  $yMenor \leftarrow$  menor valor de coordenada y do conjunto  $K$ 
5  $somaMenor \leftarrow \infty$ 
6 para  $x \leftarrow xmenor$  até  $xMaior$  faça
7   para  $y \leftarrow ymenor$  até  $yMaior$  faça
8      $soma \leftarrow 0$ 
9     para cada  $k \in K$  faça
10       $soma \leftarrow soma +$  distância entre pontos  $(k_x, k_y)$  e  $(x, y)$ 
11    fim
12    se  $soma < somaMenor$  então
13       $somaMenor \leftarrow soma$ 
14       $xSolucao \leftarrow x$ 
15       $ySolucao \leftarrow y$ 
16    fim
17  fim
18 fim
19 retorna  $(xSolucao, ySolucao)$ 

```

Percebe-se que a quantidade de operações necessárias para se encontrar a melhor solução para o problema de Weber simples é  $O(x * y * k)$ , onde  $x$  é a quantidade de colunas do plano,  $y$  é a quantidade de linhas e  $k$  é a quantidade de pontos de demanda. Assim, o problema de Weber simples possui uma solução iterativa teoricamente eficiente. Não importa o tamanho do plano e nem a quantidade de pontos de demanda: a quantidade de operações cresce de forma polinomial.

Entretanto, é mais comum que problemas reais requeiram a distribuição de mais de uma facilidade no plano. Dessa forma, uma extensão do problema de Weber simples é um problema onde a quantidade de facilidades seja maior que 1. Essa extensão é chamada de problema de Weber com múltiplas fontes (PWM).

O problema PWM consiste em distribuir  $n$  facilidades em uma determinada região geográfica com os pontos de demanda  $K$  e alocar cada ponto de demanda à uma

facilidade específica. Esse problema pode ser descrito conforme (2-2).

$$v(PWM) = \min \sum_{k \in K} \sum_{j=1}^p p_k d_k(x, y) z_{kj} \quad (2-2a)$$

sujeito a

$$\sum_{j=1}^p z_{kj} = 1 \quad \forall k \in K, \quad (2-2b)$$

$$z_{kj} \in \{0, 1\} \quad \forall k \in K, j = 1, \dots, p, \quad (2-2c)$$

$z_{kj}$  indica se o ponto de demanda  $k$  está sendo atendido pela facilidade  $j$  ( $z_{kj}$  assume valor 1) ou não ( $z_{kj}$  assume valor 0), enquanto a restrição (2-2b) garante que somente uma facilidade irá atender um ponto de demanda.

Dessa forma, avaliar todas as possibilidades de distribuição de uma facilidade em um plano resultaria na execução de  $O((x * y * k)^n)$ . Assim, analisando a quantidade de facilidades  $n$ , pode-se perceber que a complexidade é exponencial.

O problema de Weber com múltiplas fontes, em especial, é interessante para o problema abordado nesse trabalho. De fato, ele é uma variação do problema de Weber com múltiplas fontes com a adição de restrições referentes à capacidade das facilidades.

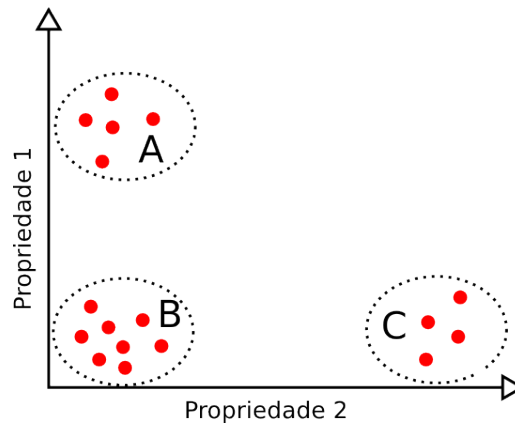
Evidentemente, uma solução que use uma abordagem determinística que avalie todas as possibilidades é inviável. Assim, algumas técnicas e abordagens são utilizadas para alcançar resultados aproximados do ótimo. Esse trabalho irá apresentar uma nova abordagem para esse tipo de problema.

## 2.2 Clusters e o método *k-means*

A mineração de dados (*data mining*) não é a única, mas certamente, a mais importante etapa na busca de conhecimento em uma base de dados [11]. Dentro da mineração de dados, existe um conjunto de funcionalidades (objetivos) que determinam qual será o tipo de informação que se deseja extrair. Uma dessas funcionalidades é a classificação.

A classificação é o objetivo utilizado na mineração quando se deseja encontrar classes nos dados. Essas classes, em geral, são definidas por padrões comuns entre os dados de grandes conjuntos de dados. Uma das técnicas utilizadas para realizar a classificação é a técnica de clusterização. Essa técnica consiste em encontrar *clusters* nos dados.

Um *cluster* é um conjunto de dados cujos registros são mais “próximos” uns dos outros do que em relação à outros registros de dados. Essa noção de distância e



**Figura 2.1:** Visualização de uma base de dados com destaque para conjuntos de dados aglomerados (*clusters*) que determinam as classes

proximidade tem mais sentido quando os dados são plotados em um gráfico. A Figura 2.1 demonstra graficamente essa noção de distância entre os dados.

Nessa figura, cada ponto no gráfico representa um registro de dados. Esse registro poderia ter vários atributos. Foram utilizados apenas dois para que fosse possível visualizar em um plano cartesiano. Alguns grupos de dados foram identificados considerando a proximidade entre os pontos de dados de cada grupo. Os dados desses agrupamentos (*clusters*) estão mais próximos entre si. Essa proximidade que torna os dados do grupo comuns entre si apontam o surgimento de uma classe. Na figura em questão, três classes estão evidenciadas. A classe *A* agrupa registros de dados cuja propriedade 1 é mais elevada, enquanto a classe *C* possui dados com a propriedade 2 em alta. A classe *B* possui registros que tenham as duas propriedades baixas. Assim, um novo registro de dados com os mesmos tipos de atributos (1 e 2) poderá ser classificado de acordo com as classes previamente encontradas durante o processo de descoberta dos *clusters*.

Existe uma relativa proximidade entre o problema de localização de *clusters* (identificação de classes) e o problema de localização de facilidades.

### 2.2.1 *Clusters* e o problema de localização

A distribuição de  $k$  facilidades em um plano é um problema de modelo contínuo. Em muitas abordagens, para resolver esse tipo de problema, um passo preliminar é encontrar os *clusters* existentes no plano. Por meio de força bruta, poderia ser definida uma abordagem para encontrar os *clusters* através de um algoritmo iterativo como o Algoritmo 2.1, gerando todas as possibilidades de localização das facilidades. Entretanto, essa atividade que avalia todas as possibilidades e encontra os  $k$  *clusters* seria impraticável, levando em consideração o grande número de comparações a serem realizadas.

Em 1967, MacQueen [20] introduziu o método *k-means*. Trata-se, na verdade, de um algoritmo classificador. O seu propósito é o de encontrar *clusters* nos dados, onde cada *cluster* encontrado representa uma classe. Esse método consiste em uma das técnicas de mineração de dados do tipo não-supervisionada. Isso significa que o *k-means* não precisa ser parametrizado e nem acompanhado. De modo geral, toda a informação necessária como entrada para esse método restringe-se ao conjunto de dados amostral. A partir desse conjunto, é possível realizar a classificação.

Mahajan, Nimbhorkar e Varadarajan [21] apresentaram uma prova de que o problema *k-means* no plano é NP-Difícil, realizando a redução através do problema 3-SAT<sup>2</sup>.

Portanto, para resolver esse problema, é necessário o uso de boas meta-heurísticas que sejam capazes de obter bons resultados em um tempo aceitável.

### 2.2.2 O método *k-means*

O método *k-means* foi concebido como uma abordagem não determinística. Essa característica se deve ao fato de que a distribuição inicial dos centroides<sup>3</sup> influencia muito no resultado.

Basicamente, para encontrar os *clusters*, o método busca minimizar a soma das distâncias entre os pontos de dados (ou pontos de demanda no escopo de problemas de localização) ao centroide mais próximo. Assim, a função de custo de modo bem simplificado sem apresentar todas as restrições poderia ser definida como (2-3).

$$\text{custo}(kmeans) = \sum_{k \in K} \sum_{j \in J} c_{kj} \quad (2-3)$$

$K$  é o conjunto de pontos de dados,  $J$  é o conjunto de centroides e  $c_{kj}$  é a distância entre os pontos de dados e o centroide mais próximo.

O método *k-means* pode ser resolvido com um algoritmo bastante simples, rápido e eficiente que busca encontrar os centroides dos *clusters* e que pode ser resumido nos passos conforme apresentado no algoritmo 2.2:

<sup>2</sup>O 3-SAT é um problema da classe NP-Completo [4].

<sup>3</sup>Um centroide  $k$  (tal que  $k \in \mathbb{R}^N$ ) é o ponto central em relação à um conjunto de outros pontos  $D$  (tal que  $D \subset \mathbb{R}^N$ ).

**Algoritmo 2.2:** *k-means***Entrada:** conjunto

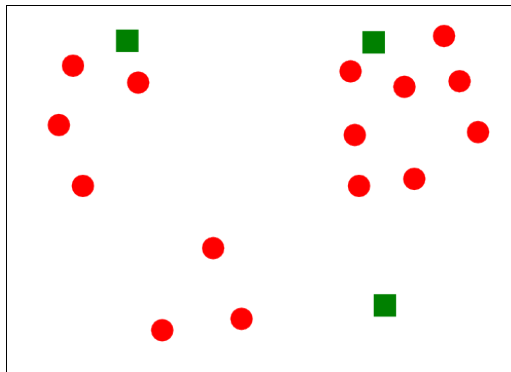
$X[(a_1, b_1, \dots, n_1), (a_2, b_2, \dots, n_2) \dots (a_x, b_x, \dots, n_x)]$  de dados dos quais deseja-se extrair as classes e inteiro  $k$  indicando a quantidade de *clusters* a serem encontrados

**Saída:** conjunto  $K[(a_1, b_1, \dots, n_1), (a_2, b_2, \dots, n_2) \dots, (a_k, b_k, \dots, n_k)]$  de centroides que determinam os *clusters*

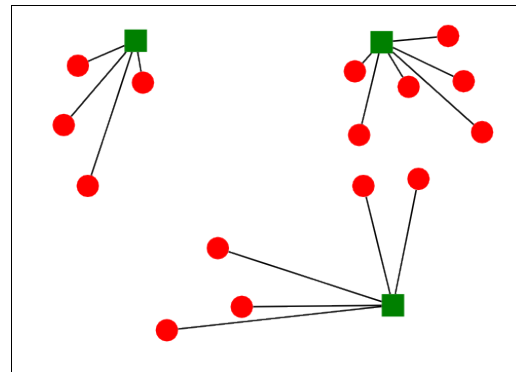
- 1 Gerar valores iniciais para os centroides. Nesse passo, cada um dos  $k$  centroides são distribuídos no espaço  $n$ -dimensional (geralmente de forma aleatória).
- 2 **repita**
- 3 Gerar uma matriz  $A_{n \times k}$ , onde  $n$  é a quantidade de registros de dados e  $A_{ij}$  tem o valor da distância (geralmente euclidiana) entre o registro  $i$  e o centroide  $j$ . Essa distância considera todos os atributos utilizados para determinar um ponto. No caso do problema de distribuição de facilidades no plano, esses atributos poderiam ser apenas dois: as coordenadas  $x$  e  $y$ .
- 4 Atribuir cada registro ao centroide mais próximo. Nesse ponto, o algoritmo termina se nenhum registro mudar de centroide em relação ao centroide a que estava atribuído anteriormente. Caso contrário, o algoritmo segue para o próximo passo.
- 5 Calcular os novos centroides (conjunto  $K$ ). Cada um dos  $k$  centroides é movimentado para o centro de cada *cluster*. Assim, cada atributo que define as coordenadas de um centroide, assume como valor, a média de todos os valores desse mesmo atributo dos registros a ele atribuídos.
- 6 **até** *Convergência*
- 7 **retorna** conjunto  $K$

A Figura 2.2 ilustra um exemplo simples do método *k-means* sendo utilizado para encontrar três *clusters* dentro do conjunto de dados.

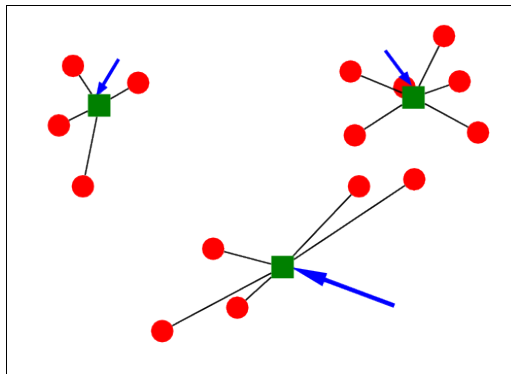
No primeiro passo (2.2(a)), os centroides são distribuídos segundo alguma estratégia. Pode ser utilizada alguma heurística referente ao problema, pode ser utilizada alguma abordagem que implemente algum algoritmo genético, ou mesmo pode ser utilizada uma distribuição simplesmente aleatória. A partir daí, todos os pontos de demanda são alocados ao centroide mais próximo 2.2(b). Em seguida, o centroide é movimentado



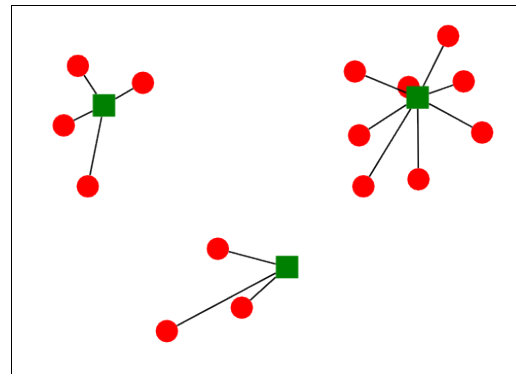
(a) Distribuição inicial - círculos vermelhos representam os pontos de demanda enquanto quadrados verdes representam os centroides



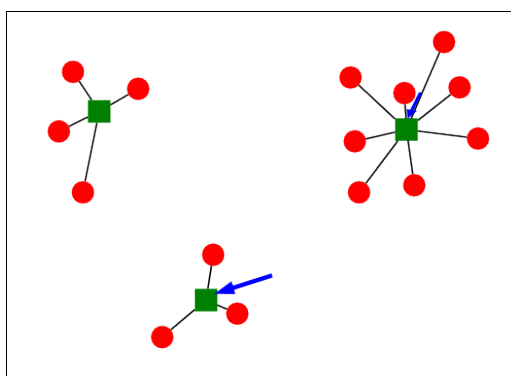
(b) Alocação dos pontos de demanda ao centroide mais próximo



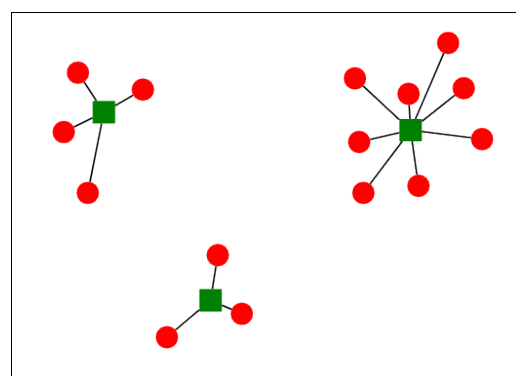
(c) Centroides se movem para o centro dos clusters



(d) Nova alocação dos pontos de demanda ao centroide mais próximo - alocação muda, logo, algoritmo continua a execução



(e) Centroides se movem para o centro dos clusters



(f) Algoritmo para, pois não existe mais mudança na alocação dos centroides aos pontos de demanda

**Figura 2.2:** Exemplo de execução do método *k*-means

para o centro do *cluster* 2.2(c). Agora os passos de alocação e movimentação do centroide vão sendo repetidos até que nenhum ponto de demanda mude de centroide na próxima alocação. Nesse caso, mais uma iteração será necessária, pois quando os centroides são realocados aos pontos de demanda, a alocação muda 2.2(d), sendo necessária, portanto, a movimentação dos centroides para o centro dos novos *clusters* 2.2(e). Por fim, uma nova alocação é realizada (pontos de demanda ao centroide mais próximo). Nesse ponto, o algoritmo percebe que não houve nenhuma mudança na alocação da iteração anterior para a atual. Assim, o método para e retorna a posição dos centroides como sendo a posição das facilidades.

Esse método converge bem rápido. Entretanto, nota-se que a qualidade do resultado está diretamente relacionada à distribuição inicial. Assim, o algoritmo *k-means* não garante a convergência para a melhor configuração de localização dos centroides. Existem muitas abordagens para encontrar a solução desse problema por meio de artifícios que busquem melhorar os resultados alcançados pelo método *k-means*. Kaveh, Zadeh e Sahraeian [25], por exemplo, sugerem um algoritmo para resolução de um problema de localização de facilidades que utiliza o *k-means* para gerar uma solução inicial e, para melhorar os resultados obtidos nessa solução inicial, o algoritmo *k-means* é executado várias vezes.

Contudo, mesmo sendo executado várias vezes, o *k-means* continua gerando soluções que convergem para um ótimo local, quando o desejado é um resultado que venha a convergir para um ótimo global. Assim, existem diversas abordagens propostas para a busca de uma convergência global. Como a distribuição inicial das facilidades é um fator que influencia diretamente no resultado da aplicação do método, várias dessas abordagens são focadas justamente nessa etapa.

Na seção 3.1 serão apresentados alguns trabalhos relacionados com essas melhorias no método *k-means*.

É muito importante deixar claro nesse ponto que o conceito de *cluster*, o método *k-means* e as abordagens estudadas para a solução desse tipo de problema classificador serviram de inspiração para a criação do algoritmo que foi desenvolvido durante a pesquisa e será apresentado nesse trabalho.

## 2.3 Processos de Poisson Modulados a Markov

Diversos problemas reais de natureza probabilística devem ser encarados considerando-se uma característica bastante recorrente: existem variações de média de ocorrências de um determinado evento de acordo com o tempo. Em geral, quando o problema atacado tem essa característica, fica inviável o uso de uma solução estatística simples.

Existe um problema voltado para a área de tráfego em redes ATM que é bastante discutido. Trata-se da análise do tráfego multimídia em um roteador, visando uma compreensão do comportamento do tráfego. Desse modo, seria possível realizar a mensuração de diversos aspectos sobre o tráfego, como o tamanho de uma demanda de fluxo de dados em um roteador ou o percentual de perda de pacotes.

Heffes e Lucantoni [10] apresentaram uma análise sobre o tráfego multimídia onde identificaram que o tráfego em roteadores tinha um comportamento que apresentava rajadas de dados. Assim, em alguns momentos, o roteador fica quase ocioso enquanto em outros momentos, grandes cargas de demanda são requeridas ao mesmo. Essas variações foram isoladas em faixas distintas e a conclusão foi de que cada tráfego medido podia ser representado por vários processos de Poisson. Dessa forma, o tráfego multimídia em roteadores teria uma distribuição de Poisson que, de acordo com o tempo, mudaria de taxa. Assim, identificar essas taxas e a correlação de mudança entre as mesmas poderia caracterizar o comportamento do tráfego.

De Cock e De Moor [3] formularam um modelo estocástico que utiliza várias taxas de Poisson para um mesmo tráfego. A probabilidade de ocorrência e transição entre taxas é descrita por uma cadeia de Markov. Esse modelo foi chamado de Processo de Poisson Modulado por Markov (*Markov Modulated Poisson Process - MMPP*).

Processos de Poisson são utilizados em diversos tipos de problemas. Um tipo em especial trata da detecção de falhas. Existem diversos trabalhos na área de detecção de falhas que apresentam abordagens com o uso de processos de Poisson. Trabalhos como os apresentados por Yamada et al. [30], Khoshgoftaar et al. [13] e Lo e Huang [18] demonstram o uso de processos de Poisson para detecção ou predição de falhas em sistemas computacionais.

Assim, devido ao fato de o gráfico de ocorrências de demanda ter apresentado um comportamento com variações de médias durante o tempo, foi suposto que essas médias pudessem ser médias de Poisson. Caso essa premissa se confirmasse, o modelo MMPP poderia ser usado. Por essa razão, o modelo MMPP, amplamente utilizado em estudos sobre tráfego multimídia, foi utilizado no contexto das ocorrências de falhas em redes elétricas, como será apresentado no Capítulo 5.

### 2.3.1 Descrição do modelo

O modelo MMPP proposto por De Cock e De Moor [3] define, basicamente que três parâmetros devem ser extraídos do conjunto de dados: a quantidade de taxas de Poisson  $N$ , o vetor de taxas médias de Poisson  $\lambda$  e a matriz de transição (matriz de Markov)  $P$ . Vale ressaltar que o tamanho do vetor  $\lambda$  é  $N$  e que o tamanho da matriz  $P$  é  $N \times N$ . Diz-se que a ordem do modelo é  $N$ .

O primeiro passo do modelo é determinar a sua ordem bem como as  $N$  taxas. Para isso, De Cock e De Moor propuseram uma formulação onde as taxas pudessem ser encontradas de acordo com a quantidade de ocorrências das mesmas. Assim, as mais representativas seriam consideradas taxas de Poisson e, por definição, a ordem do modelo também já seria encontrada.

Para tanto, utilizaram uma função de distribuição cumulativa para representar a probabilidade de ocorrência de qualquer valor no conjunto de dados medido. Na prática, isso foi representado por um vetor  $F$  de tamanho  $M$ , onde  $M$  é o valor da maior medida do conjunto de dados. Por tratar-se de uma função de distribuição cumulativa, todo elemento  $F_x$  possui a probabilidade de existir uma demanda de  $x$  ou menos. Deve-se ressaltar que, nesse modelo, os valores medidos precisam ser inteiros.

De Cock e De Moor afirmaram que, ao caracterizar o conjunto de dados medidos (no caso do trabalho deles, o tráfego multimídia em um roteador), os parâmetros encontrados deveriam ser capazes de gerar um conjunto de dados próximo ao tráfego medido novamente. Para isso, propuseram a seguinte formulação (2-4).

$$C\theta \approx F \quad (2-4)$$

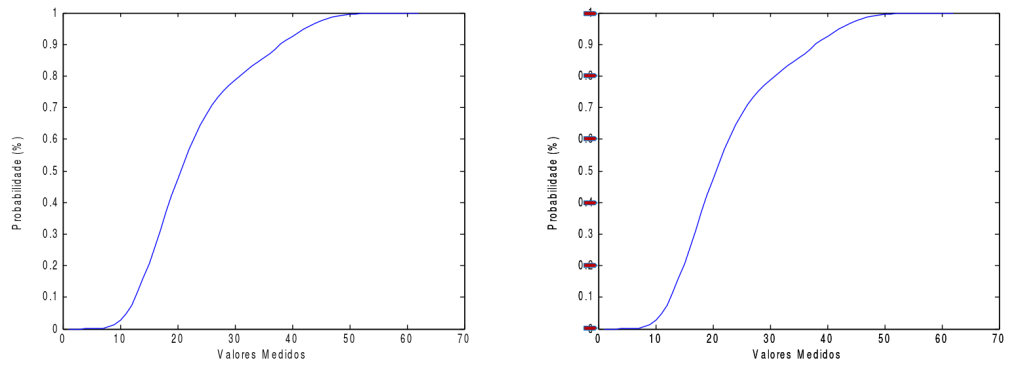
Nessa formulação,  $F$ , como já definido antes, representa a função distribuição cumulativa do tráfego medido. A matriz  $C$  proposta representa diversas funções de distribuição cumulativas calculadas, sendo cada uma definida em uma coluna da matriz. Esse cálculo é realizado considerando-se que cada coluna é um processo de Poisson com uma taxa única. Assim, cada elemento  $C_{ij}$  da matriz representa a probabilidade calculada de existir uma demanda de tamanho menor ou igual à  $i$ . Portanto, por tratar-se de um processo de Poisson, cada  $C_{ij}$  é definido pela formulação (2-5).

$$C_{ij} = e^{-\lambda_j} \sum_{l=0}^i \frac{\lambda_j^l}{l!} \quad (2-5)$$

No modelo MMPP, a matriz de Poisson é utilizada para descrever o comportamento de mudança entre um estado e outro. Cada estado dessa matriz é representado por uma taxa de Poisson do vetor  $\lambda$ . Voltando à formulação (2-4),  $\theta$  é um vetor que armazena as probabilidades de ocorrência dos estados (taxas de Poisson) calculados na matriz  $C$ , sendo que  $0 \leq \theta_j \leq 1$  para todo  $j$ . Assim, cada  $\theta_j$  representa a probabilidade de ocorrência de um estado  $j$ , ou seja, a probabilidade de a taxa ser  $\lambda_j$  a qualquer momento.

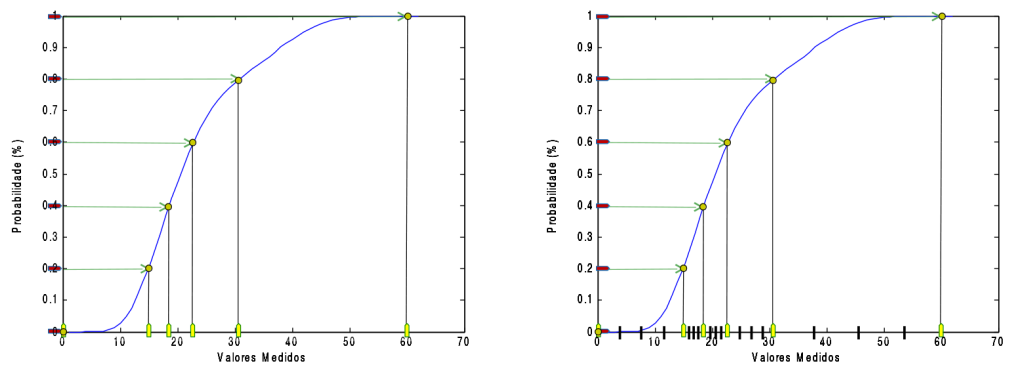
Assim, o método de De Cock e De Moor consiste em superestimar a ordem do modelo definindo prováveis taxas de Poisson. Para realizar essa superestimação, uma abordagem poderia consistir em dividir a diferença entre a maior e a menor ocorrência de demanda nos dados medidos pelo número superestimado de estados. Entretanto, essa divisão não iria considerar um fator importante que é a relevância de uma taxa.

Fazendo uma divisão em intervalos iguais, cada intervalo seria considerado com a mesma probabilidade. Para evitar isso, De Cock e De Moor sugerem o uso de uma divisão estatística. Essa divisão ocorre em intervalos iguais no eixo das probabilidades. Assim, esse valor seria mapeado para o eixo das taxas, definindo intervalos cuja probabilidade de ocorrência seria a mesma. A Figura 2.3 demonstra esse processo de divisão em uma função distribuição cumulativa.



(a) Função distribuição cumulativa dos valores medidos

(b) Divisão do eixo das probabilidades em cinco intervalos de 0,2 (20%) cada - para os cinco intervalos, são necessários seis pontos no mesmo eixo, sendo um deles com valor 0 e o outro com valor 1



(c) Mapeamento das divisões do eixo das probabilidades para pontos no eixo dos valores medidos

(d) Divisões de tamanhos iguais dentro do intervalo das taxas

**Figura 2.3:** Divisão estatística em uma função distribuição cumulativa das demandas ocorridas em janelas de tempo de uma célula da região sudoeste de Goiás no ano 2010

Nessa divisão, cada intervalo de valores medidos gerado no mapeamento (Figura 2.3(b)) possui os mesmos 20% de probabilidade de ocorrência, ainda que tenham tamanhos distintos. Dentro de cada intervalo de valores medidos, novas divisões de ta-

manhos iguais podem ser realizadas (Figura 2.3(d)) para aumentar ainda mais o tamanho superestimado da ordem do modelo. Essas divisões realizadas no eixo dos valores medidos se tornam os valores de taxas de Poisson do modelo a serem utilizadas para a criação da matriz  $C$ .

Assim, a matriz  $C$  pode ser descrita e restaria encontrar os valores de  $\theta$ . Encontrando os valores de  $\theta$ , as taxas cuja probabilidade de ocorrência fosse pouco representativa (limite parametrizado) são eliminadas, restando apenas as taxas que melhor descrevem o conjunto de dados. Assim, a ordem do modelo bem como as taxas médias de Poisson são encontradas. Entretanto, para isso, é necessário resolver o problema da formulação (2-4) dada a matriz  $C$  e o vetor  $F$ . Trata-se de um problema de mínimos quadrados não negativos (NNLS<sup>4</sup>). Existem bons métodos para a solução de problemas NNLS. Benthem e Keenan [28], por exemplo, apresentaram um algoritmo para a solução de grandes instâncias.

Uma vez resolvido o problema NNLS e as taxas pouco representativas terem sido descartadas, a ordem do modelo  $N$  e as taxas média de Poisson  $\lambda$  são obtidas. Assim, o último parâmetro da caracterização a ser obtido é a matriz de processos de Poisson.

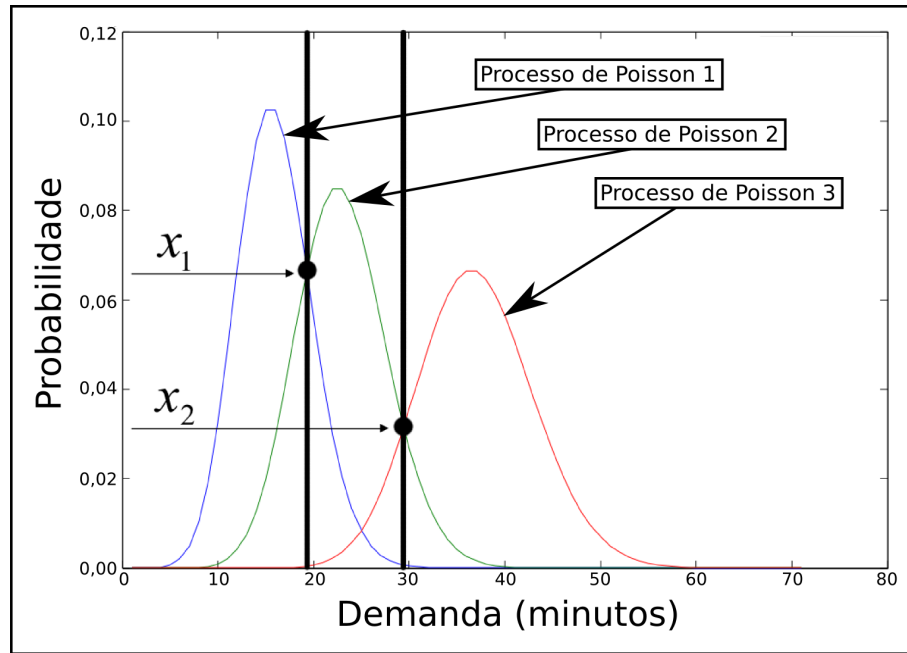
Ribeiro [26] propôs um algoritmo iterativo que varre o conjunto de dados medido de forma cronológica, contabilizando todas as mudanças de estado. Como cada elemento do conjunto de dados (janela de tempo) é um valor inteiro, resta saber em qual estado esse valor se encontra, dadas as possíveis taxas de Poisson. Assim, o primeiro passo é encontrar os limites que separam um estado do outro. Para isso, Ribeiro propôs que tais limites fossem os pontos onde os processos de Poisson vizinhos se cruzam. A Figura 2.4 explicita os limites  $x_1$  e  $x_2$  que existem para uma célula real da região sudoeste de Goiás no ano de 2010. Vale ressaltar que são dois limites, uma vez que são três estados nesse exemplo.

Assim, para esse exemplo, qualquer janela de tempo que tenha o valor abaixo de  $x_1$  estará no estado 1 do modelo caracterizado; se o valor estiver entre  $x_1$  e  $x_2$ , estará no segundo estado; e se for maior do que  $x_2$ , estará no estado 3.

A partir desses limites, o conjunto de dados passa por um processo iterativo onde o estado da janela é identificado e a matriz de Markov  $P$  é abastecida com a contagem das transições. Por exemplo, se  $F_1$  está no estado 1 e  $F_2$  está no estado 3, então o valor  $P_{1,3}$  é incrementado pois houve uma transição do estado 1 para o estado 3. Se  $F_3$  está no estado 2, logo,  $P_{3,2}$  é incrementado. Se  $F_4$  estiver no estado 2, logo  $P_{2,2}$  é incrementado, pois  $F_3$  estava no estado 2 e não houve mudança de estado na janela seguinte. Após a contagem, os valores de cada linha da matriz são transformados em probabilidades em função da soma dos valores de cada linha. Assim, a soma de cada linha da matriz  $P$  será sempre 1.

---

<sup>4</sup>NNLS - Non-negative least squares



**Figura 2.4:** Limites dos processos de Poisson

Com isso, a caracterização chega ao fim e, com os parâmetros calculados, é possível realizar uma suposição mais aproximada sobre como será o surgimento de novas ocorrências de demanda na região analisada nos cinco dias seguintes.

## 2.4 Discussão

Existem diversos modelos que podem descrever os diferentes casos de problemas de localização que as organizações enfrentam. Assim, conhecer cada um desses modelos é bastante importante antes de propor qualquer solução. Os diversos tipos de modelos não foram apresentados neste capítulo porque não são tão relevantes para a solução que será proposta. Klose e Drexel [15] apresentaram um trabalho sobre os conceitos e diversos modelos de problemas de localização.

Um ponto importante é que muitos trabalhos como os apresentados por Krishna e Narasimha [16], Lu et al. [19] e Likas et al. [17] fazem uso do método *k-means* em suas propostas de solução. Por essa razão, compreender esse método bem como seu propósito original se faz importante. Além disso, o método foi utilizado na proposta de solução apresentada neste trabalho para o problema de localização de facilidades com restrição de capacidade.

Compreender os conceitos fundamentais do assunto é o primeiro passo. Antes de buscar qualquer solução, é necessário que se atente para as soluções mais usuais que vêm sendo abordadas nos trabalhos referentes a esse tipo de problema. Assim, o capítulo seguinte pretende apresentar alguns trabalhos relacionados.

---

## Trabalhos Relacionados

---

Existem diversos tipos de problemas de localização de facilidades. Várias abordagens e algoritmos já foram propostos para resolver uma parte deles. Entretanto, é importante frisar que a solução para um problema de localização não necessariamente vai conseguir atender um outro problema de localização. As características individuais de cada um (tais como restrições, modelagem do problema, precisão da solução) podem ser decisivas na hora de criar uma abordagem.

Assim, neste capítulo serão apresentados alguns trabalhos relacionados aos problemas de localização de facilidades.

### 3.1 Melhorias no método *k-means*

Como apresentado na seção 2.2.2, o *k-means* é um dos métodos mais utilizados na clusterização. Além disso, é comumente aproveitado em diversas abordagens propostas para solucionar alguns problemas de localização.

Entretanto, devido ao fato de a clusterização ser NP-Difícil, o *k-means* não é um método determinístico. Dessa forma, suas saídas não são necessariamente as mesmas depois de executado várias vezes. De fato, o posicionamento inicial dos centroides é uma etapa determinante na qualidade dos resultados. Nas seções a seguir serão apresentadas duas abordagens para a melhoria do método.

#### 3.1.1 Algoritmos genéticos: GKA e FGKA

Como já mencionado, o *k-means* possui a característica de gerar soluções que buscam um ótimo local. Em 1999, Krishna e Murty [16] propuseram uma versão do *k-means* utilizando algoritmos genéticos, visando a geração de resultados que convergissem para um ótimo global. Nesse algoritmo, que foi batizado de GKA<sup>1</sup>, a codificação do cromossomo utilizada foi um vetor  $C$  de tamanho  $n$ , onde  $n$  é a quantidade de pontos

---

<sup>1</sup>GKA - *Genetic K-Means Algorithm*

e cada  $C_i$  assume um valor de 1 à  $k$ . Assim, cada cromossomo representa os pontos de dados e a qual centroide cada um está ligado. Essa codificação só é possível porque cada ponto só pode estar relacionado a um e somente um dos  $k$  centroides.

A população inicial é criada a partir de configurações aleatórias e execuções do *k-means* clássico.

O operador de seleção utiliza soluções aleatórias na geração atual baseando-se apenas no valor da função objetivo<sup>2</sup> selecionada.

O operador de mutação consiste na seleção de alguns pontos de uma solução que são somados ao complemento de outra solução, tendo assim, uma nova solução mutante. A partir daí, os  $k$  centroides são então recalculados novamente, baseando-se na sua nova configuração.

A Figura 3.1 ilustra um exemplo de como ocorre a mutação e como essa mutação gera uma nova solução.

Nessa figura, pode-se acompanhar o processo de mutação através do cruzamento entre dois cromossomos (Figuras 3.1(a) e 3.1(b)). Nesse caso hipotético, o cromossomo é representado com doze quadrados enfileirados. Cada um desses quadrados representa um gene, sendo que cada gene representa um dos doze pontos de dados. Cada gene é unicamente identificado com um índice que vai de 1 à 12. Dentro de cada quadrado (gene) existe um número que identifica qual é o centroide associado àquele ponto.

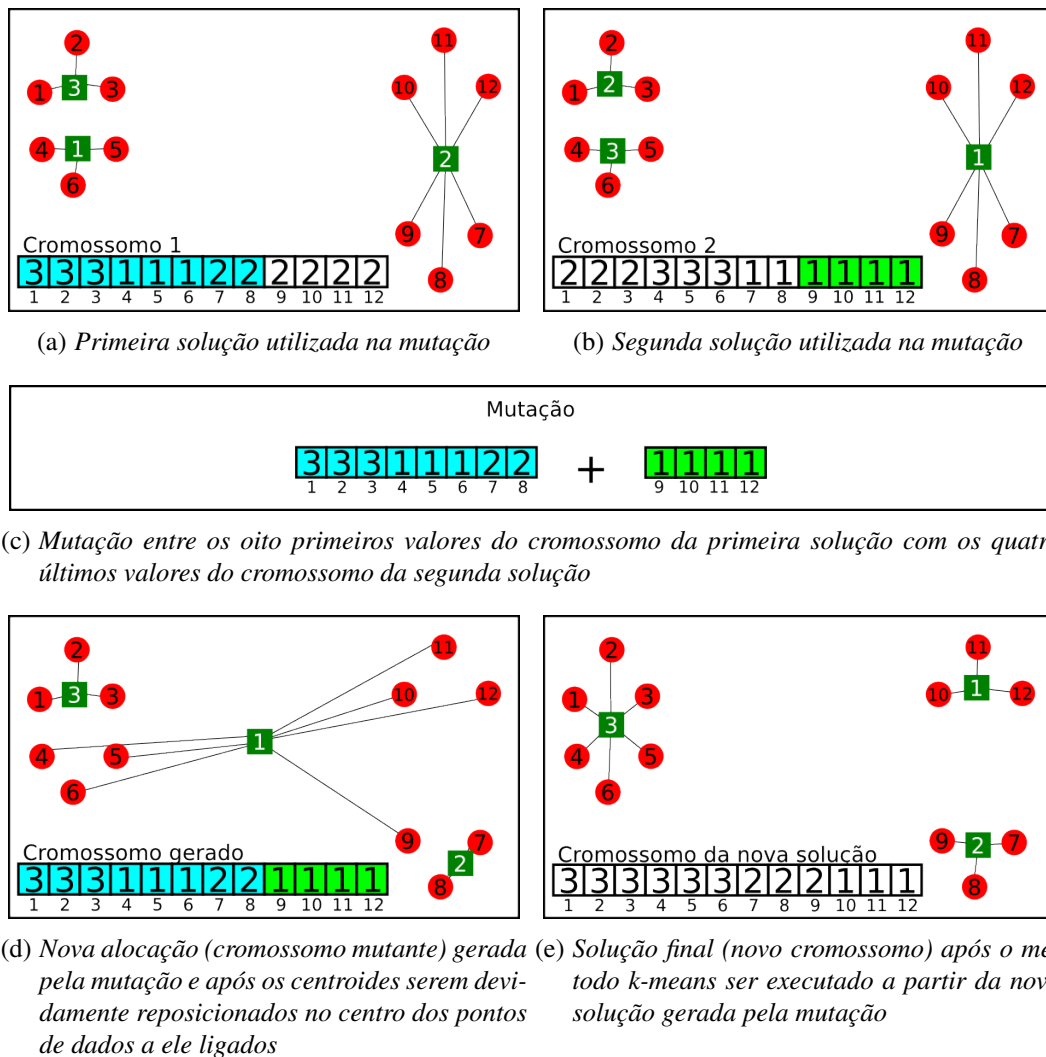
Cada cromossomo utilizado no processo de mutação deve ceder uma parte de si para compor o novo descendente. Na figura, o processo de mutação seleciona os oito primeiros genes do cromossomo 1. Para a geração de um novo indivíduo, devem ser selecionados os genes complementares do cromossomo 2. Portanto, devem ser selecionados exatos quatro genes do cromossomo 2 e nenhum deles poderá ser um dos oito primeiros. Então, os quatro últimos genes do cromossomo 2 são selecionados para a mutação (Figura 3.1(c)).

Após a criação do novo indivíduo, os centroides devem ser encontrados de acordo com a nova alocação (Figura 3.1(d)). O *k-means* é então executado e, quando converge, uma nova solução criada a partir da mutação entre duas soluções preliminares é finalmente obtida (Figura 3.1(e)).

Embora as representações visuais dessas duas soluções sejam idênticas, ambas possuem cromossomos distintos, uma vez que cada ponto de dados e cada centroide possui uma identificação única. Isso possibilitou uma mutação que acabou gerando uma nova solução. Nesse caso, a solução gerada foi melhor que as duas primeiras soluções de

---

<sup>2</sup>função objetivo é uma função que orienta a execução de um algoritmo de otimização. Pode ser usada com o intuito de maximizá-la ou minimizá-la, dependendo apenas do problema.



**Figura 3.1:** Exemplo de mutação no GKA

entrada, uma vez que a soma das distâncias entre os dados e os centroides foi diminuída. Entretanto, seria possível a geração de várias soluções piores ou equivalentes às iniciais.

A convergência do GKA é atingida depois de um número pré-definido de evoluções.

Yi Lu, Shiyong Lu, Fotouhi e DengBrown [19] criaram, em 2004, melhorias para o GKA e batizaram o novo algoritmo de FGKA<sup>3</sup>. Assim como seu antecessor, o FGKA também converge para um ótimo global mas, segundo os autores, o algoritmo faz isso muito mais rápido, graças às mudanças nos operadores de mutação e seleção.

No FGKA, o operador de seleção é semelhante ao do GKA. A diferença está na população analisada onde são selecionadas as soluções. Essa população é mantida através das gerações como sendo as melhores soluções. Assim, uma nova geração será criada a

<sup>3</sup>FGKA - *Fast Genetic K-means Algorithm*

partir de uma população que vem acumulando o melhor de cada geração. Nesse contexto, uma nova geração inteira poderá ser descartada.

O operador de mutação utiliza o mesmo tipo de cromossomo. Entretanto, os valores do cromossomo que são escolhidos para a mutação são selecionados de acordo com a existência de *clusters*, buscando inserir novos pontos de dados dentro de um outro *cluster*.

Em ambos os algoritmos (GKA e FGKA) podem ocorrer, após a mutação, a geração de uma nova solução que deixe um centroide sem nenhum ponto de dado associado. Nesse caso, uma solução assim recebe  $\infty$  como valor da função objetivo.

### 3.1.2 Método *k-means* repetitivo

Likas, Vlassis e Verbeek [17] propuseram um algoritmo com convergência global que possui um tempo de processamento consideravelmente mais longo que o *k-means* clássico, mas ainda assim, com um bom desempenho. Trata-se de um algoritmo iterativo que executará o *k-means* diversas vezes seguindo uma estratégia bem definida.

Primeiro, o *k-means* clássico é executado com  $k = 1$  para se encontrar a solução ótima para  $k = 1$ . Em seguida, o *k-means* é executado  $N$  vezes para  $k = 2$ , obedecendo a seguinte regra: a posição inicial do primeiro centroide é sempre a posição ótima para  $k = 1$  calculada anteriormente. A posição inicial do segundo centroide é aleatória para a primeira rodada do *k-means* e, a partir da segunda rodada, a posição inicial do segundo centroide passa a ser a localização final do mesmo centroide na rodada anterior.

Assim, à medida em que vão sendo adicionados centroides para que o algoritmo *k-means* execute  $N$  vezes com a nova quantidade de centroides, a estratégia de posicionamento inicial dos centroides continua a mesma: quando  $k = j$ , o algoritmo *k-means* irá ser executado  $N$  vezes e, nessas  $N$  vezes, os primeiros  $j - 1$  centroides assumem como posição inicial, a posição ótima encontrada para  $k = j - 1$ , enquanto o  $j$ -ésimo centroide inicia em uma posição aleatória na primeira rodada do *k-means* e, a partir da segunda rodada até a última, a sua posição é a posição ótima encontrada na rodada anterior.

Esse algoritmo termina sua execução quando executa a última das  $N$  iterações para a quantidade de centroides  $k$  desejada.

Uma grande vantagem desse algoritmo é que, além de resolver o problema *k-means*, encontrando uma solução que converge para um ótimo global, ainda possibilita resolver um outro problema: encontrar a quantidade de centroides necessárias para minimizar o custo total. Ou seja, encontrar o valor de  $K$ . O mesmo algoritmo aqui descrito poderia ser executado sem se limitar à quantidade  $K$ . Assim, o ponto de parada seria o valor do custo desejado. O retorno seria o valor  $K$  e o posicionamento desses centroides.

## 3.2 O problema das p-medianas capacitado

Uma restrição bastante recorrente em muitos problemas de localização de facilidades (incluindo o problema abordado nesse trabalho) é a capacidade que uma facilidade tem para atender os pontos de demanda. Os algoritmos clássicos para solução de problemas de localização geralmente não consideram a capacidade de atendimento e, dessa forma, podem gerar boas soluções, porém, inviáveis.

Dessa forma, a restrição de capacidade deve ser sempre levada em consideração, quando a mesma for uma característica do problema real. Exemplos clássicos de facilidades que possuem essa característica são as redes de telefonia e redes elétricas, onde cada facilidade (uma antena ou um transformador) tem uma capacidade que limita o tamanho da demanda que pode atender.

Portanto, para resolver problemas dessa natureza, algoritmos mais elaborados se fazem necessários.

Kaveh, Zadeh e Sahraeian [25] introduziram um algoritmo híbrido, que mescla o uso do algoritmo *k-means* com o algoritmo FNS (*Fixed Neighborhood Search*). Esse algoritmo tem o objetivo de resolver o problema das p-medianas capacitado no qual cada ponto de demanda pode ser alocado a somente uma facilidade.

O algoritmo proposto utiliza o algoritmo *k-means* para gerar uma solução inicial, que irá servir de entrada para o algoritmo FNS. Como já mencionado antes, o *k-means* é um algoritmo que não garante a solução ótima. Por essa razão, a abordagem utilizada executa o *k-means* 20 vezes iniciando o mesmo em posições aleatórias. Evidentemente, a quantidade de *clusters* que se deseja encontrar é  $p$  ( $K = p$ ). Para otimizar os resultados, garantindo que será encontrada a mediana, o *k-means* executado aqui tem a função de distância alterada. O valor utilizado para comparação é a distância elevada ao quadrado. Assim, garante-se que a mediana será encontrada.

Depois de executar o *k-means* 20 vezes, o algoritmo seleciona a melhor solução. Então, cada um dos  $k$  *clusters* (incluindo o centroide e os pontos alocados) são isolados. Esse isolamento é realizado para que seja executado um algoritmo qualquer de resolução do problema das p-medianas para  $p = 1$  em cada um dos *clusters* isolados. Esse procedimento visa encontrar o melhor posicionamento de cada centroide dentro do *cluster*. Nesse ponto, o algoritmo tem a sua solução inicial  $S$ .

No próximo passo, o algoritmo FNS é executado com a solução inicial  $S$ . Aqui, o conceito de vizinhança é definido como: “a  $k'$ -ésima vizinhança de uma solução são todas as soluções que diferem da solução corrente em exatamente  $k'$  facilidades.”

Dada uma função  $f(x)$  que retorna o valor do custo total de uma solução  $x$ , o algoritmo FNS pode ser descrito conforme os passos do Algoritmo 3.1.

**Algoritmo 3.1:**  $FNS(k', maxiter, N)$ 

**Entrada:** valor inteiro  $k'$  que vai determinar o tamanho da vizinhança;  $maxiter$  (que vai indicar quando o algoritmo irá parar); o conjunto de pontos de demanda  $N$

**Saída:** solução  $S$  contendo os pontos de localização das facilidades

```

1  $S \leftarrow$  solução inicial do problema dada pelo algoritmo  $k$ -means
2 enquanto Não convergir faça
3    $Viz_{k'}(S) \leftarrow$  a  $k'$ -ésima vizinhança de  $S$ 
4   para cada  $S' \in Viz_{k'}(S)$  faça
5     se  $f(S') < f(S)$  então
6        $S \leftarrow S'$ 
7        $r \leftarrow 1$ 
8       sai do laço “para cada”
9     fim
10    se  $f(S') > f(S)$  então
11       $r \leftarrow r + 1$ 
12      se  $r > maxiter$  então
13        retorna  $S$ 
14      fim
15    fim
16  fim
17 fim

```

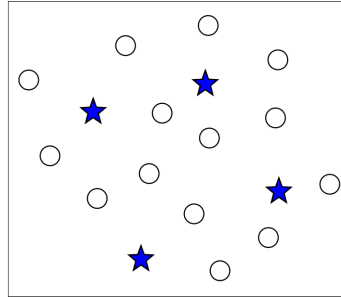
Esse algoritmo foi então adaptado para o problema das p-medianas capacitado para otimizar a escolha das novas soluções [25].

A primeira alteração foi na forma como os nós são selecionados no momento em que a  $k'$ -ésima vizinhança é gerada. No algoritmo FNS original (Algoritmo 3.1), todas as possibilidades seriam geradas. A modificação realizada exclui alguns pontos do espaço de busca. E isso pode ser feito de duas formas diferentes. Na primeira,  $k'$  pontos são removidos da solução corrente e a seleção dos novos  $k'$  pontos irá considerar todos os demais pontos que não possuem facilidade instalada, com exceção dos  $h$  nós mais próximos de cada um dos  $k'$  pontos selecionados. O valor de  $h$  deve ser parametrizado empiricamente, por tentativa e erro.

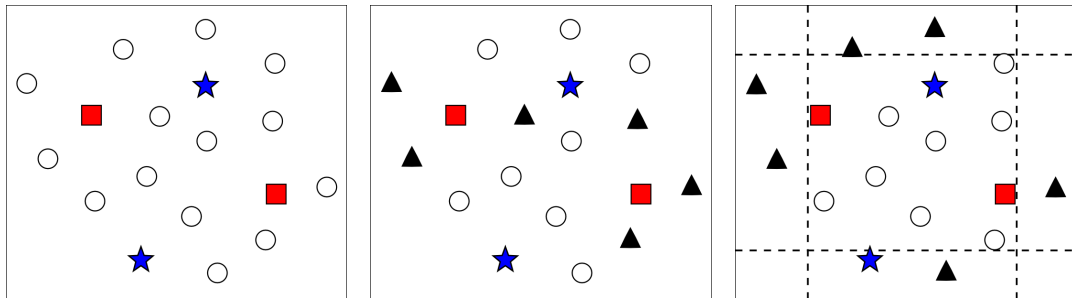
A segunda alternativa conta com uma heurística que elege os pontos mais periféricos para a exclusão, uma vez que, segundo essa heurística, esses tendem a não serem escolhidos para a solução final. Aqui, mais uma vez, os limites de periferia devem

ser parametrizados empiricamente.

A Figura 3.2 mostra o resultado da abordagem original (Figura 3.2(b)) e dessas duas alternativas propostas para a exclusão de pontos na geração da vizinhança (Figuras 3.2(c) e 3.2(d)).



(a) Solução corrente. As estrelas representam as facilidades instaladas - duas delas serão selecionadas para serem removidas



(b) Duas facilidades foram removidas (quadrados). A vizinhança passa a ser gerada a partir de todos os demais pontos (círculos)  
 (c) Para  $h = 3$ , são selecionados os três pontos mais próximos para serem excluídos do espaço de solução (triângulos)  
 (d) As linhas pontilhadas delimitam as margens periféricas, onde estarão os pontos excluídos

**Figura 3.2:** Identificação de nós candidatos à substituição na geração da vizinhança para  $k' = 2$  - são selecionadas duas facilidades para serem removidas

A segunda alteração no algoritmo FNS foi o critério de parada. Antes, o critério de parada era quando atingia a quantidade de iterações realizadas sem melhoria no custo total. Agora, todas as  $k'$  soluções da vizinhança deverão ser analisadas e o algoritmo só para se nenhuma dessas soluções for melhor que a solução corrente.

Por fim, a última alteração realizada foi a adição de memória ao algoritmo, acrescentando uma lista tabu<sup>4</sup> para armazenar soluções e evitar o recálculo de soluções já

<sup>4</sup>Lista tabu é uma estrutura utilizada em algoritmos de busca baseados em busca tabu. Na busca tabu, o algoritmo realiza a busca de soluções vizinhas à solução atual. Eventualmente, uma solução poderá ser revisitada, gerando, assim, ciclos que podem impossibilitar uma maior exploração do universo de soluções. Nesse caso, uma lista contendo algumas das últimas soluções visitadas é armazenada e consultada a cada novo passo da busca. Essa é a lista tabu.

visitadas.

### 3.3 Desenho de grafos

Quando um determinado problema de localização/alocação é estudado, comumente, são utilizadas visualizações de algumas instâncias do problema com a finalidade de descobrir padrões ou extrair informações que auxiliem no objetivo de encontrar uma solução. Existe um tipo de visualização específica e mais comum que é a de pontos espalhados no plano (representando os pontos de demanda). Esses pontos possuem localização estática, não podendo a mesma ser alterada. Adicionalmente, as facilidades podem ser representadas por outros pontos. Como a localização das facilidades é justamente o problema a ser resolvido, esses pontos podem estar em qualquer lugar dentro do espaço de soluções do problema. São, portanto, pontos dinâmicos. Ao realizar a alocação dos pontos de demanda à(s) facilidade(s) mais conveniente(s) para a melhoria da solução do problema, traçar uma reta entre facilidade e ponto de demanda é uma forma natural de se demonstrar a relação entre os dois pontos. Assim, o que se tem no final das contas como representação visual, é a imagem de um grafo.

Existe um grande interesse sobre a área de desenho de grafos que é demonstrada em uma imensa quantidade de trabalhos realizados a respeito. Di Battista et al. [5] apresentaram uma bibliografia bastante rica até o ano de 1994 reunindo diversas abordagens e algoritmos criados especificamente para o desenho de grafos. Muitos outros trabalhos foram desenvolvidos desde então. Em geral, esses trabalhos representam melhorias ou adaptações de outros trabalhos já desenvolvidos.

Diversos são os fatores que podem influenciar no modo como deseja-se desenhar um grafo, tais como: a tipografia (se é uma árvore, ou um circuito, ou uma forma geométrica), algumas restrições (como cruzamento de arestas, proximidade entre vértices, aglomerações de *clusters*, etc.).

Eades [7] propôs um método que usa um conceito de sistema de forças. Devido à sua natureza de distribuição de forças de forma equilibrada e devido à necessidade de se obter uma distribuição de facilidades onde a capacidade é distribuída de forma equilibrada, esse método também serviu de inspiração para a solução do problema de localização de facilidades capacitado desenvolvido nesse trabalho. Assim, será dada uma atenção especial a ele a seguir.

#### 3.3.1 O método *Springs* de Eades

A ideia geral do algoritmo *Springs* de Eades é a de criar um sistema de forças. Nesse sentido, é proposto no algoritmo que cada vértice seja substituído por um anel

de ferro e cada aresta por uma mola. Cada mola exerce uma força de atração e uma de repulsão sobre os vértices que ela conecta. Quanto mais os vértices estão distantes, maior é a força de atração (mola tende a se comprimir). Quanto mais próximos, maior é a força de repulsão (mola tende a se esticar). Além disso, os vértices não adjacentes se repelem. Assim, o algoritmo realiza a movimentação dos vértices de forma natural, buscando o equilíbrio entre todas essas forças.

O algoritmo Springs de Eades funciona com bom desempenho e retorna desenhos satisfatórios para grafos com um número menor do que 30 vértices. É eficiente em estruturas de *grids* regulares, árvores e grafos esparsos.

É importante notar que a quantidade de parâmetros a serem informados para esse método é pequena: bastam apenas os fatores de elasticidade e compressão das molas, bem como o fator de repulsão entre vértices não adjacentes.

As forças existentes nesse sistema proposto por Eades são específicas para atender ao propósito de desenhar grafos de forma uniforme. A elasticidade e compressão das molas são variáveis que visam um espaçamento balanceado entre os vértices adjacentes. A força de repulsão existente entre vértices não adjacentes resultam em uma explicitação visual dos grupos de vértices adjacentes.

Dessa forma, pode-se supor que o sistema de forças proposto por Eades pode ser expandido para outros propósitos. Definindo novas forças, um sistema assim poderia criar um desenho que respeitasse outras métricas e regras. E é baseada nessa suposição que a solução proposta nesse trabalho acabou se tornando um sistema de forças que desenha a solução final.

### 3.3.2 Discussão

Como já mencionado, o *k-means* é ferramenta de várias abordagens para solução de problemas de localização de facilidades. Para problemas não-capacitados (onde a capacidade não é uma restrição), encontrar os *clusters* acaba sendo a solução ideal para uma boa parte dos problemas sem essa restrição. Por essa razão, as abordagens apresentadas na seção 3.1 são muito úteis para problemas dessa natureza, uma vez que melhoram a qualidade dos resultados sem uma perda expressiva de performance.

Entretanto, para problemas de localização de facilidades capacitados, encontrar os *clusters* pode não ser suficiente. A simples busca de *clusters* para realizar a instalação de facilidades com restrição de capacidade pode resultar em *clusters* muito mais densos do que outros. Dessa forma, existem grandes chances de uma facilidade ser instalada e alocada em um *cluster* que não consegue atender, enquanto outra facilidade seria instalada em um lugar onde fosse subutilizada. Assim, abordagens como a apresentada na seção 3.2 são importantes para solucionar problemas dessa natureza.

A solução criada nesse trabalho e abordada no Capítulo 4 é voltada para o problema capacitado modelado no plano. O posicionamento de cada facilidade depende muito do valor de demanda dos pontos de demanda. Além disso, a capacidade influencia até mesmo em qual *cluster* estará localizada a facilidade. Com essa visão, o trabalho de Eades sobre o desenho de grafos se torna interessante, uma vez que a localização de facilidades pode ser visualizada como um problema de desenho de grafo.

Portanto, os trabalhos relacionados nesse capítulo, de alguma forma, tiveram influência, direta ou indireta, na criação da solução que será apresentada. O capítulo a seguir irá abordar detalhes do problema e da abordagem desenvolvida.

---

## O Problema de Localização de Facilidades com Restrição de Capacidade em um Modelo Contínuo

---

Em muitas organizações, a distribuição de recursos em uma região geográfica é uma atividade de extrema importância, que pode definir o sucesso ou o fracasso da mesma. Em alguns casos, as abordagens existentes (dentre elas, algumas já apresentadas neste trabalho) já são suficientes para garantir uma boa distribuição desses recursos. Entretanto, em muitos casos, uma variável adicional é fundamental e não pode ser ignorada durante a localização do recurso: a capacidade do mesmo.

Esse tipo de problema é conhecido como problema de localização de facilidades capacitado (PLFC). De modo geral, esse tipo de problema também lida com a alocação das facilidades aos pontos de demanda, uma vez que cada facilidade terá uma quantidade limitada de demanda que poderá atender. Assim, o problema passa a ser um problema de localização/alocação de facilidades.

Cada problema com suas peculiaridades pode ser descrito de uma forma. Alguns problemas de localização capacitados podem levar em consideração, por exemplo, o custo de instalação (que pode variar de um lugar para o outro); podem exigir que um ponto de demanda seja atendido por uma e somente uma facilidade.

O problema abordado neste trabalho pode ser descrito pelo modelo matemático (4-1).

$$v(PLFC) = \min \sum_{k \in K} \sum_{j \in J} c_{kj} z_{kj} \quad (4-1a)$$

sujeito a

$$\sum_{j \in J} z_{kj} = 1 \quad \forall k \in K, \quad (4-1b)$$

$$\sum_{k \in K} d_k z_{kj} \leq s_j \quad \forall j \in J, \quad (4-1c)$$

$$\sum_{j \in J} s_j \geq d(K) \quad (4-1d)$$

$$z_{kj} \in \{0, 1\} \quad \forall k \in K, \forall j \in J \quad (4-1e)$$

Sendo  $K$  o conjunto dos pontos de demanda,  $J$  o conjunto das facilidades, o modelo matemático (4-1) busca minimizar a soma de todos os custos referentes ao atendimento de um ponto de demanda pela facilidade que o atende. Neste trabalho, o custo  $c_{kj}$  em questão é a distância ao quadrado entre o ponto de demanda  $k$  e a facilidade  $j$  que o atende. Para tanto, cada custo  $c_{kj}$  será considerado somente quando a facilidade  $j$  estiver atendendo ao ponto de demanda  $k$ . Quando isso acontecer  $z_{kj}$  terá valor igual a 1, indicando essa relação. Caso contrário, o valor será 0.

A restrição (4-1c) estabelece que a soma de toda a demanda  $d_k$  dos pontos de demanda atendidos por uma facilidade  $j$  não pode ultrapassar a capacidade  $s_j$  da mesma facilidade.

A restrição (4-1d) existe para garantir que a demanda total do problema jamais será maior do que a capacidade de todas as facilidades juntas. Em algumas instâncias de um problema assim, pode ser que essa restrição seja quebrada. Em casos assim, faz-se necessário um ajuste nos valores da demanda de forma proporcional, de modo que a soma de toda a demanda não ultrapasse a soma das capacidades das facilidades.

Em conjunto, as restrições (4-1b) e (4-1e) indicam que um ponto de demanda deve ser atendido em sua totalidade por uma e somente uma facilidade.

## 4.1 A relaxação do problema

Com a finalidade de viabilizar a aplicação da abordagem criada durante a pesquisa e que será apresentada neste trabalho, a restrição (4-1e) será relaxada, de modo que a mesma será substituída pela restrição  $z_{kj} \in [0, 1]$ . Dessa forma, o novo problema permite que um ponto de demanda seja atendido por uma facilidade de forma parcial.

É importante perceber que a restrição (4-1c) permanece sem alteração, de modo que a capacidade de uma facilidade deve sim ser respeitada.

Assim, o novo problema de localização capacitado que é abordado neste trabalho, passa a ser descrito de acordo com o modelo matemático (4-2).

$$v(PLFC) = \min \sum_{k \in K} \sum_{j \in J} c_{kj} z_{kj} \quad (4-2a)$$

sujeito a

$$\sum_{j \in J} z_{kj} = 1 \quad \forall k \in K, \quad (4-2b)$$

$$\sum_{k \in K} d_k z_{kj} \leq s_j \quad \forall j \in J, \quad (4-2c)$$

$$\sum_{j \in J} s_j \geq d(K) \quad (4-2d)$$

$$z_{kj} \in [0, 1] \quad \forall k \in K, \forall j \in J \quad (4-2e)$$

Essa relaxação foi realizada para evitar uma variação do problema da soma de subconjuntos, que, conforme apresentado por Murty et al. [23], é NP-completo. Esse problema tornaria inviável a criação de uma solução aceitável. Se a restrição (4-1e) fosse mantida, o problema seria encontrar subconjuntos entre os pontos de demanda de modo que a demanda total de cada um desses subconjuntos fosse menor ou igual à capacidade da facilidade que os atende. Entretanto, além de ser NP-Completo a atividade de separar tais conjuntos, ainda existiria o complicador de ter que buscar subconjuntos de pontos de demanda aglomerados. Afinal de contas, se o posicionamento de uma facilidade leva em consideração a localização dos pontos de demanda por ela atendidos, tais pontos de demanda não podem estar espalhados por toda a região.

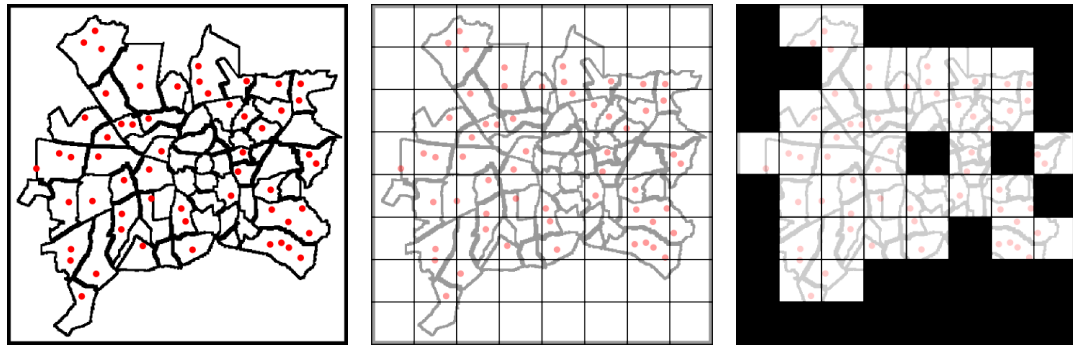
## 4.2 A Modelagem

Existem diversos modelos que podem ser empregados em problemas de localização de facilidades [15]. Neste trabalho, busca-se apresentar uma solução para problemas de modelo contínuo. Dessa forma, a distribuição de facilidades irá considerar toda a região geográfica como um plano cartesiano. Assim, uma facilidade poderá ser localizada em qualquer ponto desse plano.

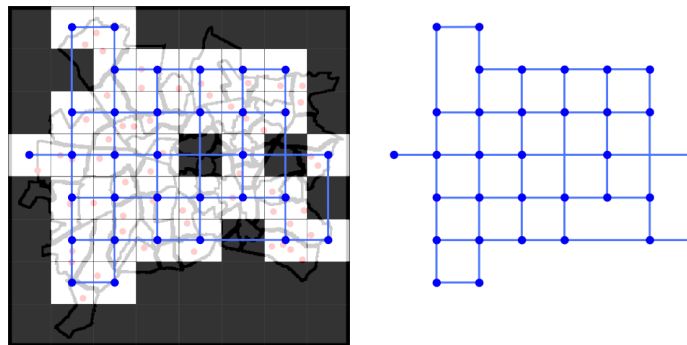
Muitos problemas podem ser mapeados para um modelo de rede. A Figura 4.1 ilustra uma forma de converter um problema genérico qualquer que existe em um espaço geográfico planar em uma representação na forma de grafo, para ser usada em um modelo de rede.

É importante notar que cada caso pode ser modelado de acordo com as características do problema. A quantidade de células bem como a restrição de tamanhos serem iguais podem ser alteradas conforme a necessidade. A política de atribuição das arestas também pode ser adaptada, permitido, por exemplo, a criação de arestas diagonais ou proibindo a criação de arestas que passem por células sem pontos de demanda.

Uma outra possível abordagem poderia considerar identificar vários *clusters* (um número relativamente bem maior que o número de facilidades) no mapa e mapear cada



- (a) Configuração inicial do problema no plano com a representação dos pontos de demanda
- (b) Mapa é dividido em células de tamanhos iguais (nesse caso, optou-se por dividir em uma grade de  $8 \times 8$ )
- (c) São identificadas as células que não possuem nenhum ponto de demanda e as mesmas são excluídas do espaço de solução



- (d) Cada célula que possui pelo menos um ponto de demanda recebe um vértice (ponto) e os vértices gerados são ligados por arestas conforme alguma regra de fronteira entre as células (nesse caso, optou-se como regra, considerar apenas fronteiras nos lados esquerdo, direito, de cima e de baixo)
- (e) O grafo gerado passa a ser o resultado da conversão, onde cada aresta tem como valor, a distância euclidiana do centro de cada célula ao centro da célula vizinha

**Figura 4.1:** Passos para um possível mapeamento de pontos de demanda em um mapa geográfico plotado em um plano cartesiano para uma representação em grafo.

centroide como um vértice. As arestas poderiam ser atribuídas conforme alguma regra estabelecida, como, por exemplo, a parametrização de uma distância máxima entre dois *clusters* para a atribuição de uma aresta entre os mesmos.

Modelos de rede utilizam soluções que, em geral, retornam o resultado em um tempo consideravelmente menor do que uma solução similar para um modelo contínuo. Isso se evidencia diante do fato que em modelos de rede existe o conceito de pontos candidatos. Para  $k$  facilidades, o conjunto de pontos candidatos  $PC$  (onde  $|PC| \leq k$ ) agrega os pontos que podem vir a se tornar a localização de alguma facilidade. Assim, com um número de possibilidades bastante reduzido, encontrar uma boa solução passa a demandar menos tempo.

Entretanto, nem todos os problemas de localização possibilitam esse tipo de mapeamento. Ou, em alguns casos, até poderia possibilitar, mas isso aconteceria em detrimento da qualidade dos resultados. Assim, neste trabalho o modelo abordado foi o contínuo.

### 4.3 Histórico de demanda

O problema de localização de facilidades em uma região geográfica será abordado na seção 4.4. Existe um problema preliminar com o qual, a maioria das organizações que desejam realizar uma distribuição de facilidades se depara: a mensuração dos pontos de demanda.

Analisar um contexto e mensurar a localização de pontos de demanda a partir de dados históricos é um problema à parte que renderia uma outra pesquisa. Assim como para o problema de localização de facilidades, não existe uma solução genérica para contemplar a tarefa de encontrar os pontos de demanda. De fato, esse tipo de problema vai sempre depender das características individuais de cada contexto.

Mas ainda que rendesse outra pesquisa, o problema de encontrar os pontos de demanda a partir de dados históricos foi tratado durante os trabalhos desta pesquisa. A razão disso é que foi realizado um estudo de caso a ser apresentado no Capítulo 5 onde foi necessária a realização da mensuração dos pontos de demanda. Entretanto, como o foco principal está na distribuição das facilidades, esse problema não foi amplamente estudado.

Encontrar esses pontos de demanda começa antes, com a definição de como eles serão definidos. A alternativa escolhida neste trabalho consiste na definição de células no plano. Cada célula se torna um ponto de demanda e, o posicionamento desse ponto de demanda é definido pelo centroide entre todas as ocorrências de registros que geraram alguma demanda. Assim, deve-se definir como cada célula será criada.

Uma alternativa seria utilizar cada bairro como uma célula. Partindo desse princípio, restaria calcular a demanda de cada bairro. Entretanto, para uma cidade com bairros muito grandes e outros muito pequenos, o mapa de pontos de demanda gerado poderia não representar bem o real contexto da organização. Assim, uma outra abordagem consiste em definir células retangulares de tamanho igual, de modo que cada célula irá representar um ponto de demanda. A localização - ponto  $(x,y)$  - do ponto de demanda vai ser o ponto centroide entre todos os pontos de demanda daquela célula.

Por fim, o cálculo do valor de demanda de cada ponto passa a ser o problema mais complexo. Isso se deve ao fato de que, ainda que dados históricos demonstrem um fiel acompanhamento de todas as ocorrências de demanda em uma determinada região, descrever formalmente o comportamento do problema e de como surgem as ocorrências não é uma atividade trivial.

Neste trabalho, duas formas foram abordadas para se encontrar o valor de demanda de uma determinada célula. A primeira consiste em uma estatística simples, utilizando a contagem de ocorrências de demanda. A segunda, e mais promissora, analisa o problema como conjuntos de processos de Poisson modulados por cadeias de Markov. Essa abordagem será explorada no Capítulo 5.

### 4.3.1 Estatística Simples

Diante do fato que o foco do trabalho está na localização de facilidades considerando o prévio conhecimento sobre os pontos de demanda, a primeira abordagem utilizada para defini-los foi o uso de estatísticas simples.

Um período de tempo previamente selecionado define as ocorrências de demanda que serão consideradas em cada célula. A partir daí, cada célula recebe como demanda, o somatório de todas as demandas das ocorrências que surgiram dentro da própria célula. Essa demanda individual de cada ocorrência poderá ser calculada conforme o interesse da organização.

Essa abordagem é muito parecida com a abordagem de conversão de um problema planar para um problema em rede apresentado na Figura 4.1. A única diferença é que, no caso da definição dos pontos de demanda, nesse caso não seria necessária a criação das arestas.

Um algoritmo iterativo simples pode realizar esses cálculos. Entretanto, o resultado final geralmente é pobre, uma vez que essas estatísticas não dizem muito a respeito do comportamento de como surgem as ocorrências de demanda de um determinado problema. Análises realizadas em dados históricos de um estudo de caso que será tratado no Capítulo 5 apontaram que, em determinados momentos, surgem mais ocorrências de demanda em uma região do que em outros momentos. Assim, essa análise que apenas con-

sidera a contagem das ocorrências de demanda acaba deixando de lado um fator bastante importante: o tempo.

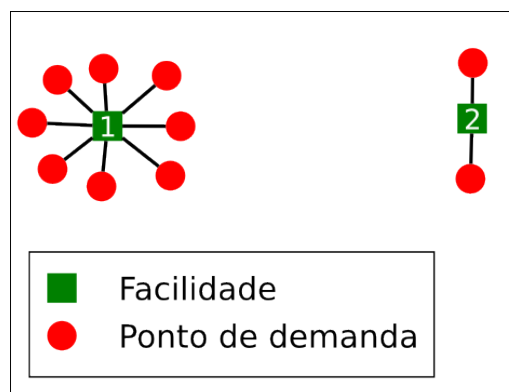
Dessa forma, uma abordagem mais aprofundada sobre os dados históricos se faz necessária para que se tenha uma melhor qualidade nos resultados.

## 4.4 A proposta de solução

Existem diversos trabalhos apresentando formas de realizar a distribuição de facilidades em um plano de modo a otimizar os parâmetros desejados. A princípio, o problema de localização de  $k$  facilidades pode se resumir em localizar os  $k$  centroides no plano. Esse problema pode ser resolvido de forma relativamente rápida pelo método  $k$ -means e suas adaptações para melhores resultados, conforme exemplificado nas seções 3.1.1 e 3.1.2.

Contudo, é comum que problemas de localização de facilidades não lidem apenas com a variável distância. Geralmente, problemas de localização levam em consideração algumas restrições, sendo a capacidade das facilidades uma das mais recorrentes.

A Figura 4.2 apresenta um cenário onde duas facilidades são localizadas no plano utilizando como função de custo apenas a soma das distâncias entre pontos de demanda até a facilidade mais próxima.



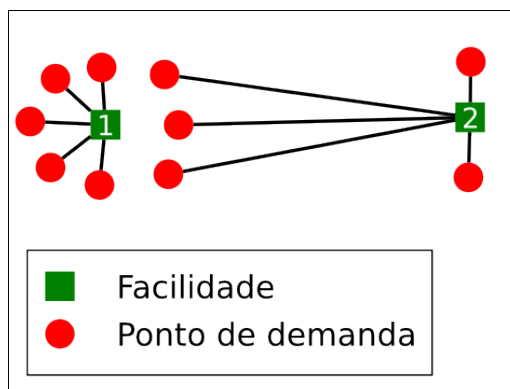
**Figura 4.2:** Cenário com duas facilidades localizadas utilizando o método  $k$ -means que encontra os centroides de  $k = 2$  clusters

Pode-se observar que, ainda que uma das facilidades esteja com uma quantidade muito maior de pontos de demanda alocados do que a outra, a função de custo apresentada em (2-3) está sendo minimizada. Isso se deve ao fato de que o  $k$ -means é um algoritmo classificador e sua função é encontrar aglomerações de dados que denotem um padrão ou uma classe.

Para um problema de localização de facilidades sem restrição de capacidade, certamente essa seria a melhor distribuição. Entretanto, o que deseja-se neste trabalho é que essa restrição não seja ignorada.

Agora, será suposto que na mesma Figura 4.2, cada um dos pontos de demanda possui valor de demanda igual a  $1ud^1$ . Assim, a facilidade  $j_1$  estaria alocada a oito pontos de demanda ( $8ud$ ) enquanto a facilidade  $j_2$  a dois pontos ( $2ud$ ). Se a capacidade da facilidade  $j_1$  for maior ou igual a  $8ud$  e a capacidade da facilidade  $j_2$  for maior ou igual a  $2ud$ , então a distribuição ainda estaria válida (e, nesse caso, seria a melhor).

Entretanto, supondo que ambas as facilidades 1 e 2 possuam capacidade de atender exatamente  $5ud$  cada uma, a facilidade  $j_1$  estaria claramente sobrecarregada ( $3ud$  a mais do que pode atender), enquanto a facilidade  $j_2$  estaria com sua capacidade subutilizada ( $3ud$  a menos do que poderia atender). Assim, parte da demanda que estava alocada à facilidade  $j_1$  ( $3ud$  nesse caso) passaria, obrigatoriamente, a ter que ser atendida pela facilidade  $j_2$ . O mesmo cenário, teria uma alocação semelhante à apresentada na Figura 4.3.

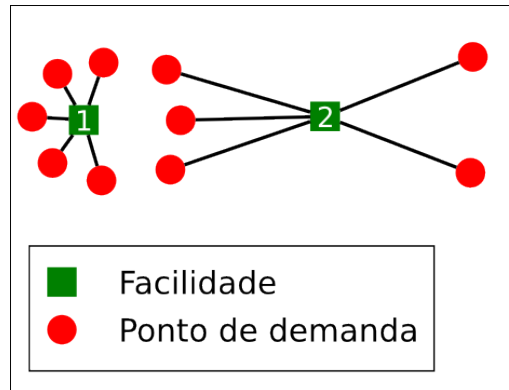


**Figura 4.3:** Cenário com duas facilidades de capacidade  $5ud$  localizadas utilizando o método *k-means* e alocadas a pontos de demanda com valor de demanda de  $1ud$  cada

Nesse novo cenário, a nova alocação das facilidades aos pontos de demanda balancearam a carga do sistema. Contudo, o posicionamento das facilidades não parece ser o ideal para essa nova alocação. Assim, uma movimentação das facilidades se faz necessária. A Figura 4.4 apresenta o estado final de localização das facilidades considerando a restrição de capacidade.

Nesse cenário final, as facilidades foram movimentadas de modo que pudessem atender melhor os pontos de demanda, considerando a capacidade de cada uma. Nota-se que a facilidade  $j_2$  movimentou-se para a direção de três pontos de demanda mais próximos que, possivelmente, ficariam sem serem atendidas pela facilidade  $j_1$ . Entretanto,

<sup>1</sup>Unidade de demanda



**Figura 4.4:** *Cenário com duas facilidades de capacidade  $5ud$  alocadas a pontos de demanda com valor de demanda de  $1ud$  cada, mas localizadas de uma forma que minimize a distância entre a facilidade e os pontos de demanda a ela alocados*

o seu movimento não pode ser realizado distanciando-se demais dos dois pontos de demanda que atendia anteriormente. Percebe-se, então, que deve-se manter um equilíbrio nessa movimentação. É com o pensamento nesse equilíbrio que a solução apresentada neste trabalho foi desenvolvida. O que aconteceu na transição do cenário da Figura 4.3 para o cenário da Figura 4.4 foi uma atração da facilidade  $j_2$  pela demanda não atendida pela facilidade  $j_1$ . Assim, o que se apresenta é uma das forças do sistema de forças que rege essa distribuição.

Na seção 4.4.1 a seguir, será apresentado o sistema de forças no contexto da solução do problema de localização de facilidades com restrição de capacidade.

#### 4.4.1 Sistema de forças

Na área de desenho de grafos, algumas abordagens partem de uma distribuição inicial dos vértices que, então, movimentam-se até que seja encontrada uma configuração (distribuição) que melhor respeite algumas restrições e otimize alguns critérios. As Figuras 4.2, 4.3 e 4.4 ilustram bem, na forma de um grafo, um cenário aleatório de um problema de localização de facilidades. De certa forma, o que se deseja é "desenhar" o grafo que melhor se adequa às restrições e critérios do problema de localização.

Como abordado na seção 3.3.1, Eades [7] apresenta um método para desenho de grafos chamado *Springs*. Nesse método, Eades propõe um sistema onde seja buscado o equilíbrio entre todas as forças existentes entre os vértices, de modo que o desenho seja o resultado desse equilíbrio.

A solução proposta neste trabalho é justamente a criação de um sistema de forças que obrigue alguns vértices (no caso, as facilidades) a se movimentarem no plano de modo que o resultado final seja um desenho que expresse a localização final de cada facilidade

no plano. Diferentemente do método de Eades, nesse sistema, os vértices que representam pontos de demanda estão presos no plano e não podem movimentar-se. Além disso, as arestas entre os vértices (arestas que ligam pontos de demanda às facilidades) não são fixas e vão sendo definidas dinamicamente. Essa dinâmica na criação das arestas é fruto do fato de que todos os pontos de demanda podem potencialmente exercer alguma força sobre todas as facilidades.

Quando as forças se estabilizarem, a função de custo total estará minimizada, uma vez que o valor de cada força é na verdade o custo existente na alocação de uma facilidade  $j$  a um ponto de demanda  $k$  dada a localização de  $j$  no plano. Isso fica evidente diante do fato que as forças devem logicamente guiar as facilidades para um posicionamento que diminua a função de custo. Assim, a função de custo total será criada de acordo com o cálculo das forças.

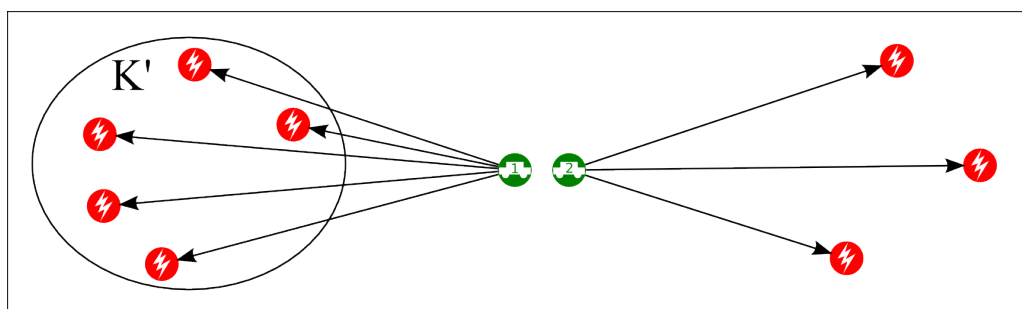
O problema passa a ser, então, definir quais são as forças que regem o sistema e como são calculadas.

### Força de atração pelos pontos de demanda à facilidade mais próxima

De modo geral, todos os pontos de demanda exercem uma força de atração sobre todas as facilidades. Entretanto, essa força de atração vai variar (podendo chegar a zero) dependendo da facilidade mais próxima ao ponto de demanda.

O sistema proposto inicia com a força principal, que continua sendo a de atração exercida pelos pontos de demanda à facilidade mais próxima. Essa é a mais forte de todas, pois sua incidência utiliza toda a demanda do ponto sobre a facilidade em questão.

Supondo duas facilidades 1 e 2 com capacidade de  $4ud$  cada a serem distribuídas em um plano onde cada ponto de demanda possui  $1ud$  de valor de demanda, a Figura 4.5 demonstra essas forças agindo sobre as duas facilidades.



**Figura 4.5:** Força de atração exercida pelos pontos de demanda (pontos marcados com um raio) sobre a facilidade (pontos marcados com um carro e um número de identificação) mais próxima

Como pode ser observado na figura, a facilidade  $j_1$ , que possui capacidade para  $4ud$  está alocada a um conjunto de pontos  $K' \subset K$  que demanda  $5ud$ . Assim, conclui-se

que ela está sobrecarregada. Entretanto, essa inconsistência será ignorada por enquanto, fazendo então com que toda a força exercida por esses pontos de demanda atraiam efetivamente a facilidade  $j_1$ .

Como não restará capacidade alguma para a facilidade  $j_1$  (uma vez que ela está com sua carga de demanda maior ou igual à sua capacidade), nesse momento, a mesma não poderá ser utilizada para atender nenhum outro ponto de demanda. Assim, uma facilidade nessa situação não sofrerá a ação de nenhuma outra força proveniente de algum ponto de demanda.

A partir do cálculo das forças que agem sobre a facilidade, a mesma se movimenta. Esse movimento irá seguir a direção e o sentido do vetor resultante da soma das forças que o atraem. Entretanto, o tamanho do passo nesse movimento é uma constante, de modo que as forças vão apenas guiando a facilidade. A razão de o tamanho do movimento ser constante (e de modo geral, relativamente curto) é que algumas forças podem existir ou deixar de existir de acordo com os movimentos das facilidades. Durante os movimentos, uma facilidade que antes atendia uma sobrecarga de demanda, pode passar a não estar mais sobrecarregada. Assim, a mesma passa a poder atender outras demandas remanescentes de outras facilidades sobrecarregadas, ou simplesmente atender a própria demanda.

Ao realizar o movimento, o valor total (incluindo todas as facilidades e pontos de demanda) da função de custo deverá ser minimizado (lembrando que a função de custo deve ser calculada a partir do cálculo de todas as forças).

O valor da força de atração exercida por um ponto de demanda  $k \in K'$  sobre a facilidade  $j$  mais próxima (ou força de atração principal - FAP) é dada por (4-3).

$$\text{custo}(FAP_{kj}) = d_k c_{kj} \quad (4-3)$$

Por outro lado, a facilidade  $j_2$  foi alocada a um conjunto de pontos que demanda  $3u_d$ , permitindo que a facilidade ainda disponha de  $1u_d$  de capacidade para atender outras demandas. Como visto na Figura 4.4, é interessante que essa facilidade leve em consideração a demanda remanescente no sistema, referente aos pontos de demanda não atendidos pela facilidade mais próxima. Isso leva à criação da próxima força: a força de atração de pontos de demanda atendidos parcialmente.

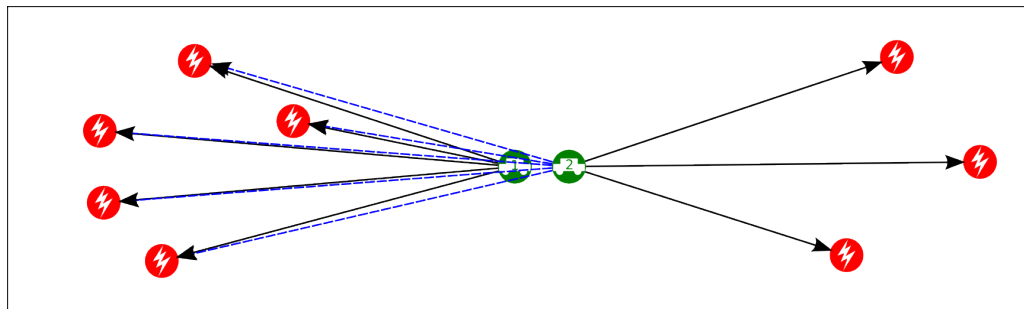
### **Força de atração pelos pontos de demanda atendidos parcialmente por outra(s) facilidade(s)**

A segunda força com mais relevância é a força proveniente da demanda remanescente. Essa demanda existe porque o ponto de demanda foi alocado a uma facilidade que não consegue atendê-lo completamente. Na prática, como mencionado no início deste

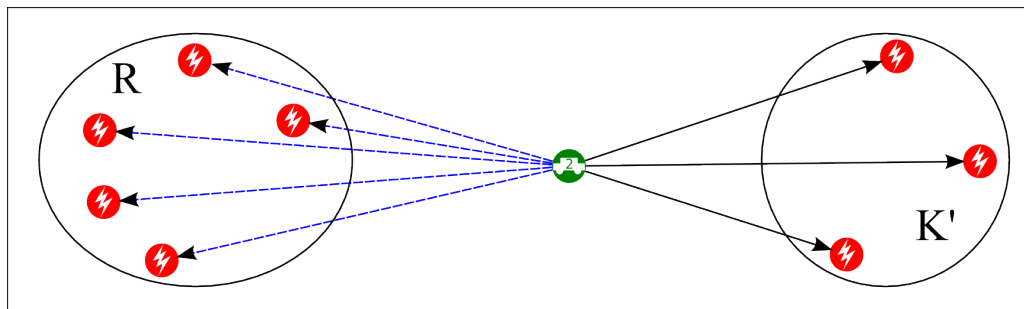
capítulo, as restrições (4-1b) e (4-1e) do problema tratado neste trabalho indicam que um ponto de demanda somente poderá ser atendido por uma e somente uma facilidade. Assim, se a alocação apresentada na Figura 4.5 for mantida, para que essas restrições não fossem quebradas, a facilidade  $j_1$  teria que atender toda a demanda dos pontos a ela alocados, incluindo a demanda que extrapola sua capacidade. Contudo, essa decisão iria quebrar a restrição (4-1d), que estabelece que a demanda alocada a uma facilidade não pode extrapolar sua capacidade.

Essa foi a razão pela qual a restrição (4-1e) foi relaxada conforme apresentado na seção 4.1. Agora, um ponto de demanda poderá ser atendido parcialmente por uma facilidade, de modo que a capacidade da mesma não seja extrapolada pela demanda dos pontos a ela alocados.

Essa demanda remanescente passa, então, a exercer uma força de atração sobre outras facilidades, conforme apresentado na Figura 4.6.



(a) Forças exercidas sobre as facilidades 1 e 2



(b) Forças exercidas pelos pontos de demanda com foco apenas na facilidade 2

**Figura 4.6:** Força de atração proveniente da demanda remanescente dos pontos (marcados com um raio) sobre facilidades (pontos marcados com um carro e um número de identificação) que ainda tenham capacidade remanescente do atendimento de demanda

A Figura 4.6(a) apresenta todas as forças exercidas pelos pontos de demanda sobre as facilidades. A Figura 4.6(b) retira a facilidade  $j_1$  apenas para melhorar a visualização das forças de atração provenientes da demanda remanescente dos pontos atendidos pela facilidade  $j_1$ .

Nessa figura, o conjunto  $K'$  destaca os pontos de demanda que são atendidos pela facilidade 2 em sua totalidade. Esses pontos são totalmente atendidos devido o fato de os mesmos estarem exercendo a força de atração pela facilidade mais próxima e essa facilidade em questão (facilidade 2) ter capacidade para atender a todos esses pontos.

O conjunto  $R \subset K$  representa todos os pontos de demanda que não foram completamente atendidos pela facilidade mais próxima a eles (nesse caso, a facilidade  $j_1$ ). Vale ressaltar que  $R \cap K' = \emptyset$ . Dessa forma, os pontos desse conjunto continuam com uma demanda remanescente que, para ser completamente atendida, necessita do atendimento de alguma facilidade que ainda tenha alguma capacidade remanescente.

É importante notar que essa força de atração só existe entre um ponto de demanda  $k$  e uma facilidade  $j$  quando:

1. o ponto de demanda  $k$  não está sendo atendido em sua totalidade pela facilidade mais próxima;
2. e a facilidade  $j$  possui uma capacidade maior do que a demanda total dos pontos a ela alocados.

Outro detalhe importante é que o cálculo do valor de cada força é feito da mesma maneira que a força de atração dos pontos de demanda pela facilidade mais próxima. A única diferença é que o valor da demanda utilizada no cálculo é a remanescente (que faltou ser atendida na alocação anterior) e o valor da capacidade utilizada é a remanescente da facilidade subutilizada (que sobrou quando foi realizada a alocação anterior).

Portanto, o valor dessa força remanescente (ou força de atração remanescente - FAR) é dada por (4-4).

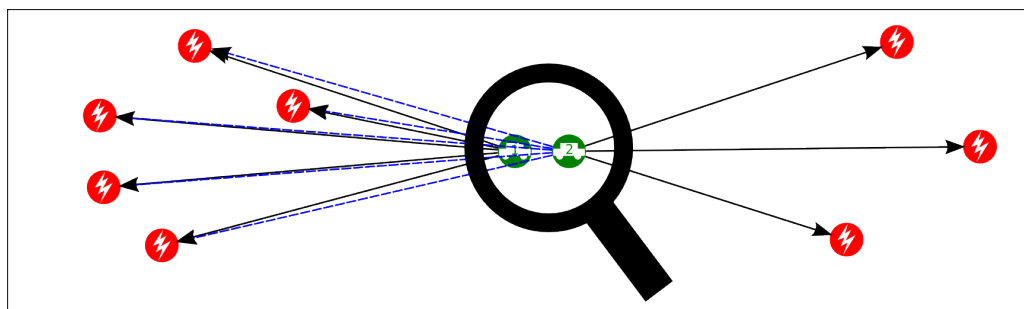
$$\text{custo}(FAR_{kj}) = r_k c_{kj} \quad \forall k \in R, \forall j \in J' \quad (4-4)$$

$J' \subset J$  é o conjunto de facilidades que possuem uma sobra de capacidade que poderá ser utilizada para atender demandas residuais de outros pontos de demanda parcialmente atendidos pela facilidade mais próxima. Além disso,  $r_k$  é o valor da demanda remanescente de  $k$ . Percebe-se que, diferentemente da força de atração principal, a força de atração remanescente age sobre todas as facilidades restantes.

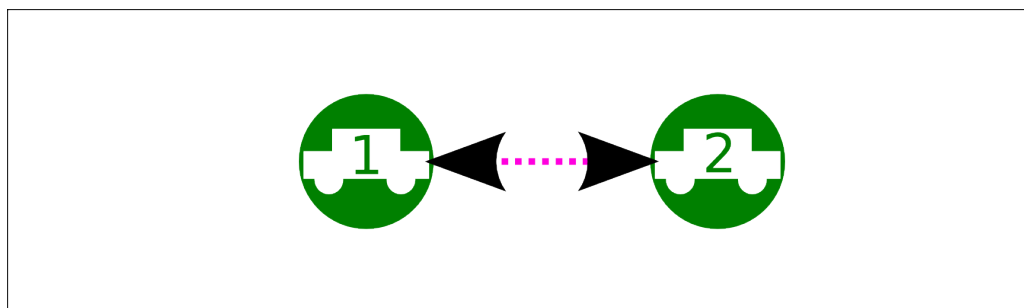
Somente essas duas forças definidas poderiam ser suficientes para bons resultados, uma vez que toda a demanda de um problema seria considerada na solução. Entretanto, uma outra característica de muitos problemas de localização é que em geral, deseja-se manter a instalação de uma facilidade sempre o mais distante possível de outra, aumentando assim a cobertura. Essa característica leva a terceira força: a força de repulsão entre duas facilidades.

### Força de repulsão entre duas facilidades

O terceira força proposta neste trabalho diz respeito à proximidade entre duas facilidades. Por uma questão logística, geralmente não é uma boa solução colocar duas facilidades muito próximas uma da outra. Pode haver casos onde isso é inevitável, sobretudo em problemas onde a capacidade das facilidades estão definidas e não podem ser modificadas. Nesses casos, é impossível substituir duas facilidades por apenas uma que tenha a capacidade das duas. A Figura 4.7 ilustra essa terceira e última força.



(a) Forças de atração exercidas sobre as facilidades 1 e 2



(b) Uma visualização mais aproximada da força de repulsão existente entre as facilidades 1 e 2

**Figura 4.7:** Conjunto das forças do sistema proposto

Quanto maior possível for a distância entre as facilidades, maior será a cobertura da região geográfica. Essa força foi definida justamente para auxiliar esse aspecto.

O valor dessa força de repulsão entre duas facilidades (FRF) é dada por (4-5).

$$\text{custo}(FRF_{ij}) = w_{ij}^{-1} \quad \forall i, j \in J, i \neq j \quad (4-5)$$

$w_{ij}$  é a distância entre duas facilidades  $i$  e  $j$ . Quanto maior a distância, melhor é a solução. Logo, o valor dessa distância é inversamente proporcional ao valor do custo. Por essa razão, a mesma é elevada à  $-1$ .

## 4.5 Discussão

Muitas organizações precisam distribuir seus recursos/facilidades em uma região geográfica de modo que possam atender a demanda da melhor forma possível. Esse problema possui diversas variações, que vão desde as necessidades específicas de cada organização até a forma de modelagem do problema.

Alguns problemas podem ser modelados na forma de rede. De modo geral, esse tipo de modelagem pode ser bastante interessante, uma vez que o número de possíveis locais para a instalação de uma facilidade fica consideravelmente menor. Contudo, alguns problemas permitem que a facilidade seja instalada em qualquer lugar. Eventualmente, mesmo problemas assim podem ser modelados em rede. Mas, quando a organização necessita de bastante precisão na localização das facilidades, restringir os possíveis locais de instalação acaba não sendo uma boa alternativa.

Assim, a abordagem apresentada nesse capítulo foi motivada pela dificuldade em resolver problemas de localização planares de forma eficiente. O sistema de forças apresentado é apenas uma ideia seguida da implementação de três forças. A proposta é que esse sistema possa ser aplicado com as forças que se fizerem necessárias de serem implementadas em um problema específico.

Esse sistema de forças foi utilizado em um estudo de caso que será detalhado no capítulo seguinte.

## Estudo de Caso e Resultados

---

O estudo realizado nesse trabalho foi aplicado em um caso real onde o problema de localização de recursos é crucial para a melhoria do atendimento de demandas além da redução de custos operacionais. Trata-se do problema de localização de viaturas da CELG-D<sup>1</sup>.

Quando ocorrem falhas na rede elétrica que atende o Estado de Goiás, a CELG-D deve prover equipes qualificadas para avaliar e reparar os pontos onde tais falhas surgem. A CELG-D não possui concorrência no fornecimento de energia elétrica. Contudo, com a finalidade de fazer com que a qualidade de serviço prestado ao consumidor seja aceitável, a ANEEL<sup>2</sup> regulamenta os indicadores que determinam a qualidade do serviço. Um desses parâmetros é o tempo que um consumidor fica sem energia elétrica por falha na rede. Existe um limite de tempo estabelecido que pode gerar multas para a empresa caso tal limite seja ultrapassado.

Com a finalidade de agilizar o atendimento de ocorrências de falhas elétricas, viaturas com equipes especializadas<sup>3</sup> são espalhadas por todo o território de atendimento de modo que, quando ocorrer uma falha, espera-se que haja uma viatura próxima ao local. Essa distribuição ajuda a manter o tempo que um consumidor fica sem energia elétrica abaixo do limite imposto pela ANEEL. Todavia, a quantidade de recursos (viaturas) não é ilimitada, não sendo possível localizar uma em cada esquina. Além disso, o deslocamento de uma viatura até um ponto de falha na rede gera um custo de tempo da equipe especializada bem como o próprio combustível utilizado pela viatura para realizar esse trajeto. Assim, localizar as viaturas de modo que os pontos de demanda possam ser atendidos com o menor deslocamento possível acaba sendo crucial para a qualidade do serviço bem como para a economia de recursos da CELG-D.

Entretanto, a localização dessas viaturas é realizada arbitrariamente sem o uso de nenhuma abordagem que assegure boas distribuições dessas viaturas no território

---

<sup>1</sup>CELG-D (CELG Distribuição S.A.) é a empresa responsável por administrar e distribuir a energia elétrica em todo o Estado de Goiás

<sup>2</sup>ANEEL - Agência Nacional de Energia Elétrica

<sup>3</sup>Equipes especializadas em geral são compostas por, pelo menos, um eletricista

geográfico. Assim, a pesquisa realizada nesse trabalho foi aplicada a esse caso, onde as viaturas representam as facilidades e as falhas na rede elétrica irão gerar os pontos de demanda.

No capítulo 4 foi apresentado o problema bem como uma proposta de solução para o modelo contínuo do problema de localização. Entretanto, a localização das viaturas vai depender diretamente dos pontos de demanda. Portanto, esse estudo de caso será dividido em duas etapas distintas. A primeira etapa é a determinação dos pontos de demanda e será apresentada na seção 5.1. A segunda etapa é a distribuição das viaturas no espaço geográfico, considerando, obviamente, os pontos de demanda encontrados na primeira etapa. Essa segunda etapa será explorada na seção 5.2.

## 5.1 Determinação dos pontos de demanda

Com o foco voltado para o problema de distribuição de facilidades em um plano, esse trabalho não se aprofundou em uma solução para o problema de determinar os pontos de demanda de um dado escopo. De fato, a forma como tais pontos são determinados sofre influência direta do problema real.

O caso que está sendo analisado nesse capítulo lida com o fato de que as demandas podem ou não ocorrer. Trata-se de um ambiente onde uma falha na rede elétrica gera uma demanda. Assim, determinar os pontos de demanda consiste basicamente em buscar um modelo matemático que seja capaz de caracterizar o comportamento do sistema, de modo que seja possível identificar com maior precisão, os prováveis locais onde irão ocorrer falhas na rede elétrica.

A seção 4.3 apresentou, de forma resumida, como foi realizada a construção dos pontos de demanda. O território coberto pela CELG-D foi dividido em células quadradas de mesmo tamanho (parametrizado). Cada célula passou a ser um ponto de demanda em potencial. Assim, calcular o valor desses pontos de demanda passou a ser o problema.

Como nesse estudo de caso da CELG-D, os recursos são trabalhadores em viaturas, a capacidade de cada facilidade (viatura) será representada pelo tempo de turno que cada viatura tem. Assim, se o turno de uma viatura for de oito horas, a capacidade daquela viatura será de 8 horas. Como a unidade de medida da capacidade é dada em unidade de tempo, então a demanda também deverá ser representada nessa mesma unidade. Na base de dados históricos da CELG-D existe a informação do tempo gasto na conclusão da manutenção realizada no ponto onde ocorreu a falha. Assim, o tempo gasto para se corrigir alguma falha passa a ser o valor da demanda.

Um estudo sobre o uso de redes bayesianas no problema da CELG-D para determinar os pontos de demanda seria bastante relevante, por se tratar de uma abordagem que consegue contemplar diversos parâmetros de um problema. Entretanto, por ser

considerado um problema fora do escopo desse trabalho, foram utilizadas duas diferentes abordagens. A primeira e mais simples realiza uma contagem de todas as ocorrências e faz uma média no tempo. A segunda utiliza um modelo estocástico bastante utilizado em problemas de caracterização de tráfego em redes diversas.

### 5.1.1 Estatística Simples com Média nas Janelas de Tempo

Na seção 4.3.1 foi apresentada uma forma simples, porém pobre, de determinar o valor de demanda de cada célula. Nessa abordagem, cada célula acumula toda a demanda ocorrida dentro dela no período especificado. O grande problema dessa abordagem é o fato de que ela não considera o tempo em seus cálculos. A Figura 5.1 ilustra esse tipo de problema.

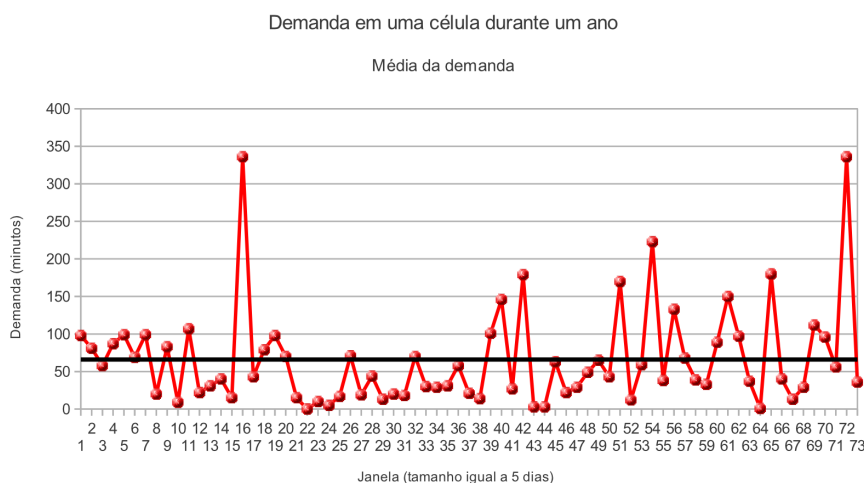


**Figura 5.1:** Comparação entre demandas que ocorrem em duas células distintas em um período de vinte dias utilizando dados hipotéticos

Nessa figura, as células 1 e 2 tiveram a mesma quantidade de demanda no mesmo período de vinte dias: 300 minutos. Entretanto, na célula 1, ocorreu apenas uma única falha no dia 3 que consumiu os 300 minutos de uma só vez. Em contrapartida, na célula 2, as falhas na rede elétrica ocorreram em diversos dias. Assim, em uma célula onde ocorresse uma única falha com demanda de 300 minutos em um período de 20 dias teria a mesma representatividade que uma célula que tivesse a mesma demanda, mas distribuída em todos os dias do mesmo período. Com a simples quantidade de demanda que ocorreu em uma célula no período analisado, ambas as células seriam equivalentes. Entretanto, a falha que demandou sozinha 300 minutos poderia ser uma falha esporádica que provavelmente não vai se repetir. Dessa forma, uma célula onde ocorre uma falha por dia teria muito mais relevância do que uma célula onde ocorre, por acaso, uma única falha com demanda tão elevada.

Com a finalidade de diminuir o efeito desse tipo de situação, foi desenvolvida uma estrutura de janelas de tempo. Assim, o período analisado foi dividido em janelas de tempo iguais. Um período de um ano (365 dias), por exemplo, poderia ser dividido em janelas de cinco dias (73 janelas por ano), por exemplo. Com essas janelas de tempo, é possível então obter uma média de ocorrências de falhas a cada cinco dias. Devido ao fato de que a cada cinco ou sete dias o gerente de turno deve refazer a distribuição de viaturas no mapa, essa abordagem torna-se particularmente interessante para o problema da CELG-D. É possível definir outros tamanhos para essa janela de tempo. De fato, essa definição deve ser realizada de forma empírica, de modo que o tamanho da janela não deva ser muito grande para que não reduza a amostragem (quantidade de janelas) e nem muito pequeno para que as janelas de tempo tenham valores relevantes.

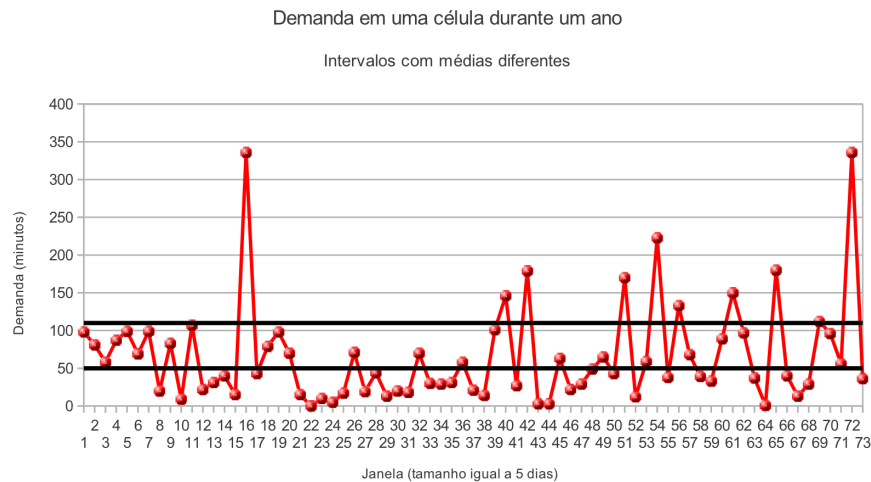
A Figura 5.2 exibe todas as janelas de uma célula no ano de 2010 com destaque para a média de demanda entre essas janelas.



**Figura 5.2:** Demanda de uma célula representada em janelas de 5 dias com destaque para a média. Os valores são de uma célula da região sudoeste de Goiás com tamanho de 2.5 km em cada lado no período entre 01/01/2010 e 31/12/2010.

Percebe-se que a média está muito distante de diversos valores. Além disso ainda que os intervalos sejam definidos, uma célula com uma ocorrência esporádica com demanda bem alta continuará a ter relevância equivalente a uma célula que tenha a mesma demanda dividida em várias janelas. Assim, utilizar a média seria uma forma pouco representativa de caracterizar o comportamento de surgimento de ocorrências.

Observando com mais atenção, de forma arbitrária, seria possível identificar nesse gráfico, ao menos três faixas distintas de valores. Valores que vão de 0 à 50 minutos de demanda, de 50 à 110 e acima de 110. A Figura 5.3 destaca essas três faixas separadas.



**Figura 5.3:** *Demanda de uma célula representada em janelas de 5 dias com destaque para faixas definidas arbitrariamente e que possuem médias próprias.*

Nessa divisão arbitrária, cada faixa possui uma média própria. A divisão foi criada com a finalidade de identificar intervalos onde as janelas de cada intervalo possuem uma proximidade muito maior com a demanda média das janelas do mesmo intervalo. A exceção, nesse caso, é a faixa que abrange os pontos com demanda acima de 110 minutos. Essa faixa é a que contém menos janelas e, portanto, possuem uma probabilidade menor de ocorrer.

O ponto é que, em algumas faixas, o comportamento da demanda se mantém semelhante. Portanto, utilizar algum modelo estocástico que aproveite esse comportamento que lida com múltiplas médias torna-se uma alternativa interessante. Com base nessa premissa, a seção 5.1.2 a seguir mostrará como foi implementada uma solução usando o modelo MMPP apresentado na seção 2.3, que é um modelo estocástico que lida com múltiplas médias e que, curiosamente, apresentou bons resultados para o estudo de caso da CELG-D.

### 5.1.2 Usando os parâmetros de caracterização do MMPP no problema da CELG-D

O modelo MMPP, apresentado na seção 2.3, foi utilizado no problema de determinar os pontos de demanda (ocorrências de defeitos) da CELG-D. Cada célula foi adaptada como um roteador e, ao invés de pacotes de dados, a demanda das janelas de tempo (5 dias por janela neste estudo de caso) compuseram o conjunto de dados medidos.

Assim, após a aplicação do método proposto por De Cock e De Moor associado ao método proposto por Ribeiro, os parâmetros de caracterização foram encontrados e

podem ser utilizados para realizar uma aproximação de surgimento de ocorrências nos cinco dias seguintes.

Para isso, foram realizados os seguintes passos:

1. Identificar o estado da última janela de tempo antes da janela de tempo que se deseja mensurar (no caso, os cinco dias seguintes ao período medido).
2. A partir da matriz de transição  $P$ , cada elemento da linha que corresponde ao estado dessa última janela será multiplicado pela sua taxa média de Poisson correspondente.
3. O resultado dessa multiplicação será a demanda final da célula.

Esses três passos garantem que, a partir do estado atual, cada possível taxa média será considerada de acordo com a probabilidade de o estado da célula mudar para essa taxa.

Por fim, todas as células que tenham ao menos uma janela de tempo com demanda passarão pelo processo de caracterização dos parâmetros MMPP. E a demanda em cada uma dessas células será calculada de acordo com os passos descritos nessa seção.

### 5.1.3 Aplicação e resultados da mensuração dos pontos de demanda

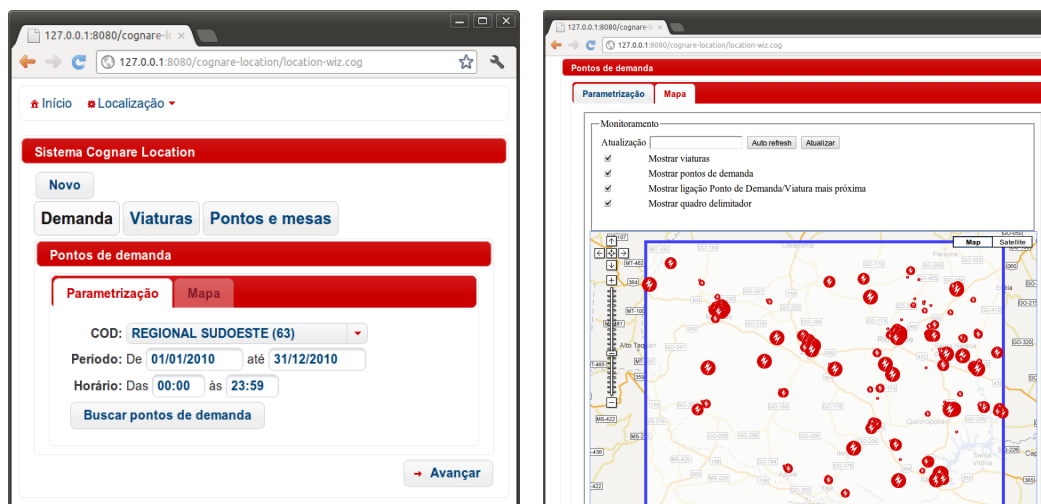
Foi desenvolvida uma aplicação para realizar a determinação dos pontos de demanda bem como a distribuição das viaturas no mapa. A parte da aplicação que determina os pontos de demanda exige como informação do usuário o período (data e horário) e o COD<sup>4</sup>, que representa a região que deve ser analisada. A mesma é ilustrada na Figura 5.4.

Após realizar essa parametrização, ilustrada na Figura 5.4(a), o usuário aciona o botão “Buscar pontos de demanda” e a aplicação seleciona todas as ocorrências do período informado e separa as mesmas em janelas de acordo com a sua data. Cada janela tem um tamanho parametrizado na aplicação. Nos testes foram usado 5 dias. A demanda de cada janela é o tempo acumulado gasto para atender cada uma das ocorrências. Nem todas as janelas terão obrigatoriamente alguma ocorrência, uma vez que num período de cinco ou mais dias consecutivos é possível não haver nenhuma ocorrência em alguma célula. Assim, o valor de demanda dessas janelas sem ocorrências será zero.

Em seguida, para cada célula que tenha ao menos uma janela com valor de demanda maior que zero, é aplicado o método de determinação da demanda utilizando o modelo MMPP. Com isso, os pontos de demanda são encontrados e podem ser visualizados no mapa (Figura 5.4(b)).

---

<sup>4</sup>Centro de Operação e Distribuição



(a) *Parametrização do processo de determinação dos pontos de demanda* (b) *Mapa mostrando os pontos de demanda encontrados*

**Figura 5.4:** *Aplicação que determina os pontos de demanda a partir de um período e uma região*

Foram implementados o método simples (que não considera o fator tempo) e o método utilizando o modelo MMPP. Em um ano, surgem muitas ocorrências de falhas em diversos pontos da rede elétrica. Entre esses diversos pontos, alguns manifestam problemas na rede elétrica de forma esporádica. Existem grandes regiões onde surgem uma ou duas ocorrências de falhas na rede elétrica. Essas regiões deveriam representar menor relevância como ponto de demanda ou mesmo desaparecer. A Figura 5.5 ilustra como ficariam os pontos de demanda da região sudoeste de Goiás no ano de 2010 caso fosse utilizado o método simples ou algo parecido que não considerasse o fator tempo.

Com esse método, centenas ou milhares de pontos de demanda seriam considerados no problema. Entretanto, alguns desses pontos deveriam ser desconsiderados, uma vez que representam apenas uma ou duas ocorrências que surgiram durante todo o ano de 2010. Em contrapartida, células com uma frequência maior de ocorrências deveriam ser melhor consideradas. Assim, pontos de demanda errados poderão resultar em uma distribuição equivocada das viaturas. Além disso, com uma quantidade muito grande de pontos de demanda, o algoritmo de localização de facilidades iria demorar muito mais a convergir.

O método utilizando o modelo MMPP gerou uma quantidade consideravelmente menor de pontos de demanda. Diversas regiões ficaram sem pontos de demanda que a representassem. Entretanto, os pontos de demanda gerados foram mais representativos.

Para efeito de comparação, foi criada uma janela de tempo medida que compreendeu o intervalo entre os dias 01/01/2011 e 05/01/2011. Esse intervalo de dias medido foi comparado visualmente com o mesmo intervalo inferido. Os resultados do período medido e do período calculado diferiram, mas se aproximaram muito mais um do outro



Espera-se que os resultados sejam de boa qualidade. A razão dessa expectativa é que, após a geração dos pontos de demanda, foi realizada, para cada célula, uma verificação que avaliou se cada uma das taxas médias de Poisson obtidas no MMPP representavam realmente processos de Poisson que descreviam o surgimento de ocorrências. Essa avaliação foi realizada da seguinte forma:

- A partir dos intervalos de valores que definem o estado de uma janela, os dados de ocorrências foram separadas em função de seu estado.
- Em seguida, para conjunto de amostras, foi calculada a média e a variância.
- Por fim, para que os valores da distribuição representem um processo de Poisson, média tem que ser igual à variância.

A Tabela 5.1 apresenta o resultado desse processo para uma célula aleatória.

Média calculada em <b>minutos</b>		
Média	Variância	Janelas
1,30	1,66	31
11,60	13,18	20
32,80	35,31	22

**Tabela 5.1:** Média e variância calculadas para cada um dos processos de Poisson encontrados em uma célula. Nesse caso foram três processos de Poisson.

Como pode-se perceber, a média calculada para os processos dessa célula se aproximam bastante da variância. Assim, pode-se dizer que o MMPP conseguiu caracterizar essa célula.

Uma outra possibilidade de análise dos dados avalia a quantidade de ocorrências ao invés da quantidade de tempo gasto no atendimento. Essa análise também foi realizada e a Tabela 5.2 demonstra o cálculo das médias e variâncias da mesma célula.

Média calculada em <b>ocorrências</b>		
Média	Variância	Janelas
8,65	7,68	54
18,00	18,11	19

**Tabela 5.2:** Média e variância calculadas para cada um dos processos de Poisson encontrados em uma célula. Nesse caso foram dois processos de Poisson.

Os valores de média e variância, nesse caso, continuam bastante parecidos. A diferença é que, apenas dois estados foram caracterizados. Isso se deve ao fato de que

a quantidade de ocorrências variou menos do que o tempo gasto para atendê-las. De qualquer maneira, o MMPP pôde ser utilizado em ambos os casos.

Evidentemente, não foram todas as células que retornaram resultados realmente coerentes. Em cerca de 80% dos casos, média e variância da mesma distribuição ficaram muito próximas uma da outra. Em contrapartida, aproximadamente 20% dos casos foram células com valor de demanda incoerente, onde média e variância ficaram muito distantes entre si.

A quantidade de células com valores coerentes é consideravelmente maior. Mas os resultados da heurística de localização de facilidades é que irá validar ou não esses valores de demanda mensurados.

## 5.2 Localização das Viaturas

Uma vez que os pontos de demanda estão definidos, torna-se possível aplicar a abordagem proposta na seção 4.4. Em teoria, essa proposta de um sistema de forças é simples de se entender. Entretanto, na prática, existe um problema. No mundo real, em qualquer sistema de forças, as forças trabalham constantemente e paralelamente. Em contrapartida, não é possível realizar um processamento paralelo de modo que todas as forças atuem ao mesmo tempo sobre as facilidades e pontos de demanda. Portanto, implementar o sistema de forças exige um método que permita a simulação desse paralelismo.

O método encontrado baseou-se em uma heurística onde os pontos de demanda com demanda remanescente atraem as facilidades mais próximas que ainda disponham de alguma capacidade remanescente. Nesse método, a cada iteração, todas as viaturas tentam realizar algum movimento. O movimento realizado tem sempre o mesmo tamanho de  $x$  metros e pode ser feito em oito direções: para norte, nordeste, leste, sudeste, sul, sudoeste, oeste ou noroeste. Se nenhum desses movimentos resultar em uma solução melhor, a viatura permanece no mesmo lugar. O algoritmo finaliza a busca da solução quando não existe nenhuma mudança da solução anterior para a solução atual. O método é descrito pelo Algoritmo 5.1.

---

**Algoritmo 5.1:** *Location(K, J)*

---

**Entrada:** Lista de pontos de demanda  $K$  e lista de facilidades  $J$ **Saída:** Localização de todas as facilidades  $J$  de acordo com a demanda de  $K$ 

```

1 Gerar solução inicial
2  $custoTotal \leftarrow \text{CalcularCusto}(K, J)$ 
3 enquanto Não convergir faça
4   para cada  $j \in J$  faça
5      $novoPontoXY \leftarrow j.pontoXY$ 
6     para cada Movimento  $m$  de tamanho  $s$  entre norte,
       nordeste, leste, sudeste, sul, sudoeste, oeste, noroeste faça
7        $j.pontoXY \leftarrow m.pontoXY$ 
8        $custoAtual \leftarrow \text{CalcularCusto}(K, J)$ 
9       se  $custoAtual < custoTotal$  então
10         $custoTotal \leftarrow custoAtual$ 
11         $novoPontoXY \leftarrow m.pontoXY$ 
12      fim
13    fim
14     $j.pontoXY \leftarrow novoPontoXY$ 
15  fim
16 fim
17 retorna Lista de facilidades  $J$  com as coordenadas de
    posicionamento

```

---

Nesse algoritmo, a solução inicial pode obedecer qualquer estratégia que gere uma solução rápida sem compromisso de ser necessariamente relevante. De fato, o uso de alguma implementação do método *k-means* é o mais comum. Entretanto, no problema da CELG-D foi utilizada uma estratégia que torna todas as facilidades potencialmente iguais. Nessa estratégia, todas elas são inicializadas exatamente no ponto central de todos os pontos de demanda. Assim, todas as forças irão incidir de forma semelhante em todas as facilidades, de modo que a distribuição de demanda seja mais equilibrada. Como consequência dessa disposição inicial fixa e do fato de que todos os passos do algoritmo são iterativos e não dependem de nenhum fator randômico, o resultado sempre será determinístico. Isso não quer dizer que o resultado será necessariamente a solução ótima. Mas, como será apresentado na seção 5.2.2, os resultados atingem os objetivos.

No algoritmo em questão, existe a presença ainda de um elemento não detalhado que é a função de custo  $\text{CalcularCusto}(K, J)$ . Essa função de custo terá que abranger as regras necessárias para modelar a movimentação de cada facilidade no espaço de

soluções. Essa modelagem é feita em função dos objetivos de distribuição de facilidades. No caso de estudo do problema da CELG-D, essas facilidades são as viaturas e os objetivos almejados para uma boa distribuição são:

1. Cada ponto de demanda precisa ser atendido na íntegra pela capacidade de uma ou mais viaturas, onde tais viaturas devem ser as mais próximas possíveis do ponto de demanda.
2. Cada viatura deverá estar posicionada exatamente no centro dos pontos de demanda que ela atende, considerando para isso, a carga de demanda que atende em cada ponto. Dessa forma, haverá sempre um peso a ser multiplicado pela distância entre a viatura e o ponto de demanda, que é o valor de demanda atendido pela viatura.
3. Toda viatura que estiver com sobra de capacidade deverá sofrer influência dos pontos de demanda que ainda possuem demanda não atendida.
4. Quaisquer viaturas deverão estar o mais distante possível umas das outras, desde que esse distanciamento não interfira no posicionamento da viatura em relação ao centro dos pontos de demanda que ela atende.

Para atingir esses objetivos, foi implementado na função de custo um algoritmo que simula a ação de três forças: atração entre pontos de demanda à facilidade mais próxima, atração entre pontos de demanda que não foram completamente atendidos às demais facilidades e repulsão entre as viaturas.

Os pontos de demanda deverão ser atendidos sempre pela(s) viatura(s) mais próxima(s). Entretanto, uma viatura poderá ser a mais próxima de uma quantidade de pontos de demanda cuja demanda total seja maior que a sua capacidade de atendê-las. Em contrapartida, quando isso acontecer, é certo que haverá uma ou mais viaturas cuja capacidade não fora totalmente alocada. Assim, essas facilidades passarão a sentir a força de atração desses pontos de demanda parcialmente atendidos. Entretanto, ainda assim, duas viaturas poderão não ser suficientes. Por essa razão, foi desenvolvido um algoritmo iterativo que realiza a alocação de toda a capacidade das viaturas à demanda, de acordo com um determinado posicionamento. Como a intensidade das forças dependem do quanto um ponto de demanda foi ou não atendido e do quanto de capacidade uma viatura ainda possui, o algoritmo desenvolvido prevê  $|J|$  níveis, onde  $|J|$  é a quantidade de facilidades. O Algoritmo 5.2 apresenta a função de custo implementando essa força de atração em níveis.

---

**Algoritmo 5.2:** *CalcularCusto(K,J)*

---

**Entrada:** Solução contendo a lista de pontos de demanda  $K$  e  
lista de facilidades  $J$

**Saída:** Custo da solução

```

1  para cada  $k \in K$  faça
2  |    $k.facilidades \leftarrow$  lista com todas as facilidades  $J$  com
   |   ordenação crescente por distância
3  |    $k.demandaRestante \leftarrow k.demandaTotal$ 
4  fim
5  para cada  $j \in J$  faça
6  |    $j.capacidadeRemanescente \leftarrow j.capacidadeTotal$ 
7  fim
8   $custo \leftarrow 0$ 
9  enquanto Existir um  $k \in K$  tal que  $k.demandaRestante > 0$  faça
10 |   para cada  $j \in J$  faça
11 |   |   Limpar lista  $j.pontosDemandaNivel$ 
12 |   |    $j.demandaNivel \leftarrow 0$ 
13 |   fim
14 |   para cada  $k \in K$  faça
15 |   |   se  $k.demandaRestante > 0$  então
16 |   |   |    $j \leftarrow$  facilidade mais próxima de  $k$  onde
   |   |   |    $j.capacidadeRemanescente > 0$ 
17 |   |   |   Adicionar  $k$  à lista  $j.pontosDemandaNivel$ 
18 |   |   |    $j.demandaNivel \leftarrow$ 
   |   |   |    $j.demandaNivel + k.demandaRestante$ 
19 |   |   fim
20 |   fim
21 |    $custo \leftarrow custo + CalcularCustoNivel(K,J)$ 
22 fim
23 retorna  $custo$ 

```

---

O Algoritmo 5.3 descreve como é calculado o custo de cada nível.

**Algoritmo 5.3:** *CalcularCustoNivel(K, J)*

**Entrada:** Solução contendo a lista de pontos de demanda  $K$  e lista de facilidades  $J$

**Saída:** Custo do nível

```

1 custoNivel ← 0
2 para cada  $j \in J$  faça
3   capacidadePorUnidadeDemanda ←  $\frac{j.\text{capacidadeRemanescente}}{j.\text{demandaNivel}}$ 
4   para cada  $k \in j.\text{pontosDemandaNivel}$  faça
5     se capacidadePorUnidadeDemanda  $\geq 1$  então
6       | demandaAtendida ←  $k.\text{demandaRestante}$ 
7     senão
8       | demandaAtendida ←  $k.\text{demandaRestante} \times$ 
9         | capacidadePorUnidadeDemanda
10      fim
11     custoNivel ←
12       custoNivel +  $\text{distancia}(j, k)^2 \times \text{demandaAtendida}$ 
13     k.demandaRestante ←
14       k.demandaRestante – demandaAtendida
15     j.capacidadeRemanescente ←
16       j.capacidadeRemanescente – demandaAtendida
17   fim
18 fim
19 retorna custoNivel

```

Os dados de entrada para o Algoritmo 5.2 são o conjunto de pontos de demanda  $K$  e o conjunto de facilidades (viaturas)  $J$ . No laço que vai da linha 1 à linha 4, os pontos de demanda de  $K$  tem seu campo *demandaRestante* inicializado com o valor total de sua demanda (linha 3). Como até esse momento nenhuma facilidade foi alocada ao ponto de demanda, sua demanda restante é toda a demanda que possui. Além disso, a lista *facilidades* é inicializada com uma lista de todas as facilidades com ordenação crescente pela distância das facilidades ao ponto de demanda  $k$ . Essa lista será utilizada para o algoritmo alocar a demanda dos pontos sempre à facilidade mais próxima.

O laço que vai da linha 5 à 7 inicializa o campo *capacidadeRemanescente* das facilidades de  $J$  com o valor da capacidade total de cada facilidade. Assim como ocorre nos pontos de demanda, ainda não foi realizada nenhuma alocação. Logo, a capacidade total das facilidades é a capacidade remanescente.

Cada iteração do laço que compreende o intervalo entre as linhas 9 e 22 representa um nível. Em cada um desses níveis, sempre será realizada a alocação de pontos

de demanda com demanda restante maior que zero à facilidade mais próxima que possua capacidade remanescente. O laço termina quando todos os pontos de demanda estão completamente atendidos.

Em cada nível, serão armazenadas em cada facilidade, a lista de pontos de demanda (*pontosDemandaNivel*) a ela alocados naquela iteração. Assim, no laço entre as linhas 10 e 13, para todas as facilidades, essa lista será previamente limpada e a soma da demanda alocada à facilidade no nível em questão será zerada (linha 12). Ainda dentro de cada nível, todos os pontos de demanda que ainda tiverem demanda restante maior que zero deverão ter sua demanda restante alocada à facilidade mais próxima do ponto de demanda e com capacidade maior que zero. Cada ponto de demanda irá ser armazenado na lista de pontos de demanda da facilidade a ele alocada dentro do nível. A demanda alocada (*demandaNivel*) à facilidade será a soma de toda a demanda restante que for alocada à facilidade.

Por fim, com essa nova alocação e configuração de demandas e capacidades restantes, é realizado o cálculo do custo de alocação daquele nível. O Algoritmo 5.3 modularizou essa etapa. Esse algoritmo também recebe os conjuntos  $K$  e  $J$ . No laço entre as linhas 2 e 14 será calculada a proporção de atendimento que cada facilidade de  $J$  possui (*capacidadePorUnidadeDemanda*). Se a capacidade remanescente da facilidade for menor que a demanda alocada à facilidade no nível em questão, então essa proporção será menor do que 1. Logo, a demanda atendida em cada ponto de demanda alocado a essa facilidade será calculado em função dessa proporção (linha 8). Em contrapartida, se essa proporção for maior ou igual a 1 (capacidade remanescente maior ou igual à demanda alocada no nível) a demanda atendida em cada ponto de demanda alocado será o total da demanda restante de cada ponto (linha 6). Finalmente, o custo irá acumular todo o somatório das distâncias quadradas entre facilidade e ponto de demanda multiplicado pela demanda do ponto atendida pela facilidade naquele nível (linha 10). Além disso, a demanda restante dos pontos de demanda e a capacidade remanescente das facilidades são atualizados em função da demanda atendida pela facilidade (linhas 11 e 12).

Voltando ao Algoritmo 5.1, ao final da execução da função de custo que é chamada na linha 8, o custo será avaliado e, se for menor que o custo atual, então o novo posicionamento da facilidade  $j$  será considerada a nova melhor solução. A convergência será atingida quando não houver mais nenhuma mudança de solução após o laço entre as linhas 4 e 15.

### 5.2.1 O uso de temperatura e resfriamento no algoritmo

Uma vez que a solução local de mais baixo custo é sempre selecionada como única solução atualmente válida, a abordagem mostra-se gulosa. Dessa forma, durante

alguns testes preliminares, percebeu-se que, com frequência, os resultados convergiam de forma precoce, prendendo-se em mínimos locais. O uso de uma lista tabu foi considerado, mas dado o tamanho do espaço de busca, tal lista poderia ter um tamanho muito grande para poder gerar bons resultados.

Para viabilizar a busca por soluções mais distantes do ótimo local, foram inseridos no algoritmo, dois conceitos de abordagens que implementam o *simulated annealing* [14]: a temperatura e o resfriamento. Assim, o tamanho  $s$  do movimento  $m$  mencionado na linha 6 do Algoritmo 5.1 passa a ser a temperatura e irá variar conforme alguma regra.

A princípio esse tamanho  $s$  será inicializado com um valor alto, para que as viaturas explorem lugares bem distantes de seu local de origem. Esse tamanho irá diminuir quando o algoritmo encontrar a convergência para aquele tamanho de movimento. Assim, seguindo alguma estratégia de resfriamento (diminuição do tamanho do movimento), gradativamente, à medida em que o algoritmo for encontrando a convergência, o mesmo é reiniciado a partir da solução encontrada mas com um novo tamanho de movimento  $s$ . Esse tamanho irá diminuir até chegar a um limite, quando finalmente a temperatura não irá mais resfriar e o algoritmo para.

Assim, o Algoritmo 5.4 descreve como o Algoritmo 5.1 será usado em uma estratégia com tamanho de movimento variável.

---

**Algoritmo 5.4:** *LocationChangeMoveSize(K,J)*

---

**Entrada:** Lista de pontos de demanda  $K$  e lista de facilidades  $J$

**Saída:** Localização de todas as facilidades  $J$  de acordo com a demanda de  $K$

```

1 Gerar solução inicial
2  $s \leftarrow$  um valor alto como temperatura (tamanho do movimento)
3 enquanto Não convergir faça
4   | Location(K,J) se  $s$  ainda puder ser diminuído então
5   |   |  $s \leftarrow$  um valor menor do que ele mesmo (resfriamento)
6   | senão
7   |   | Algoritmo converge
8   | fim
9 fim
10 retorna Lista de facilidades J com as coordenadas de
    posicionamento
```

---

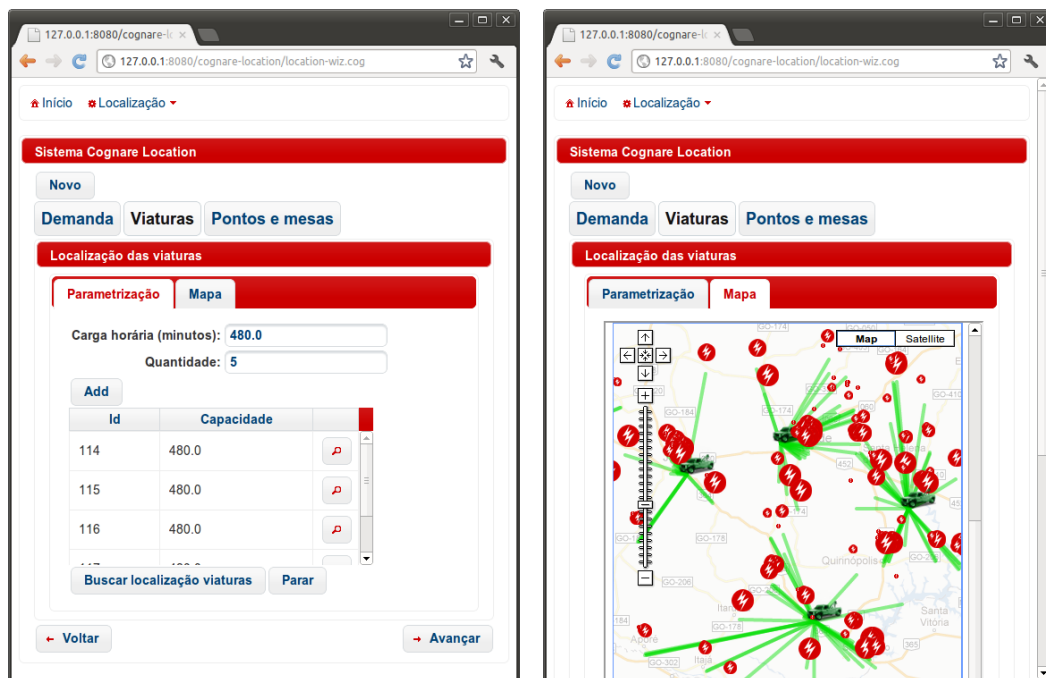
Vale ressaltar que esse algoritmo considera que a variável  $s$  é global, podendo ser enxergada dentro do método *Location(K,J)* descrito no Algoritmo 5.1.

Além de garantir uma busca global, a estratégia usando temperatura e resfria-

mento ainda garante um melhor desempenho, uma vez que grandes movimentos podem ser realizados em detrimento de pequenos movimentos que deveriam chegar no mesmo lugar, mas com passos menores.

## 5.2.2 Aplicação e resultados da localização das viaturas

Na aplicação desenvolvida neste trabalho para o estudo de caso da CELG-D, depois da etapa de mensuração dos pontos de demanda, o próximo passo é a localização das viaturas no mapa levando em consideração os pontos de demanda encontrados na etapa anterior. A parte da aplicação que busca a localização das viaturas exige, como informação do usuário, apenas o conjunto de viaturas (incluindo sua capacidade). Vale ressaltar, mais uma vez, que essa capacidade diz respeito à carga horária do turno de cada viatura que atenderá aquele período. Essa carga horária é dada nessa tela em minutos. Essa tela de parametrização é ilustrada na Figura 5.7.



(a) *Parametrização do algoritmo de determinação da localização das viaturas* (b) *Mapa mostrando os pontos de demanda e as viaturas em movimento*

**Figura 5.7:** *Aplicação que determina a localização das viaturas da CELG-D em função dos pontos de demanda pré-definidos*

Comumente, as viaturas terão o mesmo tamanho de turno. Por essa razão várias viaturas com a mesma capacidade podem ser inseridas de uma só vez. Embora nessa parametrização a capacidade possa ser informada, antes de iniciar a execução do algoritmo de localização todos os valores das demandas são proporcionalmente alterados

para que a soma da demanda de todos os pontos seja exatamente igual à soma da capacidade de todas as viaturas. Essa medida é tomada para que dois problemas sejam evitados. O primeiro ocorre quando a soma da capacidade é maior que a soma da demanda. Nesse caso, algumas viaturas seriam subutilizadas e, em alguns casos, até mesmo descartadas, uma vez que a capacidade pode ser muito maior. O segundo problema ocorre quando a soma da demanda é maior que a soma da capacidade. Esse problema é ainda pior porque fere diretamente a restrição (4-2d) do problema de localização abordado neste trabalho que diz que a soma da capacidade tem que ser maior ou igual à soma da demanda. Contudo, a opção de informar a capacidade foi disponibilizada para que seja possível estabelecer um cenário onde houvessem viaturas com capacidades distintas.

Após informar todas as viaturas na lista parametrização (Figura 5.7(a)), o usuário deve clicar no botão “Buscar localização viaturas”. A aplicação então irá executar o algoritmo de localização das facilidades (viaturas). Dependendo da quantidade de pontos de demanda e de viaturas, o algoritmo pode demorar mais ou menos para convergir. De qualquer maneira, o mesmo é executado em *background* e pode ser acompanhado pelo mapa, onde pode ser visualizada toda a movimentação das facilidades rumo ao melhor posicionamento (Figura 5.7(b)). As linhas que ligam as viaturas (carros) aos pontos de demanda (círculos) denotam apenas a alocação do primeiro nível do algoritmo de cálculo do custo. Não é viável colocar todas as forças que existem no sistema, pois como um ponto de demanda pode ser atendido por mais de uma viatura, as linhas que demonstrariam tais forças poluiriam muito a visualização.

Quando ocorrer a convergência, o resultado final será justamente o posicionamento de cada viatura no plano. Esse posicionamento poderá ainda ser conferido no mapa da aplicação.

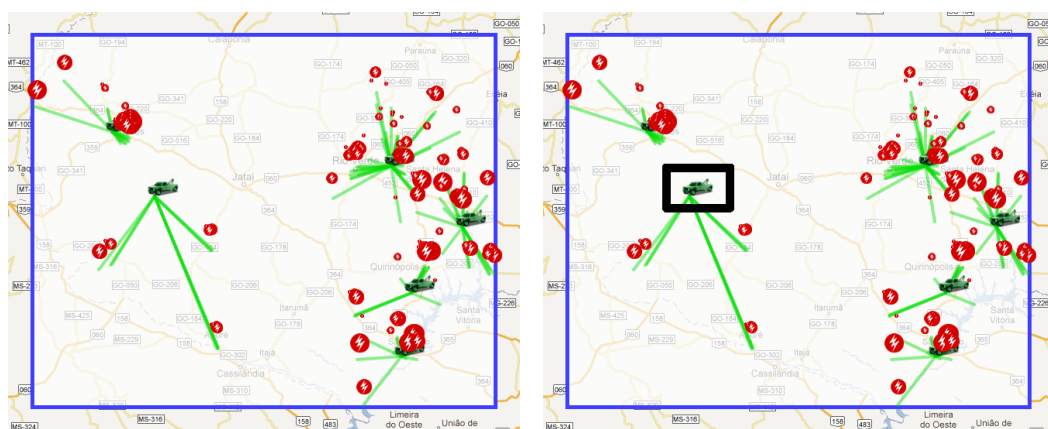
### **Objetivos atingidos**

Avaliar os resultados do algoritmo de localização, assim como os resultados do algoritmo de mensuração de pontos de demanda, não é uma atividade trivial. De fato, os objetivos foram estabelecidos para a criação da função de custo e, a partir desses objetivos, a validação dos resultados será realizada pela checagem desses objetivos.

O objetivo de cada ponto de demanda ser atendido na íntegra foi explicitamente garantido dentro do Algoritmo 5.2 na condição de parada do laço que começa na linha 9.

O objetivo de fazer com que facilidades com capacidade remanescente maior que zero sofram influência de outros pontos de demanda cuja demanda não foi totalmente atendida é assegurada dentro do Algoritmo 5.2 na linha 16. Esse ponto do algoritmo é importante, pois ele também ajuda a atingir o objetivo de alocar a demanda à facilidade mais próxima.

A linha 10 do Algoritmo 5.3 explicita a forma como o custo é finalmente calculado em função das alocações previamente realizadas. Nesse cálculo, a distância quadrada ao invés da simples distância euclidiana assegura que se pontos de demanda com forças de intensidade igual e de lados opostos atraírem uma viatura, o ponto médio entre os dois pontos de demanda é onde o custo será mais baixo. Entretanto, a intensidade da força nem sempre será a mesma. E o fator que influencia nessa intensidade é a demanda dos pontos. Assim, a multiplicação dessa distância quadrada pela demanda atendida pela viatura é o que garante o equilíbrio e a localização da viatura no exato ponto médio de todas as forças. A Figura 5.8 apresenta o resultado da execução do algoritmo de localização com pontos de demanda hipotéticos.



(a) *Viaturas distribuídas segundo objetivos pré-estabelecidos* (b) *Destaque na viatura que está sendo atraída por outros pontos de demanda distantes dos pontos que atende por estar mais próxima*

**Figura 5.8:** *Visualização de viaturas localizadas no plano de acordo com os pontos de demanda existentes*

A viatura em destaque (5.8(b)) está atendendo sozinha na íntegra aos pontos de demanda cuja viatura mais próxima é ela. Entretanto, como esses pontos de demanda não consumiram toda a capacidade dessa viatura em questão, a mesma possui uma considerável capacidade remanescente. Essa capacidade é atraída pelos pontos de demanda que ainda não foram completamente atendidos. Como a quantidade de demanda não atendida é aparentemente grande, a viatura se afasta dos pontos de demanda que atende na íntegra. Assim, essa visualização atesta o fato de que o objetivo de colocar a viatura no centro das forças que a atraem foi atingido.

O objetivo das viaturas estarem posicionadas o mais distante possível não precisou ser explicitamente implementado, uma vez que o objetivo de posicionar a viatura no centro das forças que a atraem fora atingido.

### **Validação da distribuição de pontos de demanda com a localização de viaturas**

Uma vez que os objetivos foram atingidos com a proposta do sistema de forças para realizar a localização das viaturas no plano, foi utilizada uma estratégia para realizar uma validação dos resultados obtidos na determinação dos pontos de demanda utilizando a solução para localização das viaturas.

Essa estratégia de testes consiste nos seguintes passos:

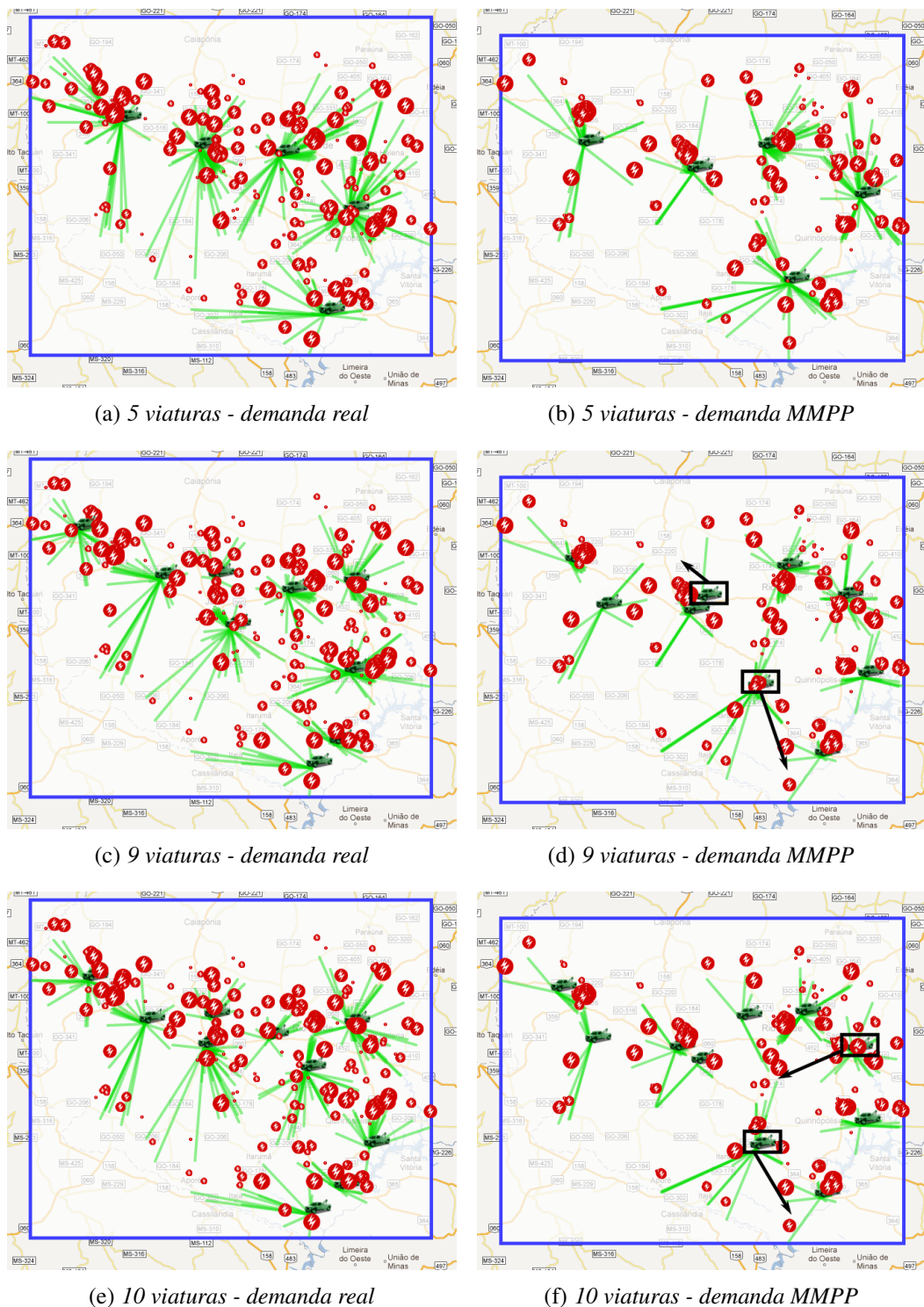
1. Definir uma janela (cinco dias seguidos) que se deseje mensurar a demanda.
2. Realizar uma contagem de toda a demanda real que ocorreu nesse período de cinco dias em um determinado COD e separar em células para serem apresentadas em um mapa como pontos de demanda. Essa será a demanda medida.
3. Realizar o processo de mensuração da demanda desse mesmo período de cinco dias no mesmo COD utilizando o método MMPP. Essa será considerada a demanda mensurada.
4. Executar o algoritmo de localização de facilidades para os dois ambientes: o medido e o mensurado.
5. Comparar a localização final das viaturas.

O objetivo dessa estratégia é comparar o resultado da localização de viaturas para pontos de demanda de uma janela de cinco dias medida com uma janela do mesmo período mensurada. Como trata-se do mesmo período, espera-se que a localização das viaturas em ambas configurações de demanda sejam aproximadas. A Figura 5.9 apresenta comparações de alguns resultados obtidos nesses testes.

Nos testes realizados, foram utilizadas poucas viaturas para facilitar a visualização. Com cinco viaturas (figuras 5.9(a) e 5.9(b)) o posicionamento ficou muito semelhante, tanto para o ambiente medido como para o ambiente mensurado. Com os demais testes apresentados na figura, houveram poucas diferenças notáveis. As viaturas foram circuladas com um quadrado ligado à uma aresta indicando onde deveria estar localizada. De qualquer forma, isso aconteceu apenas duas vezes para 9 viaturas e duas vezes para 10 viaturas. Assim, com tamanha semelhança (que também aconteceu para outros testes com maior quantidade de viaturas) entre ambiente real e mensurado, é coerente dizer que o método de determinação dos pontos de demanda utilizando o modelo MMPP retorna bons resultados para serem utilizados para o algoritmo de localização de facilidades. E uma vez que esse algoritmo atingiu os objetivos almejados, a solução como um todo se mostrou satisfatória.

### **Desempenho**

Foram realizados testes com diversas quantidades de pontos de demanda e de facilidades. A máquina utilizada possui como configuração: AMD Phenom(tm) II X4 800

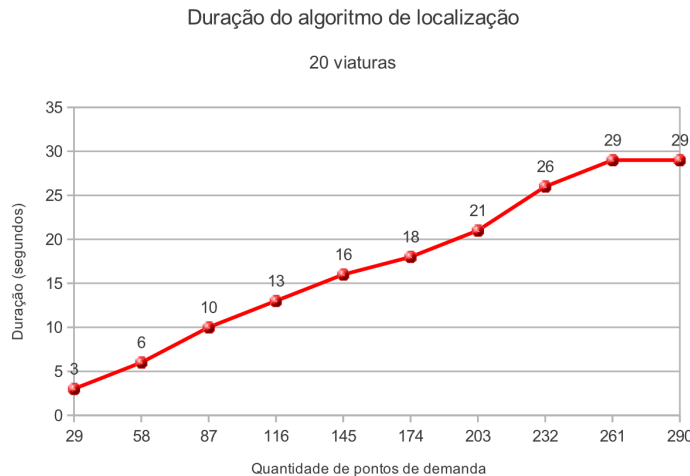


**Figura 5.9:** Comparação entre localizações de facilidades realizadas a partir de pontos de demanda reais medidos e a partir de pontos de demanda mensurados pelo modelo MMPP. A janela apresentada compreende ocorrências de falhas entre os dias 01/01/2011 e 05/01/2011 na região sudoeste do Estado de Goiás.

Processor 64 bits 8 GB RAM.

Um comportamento curioso, porém esperado, é que o aumento da quantidade de viaturas aumenta mais o tempo para convergência do algoritmo do que o aumento da quantidade de pontos de demanda.

A Figura 5.10 mostra o aumento do tempo gasto para a convergência do algoritmo de localização de facilidades enquanto a demanda é gradativamente aumentada.



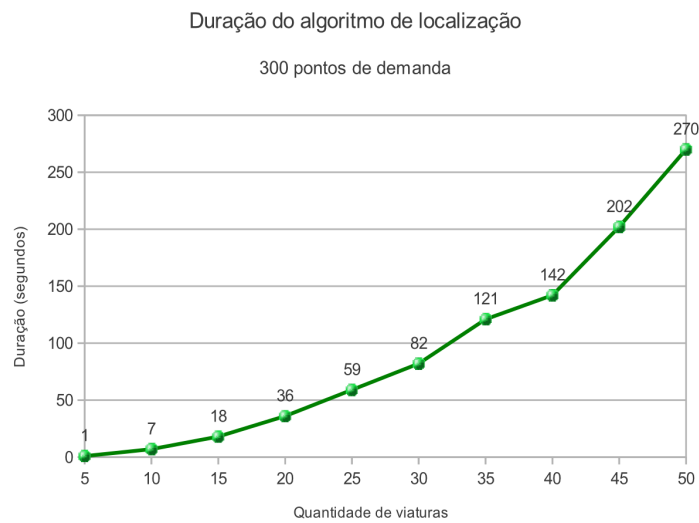
**Figura 5.10:** *Desempenho do algoritmo de localização de facilidades com 20 viaturas enquanto a quantidade de pontos de demanda é aumentada*

Ainda que esse aumento de tempo gasto seja consideravelmente alto, a queda de desempenho ocorre, em geral, de forma linear. No gráfico, pode ser visto que na maior parte das vezes, a quantidade de tempo aumentada é constante (acompanhando o aumento da quantidade de pontos de demanda que também é constante no gráfico), não apresentando nenhum indício de crescimento exponencial.

Em seguida, a Figura 5.11 demonstra o desempenho do algoritmo conforme a quantidade de viaturas é aumentado.

Nesse caso, a queda de desempenho é mais problemática. Embora, em alguns pontos, o aumento do tempo gasto para a convergência se mantenha constante, na maioria das vezes, a queda de desempenho aumenta de forma inconstante. Isso pode ser percebido com uma breve análise sobre o gráfico. Entre 5 e 10 viaturas, a diferença de tempo foi de 6 segundos. Entre 10 e 15 viaturas, a diferença aumentou para 11 segundos. Entre 15 e 20 a diferença foi para 18 e em seguida para 23. Assim, nota-se que o aumento das viaturas tem um impacto maior sobre o desempenho do que o aumento de pontos de demanda.

Como havia sido mencionado antes, embora esse comportamento seja curioso, o mesmo era previsível. Analisando o algoritmo apresentado, nota-se que não existe nenhum laço aninhado que percorra a lista de pontos de demanda duas vezes. Entretanto,



**Figura 5.11:** *Desempenho do algoritmo de localização de facilidades com 300 pontos de demanda enquanto a quantidade de viaturas é aumentada*

existe não apenas um aninhamento de dois laços que percorrem a lista de viaturas, mas um aninhamento com três laços que percorrem tal lista. Os aninhamentos com dois laços estão explícitos no algoritmo. Entretanto, o laço que começa na linha 9 do Algoritmo 5.2 esconde um aninhamento implícito. A condição de parada desse laço é quando nenhum ponto de demanda tiver demanda sem alguma viatura alocada. Entretanto, cada iteração nesse laço representa um nível. A quantidade de níveis está vinculada à quantidade de viaturas, uma vez que um ponto de demanda pode ser atendido por uma ou por mais viaturas, podendo, inclusive, ser atendida parcialmente por todas as viaturas. Quando isso ocorrer, esse laço irá se repetir na exata quantidade de facilidades.

Assim, faz sentido que, para esse algoritmo proposto, a quantidade de viaturas seja mais impactante no desempenho do que a quantidade de pontos de demanda.

## 5.3 Discussão

O modelo que usa cadeias de Markov para modular processos de Poisson (MMPP) é bastante utilizado em diversos problemas de caracterização de tráfego multimídia. Embora seja mais recomendável para problemas dessa área, ele apresentou bons resultados na geração de pontos de demanda que denotam falhas na rede elétrica. Vale ressaltar que esse não é o único modelo que poderia ter sido usado.

Existem outros modelos estocásticos que poderiam ter sido considerados e mesmo utilizados neste trabalho. O Modelo de regressão Weibull é um deles. Geist et al. [9] aplicaram o modelo em um sistema com tolerância a falhas. Dubuisson e Hassan [6] utilizaram a distribuição de Weibull para avaliação de falhas durante o tempo. Embora

existam trabalhos com bons resultados usando distribuições de Poisson e de Weibull, a razão pela qual um modelo que utiliza processos de Poisson foi escolhido se deve ao fato de que o mesmo é bem mais simples, uma vez que o foco deste trabalho está na localização das facilidades e não na determinação dos pontos de demanda.

Além disso, problemas de predição de eventos podem ser modelados com redes bayesianas (como nos trabalhos de Park et al. [24] e Bradford et al. [2]) ou redes neurais (como nos trabalhos de Kalayci et al. [12] e Rowley et al. [27]). Contudo, o foco desse trabalho não foi na atividade de inferir os pontos de demanda, mas realizar a localização das viaturas.

Na etapa de localização de viaturas foi necessário fazer algumas adaptações na implementação do sistema de forças proposto. Essas adaptações foram feitas para que fosse possível simular a ação de todas as forças ao mesmo tempo. A alternativa encontrada tratou todo o processo de alocação em níveis, de modo que todos os pontos de demanda atraem sempre a facilidade mais próxima para atender sua demanda remanescente. Essa abordagem mostrou-se natural na atividade de encontrar uma solução. Entretanto, foi necessário tomar emprestado os conceitos de temperatura e resfriamento do modelo *simulated annealing* para fugir de ótimos locais e buscar por soluções melhores globalmente. Mas, como bônus, a inclusão desses conceitos no algoritmo resultou em uma solução com melhor desempenho.

Nessa etapa, foi necessário fazer uma avaliação dos resultados obtidos com a implementação da solução proposta. Infelizmente, avaliar resultados de localização de facilidades não é uma atividade trivial. Após a obtenção de uma distribuição, dizer se ela ficou boa ou ruim é uma atividade que só é possível se existirem objetivos bem definidos. A partir desses objetivos, torna-se viável a análise dos resultados verificando se tais objetivos foram atingidos. Especialistas da CELG-D, responsáveis pela distribuição de viaturas e turnos, definiram os objetivos apresentados neste capítulo. Com a evidenciação do cumprimento dos mesmos, a um custo computacional aceitável, é coerente dizer que os resultados obtidos foram satisfatórios.

Entretanto, embora a implementação apresentada consiga atender aos requisitos da CELG-D, a solução proposta pode não ser eficiente o suficiente para instâncias de tamanho muito grandes. Isso porque constatou-se que, embora a quantidade de pontos de demanda não interfira tanto no desempenho, a quantidade de facilidades interfere muito. Além disso, devido ao fato de que cada nível de alocação de demanda ser devidamente controlado, o algoritmo não permite paralelismo, não aproveitando todos os núcleos de um processador *multicore* ou de uma estrutura em grade. Assim, melhorias nessa abordagem ou mesmo na implementação certamente poderão valorizar ainda mais o trabalho.

Por fim, embora existam problemas de desempenho provenientes do grande

aumento no número de facilidades, a solução apresentada mostrou-se útil, uma vez que a CELG-D posiciona as viaturas no mapa em uma quantidade que, em geral, não compromete o desempenho do algoritmo. Dessa forma, os resultados demonstraram que a aplicação pode ser usada para auxiliar na atividade de localização de viaturas da CELG-D.

---

## Considerações Finais

---

A pesquisa realizada neste trabalho sobre problemas de localização de facilidades contínuos resultou em uma nova proposta para trabalhar-se com problemas dessa natureza. As conclusões sobre todo o trabalho realizado serão apresentadas na seção 6.1. Além disso, na seção 6.2, serão sugeridos trabalhos relacionados que podem merecer atenção em futuras pesquisas.

### 6.1 Conclusões

Diversas organizações precisam distribuir recursos em uma região geográfica para atender à demanda existente. Recursos esses que podem ser hospitais, viaturas policiais, ambulâncias, pontos de parada para ônibus, dentre outros. O problema de localização e alocação de facilidades é constantemente enfrentado por organizações. E cada uma tem suas peculiaridades e especialidades que devem ser tratadas individualmente, visando sempre os melhores resultados nessa distribuição.

Realizar essa distribuição e a alocação dessas facilidades de forma manual não é uma tarefa trivial. De fato, para instâncias não muito pequenas, tal distribuição só pode ser feita de forma aleatória ou sem considerar de forma coerente a real demanda existente. Assim, este trabalho visou, prioritariamente, desenvolver uma proposta de solução que pudesse realizar a localização de facilidades em um plano, considerando todas as coordenadas do plano como potenciais pontos onde a facilidade seria instalada. Paralelamente, a mesma solução realizou a tarefa de alocação, uma vez que o problema de localização abordado considera que as facilidades possuem uma capacidade limitada.

Os resultados obtidos foram bastante satisfatórios, tendo em vista que foi realizado um estudo de caso real, como apresentado no Capítulo 5, onde as distribuições encontradas foram consideradas próximas do ótimo. Além disso, o tempo gasto na geração da solução foi consideravelmente baixo e, portanto, aceitável para problemas com instâncias grandes.

Outros problemas contínuos podem ter restrições diferentes, como prioridades para certos pontos de demanda ou o desejo que uma certa facilidade busque se instalar em

uma determinada área da região geográfica. Para essas e outras restrições específicas que o problema real possa ter, a abordagem proposta permite que sejam implementadas novas forças que façam essa movimentação esperada nas facilidades.

O problema de determinar os pontos de demanda a partir de dados históricos não foi fortemente estudado por este trabalho, uma vez que o foco foi direcionado à localização das facilidades. De fato, caracterizar o comportamento do surgimento de pontos de demanda depende do problema analisado. No estudo de caso realizado, foi utilizado um método muito usado para caracterizar tráfegos multimídia. O ideal seria direcionar mais tempo de pesquisa para esse problema específico. Entretanto, a caracterização utilizando Processos de Poisson Modulados por Markov mostrou-se mais efetiva do que uma simples contagem ou uma média de ocorrências. Por essa razão, esse método foi utilizado.

Vale ressaltar que tanto os pontos de demanda quanto a localização das facilidades foram localizados considerando todos os pontos do plano como pontos em potencial. Dessa forma, o modelo contínuo não foi convertido em um modelo discreto. Esse aspecto é importante e foi um desafio a parte, tendo em vista que grande parte dos trabalhos existentes que abordam tais problemas realizam a conversão, restringindo a alguns pontos candidatos para a instalação das facilidades.

Por fim, conclui-se que, embora muitos problemas de localização contínuos com restrição de capacidade possam ser convertidos para um modelo discreto (podendo-se assim, utilizar algumas das tantas abordagens existentes), a criação de uma abordagem completamente contínua abre espaço para que tais problemas possam ser tratados conforme suas reais características. Assim, os resultados podem ser mais exatos quando for possível a instalação da facilidade em qualquer lugar do plano. Além disso, mesmo os problemas discretos poderiam fazer uso do sistema de forças apresentado para gerar a localização nos nós candidatos. Conclui-se ainda que com os bons resultados obtidos com pouco tempo de processamento, a aplicação da abordagem proposta mostrou-se interessante para o estudo de caso da CELG-D. Dessa forma, a CELG Distribuição S.A. irá integrar tal solução aos seus processos de gerência das viaturas disponíveis, comprovando os bons resultados alcançados.

## 6.2 Trabalhos Futuros

Durante a análise do estudo de caso e durante a pesquisa realizada neste trabalho foram identificados diversos temas que certamente podem ser objeto de futuras pesquisas.

A abordagem apresentada mostrou-se eficiente no problema estudado. Seria interessante aplicá-la a outros problemas com diferenças nas restrições, como a localização da facilidade em alguma subárea da região total.

O algoritmo implementado para a movimentação segue uma estratégia gulosa. Uma sugestão seria adicionar recursos de busca Tabu para que fossem realizados movimentos que não gerassem soluções melhores do que a atual, ampliando ainda o alcance do algoritmo. Ainda no algoritmo de movimentação, uma estratégia menos iterativa e mais inteligente poderia ser utilizada para evitar tentativas de movimentos que não chegam a uma boa solução ao invés de tentar movimentos em oito direções. Uma estratégia sugerida seria realizar uma soma vetorial que guiaria a direção e sentido.

Existem problemas de localização que podem ser modelados de forma discreta. Uma sugestão adicional seria adaptar o conceito do sistema de forças proposto em modelos discretos com a finalidade de avaliar a possibilidade de usar tal sistema em qualquer problema de localização com restrição de capacidade.

Outra sugestão concentra-se em estabelecer um modelo mais específico e melhor estudado que caracterize o surgimento de ocorrências de demanda durante o tempo para o estudo de caso realizado no escopo das companhias de distribuição de energia elétrica. É fato que o modelo MMPP utilizado para realizar essa caracterização melhorou consideravelmente os resultados em relação a uma simples média de ocorrências por célula da região geográfica. Contudo, o comportamento do surgimento das ocorrências não foi exaustivamente estudado e, certamente, um modelo mais específico, que considere características da rede elétrica, resultaria em inferências de pontos de demanda mais próximas da realidade. Características como, data da última manutenção de cada ponto elétrico na rede, condições climáticas atuais, eventos de grande porte que aumentem a demanda de energia elétrica, dentre outras características que podem influenciar diretamente no surgimento de falhas na rede elétrica.

Por fim, uma vez que cada facilidade se movimenta de forma autônoma mas esperando por sua vez no processo iterativo, sugere-se que seja estudada e adaptada uma maneira de paralelizar o processo de movimentação das facilidades, de modo que sejam aproveitadas as arquiteturas de processadores *multicore* ou mesmo as estruturas de *grids* computacionais. Sugere-se ainda o desenvolvimento de uma solução multi-agente onde cada facilidade seria um agente que busca seu melhor posicionamento considerando as forças que o atraem e o posicionamento dos demais agentes.

---

## Referências Bibliográficas

---

- [1] BENATI, S.; LAPORTE, G. **Tabu search algorithms for the (r|xp)-medianoid and (r|p)-centroid problems.** *Location Science* 2, 1994.
- [2] BRADFORD, J. R.; NEEDHAM, C. J.; BULPITT, A. J.; WESTHEAD, D. R. **Insights into protein-protein interfaces using a bayesian network prediction method.** *Journal of Molecular Biology*, 362(2):365 – 386, 2006.
- [3] COCK, K. D.; MOOR, B. D. **Stochastic System Identification for ATM Network Traffic Models: a Time Domain Approach.** *Internal Report K. U.*, 1998.
- [4] COOK, S. A. **The complexity of theorem-proving procedures.** In: *Proceedings of the third annual ACM symposium on Theory of computing*, STOC '71, p. 151–158, New York, NY, USA, 1971. ACM.
- [5] DI BATTISTA, G.; EADES, P.; TAMASSIA, R.; TOLLIS, I. G. **Algorithms for drawing graphs: an annotated bibliography.** *Comput. Geom. Theory Appl.*, 4:235–282, October 1994.
- [6] DUBUISSON, B.; HASSAN, M. **A simplified method for evaluating time dependent fault-trees, using weibull distribution.** *Microelectronics Reliability*, 20(3):347 – 350, 1980.
- [7] EADES, P. **A Heuristic for Graph Drawing.** *Congressus Numerantium*, 42:149–160, 1984.
- [8] FARAHANI, R. Z.; STEADIESEIFI, M.; ASGARI, N. **Multiple criteria facility location problems: A survey.** *Applied Mathematical Modelling*, 34(7):1689 – 1709, 2010.
- [9] GEIST, R.; SMOTHERMAN, M.; TRIVEDI, K.; DUGAN, B. J. **On the Use of Weibull Fault Processes in Modeling Fault Tolerant Systems.** Technical report, Durham, NC, USA, 1985.
- [10] H. HEFFES, D. M. L. **A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance.** *IEEE Journal*, 1986.

- [11] HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. The Morgan Kaufmann series in data management systems. Elsevier, 2006.
- [12] KALAYCI, T.; OZDAMAR, O. **Wavelet preprocessing for automated neural network detection of eeg spikes**. *Engineering in Medicine and Biology Magazine, IEEE*, 14(2):160–166, mar/apr 1995.
- [13] KHOSHGOFTAAR, T.; GAO, K.; SZABO, R. **An application of zero-inflated Poisson regression for software fault prediction**. In: *Software Reliability Engineering, 2001. ISSRE 2001. Proceedings. 12th International Symposium on*, p. 66–73, nov. 2001.
- [14] KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. **Optimization by Simulated Annealing**. *Science, Number 4598, 13 May 1983*, 220, 4598:671–680, 1983.
- [15] KLOSE, A.; DREXL, A. **Facility location models for distribution system design**. *European Journal of Operational Research*, 162(1):4–29, 2005. *Logistics: From Theory to Application*.
- [16] KRISHNA, K.; NARASIMHA MURTY, M. **Genetic K-means algorithm**. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 29(3):433–439, jun 1999.
- [17] LIKAS, A.; VLASSIS, N.; VERBEEK, J. J. **The global k-means clustering algorithm**. *Pattern Recognition*, 36(2):451–461, 2003.
- [18] LO, J.-H.; HUANG, C.-Y. **An integration of fault detection and correction processes in software reliability analysis**. *Journal of Systems and Software*, 79(9):1312–1323, 2006.
- [19] LU, Y.; LU, S.; FOTOUHI, F.; DENG, Y.; BROWN, S. J. **FGKA: a Fast Genetic K-means Clustering Algorithm**. In: *Proceedings of the 2004 ACM symposium on Applied computing, SAC '04*, p. 622–623, New York, NY, USA, 2004. ACM.
- [20] MACQUEEN, J. B. **Some Methods for Classification and Analysis of Multivariate Observations**. In: Cam, L. M. L.; Neyman, J., editors, *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, p. 281–297. University of California Press, 1967.
- [21] MAHAJAN, M.; NIMBHORKAR, P.; VARADARAJAN, K. **The Planar k-Means Problem is NP-Hard**. In: Das, S.; Uehara, R., editors, *WALCOM: Algorithms and Computation*, volume 5431 de **Lecture Notes in Computer Science**, p. 274–285. Springer Berlin / Heidelberg, 2009.

- [22] MELKOTE, S.; DASKIN, M. S. **Capacitated facility location/network design problems**. *European Journal of Operational Research*, 129(3):481 – 495, 2001.
- [23] MURTY, K.; KABADI, S. **Some np-complete problems in quadratic and nonlinear programming**. *Mathematical Programming*, 39:117–129, 1987.
- [24] PARK, S.; AGGARWAL, J. **A hierarchical bayesian network for event recognition of human actions and interactions**. *Multimedia Systems*, 10:164–179, 2004.
- [25] PAYMAN KAVEH, A. S. Z.; SAHRAEIAN, R. **Solving Capacitated P-median Problem by Hybrid K-means Clustering and FNS Algorithm**. *International Journal of Innovation, Management and Technology*, 2010.
- [26] RIBEIRO, A. **Caracterização da Matriz Estocástica em Tráfego Multimídia MMPP**. Master's thesis, Universidade Federal de Goiás - Escola de Engenharia Elétrica, 2000.
- [27] ROWLEY, H.; BALUJA, S.; KANADE, T. **Neural network-based face detection**. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(1):23 –38, jan 1998.
- [28] VAN BENTHEM, M. H.; KEENAN, M. R. **Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems**. *Journal of Chemometrics*, 18(10):441–450, 2004.
- [29] WEBER, A.; PICK, G. **Ueber den Standort der Industrien**. Ueber den Standort der Industrien. J.C.B. Mohr (Paul Siebeck), 1909.
- [30] YAMADA, S.; OHBA, M.; OSAKI, S. **S-Shaped Reliability Growth Modeling for Software Error Detection**. *Reliability, IEEE Transactions on*, R-32(5):475 –484, dec. 1983.