



Universidade Federal do Piauí
Centro de Ciências da Natureza
Programa de Pós-Graduação em Ciência da Computação

Método para Localização da Fonte de Afundamento de Tensão por Algoritmo de Clusterização e Rotulação por Regra de Decisão

José Carlos Correia Lima da Silva Filho

Número de Ordem PPGCC: M001

Teresina-PI, 09 de abril de 2019

José Carlos Correia Lima da Silva Filho

Método para Localização da Fonte de Afundamento de Tensão por Algoritmo de Clusterização e Rotulação por Regra de Decisão

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação

Orientador: Ivan Saraiva Silva

Coorientador: Ricardo de Andrade Lira Rabelo

Teresina-PI

09 de abril de 2019

José Carlos Correia Lima da Silva Filho

Método para Localização da Fonte de Afundamento de Tensão por Algoritmo de Clusterização e Rotulação por Regra de Decisão/ José Carlos Correia Lima da Silva Filho. – Teresina-PI, 09 de abril de 2019-

62 p. : il. (algumas color.) ; 30 cm.

Orientador: Ivan Saraiva Silva

Dissertação (Mestrado) – Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação, 09 de abril de 2019.

1. Afundamento de Tensão. 2. Algoritmo de Clusterização. 3. Localização de Afundamento de Tensão. 4. Qualidade de Energia Elétrica. 5. Smart Grids. I. Orientador: Ivan Saraiva Silva II. Universidade Federal do Piauí. III. Título

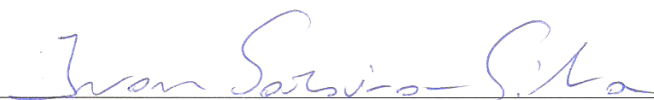
CDU 02:141:005.7

“Método para Localização da Fonte de Afundamento de Tensão por Algoritmo de Clusterização e Rotulação por Regra de Decisão”

JOSÉ CARLOS CORREIA LIMA DA SILVA FILHO

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Natureza da Universidade Federal do Piauí, como parte integrante dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

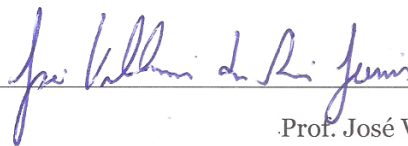
Aprovada por:



Prof. Ivan Saraiva Silva
(Presidente da Banca Examinadora)



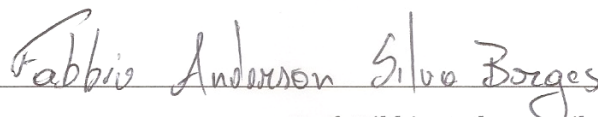
Prof. Ricardo de Andrade Lira Rabêlo
(Coorientador e Examinador Interno)



Prof. José Valdemir dos Reis Júnior
(Examinador Interno)



Prof. Daniel Barbosa
(Examinador Externo à Instituição)



Prof. Fábio Anderson Silva Borges
(Examinador Externo ao Programa)

Teresina, 09 de abril de 2019

*Aos meus pais José Carlos e Ana Maria, a minha esposa Ana Paula
por sempre estarem comigo em todos os momentos e aos meus filhos Luana Lima e
Gabriel Lima por suportarem minhas ausências e muitas vezes meu stress exagerado.*

Agradecimentos

À Deus e a minha família.

À Universidade Federal do Piauí, pelo apoio ao Programa de Pós Graduação em Ciência da Computação – PPGCC.

Ao Prof. Orientador Dr. Ivan Saraiva Silva, pela parceria, pelo conhecimento repassado e pelas orientações

Agradeço ao professor Dr. Ricardo de Andrade Lira Rabêlo, pessoa humana e de extrema sabedoria, que soube conduzir-me por meio de seu ofício de ensinar.

Ao Prof. Dr. Harilton da Silva Araujo por fornecer seu apoio científico;

Ao Prof. Dr. Fabbio Anderson Silva Borges pelo imenso apoio e contribuições no desenvolvimento deste trabalho

Aos Professores do Programa de Pós Graduação em Ciência da Computação – PGCC, agradeço pela dedicação e incentivo;

Aos membros da Banca de Qualificação e da Defesa da Dissertação pela disposição em colaborar com o aperfeiçoamento desse estudo.

À toda a família do Laboratório de Oasis (Optimization, Autonomous Solutions and Intelligent Systems), pela ajuda, trocas de experiências e tornar o ambiente de trabalho bem mais agradável.

Agradeço a todas as pessoas que direta ou indiretamente participaram deste trabalho.

*“Tenha coragem de seguir
o que seu coração
e sua intuição dizem.
Eles já sabem o que você
realmente deseja.
Todo resto é secundário.”
(Steve Jobs)*

Resumo

O distúrbio de afundamento de tensão corresponde a uma redução entre 0,9 e 0,1 p.u. do valor eficaz da tensão nominal por um curto período de tempo e são causados principalmente por curto-circuito no sistema de energia. Tais distúrbios, destacam-se como sendo a alteração da forma de onda que mais se manifestam nas redes elétricas, sendo que sua ocorrência ocasiona danos aos consumidores. Desta forma, apontar o local no sistema de distribuição de energia que está conectada a fonte causadora do distúrbio de afundamento de tensão é o primeiro passo para corrigir o problema. No entanto, esta tarefa não é trivial, uma vez que os afundamentos de tensão ocorrem em um curto intervalo de tempo e se propagam ao longo de todo sistema. Este trabalho tem como objetivo principal apresentar uma metodologia baseada em algoritmo de clusterização combinada com regra de decisão para apontar a região (cluster) de origem do afundamento de tensão. O algoritmo de clusterização é responsável por analisar os dados do sinal de tensão advindos de diferentes pontos de medição e separar esses dados em grupos, onde um deles agrega os locais próximos da região de ocorrência do distúrbio. Para tanto, foram avaliados sobre diferentes contextos a performance de 5 (cinco) algoritmos de clusterização, são eles: *K-Means Clustering*, *Gaussian Mixture Models (GMM)*, *Density Based Spatial Clustering of Applications with Noise (DBSCAN)*, *Fuzzy C-means (FCM)* e *Agglomerative Hierarchical Clustering*. Em seguida o algoritmo *Partial Decision Trees (PART)* defini um conjunto de regra de decisão, responsável por confrontar as características de cada clusters e rotular qual grupo (cluster) agrega o local de origem do distúrbio. A metodologia proposta foi validada nos dados obtidos a partir de simulações de curto-circuito trifásico no Sistema IEEE 34 barras. Os resultados mostram que a abordagem proposta neste trabalho é promissora para ser empregada em *smart grids*, uma vez que apresentou resultados satisfatórios, apresentando em alguns cenários uma taxa de acerto maior que 90%. Portanto, a metodologia pode apoiar a decisão do corpo técnico para que ações corretivas venham a ser estabelecidas de forma assertiva.

Palavras-chave: Afundamento de Tensão, Algoritmo de Clusterização, Localização de Afundamento de Tensão, Qualidade de Energia Elétrica, *Smart Grids*.

Abstract

Voltage sag disturbances corresponds to a reduction between 0.9 and 0.1 p.u. of the rated voltage effective value for a short period of time and are mainly caused by short circuit in the power system. These disturbances include the alteration of the wave form that manifested most in the electrical networks, since its occurrence causes damage to the consumers. In this way, pinpointing the location in the power distribution system that is connected to the source causing the voltage sag disturbance is the first step in correcting the problem. However, this task is not trivial, since voltage sags occur during a short period of time and propagate throughout the system. This work aims to present a methodology based on clustering algorithm combined with decision rule to point out the region (cluster) of origin of the voltage sag. The clustering algorithm is responsible for analyzing the voltage signal data from different measurement points and separating that data into groups, where one of them aggregates the locations close to the region of occurrence of the disturbance. In order to do this, the performance of 5 (five) clustering algorithms was evaluated on different contexts, they include: K-Means Clustering, Gaussian Mixture Models (GMM), Density Based Spatial Clustering of Applications with Noise (DBSCAN), Fuzzy C-means (FCM) and Agglomerative Hierarchical Clustering. Then the Partial Decision Trees (PART) algorithm defined a set of decision rule, responsible for confronting the characteristics of each clusters and labeling which cluster (cluster) aggregates the source location of the disturbance. The proposed methodology was validated in the data obtained from three-phase short-circuit simulations in the IEEE 34 bus system. The results show that the approach proposed in this work is promising and can be used in smart grids, since it showed satisfactory results, demonstrating in some scenarios a hit rate greater than 90%. Therefore, this methodology can support the decision making of the technical staff so that corrective actions can be established assertively.

Keywords: Voltage Sag, Clustering Algorithm, Voltage Sag Location, Electric Power Quality, Smart Grids.

Lista de ilustrações

Figura 1 – Principais distúrbios de energia	7
Figura 2 – Abordagem para análise dos algoritmos de clusterização - fase 1.	14
Figura 3 – Criação do conjunto de métricas para validar acurácia dos clusters formados - fase 2.	15
Figura 4 – Conjunto de vetores	24
Figura 5 – Indicação das barras selecionadas no diagrama do Sistema IEEE 34 <i>Node Test Feeder</i>	26
Figura 6 – Silhueta média do algoritmo k-means com $k = 2$ (a), $k = 3$ (b), $k = 4$ (c), $k = 5$ (d), $k = 6$ (e) de curto-circuito na barra 846	29
Figura 7 – Comparação da silhueta do algoritmo k-means com $k = 2$ (a), $k = 3$ (b), $k = 4$ (c), $k = 5$ (d), $k = 6$ (e) e fuzzy c-means $k = 2$ (f), $k = 3$ (g), $k = 4$ (h), $k = 5$ (i), $k = 6$ (j) de curto-circuito na barra 846	30

Lista de tabelas

Tabela 1 – Classificação das Variações de Tensão de Curta Duração (PRODIST, 2014)	9
Tabela 2 – Ralação entre os dados analisados, características do sinal e métricas estatísticas para criação das características analisadas para induzir o conjunto de Regras de Decisão.	24
Tabela 3 – Comparação entre algoritmos de clusterização Agglomerative Hierarchical Clustering (HAC), k -Means (KM), Gaussian Mixture Models (GMM) e Fuzzy c -means (FCM) através do valor da Largura da silhueta (LS) para as barras 802, 846, 850 e 852 com resistência 10Ω	27
Tabela 4 – Comparação do valor da largura da silhueta entre o algoritmo DBSCAN em relação ao melhor resultado obtido nos outros algoritmos para as barras 802, 846, 850 e 852 com resistência 10Ω	28
Tabela 5 – Quantidade de simulações com indicação do melhor k para cada algoritmo	30
Tabela 6 – Quantidade de ocorrência do valor de k para o algoritmo de clusterização DBSCAN	30
Tabela 7 – Base de Regras para localização dos distúrbios de Afundamento de Tensão induzida a partir dos parâmetros definidos no vetor das características apresentado na Tabela 2	32
Tabela 8 – Percentual das instâncias classificadas corretamente obtido por regra de decisão PART para cada conjunto de características gerados pelos algoritmos de clusterização avaliado	35
Tabela 9 – Comparação entre algoritmos de clusterização Agglomerative Hierarchical Clustering (HAC), k -Means (KM), Gaussian Mixture Models (GMM) e Fuzzy c -means (FCM) através do valor da Largura da silhueta (LS) para as todas as barras com as resistência $10, 20, 30, 40$ e 50Ω . .	47

Lista de abreviaturas e siglas

AMT	Afundamento Momentâneo de Tensão
DBSCAN	<i>Density-Based Spatial Clustering of Applications with Noise</i>
EM	Expectativa-Máxima
FCM	<i>Fuzzy C-means clustering</i>
GMM	<i>Gaussian Mixture Models</i>
HAC	<i>clustering aglomerativo hierárquico</i>
IEEE	<i>Institute of Electrical and Electronic Engineers</i>
LDA	análise linear de discriminante
PART	Partial Decision Trees
PCA	Análise de Componentes Principais
PQ	Qualidade da Energia
PV	sistema solar fotovoltaico
QEE	Qualidade da Energia Elétrica
RBF	<i>Radial Basis Function</i>
RMS	<i>Root Mean Square</i>
SAI	<i>Sag Activity Index</i>
SVM	<i>Support Vector Machine</i>
VTCD	Variações de Tensão de Curta Duração

Sumário

	Introdução	1
	Definição do Problema	2
	Justificativa	2
	Objetivos	4
	Publicações	5
1	ESTADO DA ARTE	7
1.1	Variações de tensão de curta duração (VTCDs)	7
1.2	Métodos para localização da origem dos distúrbios de QEE	8
2	METODOLOGIA	13
2.1	Abordagem geral do Método para Localização da Fonte de Afundamento de Tensão por Algoritmo de Clusterização e Rotulação por Regra de Decisão	13
2.2	Características Analisadas	15
2.2.1	<i>Root Mean Square (RMS)</i>	16
2.2.2	Valor de Pico	16
2.2.3	Fator de Pico	16
2.2.4	Valor Médio	16
2.2.5	Normalização	17
2.3	Algoritmos de Clusterização	18
2.3.1	<i>K-Means Clustering</i>	19
2.3.2	<i>Fuzzy c-means</i>	20
2.3.3	<i>Gaussian Mixture Models (GMM)</i>	20
2.3.4	<i>Density-Based Spatial Clustering of Applications with Noise (DBSCAN)</i>	20
2.3.5	<i>Agglomerative Hierarchical Clustering</i>	21
2.4	Análise dos Algoritmos de Clusterização	21
2.5	Regra de decisão	22
2.6	Análise dos dados	23
3	RESULTADOS E DISCUSSÃO	25
3.1	Descrição dos Cenários	26
3.1.1	Resultados da Análise dos algoritmos de clusterização	26

3.2	Resultado do Método de Localização dos Distúrbios de Afundamento de Tensão	31
4	CONCLUSÕES E TRABALHOS FUTUROS	37
4.1	Conclusões	37
4.1.1	Limitações e Trabalhos Futuros	37
	REFERÊNCIAS	39
	APÊNDICES	45
	APÊNDICE A – RESULTADO DA ANÁLISE DOS ALGORITMOS DE CLUSTERIZAÇÃO	47

Introdução

O termo qualidade da energia elétrica (QEE) está relacionado ao conjunto de alterações que se refletem no sinal de tensão e de corrente entregue pelas concessionárias de energia aos consumidores (DUGAN et al., 1996; FILHO et al., 2016). Dentre essas alterações, o distúrbio de afundamento de tensão destaca-se como sendo a alteração da forma de onda que mais se manifesta nas redes elétricas (CHANG et al., 2007; ROMERO; GALLEGO; PAVAS, 2012). Esse distúrbio corresponde a uma redução da amplitude da tensão, entre 0,9 e 0,1 p.u. do valor eficaz da tensão nominal (BHATTACHARYYA, 2011). Este tipo de distúrbio é classificado como uma Variação de Tensão de Curta Duração (VTCD) e é ocasionado por curtos-circuitos ocorridos nas redes de distribuição de energia, energização de grandes cargas e magnetização de transformadores (MOHAMMADI; MORADI; LEBORGNE, 2017).

A presença do evento de distúrbio de afundamento de tensão na rede pode ocasionar danos aos consumidores, uma vez que as cargas possuem restrita tolerância em relação à tensão de alimentação para seu funcionamento correto e seguro. Com o desenvolvimento de novas tecnologias, que na maioria dos casos utiliza dispositivos eletrônicos, as cargas eletroeletrônicas estão prevalecendo nas indústrias em substituição às antigas cargas eletromecânicas, promovendo aumentos na produção e na eficiência no âmbito industrial.

Porém, os equipamentos modernos em geral, possuem maior sensibilidade frente aos distúrbios de qualidade da energia elétrica, principalmente, aqueles relacionados aos eventos de afundamento de tensão (DUAN et al., 2015). Neste contexto é importante garantir uma boa qualidade da energia elétrica por meio da eliminação dos distúrbios de energia elétrica. Indicar o local no sistema de distribuição que está conectado a fonte causadora do distúrbio é o primeiro passo para se diagnosticar o problema de localização do evento de afundamento de tensão.

Além disso, a detecção do local onde ocorreu o distúrbio desempenha também um papel fundamental nas decisões a serem tomadas a respeito dos responsáveis pela ocorrência do evento na rede (MORADI; MOHAMMADI, 2012). Conseqüentemente, o afundamento de tensão torna-se um fenômeno típico de qualidade da energia elétrica, e localizar a fonte causadora de afundamento de tensão para identificar os eventos que causam o problema e melhorar as estratégias de mitigação torna-se uma questão importante para o monitoramento e gerenciamento dos afundamentos.

Definição do Problema

O problema de localização do evento de afundamento de tensão visa indicar o menor local ou região ou distância em relação à subestação no sistema de distribuição que está conectada a fonte causadora do distúrbio. Os principais fenômenos que acarretam em distúrbio de afundamento de tensão são: curto-circuito, energização de grandes cargas ou magnetização de transformadores (SMITH; HENSLEY; RAY, 2009). Dentre estes, o curto-circuito, também denominado de falta no sistema, é a fonte mais comum. No entanto, a localização dessas fontes causadoras do distúrbio não é uma tarefa trivial, pelo fato de que os afundamentos de tensão ocorrem em um curto intervalo de tempo e se propagam ao longo de todo sistema, somando-se ao fato de que existem poucos medidores que conseguem transmitir informações a respeito deste tipo de distúrbio (CHANG et al., 2007; FILHO et al., 2016).

A propagação do evento em relação à sua origem depende de vários fatores como: topologia da rede, intensidade do distúrbio e localização da fonte causadora em relação a subestação, sendo que a cada nova configuração um novo padrão de propagação é evidenciado. Portanto, nem sempre a região de localização será determinada pelo local, no sistema de distribuição, onde o distúrbio foi evidenciado de forma mais severa.

Desta forma, o problema de localização é complexo, visto que estes distúrbios são evidenciados em um curto intervalo de tempo e disseminados com características diferentes ao longo do sistema de distribuição.

Justificativa

Estudos relacionados com as variações de tensão de curta duração têm sido conduzidos, ao longo do tempo, do Monitoramento do sistema elétrico e/ou por meio da utilização de metodologias de simulação. Observa-se que os métodos tipicamente usam informações obtidas de apenas um ponto de monitoramento para determinar a localização relativa da fonte de tensão (a montante ou a jusante) com base em vários critérios, tais como energia de perturbação e energia (PARSONS et al., 1998; FILHO et al., 2016); impedância pelo relé de distância e seu ângulo (PRADHAN; ROUTRAY, 2005). Tais métodos não alcançam uma precisão satisfatória, pois realizam os testes utilizando somente um ponto de medição e analisam a troca de informação em vários pontos de medição (MOHAMMADI; MORADI; LEBORGNE, 2017).

Para contornar tais problemas, grandes avanços podem ser alcançados a partir da utilização da estrutura de comunicação provida pelas redes *smart grids* que oferece dados em tempo real e que podem ser analisados por ferramentas computacionais, garantindo altos índices de confiabilidade, continuidade e segurança. (YAN et al., 2013), (FANG

et al., 2012). Desta forma, os dados advindos do sistema de monitoramento distribuído, disponibilizados pela *smart grid*, podem ser analisados para determinar padrões e inferir na localização dos distúrbios.

Em (FAISAL; MOHAMED, 2009), a Transformada-S é utilizada para extrair as características do sinal elétrico medidos em vários pontos, enquanto o SVM (*Support Vector Machine*) é responsável por definir a zona de ocorrência do distúrbio. Em (SHAREEF; MOHAMED, 2013), propõe um novo método para identificar a localização da fonte de tensão usando a rede neural artificial, no qual as redes de função de base radial (RBF – *Radial Basis Function*) são inicialmente treinadas para estimar as tensões de barramento não monitoradas durante vários afundamentos causados por falhas. Em seguida, o desvio de tensão das barras do sistema é calculado para identificar a localização da queda de tensão. Em (ALI et al., 2016) os autores utilizam redes *bayesianas* para identificar os locais mais adequados para instalação dos medidores de qualidade de energia.

Neste contexto, nota-se que a maioria dos trabalhos utiliza algoritmos supervisionados, os quais dependem de um conjunto de treinamento e não analisam as influências de vários fatores como: topologia da rede, intensidade do distúrbio e localização da fonte causadora em relação à subestação. De forma que a cada nova configuração um novo padrão de propagação é obtido, o que dificulta o desenvolvimento de um algoritmo simples que consiga generalizar todos esses fatores.

Em (SMITH; HENSLEY; RAY, 2009) destaca-se que na ocorrência do evento de afundamento de tensão em um local cuja influência é refletido em toda a rede de distribuição com características diferentes, as quais são influenciadas principalmente pelo local de origem e topologia do sistema. Desse modo, os distúrbios refletidos em pontos próximos apresentam características semelhantes entre si e distintas dos eventos refletidos longe do local da falta.

Algoritmo de agrupamento corresponde a classe de aprendizagem não supervisionada, onde o indutor analisa os exemplos de entrada e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou *clusters* (CHEESEMAN; STUTZ, 1996; MOHAMMADI; MORADI; LEBORGNE, 2017).

Nessa perspectiva, os algoritmos de agrupamento podem ser empregado para agrupar os dados advindos dos *smart meters* de acordo com suas características, evidenciando melhor a propagação do distúrbio, independente da topologia do tamanho da rede de distribuição. Com isto, na busca por informação, justifica-se o emprego de métodos e ferramentas eficientes de algoritmo de agrupamento para a análise e a organização dos dados no instante de ocorrência do evento.

O algoritmo de agrupamento pode ser responsável por destacar melhor a propagação do evento de afundamento de tensão na rede de distribuição, dividindo os dados de entrada

não rotulados em grupos de acordo com a similaridade entre os distúrbios evidenciados na rede e independentes da topologia do sistema.

Nas diversas técnicas de análise de dados, a indução de regras de decisão e de árvores de decisão recebem especial atenção quando o foco é descoberta de conhecimento (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996), a utilizarem representações simbólicas que permitem explicitar o conhecimento de forma compreensível ao ser humano. As representações simbólicas possibilitam o entendimento das fronteiras de decisão presentes nos dados e também da lógica implícita neles (APTÉ; WEISS, 1997).

O objetivo básico da indução de uma árvore de decisão é produzir um modelo de predição ou compreender quais e como as variáveis e interações entre variáveis interferem no fenômeno estudado (MEIRA et al., 2008). Por sua vez, a indução de regras de decisão tem foco voltado à criação de modelos descritivos que são mais fáceis de interpretar que os gerados pelas árvores de decisão (HAN; PEI; KAMBER, 2011).

Partindo dessas perspectivas, a motivação fundamental desta dissertação é apresentar uma metodologia baseada em algoritmo de clusterização combinada com regra de decisão para apontar a região (cluster) de origem do afundamento de tensão em sistemas de distribuição de energia sobre o contexto de *smart grid*. Devido a característica dos algoritmos de clusterização, é importante investigar diferentes tipos, de modo a encontrar o que adapte ao processo de localização da região, no sistema de distribuição, onde o distúrbio de afundamento de tensão tenha sido originado.

Objetivos

O objetivo deste trabalho é apresentar um método capaz de identificar a localização da menor região do distúrbio de afundamento de tensão em sistemas de distribuição de energia, a partir da aplicação de algoritmos de clusterização e regra de decisão.

Os objetivos específicos desse trabalho podem ser pontuados da seguinte maneira:

- Analisar um conjunto de características e informações das formas de onda de tensão que permitam identificar os melhores grupos e por consequência inferir como ocorre a propagação do evento presente no sistema;
- Analisar diferentes algoritmos de clusterização que melhor organizem o conjunto de objetos de entrada em grupos de acordo com medidas de similaridades.
- Extrair um conjunto de dados que servirá de dados de entrada como conjunto de treinamento para o algoritmo de regra de decisão.
- Induzir o conjunto de regras a partir de uma entrada que representa o conjunto de treinamento e os parâmetros de configurações do algoritmo PART.

Publicações

Durante este trabalho de pesquisas, vários artigos foram preparados. Eles estão listados abaixo.

- **José Carlos C. L. da Silva Filho**, Fabbio Anderson Silva Borges, Ricardo de Andrade L.Rabelo e Ivan Saraiva Silva, "A Method for Voltage Sag Source Location Using Clustering Algorithm and Decision Rule Labeling", aceito para publicação em International Joint Conference on Neural Networks 2019 (Março de 2019).
- **José Carlos C. L. da Silva Filho**, Fabbio Anderson Silva Borges, Ricardo de Andrade L.Rabelo, Ivan Saraiva Silva, Ricardo Pinto Teixeira Júnior e Antônio Oséas de Carvalho Filho, "Methods for voltage sag source location by Cluster Algorithm and Decision Rule Labeling with a Comparative Approach of K-means and DBSCAN Clustering Algorithms", Submetido para IEEE International Conference on Systems, Man, and Cybernetics 2019 (Março de 2019).

1 Estado da Arte

Neste capítulo são apresentados conceitos associados a localização das fontes de afundamento de tensão. Inicialmente, descreve-se o que é a classe dos distúrbios de afundamento de tensão. Em seguida, discute-se os métodos de localização dos distúrbios de qualidade de energia elétrica.

1.1 Variações de tensão de curta duração (VTCDs)

As VTCDs são os distúrbios de Qualidade da Energia Elétrica (QEE) que mais afetam os sistemas industriais, podendo resultar de partida de grandes motores de energização de transformadores e, sobretudo, de ocorrência de curtos-circuitos em extensas áreas, mesmo que localizadas em pontos remotos da onde se está analisando no sistema elétrico (ROMERO; GALLEGO; PAVAS, 2012). No Brasil, de acordo com os Procedimentos de Rede elaborados pelo Operador Nacional do Sistema (ONS, 2011), no item gerenciamento dos indicadores de desempenho da rede básica e de seus componentes, é definida VTCD como um evento aleatório de tensão caracterizado por desvio significativo, por curto intervalo de tempo, do valor eficaz da tensão. Os prejuízos associados a estes distúrbios são quantificados pelas perdas de produção devido às interrupções de processos, e custos relacionados à mão de obra e reparos ou substituição de equipamentos danificados.

Desta forma, as VTCDs se destacam por corresponder aos eventos mais comuns de QEE (DUGAN et al., 1996), (FILHO et al., 2016), sendo que sua ocorrência pode acarretar grandes prejuízos para o consumidor. A norma *IEEE Std 1159-2009, smith1995ieee* define as VTCDs como sendo desvios no valor eficaz da tensão em curtos intervalos de tempo com duração entre 1 ciclo e 1 minuto, sendo classificadas de acordo com a duração de ocorrência e da amplitude do sinal, como apresentado na Tabela 1. Com relação à amplitude, as VTCDs são classificadas como interrupção de tensão, elevação de tensão e afundamento de tensão.

Os distúrbios citados são ilustrados em relação a uma forma de onda senoidal na Figura 1.

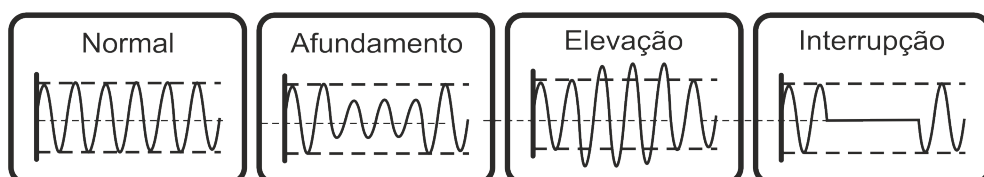


Figura 1 – Principais distúrbios de energia

O distúrbio de interrupção corresponde a uma variação da tensão abaixo de 0,1 p.u. do valor eficaz da tensão nominal e são resultados de curtos-circuitos no sistema de energia elétrica. Já os sinais que apresentam uma variação acima de 1,1 p.u. são denominados como elevações de tensão. A elevação está associada às abruptas falhas no sistema ou decréscimo na absorção de correntes por cargas indutivas, assim como, chaveamento de grandes bancos de capacitores.

Quanto ao afundamento de tensão, este corresponde a uma redução da amplitude da tensão, entre 0,9 e 0,1 p.u. do valor eficaz da tensão nominal. Este tipo de distúrbio é ocasionado por curtos-circuitos ocorridos nas redes de distribuição de energia elétrica, energização de grandes cargas e magnetização de transformadores (SMITH; HENSLEY; RAY, 2009).

Dos três tipos de distúrbios, o afundamento de tensão destaca-se como sendo a alteração da forma de onda que mais se manifesta nas redes elétricas (MALANGE, 2010; PASSOS et al., 2015), sendo que a presença desses eventos na rede ocasiona danos aos consumidores, uma vez que as cargas pode se possuir restrita tolerância em relação à tensão de alimentação. Portanto é importante garantir uma boa qualidade de energia elétrica por meio da eliminação desse distúrbio.

Como se observa na Tabela 1, o afundamento de tensão se insere em uma das categorias de VTCDs e de acordo com sua duração pode ser classificado como sendo momentâneo ou temporário. Conseqüentemente, o afundamento de tensão torna-se um fenômeno típico de qualidade da energia elétrica, e localizar a fonte causadora de afundamento de tensão para identificar os eventos que causam o problema e melhorar as estratégias de mitigação torna-se uma questão importante para o monitoramento e gerenciamento dos afundamentos, com a qual podem ser utilizados algoritmos de aprendizagem.

1.2 Métodos para localização da origem dos distúrbios de QEE

Os métodos para localização da origem dos distúrbios de qualidade de energia elétrica têm como finalidade indicar o local no sistema de distribuição que está conectado a fonte que originou o distúrbio. A correta localização da fonte e a detecção do tipo de distúrbio é o primeiro passo para se diagnosticar o problema. Como resultado, medidas de controle podem ser executadas de modo a aumentar a QEE. De modo suplementar, a detecção do local desempenha também um papel fundamental nas decisões a serem tomadas a respeito dos responsáveis pela ocorrência do evento na rede. Deste modo, a localização precisa das fontes de distúrbios de QEE é muito importante e vem sendo alvo de pesquisas.

Foi realizado um levantamento bibliográfico com foco principal a localização dos eventos de afundamento de tensão, isto ocorre devido ao fato de que tal distúrbio ocorre

Tabela 1 – Classificação das Variações de Tensão de Curta Duração (PRODIST, 2014)

Classificação	Denominação	Duração da Variação	Amplitude da tensão (valor eficaz) em relação à tensão de referência
Variação Momentânea de Tensão	Interrupção Momentânea de Tensão	Inferior ou igual a três segundos	Inferior a 0,1 p.u
	Afundamento Momentâneo de Tensão	Superior ou igual a um ciclo e inferior ou igual a três segundos	Superior a 0,1 p.u e inferior a 0,9 p.u
	Elevação Momentânea de Tensão	Superior ou igual a um ciclo e inferior ou igual a três segundos	Superior a 1,1 p.u
Variação Temporária de Tensão	Interrupção Temporária de Tensão	Superior a três segundos e inferior a um minutos	Inferior a 0,1 p.u
	Afundamento Temporária de Tensão	Superior a três segundos e inferior a um minutos	Superior a 0,1 p.u e inferior a 0,9 p.u
	Elevação Temporária de Tensão	Superior a três segundos e inferior a um minutos	Superior a 1,1 p.u

com mais frequência na rede. Depois de selecionar os artigos que mais estavam relacionados com o tema desta dissertação, foi feito um breve resumo de cada um destes trabalhos e os mais relevantes serão apresentados a seguir.

É apresentado por (ROMERO; GALLEGO; PAVAS, 2012) uma nova metodologia para estimação de padrões de afundamento de tensão e de agrupamento das zonas de falta em sistemas de alta e média tensão. A metodologia proposta compreende três estágios. Primeiro, modelagem da rede elétrica e simulação de faltas foram realizadas a fim de obter informações sobre os afundamentos de tensão causados por falhas no sistema de transmissão. Os padrões para AMTs (Afundamento Momentâneo de Tensão) foram identificados no segundo estágio, pelas médias do algoritmo de formação de conjuntos *k-médias*, permitindo a determinação das zonas de faltas. Usando a base de dados de medições de QEE de uma grande companhia de eletricidade de Bogotá, os AMTs foram classificados de acordo com os padrões previamente estabelecidos. No terceiro estágio da metodologia uma comparação entre simulação e medição foi realizada, permitindo a identificação dos afundamentos causados por faltas.

Foi desenvolvido por (BALOUJI; SALOR, 2014) um método para o agrupamento e classificação dos eventos de qualidade da energia elétrica, baseado na análise de autovalores. A pretensão dos autores é a de que tal método sirva para suprir as necessidades das aplicações futuras das redes elétricas inteligentes. O procedimento é fundamentado na

técnica de agrupamento *k*-médias, que usa os valores de tensão RMS, como sugerido pela norma IEC 61000-4-30. Para reduzir o número de conjuntos formados e para aumentar a eficiência do agrupamento, foram aplicadas duas *EingenAnalysis*, análise de componentes principais (PCA) e análise linear de discriminante (LDA) antes do processo de formação de conjuntos pelo método das *k*-médias.

Para resolver o problema de identificação da fonte causadora do distúrbio de afundamento de tensão, pois devido à correlatividade e à redundância das características, os métodos de identificação das fontes dos distúrbios são pouco precisos (WANG; WANG; JIA, 2014), propõem um método para redução das características dos eventos de afundamentos de tensão, baseado em análise de componentes principais (PCA). Foram analisadas as oscilografias de eventos de afundamentos de tensão causadas por faltas no sistema, partidas de motores de indução, energização de transformadores e três tipos de eventos resultantes do efeito combinado de fontes de AMTs. Os índices das características relacionadas aos afundamentos de tensão foram reduzidos através de PCA, que é uma técnica que permite transformar o espaço das características de muitas dimensões para poucas dimensões, mantendo a informação das características e eliminando as correlações entre as características. Os resultados de simulação mostraram que o método proposto reduz efetivamente o vetor de características iniciais, e a precisão da classificação e identificação das fontes causadoras de AMTs obtida com este método é melhor quando comparada aos métodos que não utilizam a técnica PCA.

Outra pesquisa relacionada com o tema em questão foi realizada por (HARIYANTO; ANGGORO; NOEGROHO, 2014), onde foi introduzida uma nova maneira de identificar o perfil de severidade da tensão de um barramento devido a curto-circuito ocorrido em determinado ponto do sistema de distribuição de energia. Um curto-circuito causa um afundamento na tensão por um período relacionado ao tempo de operação da proteção. Os dados contendo as duas variáveis (amplitude e duração) do afundamento de tensão devido ao curto-circuito foram gerados para cada barramento do sistema. Posteriormente, estes dados de todas as barras foram separados em conjuntos pelo método de *k-means*. E este processo produzirá barramentos de referência (que serão os centros dos conjuntos) e mostrará os barramentos associados a cada referência. Para ser capaz de perceber a severidade do afundamento de tensão, os centros dos conjuntos serão convertidos para o índice de severidade do evento, que explica a severidade do AMT, com base na curva CBEMA-ITIC. Dados de um determinado barramento que está sujeito a eventos de afundamento de tensão serão classificados com base no conjunto que está associado. Assim, será obtido a frequência de eventos que são classificados em conjuntos particulares. Para melhor observar os dados recomenda-se o uso de histogramas.

Em 2015, (DUAN et al., 2015) apresentaram uma abordagem de algoritmo de agrupamento *k-means* para analisar, classificar e reconhecer afundamentos de tensão de

dados históricos de medição de rede em grande escala em Shenzhen, China. São calculadas as distâncias entre diferentes eventos de afundamentos em um diagrama de distribuição, e quando alguns deles estão próximos, um centro de conjunto denominado *centroide* pode ser fixado para representar esses eventos. A amplitude e duração refletidas por esses *centroides* podem ser consideradas como as características de afundamentos de tensão de subestações semelhantes, que representarão a condição de operação e identificarão o elo mais fraco do sistema elétrico.

No trabalho realizado por (MAHELA; SHAIK, 2017) é apresentada uma técnica para reconhecer os distúrbios de qualidade da energia elétrica (QEE) associados à penetração da energia solar na rede de distribuição. O sinal de tensão é analisado com a ajuda da transformação de *Stockwell* (*S-transform*), em seguida, seus recursos são utilizados para reconhecer eventos de qualidade de energia elétrica com a ajuda do *Fuzzy C-means clustering* (FCM). As operações consideradas para a avaliação do QEE incluem a sincronização da rede e a interrupção do sistema solar fotovoltaico (PV). Os eventos de QEE associados à mudança súbita na insolação solar também foram investigados.

Em (HAMZAH; MOHAMED; HUSSAIN, 2004) propõem um método para localizar a fonte de queda de tensão em um sistema de distribuição de energia. O método proposto usa a polaridade do componente de corrente real para determinar a localização da depressão em relação ao ponto de monitoramento. O produto da corrente RMS e o ângulo do fator de potência no ponto de monitoramento são empregados para a localização da fonte de afundamento.

O método proposto em (SHAREEF; MOHAMED, 2013) identifica o local de origem do afundamento de tensão a partir do maior desvio de tensão caracterizado em cada barra, o qual é estimado por meio de uma rede neural com função de base radial. A metodologia é validada no sistema IEEE 30 barras.

Em (AHN et al., 2008), a localização do afundamento de tensão é realizada utilizando vários medidores, em que cada um é capaz de determinar se o fluxo do evento está a montante ou a jusante. Cada medidor tenta identificar a direção da fonte de acordo com a causa que originou o distúrbio, pois para cada dependendo da fonte que originou o distúrbio ele é evidenciado com características diferentes. O algoritmo proposto analisa a amplitude da corrente de falta para distúrbios causados por saturação de transformadores, o aumento da potência ativa para distúrbios ocasionados pela partida do motor de indução e a amplitude do segundo componente harmônico para distúrbios causados por curto-circuito. O método é testado em um sistema prático da Coreia e no sistema IEEE 13-nós.

No trabalho de (WON et al., 2006), é apresentado um algoritmo sistemático para localizar a fonte de evento da qualidade da energia elétrica (QEE) no sistema de monitoramento distribuído. O algoritmo utiliza a topologia do sistema de potência e a direção dos eventos de QEE determina as áreas candidatas para uma fonte de eventos, manipulando

essas duas matrizes. O local da fonte é determinado pelo índice do elemento que apresenta um valor igual ao número de monitores instalados no sistema, significando assim, que o fluxo de energia elétrica de todos os monitores foi dirigido para o monitor localizado nessa posição.

Em (TAYJASANANT; LI; XU, 2005), a localização é determinada pelo princípio de estimar a impedância equivalente do lado não perturbado, utilizando as mudanças de tensão e de corrente causadas pela perturbação. O sinal da parte real da impedância estimada pode revelar se a perturbação é de montante ou jusante. A dificuldade deste método está na escolha do período de medição, antes e depois do distúrbio, que pode afetar a exatidão do resultado

2 Metodologia

Neste capítulo, é apresentada a metodologia proposta que é baseada em algoritmo de clusterização combinado com regra de decisão para localizar a origem do distúrbio de afundamento de tensão em sistema de distribuição de energia elétrica. Faz-se inicialmente um resumo da abordagem geral do método proposto. Posteriormente, detalha-se cada etapa na geração dos resultados, algoritmos de clusterização, análise dos resultados, obtenção da base de treinamento e aplicação da Regra de decisão.

2.1 Abordagem geral do Método para Localização da Fonte de Afundamento de Tensão por Algoritmo de Clusterização e Rotulação por Regra de Decisão

A metodologia utilizada é dividida em duas fases. A primeira, é responsável pela obtenção e análise dos clusters gerados a partir das características obtidas de uma falha em um sistema de distribuição de energia. Para tanto, é necessário investigar um algoritmo de clusterização que seja capaz de inferir sobre a propagação do evento de afundamento de tensão na rede de distribuição, dividindo os dados de entrada não rotulados em grupos de acordo com a similaridade entre os distúrbios evidenciados na rede e independentes da topologia do sistema. A segunda fase corresponde na aplicação do sistema baseado em regras de decisões, a qual é responsável por analisar as características de cada um dos clusters gerados e identificar qual deles fornece a região de localização do distúrbio de afundamento de tensão.

A primeira etapa se inicia com a base de dados contendo as características extraídas do distúrbio que são mensurados em cada ponto de medição do sistema de distribuição, as quais evidenciam as diferenças entre os eventos. Faz-se um detalhamento dessas características na sessão 2.2 deste capítulo. Em seguida esses dados sofrem um processo de normalização descrito na sessão 2.2.5.

Com os dados já normalizados eles são repassados como variáveis de entrada para os algoritmos de clusterização, como mostra a Figura 2. Existem na literatura diferentes algoritmos de clusterização, os quais possuem particularidades importantes e agrupam os dados de entrada sobre diferentes heurísticas. Dessa forma, a metodologia é implementada considerando diferentes algoritmos de clusterização, baseados em método hierárquico, algoritmos particionais, baseado em densidade, baseados em Modelo e baseados em Lógica Fuzzy. Os algoritmos são apresentados na sessão 2.3.

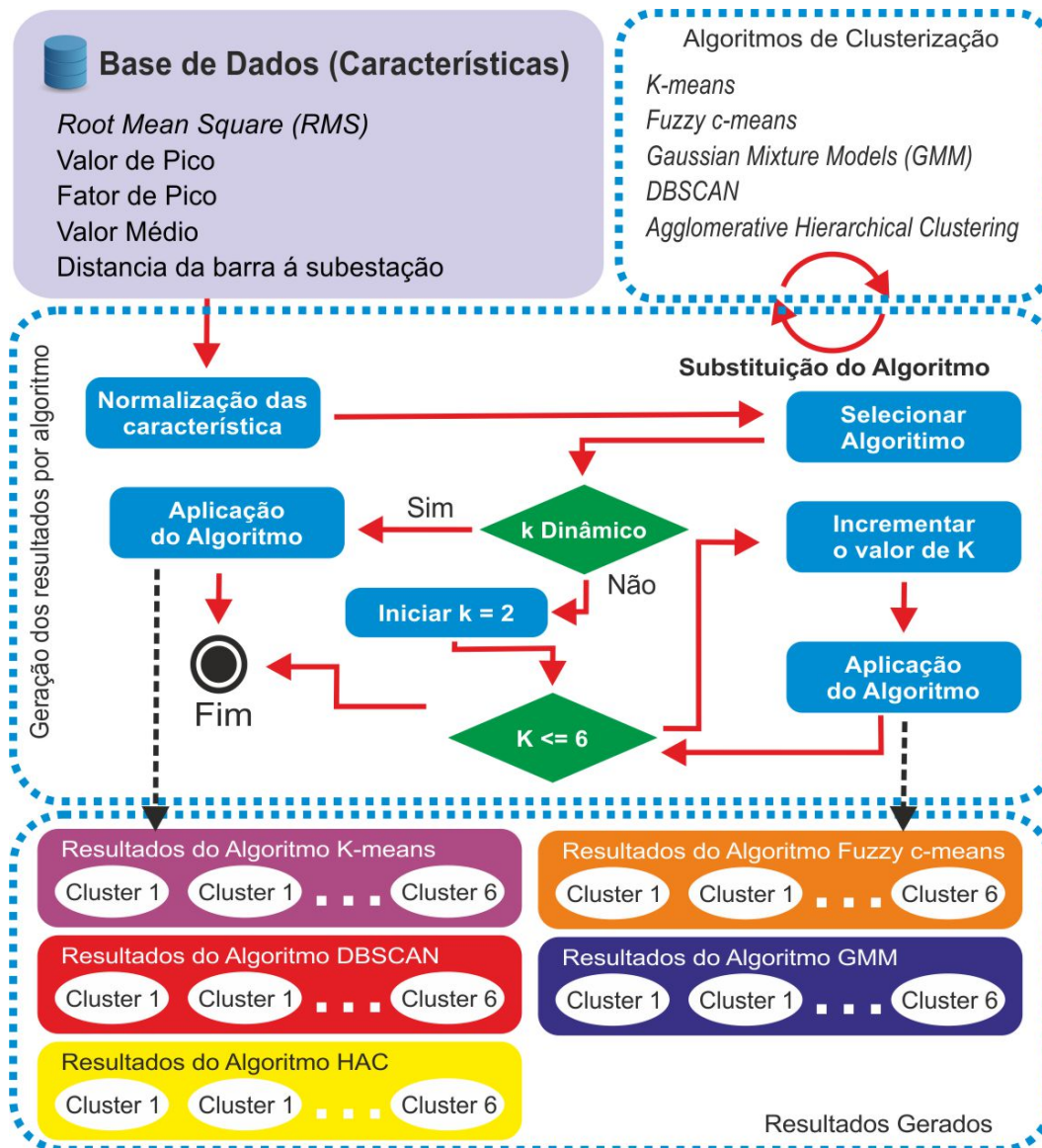


Figura 2 – Abordagem para análise dos algoritmos de clusterização - fase 1.

Após a execução do algoritmo de clusterização os dados de entrada são representados de forma mais simples ou mais organizados, ao passo que eles podem ser analisados, de modo a obter padrões em objetos pertencentes ao mesmo cluster e diferenças entre objetos pertencentes a cluster diferentes (FAHAD et al., 2014). Os dados são processados pelos algoritmos de clusterização, para posterior análise e localização.

Em seguida é necessário determinar em qual dos clusters está localizado a origem do distúrbio. Desse modo, o sistema baseado em regra irá utilizar como entrada as medidas que caracterizam cada grupo de modo a rotular a partir da sua base de regras o grupo que agrega a região de origem, como mostra a figura 3. As características RMS, Valor de pico, Fator de Pico, Valor Médio e Distância extraídas de cada distúrbio medido e pertencente à um cluster são avaliadas através de cinco métricas estatísticas: média, desvio padrão, amplitude, assimetria e curtoses. Adicionalmente, foi averiguada métrica para quantificar

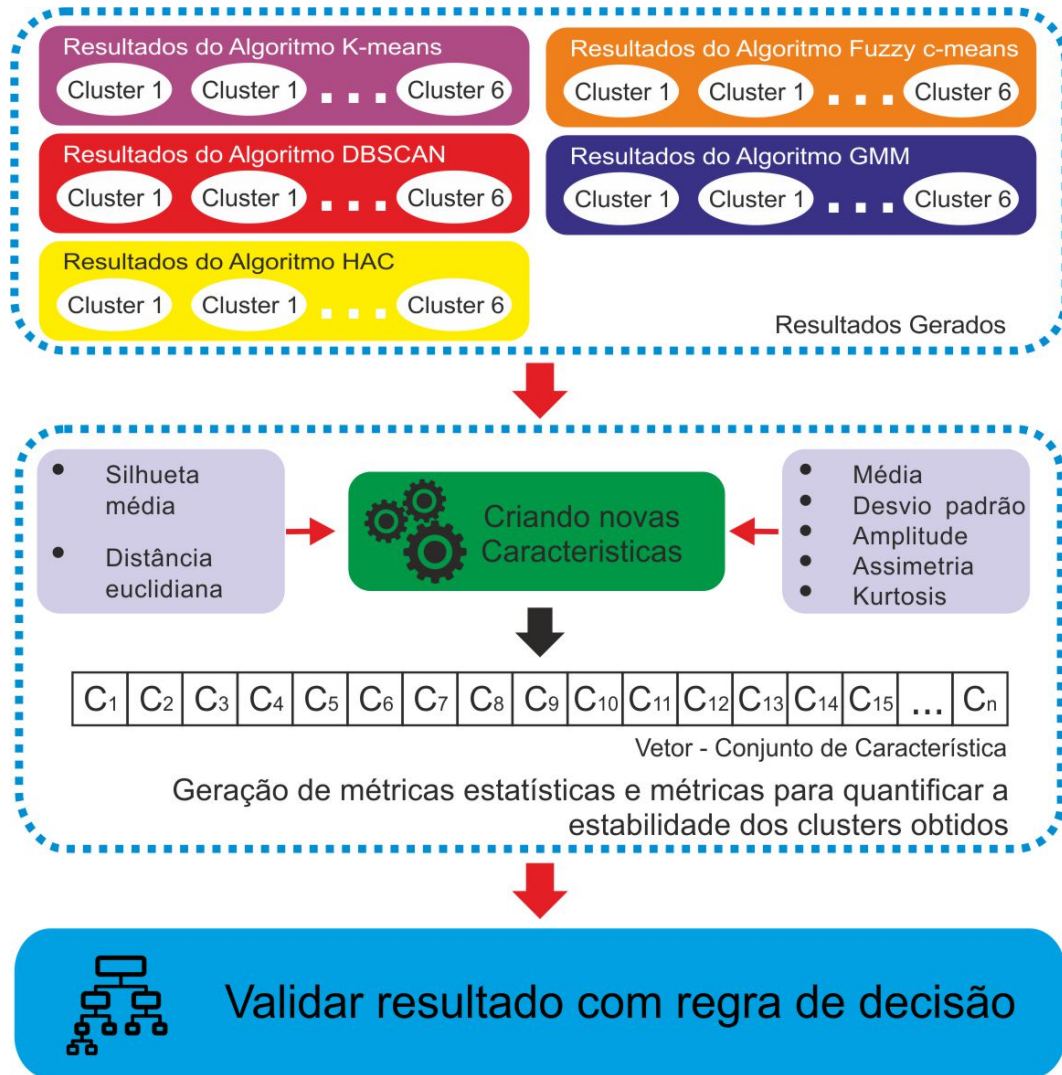


Figura 3 – Criação do conjunto de métricas para validar acurácia dos clusters formados - fase 2.

a estabilidade dos clusters obtidos: silhueta média. Esse conjunto de métricas é detalhado na Tabela 2 da sessão 2.6.

2.2 Características Analisadas

Para analisar cada distúrbio evidenciado no sinal de tensão, é importante primeiro extrair características desses sinais que evidenciam as diferenças entre os eventos medidos em cada barra do sistema de distribuição. Neste estudo são investigadas, para formação dos clusters, cinco características que são: RMS, Valor de Pico, Fator de Pico, Valor Médio e Distância da Barra à Subestação (BORGES, 2017).

As quatro primeiras medidas são obtidas no domínio do tempo e tem relação com o fato que os distúrbios de VTCDs alteram os valores de tensão da forma de onda senoidal pura. Assim, elas destacam as características de amplitude do evento de afundamento

de tensão. A quinta característica, distância da barra à subestação, tem como objetivo destacar a influencia que a topologia pode causar na propagação do distúrbio.

2.2.1 Root Mean Square (RMS)

O RMS ou valor quadrático médio corresponde ao cálculo médio estatístico de uma grandeza variável que oscila entre valores positivos e negativos em relação à grandeza elétrica (LEONARDO; OLIVEIRA, 2016). O valor RMS é expresso por:

$$RMS = \sqrt{\frac{1}{N} * \sum_{j=1}^N d_j^2} \quad (2.1)$$

Na qual N é o tamanho do distúrbio discretizado durante o tempo de ocorrência do distúrbio e d_j equivale ao j -ésimo ponto do distúrbio segmentado de tamanho N .

2.2.2 Valor de Pico

O Valor de Pico de uma onda senoidal corresponde ao máximo valor absoluto dentre os valores que compõem a forma de onda que representa o distúrbio, sendo numericamente igual à amplitude de sua senóide e tem seu valor expresso por:

$$ValorDePico = \max(d_1, d_2, \dots, d_j, \dots, d_n) \quad (2.2)$$

Sendo, d_j equivale ao j -ésimo ponto do distúrbio segmentado de tamanho N .

2.2.3 Fator de Pico

O Fator de Pico calcula a relação entre o Valor de Pico e o valor RMS, sendo utilizado para identificar a presença de valores atípicos ou pontuais presentes no sinal senoidal em relação ao valor eficaz (LACERDA, 2017). Sendo expresso por:

$$FatorDePico = ValorDePico/RMS. \quad (2.3)$$

Sendo, ValorDePico descrito na equação 2.2 e RMS pela equação 2.1.

2.2.4 Valor Médio

O Valor Médio foi utilizado para fim de extrair características do transitório oscilatório (LUZ, 2018) presente no início da ocorrência de alguns distúrbios. Para uma onda senoidal sem distúrbio o valor médio é igual a 0. Porém, ondas senoidais distorcidas por transitórios oscilatórios apresentam variações do valor médio.

Mesmo existindo outras ferramentas como Transformada Rápida de *Fourier* ou Transformada *Wavelet* que extraem informações mais precisas do transitório oscilatório, o valor médio foi escolhido devido ao seu baixo poder computacional exigido.

O valor médio é expresso por:

$$ValorMedio = \frac{\sum_{j=1}^N d^j}{N} \quad (2.4)$$

Na qual N é o tamanho do distúrbio discretizado durante o tempo de ocorrência do distúrbio e d_j equivale ao j -ésimo ponto do distúrbio segmentado de tamanho N .

2.2.5 Normalização

Os algoritmos de agrupamento calculam a similaridade ou dissimilaridade entre as características de entrada. Assim, é importante garantir que cada variável exerça a mesma influência para o modelo, para tanto utiliza a normalização para escalonar as características de entrada no mesmo intervalo, de modo que exista uma comparação justa entre eles.

Destaca-se que para evento de afundamento de tensão quanto menor for o valor da característica mais severo é a consequência do evento. Desse modo, a normalização dos valores agregados das características *Root Mean Square* (RMS), Valor de Pico, Fator de Pico e Valor Médio tem uma particularidade na formulação de suas equações, onde o resultado da normalização é atribuído de forma inversa sendo que menor valor do conjunto de entrada é normalizado para 1 e o valor máximo é normalizado para 0, com suas equações expressas, respectivamente, por:

$$normalizacaoRMS_i = \frac{RMS_{max} - RMS_i}{RMS_{max} - RMS_{min}} \quad (2.5)$$

$$normalizacaValorPico_i = \frac{ValorPico_{max} - ValorPico_i}{ValorPico_{max} - ValorPico_{min}} \quad (2.6)$$

$$normalizacaFatorPico_i = \frac{FatorPico_{max} - FatorPico_i}{FatorPico_{max} - FatorPico_{min}} \quad (2.7)$$

$$normalizacaValorMedio_i = \frac{ValorMedio_{max} - ValorMedio_i}{ValorMedio_{max} - ValorMedio_{min}} \quad (2.8)$$

Em relação a normalização da distância da barra à subestação segue a equação padrão. Com o máximo valor da distância da barra à subestação sendo normalizado com valor 1 e a menor distância da barra à subestação recebe o valor 0.

$$normalizacaDistancia_i = \frac{Distancia_i - Distancia_{min}}{Distancia_{max} - Distancia_{min}} \quad (2.9)$$

Para as equações 2.5, 2.6, 2.7, 2.8 e 2.9, i representa cada um dos indicadores mensurados dentro do intervalo normalizado, max o maior valor encontrado dentro do intervalo normalizado e min o menor valor encontrado dentro do intervalo normalizado.

2.3 Algoritmos de Clusterização

A análise de agrupamentos, também conhecida como análise de conglomerados ou *cluster*, tem como objetivo particionar os elementos da amostra, ou população, em grupos de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com respeito às variáveis (características) que neles foram medidas, e os elementos em grupos diferentes sejam heterogêneos em relação a estas mesmas características (MINGOTI, 2005).

Mesmo com o auxílio de um computador com grande capacidade de processamento, é impossível verificar todas as possibilidades de agrupamentos em problemas onde o número de objetos e variáveis são consideráveis. Devido a este problema, foram criados uma variedade de algoritmos para formação de conjuntos que, sem olhar para todas as configurações possíveis de agrupamento, permitem encontrar agrupamentos razoáveis (SOUZA, 2016).

Os métodos de clusterização mais utilizados podem ser divididos de acordo com os seguintes aspectos (XU; WUNSCH, 2005), (FAHAD et al., 2014):

- Método hierárquico: Organizam um conjunto de dados em uma estrutura hierárquica de acordo com a proximidade entre os indivíduos. Os resultados de um algoritmo hierárquico são normalmente mostrados como uma árvore binária ou dendograma, que é uma árvore que iterativamente divide a base de dados em subconjuntos menores (DONI, 2004).
- Os algoritmos particionais: dividem a base de dados em k -grupos, onde o número k é dado pelo usuário. Inicialmente, o algoritmo escolhe k objetos como sendo os centros dos k *clusters*. Os objetos são divididos entre os k *clusters* de acordo com a medida de similaridade adotada, de modo que cada objeto fique no *cluster* que forneça o menor valor de distância entre o objeto e o centro do mesmo. Então, o algoritmo utiliza uma estratégia iterativa de controle para determinar que objetos devem mudar de *cluster*, de forma que a função objetivo usada seja otimizada (DONI, 2004).
- Método baseado em densidade: Os *clusters* são definidos como regiões densas, separadas por regiões menos densas que representam os ruídos. As regiões densas podem ter uma forma arbitrária e os pontos dentro de uma região podem também estar distribuídos arbitrariamente e, por isso, os métodos baseados em densidade são adequados para descobrir *clusters* com forma arbitrária, tais como elíptica, cilíndrica,

espiralada, etc. até os completamente cercados por outro “cluster” e são especialistas em identificar e filtrar ruídos (HAN; PEI; KAMBER, 2011).

- Algoritmos de clusterização baseados em Modelo são métodos que usam um modelo de referência para cada cluster. Eles tentam otimizar a curva entre os objetos dados e algum modelo matemático. Um algoritmo baseado em modelo pode descobrir "clusters" construindo uma função de densidade que reflete a distribuição espacial dos pontos de dados. Ele também conduz a um modo de determinar automaticamente o número de clusters baseado na estatística padrão, identificando ruídos no relatório e assim produzindo métodos de Clusterização robustos. Tais modelos são, frequentemente, baseados na suposição que os dados são gerados por uma mistura de distribuições de probabilidades (HAN; PEI; KAMBER, 2011).
- Os métodos de clusterização baseados em Lógica Fuzzy são métodos não ‘hard’, que permitem associar um indivíduo a de todos os clusters usando uma função de pertinência (ZADEH, 1965). A restrição adotada nesta metodologia é que a soma dos graus de pertinência de um indivíduo aos clusters seja igual a 1. Em um algoritmo de clusterização fuzzy, cada cluster é um conjunto fuzzy de todos os indivíduos (HAN; PEI; KAMBER, 2011).

Neste contexto foram escolhidos 5 algoritmos de clusterização para realização da análise de comparação, o critério de seleção teve como base selecionar um algoritmo para cada um dos métodos: hierárquico, particional, baseado em densidade, baseado em modelo e Método baseado em agrupamento fuzzy. Também foram verificadas suas vantagens e desvantagens, que serão descritas nas sub-sessões 2.3.1, 2.3.2, 2.3.3, 2.3.4 e 2.3.5.

2.3.1 *K-Means Clustering*

O algoritmo *k*-Means foi primeiramente descrito por (HARTIGAN, 1975) um dos objetivos do *k*-means é encontrar similaridades entre os dados da amostra e agrupá-los conforme o número de cluster. O *k*, de *k*-Means, é a quantidade de centróides (pontos centrais dos grupos) que serão criados e auxiliará no objetivo de encontrar a similaridade dos dados. Tem como método algoritmos particionais.

O *k*-Means tem a vantagem de ser bem rápido, já que são calculados apenas as distâncias entre pontos e centros de grupo. Assim, tem uma complexidade linear $O(n)$. Por outro lado, *k*-Means tem algumas desvantagens. Em primeiro lugar, é necessário indicar a quantidade de grupos, ou seja, o número de *k*. Além disso o algoritmo de *clustering k-means* começa com uma escolha aleatória de centros de *cluster* e, portanto, pode gerar resultados de *cluster* diferentes em diferentes execuções do algoritmo. Assim, os resultados podem não ser consistentes.

2.3.2 Fuzzy *c*-means

Fuzzy c-means (FCM) é uma técnica de algoritmo baseado em lógica *fuzzy*. Esta técnica foi originalmente introduzida por *Jim Bezdek* em 1981 como uma evolução das técnicas de agrupamento de dados mais recentes e fornece um método que mostra como agrupar padrões que pertencem a um espaço multidimensional em um número específico de diferentes agrupamentos (VARGAS, 2012).

Como vantagem o algoritmo FCM tem o objetivo de encontrar grupos *fuzzy* para um conjunto de dados. Para alcançar este objetivo, o algoritmo precisa minimizar uma função que diz respeito à minimização das distâncias entre os dados e os centros dos grupos aos quais tais dados pertencem com algum grau de pertinência. A minimização da função citada deve produzir grupos mais adequados (melhores ou mais naturais) do que aqueles produzidos pelo algoritmo *c-Means* clássico (PERES et al., 2012). O FCM é uma extensão do *k-means* baseada em lógica *fuzzy*. Por consequência tem a mesma desvantagem do *k-means* em relação a selecionar quantos grupos existem.

2.3.3 Gaussian Mixture Models (GMM)

O Algoritmo GMM é uma abordagem baseada em modelo, que consiste em usar determinados modelos para *clusters* e tentar otimizar o ajuste entre os dados e o modelo. Na prática, cada *cluster* pode ser representado matematicamente por uma distribuição paramétrica, como um gaussiano (contínuo) ou um Poisson (discreto). O conjunto de dados inteiro é, portanto, modelado por uma mistura dessas distribuições. Uma distribuição individual usada para modelar um *cluster* específico é geralmente chamada de distribuição de componentes (HE et al., 2011).

Enquanto o algoritmo K-means executa uma atribuição difícil de pontos de dados a clusters, nos quais cada ponto de dados é associado unicamente a um cluster, o algoritmo GMM com EM (Expectativa-Máxima) faz uma atribuição leve com base nas probabilidades posteriores (BISHOP, 2006).

2.3.4 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

O modelo de clustering baseado na densidade define os conjuntos de objetos ou clusters, criando relações entre os objetos que se encontram nas áreas mais densas da região. O algoritmo DBSCAN é um algoritmo de baseado no modelo de densidade (ESTER et al., 1996; REHMAN; MEHDI, 2005).

Sua utilização é vantajosa por não é necessário especificar o número de clusters *a priori*, consegue definir clusters de diferentes formas, como por exemplo criar um cluster *à volta* de outro e Utiliza a noção de ruído. A principal desvantagem do DBSCAN é que

ele não funciona tão bem quanto os *clusters* são de densidade variável ou com dados de dimensões muito altas.

2.3.5 Agglomerative Hierarchical Clustering

O agrupamento hierárquico ascendente é, portanto, chamado de *clustering* aglomerativo hierárquico ou HAC. Essa hierarquia de *clusters* é representada como uma árvore (ou dendrograma). A raiz da árvore é o *cluster* exclusivo que reúne todas as amostras, sendo as folhas os *clusters* com apenas uma amostra (DAVIDSON; RAVI, 2009).

O armazenamento em *cluster* hierárquico não exige que especifiquemos o número de *clusters* e podemos até selecionar qual número de *clusters* parece melhor, já que estamos construindo uma árvore. Além disso, o algoritmo não é sensível à escolha da métrica de distância; todos eles tendem a funcionar igualmente bem, enquanto que com outros algoritmos de agrupamento, a escolha da métrica de distância é crítica. Um caso de uso particularmente bom de métodos de agrupamento hierárquico é quando os dados subjacentes têm uma estrutura hierárquica e você deseja recuperar a hierarquia; outros algoritmos de *clustering* não podem fazer isso. Essas vantagens do *clustering* hierárquico gera uma desvantagem, já que tem uma complexidade temporal de $O(n^3)$ (DAY; EDELSBRUNNER, 1984), ao contrário da complexidade linear de *k-Means*.

2.4 Análise dos Algoritmos de Clusterização

Existem muitos critérios de validade de agrupamentos relativos diferentes que são muito úteis na prática como medidas quantitativas para avaliar a qualidade das partições de dados. Esses critérios são dotados de características particulares que podem tornar cada um capaz de superar outros em classes específicas de problemas. Além disso, eles podem ter requisitos computacionais completamente diferentes. Então, é uma tarefa difícil para o usuário escolher um critério específico quando ele enfrenta uma variedade de possibilidades.

Independentemente do método de análise de agrupamentos utilizado, os *clusters* resultantes devem ser compactos, bem separados e estáveis (PESSANHA et al., 2015). Contudo, há medidas mais gerais, denominadas por medidas de validação, que permitem avaliar a qualidade dos agrupamentos identificados.

Em geral, os melhores desempenhos têm sido associados às silhuetas, PBM, VRC e point-biserial. Entre eles, as silhuetas aparentemente mostraram os desempenhos mais robustos em relação aos diferentes cenários de avaliação (VENDRAMIN; CAMPELLO; HRUSCHKA, 2010).

A Largura de Silhueta (LS) baseia-se nos conceitos de compactação intra-grupo e separação inter-grupos. A silhueta de um objeto individual x_i pertencente a um grupo

$C_l(S(x_i))$ é definida por:

$$S(x_i) = \frac{b(x_i) - a(x_i)}{\max(b(x_i), a(x_i))} \quad (2.10)$$

sendo $a(x_i)$ e $b(x_i)$ calculados como:

$$a(x_i) = \frac{1}{|C_l| - 1} \sum_{x_j \in C_l} d(x_1, x_j), \quad (2.11)$$

$$b(x_i) = \min_{\substack{0 \leq h \leq k \\ h \neq 1}} \left\{ \frac{1}{|C_h|} \sum_{x_j \in C_h} d(x_1, x_j) \right\} \quad (2.12)$$

Com isso, a LS é definida como a silhueta média entre todos os objetos do conjunto de dados (GIL; FERRARI; EMMENDORFER, 2015). Ou seja:

$$LS(C) = \frac{1}{n} \sum_{i=1}^n S(x_i) \quad (2.13)$$

A LS assume valores no intervalo $[1,1]$. A melhor partição possível segundo esse critério é aquela que atinge $LS(C) = 1$.

Nesse contexto, a avaliação dos algoritmos de clusterização utilizado nesse trabalho terá como medida de avaliação a Largura de silhueta.

2.5 Regra de decisão

Após o agrupamento dos dados, torna-se possível ressaltar os padrões similares que estavam intrínsecos nos dados de entrada. Fundamentado nesse raciocínio, é possível confrontar as métricas que resumem as particularidades e validam cada cluster, com a finalidade de apontar em qual grupo o nó de origem do distúrbio está localizado.

Para isto, é aplicado um sistema baseado em regras de decisões, o qual faz uso de um conjunto de estruturas do tipo “se...então”, sendo capaz de inferir sobre a região de interesse reduzida por meio de regras compacta e de fácil interpretação. Adicionalmente, é possível expressar o conhecimento do especialista de forma bem próxima à linguagem natural.

Para a construção automática do conjunto de regras, optou-se por aplicar algoritmo Partial Decision Trees (PART), devido a sua simplicidade e por retornar o conjunto de regra otimizado (FRANK; WITTEN, 1998). O algoritmo PART é baseado no algoritmo j48 (JÚNIOR, 2018) e, portanto, cada regra é induzida com base na análise da entropia do conjunto de dados, o qual mede o grau de informação/pureza de um conjunto de dados de entrada.

O algoritmo PART obtém regras de decisão a partir dos nós de uma árvore, sendo que a cada iteração o algoritmo PART constrói uma árvore de decisão e transforma em uma regra o caminho do nó raiz da árvore até o nó folha que melhor separa o conjunto dos dados de entrada (FRANK; WITTEN, 1998). À medida que uma regra é induzida as instâncias cobertas por esta regra são removidas, continuando a criar regras recursivamente para as instâncias restantes. O passo seguinte consiste em construir uma nova árvore a partir dos dados restantes que o nó folha anterior não conseguiu analisar. As regras são induzidas até que todas as instâncias sejam analisadas. Ao final da fase de construção, têm-se as melhores regras que separam os dados de entrada.

Para o problema de rotulação dos grupos, as regras devem confrontar as características dos dados que compõem o cluster 1 com as características dos dados que compõem os outros agrupamentos, depois deve confrontar as características do cluster 2 com os demais, e assim por diante até confrontar todas as possibilidades. Assim, o algoritmo PART fornece regras automáticas que tem a capacidade de inferir em qual cluster agrupa a região da falta. O conjunto para indução das regras é formado pelas características apresentadas na Tabela 2 e obtidos após a aplicação do algoritmo de clusterização sobre os dados de simulações que ocasionam distúrbios de afundamento de tensão.

2.6 Análise dos dados

A estatística descritiva fornece meios para descrever, analisar e interpretar quantitativamente a tendência, dispersão e distribuição de um conjunto de dados, dando suporte à tomada de decisões (CORREIA, 2003). Nessa perspectiva, tais métricas podem ser utilizadas para resumir o conjunto de características extraídas dos distúrbios detectados em um determinado cenário de simulação, ao passo que estas métricas podem ser então confrontadas, para fim de obter conclusões sobre o comportamento do sistema (BORGES, 2017).

Para análise dos dados, os resultados obtidos em uma simulação podem ser agrupados e representados por meio de um vetor de dados. As linhas podem ser relacionadas com o tipo de característica extraída do distúrbio e cada vetor identifica uma barra do sistema que o distúrbio foi medido, como mostra a figura 4.

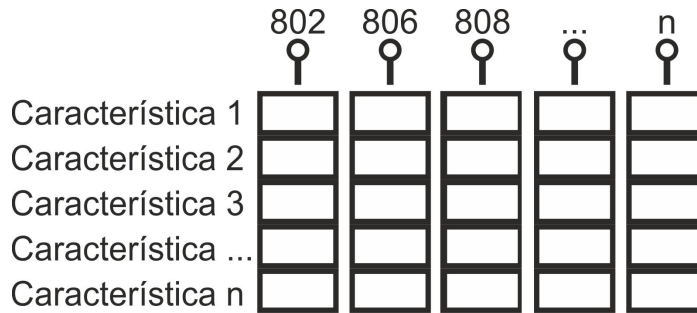


Figura 4 – Conjunto de vetores

Assim, em cada um desse vetor, torna-se possível sintetizar os dados numéricos relacionados a cada característica (linhas) por meio de aplicações de métricas estatísticas, como: Média aritmética, desvio padrão, amplitude do conjunto, curtose e assimetria, descrito na Tabela 2 .

Tabela 2 – Ralação entre os dados analisados, características do sinal e métricas estatísticas para criação das características analisadas para induzir o conjunto de Regras de Decisão.

Dados Analisados	Característica do sinal	metrica
Dados formados pela união de todos os clusters	RMS	Média
	Valor de Pico	Desvio Padrão
	Fator de Pico	Amplitude
Dados compõem os Cluster de 1 a 6	Valor Médio	Assimetria
	Distância da Barra à Subestação	Curtose
		Silhueta

A Tabela 2 mostra as características que se relacionam para compor o vetor das métricas avaliadas pela regra de decisão. Como os algoritmos de clusterização o valor de k vai de 2 a 6 os clusters formado são rotulados de 1 a 5, quando o valor de k em execução for igual a 5 as características dos dados que compõem o Cluster 6 serão nulas, quando o valor de k em execução for igual a 4 as características dos dados que compõem o Cluster 6 e Cluster 5 serão nulas, quando o valor de k em execução for igual a 3 as características dos dados que compõem o Cluster 6, Cluster 5 e Cluster 4 serão nulas, quando o valor de k em execução for igual a 2 as características dos dados que compõem o Cluster 6, Cluster 5, Cluster 4 e Cluster 3 serão nulas. Os valores nulos não interferem no resultado da regra de decisão.

3 Resultados e Discussão

Para a primeira etapa da metodologia é realizado uma análise entre diferentes algoritmos de clusterização. Segundo (STARCZEWSKI; KRZYŻAK, 2015) para analisar a validade e eficácia dos grupos formados existem vários índices de validade que podem ser empregados. Por exemplo, os índices mais populares incluem Dunn (DUNN, 1974), Davies-Bouldin (DB) (DAVIES; BOULDIN, 1979) ou índice de silhueta (ROUSSEEUW, 1987).

Neste trabalho, a avaliação foi realizada por meio do índice de silhueta, que é uma medida que combina a coesão dos agrupamentos, ou seja, quão semelhante é uma observação ao seu próprio agrupamento e a sua separação. Desta maneira a utilização do índice de silhueta auxilia no fornecimento das informações na forma gráfica e numérica (VANDELEENE et al., 2018). Além disso o índice de silhueta média é vantajosa porque pode ser usado com qualquer rotina de agrupamento e quaisquer métricas de dissimilaridade (POLLARD; LAAN, 2005).

Os coeficientes (valor da silhueta) estão compreendidos entre -1 e 1. Os coeficientes de silhueta perto de 1 indicam que a amostra está longe dos *clusters* vizinhos, ou seja, condizente com o cluster associado a ela. Um valor de 0 indica que a amostra está ligada ou muito próxima do limite de decisão entre dois *clusters* vizinhos e os valores negativos indicam que essas amostras podem ter sido atribuídas ao *cluster* errado.

O valor da silhueta calculada em cada simulação é comparada por meio do valor das medidas globais que são dadas ou pela média por agrupamento ou por todo o conjunto de dados (JAIN; MURTY; FLYNN, 2011). Para melhor visualização a silhueta é confrontada graficamente. Assim, um gráfico de barras apresenta uma maneira de avaliar parâmetros como número de *clusters*, valor silhueta média e o tamanho do *cluster* pela espessura de cada grupo.

Na segunda etapa é validado a acurácia dos grupos gerados pelos algoritmos de clusterização em um sistema baseado em regra que irá utilizar como entrada as medidas que caracterizam cada grupo de modo a rotular a partir da sua base de regras o grupo que agrega a região de origem. Assim, é possível confrontar a eficácia de cada algoritmo de clusterização comparando a acurácia do algoritmo baseado em regra. Por fim, a metodologia proposta pode ser validada.

3.1 Descrição dos Cenários

Para a realização das simulações foram utilizadas 23 barras trifásicas, dentro do sistema de distribuição IEEE 34-nós, para cada barra foi aplicado um curto-circuito trifásico e os valores de tensão foram medidos na barra que está localizado o curto-circuito e nas outras barras do sistema que sentiram o afundamento de tensão. A situação modelada consiste em um curto-circuito trifásico com duração de falta de 4 ciclos e impedância de 10, 20, 30, 40 e 50 Ω , com ângulo de incidência de 0, gerando um total de 115 casos de curto-circuito, com diferentes características. Para a modelagem e simulação de um Sistema Elétrico de Potência (SEP), implementada por meio do software ATP (Alternative Transients Program). O sistema utilizado corresponde ao IEEE 34 Node Test Feeder.

Para demonstrar os resultados foram selecionadas 4 barras com impedância de 10 Ω (que exerce uma falta mais severa) tendo como parâmetro para suas escolhas a localização dentro do sistema de distribuição, com isso foram selecionadas as barras 802, 850, 852 e 846, sendo a primeira próxima da subestação, outras duas na zona intermediária do sistema e próxima a um regulador de tensão, e a última localizada no final do sistema, como mostra a Figura 5. A análise dos clusters formados para situação de curto-circuito em outras barras podem ser visualizados no Apêndice A.

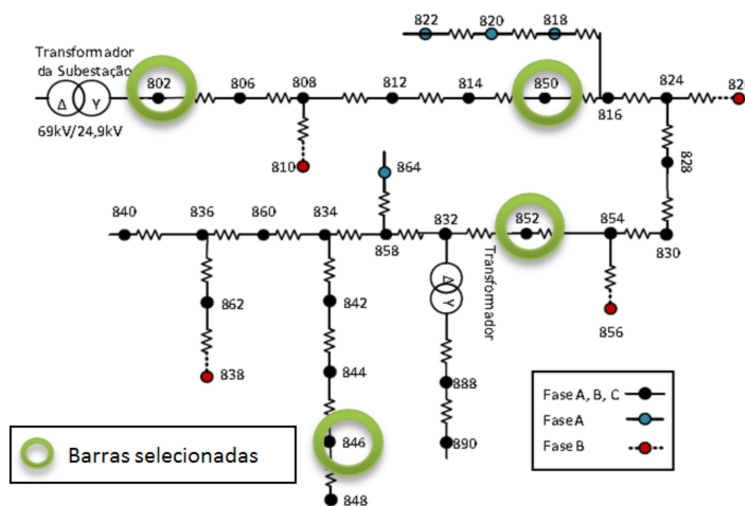


Figura 5 – Indicação das barras selecionadas no diagrama do Sistema IEEE 34 *Node Test Feeder*

Fonte: Adaptado de (FEEDERS, 2013)

3.1.1 Resultados da Análise dos algoritmos de clusterização

Nesta sessão são demonstrados os resultados para validação dos algoritmos de clusterização através das simulações, que originam afundamento de tensão. As simulações

são formadas da seguinte forma, para os algoritmos de Clusterização Agglomerative Hierarchical Clustering (HAC), k -Means (KM), Gaussian Mixture Models (GMM) e Fuzzy c-means (FCM) foi aplicado em cada simulação de falta. Como é necessário indicar a priori o valor de K , em cada simulação variou-se o entre 2 a 6, gerando um total de 575 simulações. Para cada uma das simulações a metodologia retorna o valor da largura da silhueta. Já para o algoritmo de Clusterização DBSCAN, como a identificação do valor de k é gerada dinamicamente pelo próprio algoritmo as simulações realizadas foram realizadas para as 23 barras trifásicas e impedância de 10, 20, 30, 40 e 50 Ω no total de 115 simulações.

Como já mencionado, os resultados serão demonstrado com os valores das barras 802, 846, 850 e 852 com impedância de 10 Ω , como mostra a Tabela 3 contendo os valores das Larguras das silhuetas.

Tabela 3 – Comparação entre algoritmos de clusterização Agglomerative Hierarchical Clustering (HAC), k -Means (KM), Gaussian Mixture Models (GMM) e Fuzzy c-means (FCM) através do valor da Largura da silhueta (LS) para as barras 802, 846, 850 e 852 com resistência 10 Ω

Barra	K	HAC	KM	GMM	FCM	Algoritmo com maior silhueta	Melhor K
802	2	0,706	0,706	0,706	0,706	TODOS	6
802	3	0,829	0,829	0,829	0,829	TODOS	
802	4	0,760	0,760	0,760	0,760	TODOS	
802	5	0,811	0,811	0,811	0,782	HAC-KM-GMM	
802	6	0,831	0,831	0,831	0,678	HAC-KM-GMM	
846	2	0,798	0,798	0,798	0,798	TODOS	3-5
846	3	0,855	0,892	0,892	0,892	KM-GMM-FCM	
846	4	0,806	0,840	0,793	0,793	KM	
846	5	0,819	0,892	0,892	0,714	KM-GMM	
846	6	0,856	0,890	0,890	0,704	KM-GMM	
850	2	0,810	0,858	0,858	0,570	KM-GMM	6
850	3	0,788	0,838	0,822	0,838	KM-FCM	
850	4	0,795	0,878	0,795	0,801	KM	
850	5	0,866	0,866	0,912	0,885	GMM	
850	6	0,952	0,952	0,952	0,779	HAC-KM-GMM	
852	2	0,789	0,789	0,789	0,789	TODOS	3
852	3	0,855	0,898	0,898	0,898	KM-GMM-FCM	
852	4	0,810	0,822	0,822	0,822	KM-GMM-FCM	
852	5	0,790	0,892	0,892	0,761	KM-GMM	
852	6	0,858	0,891	0,890	0,760	KM	

No resultado apresentado na Tabela 3 na barra 802 para o k informado de 2 a 3 não apresentou diferença nos valores de silhueta média entre os algoritmos analisados com valor de 0.706, já para o k 5 e 6 o algoritmo FCM apresentou o pior resultado com 0.782 e 0.678, respectivamente, não tendo diferença entre os valores dos outros algoritmos analisados com valor de 0.831. Também é possível verificar que o maior valor da silhueta média é indicado no k igual a 6 para os algoritmos HAC, KM e GMM. Para barra 846 o

melhor resultado é representado pelo valor 0.892 e se apresentam nos valores de k igual a 3 e 5, sendo que para $k = 3$, o valor de 0.892 foi obtido a partir dos resultados dos algoritmos de clusterização KM, GMM e FCM, quando o k for igual a 5 o melhor índice está na formação dos grupos obtidos pelos algoritmos KM e GMM. Na barra 850 o melhor valor de k é igual a 6, onde a silhueta média apresenta o valor e 0.952 nos algoritmos HAC, KM e GMM. E por fim a Barra 852, com melhor k igual a 3 nos algoritmos KM, GMM e FCM com valor de silhueta igual a 0.898.

Com o algoritmo DBSCAN o seu resultado é comparado com o que foi obtido como melhor resultado através dos maiores valores das silhueta média dos algoritmos analisados na Tabela 3.

Tabela 4 – Comparação do valor da largura da silhueta entre o algoritmo DBSCAN em relação ao melhor resultado obtido nos outros algoritmos para as barras 802, 846, 850 e 852 com resistência 10

Barra	DBSCAN		Outros Algoritmo		
	K	Silhueta Média	K	Silhueta Média	Algoritmos
802	4	0,700	6	0,831	HAC-KM-GMM
846	2	0,798	5	0,798	KM-GMM
850	3	0,677	6	0,952	HAC-KM-GMM
852	2	0,789	3	0,898	KM-GMM-FCM

Como mostra a Tabela 4 o DBSCAN se mostra compacto e bem separado na barra 846 com valor da silhueta média 0.798. Entretanto ele não se sobressai em relação a todos os outros algoritmos analisados, seu valor da silhueta média é igual aos algoritmos k -Means (KM), Gaussian Mixture Models (GMM). Nas barras 802, 850 e 852 tem o valor da silhueta média abaixo do melhor resultado dos outros algoritmos analisados.

Neste trabalho a silhueta de cada amostra de entrada é apresentada graficamente a partir de um gráfico de barras, adicionalmente é traçada uma linha vermelha que destaca a silhueta média. Assim, o gráfico de barras apresenta uma maneira de avaliar parâmetros como número de *clusters* visual comparando cada valor da silhueta com o valor silhueta média. Quanto mais pontos próximos da silhueta média melhor a definição do número de *clusters*. Também a partir da espessura do gráfico de silhueta, o tamanho do *cluster* pode ser visualizado nas Figuras 6 e 7, onde pode-se demonstrado a silhueta de cada barra em cada grupo em relação a silhueta média do resultado do agrupamento, para visualizar a análise foi selecionado o resultado da barra 846 por apresentar o mesmo valor da largura da silhueta nos resultado do DBSCAN em relação ao maior valor da largura da silhueta dos outros algoritmos.

A Figura 6 mostra todos os resultado do algoritmo k -means que foi o algoritmo que apresentou maior valor de silhueta em todos os valores de k na barra 846.

No Gráfico 6a podemos visualizar uma boa divisão das barras entre os grupos,

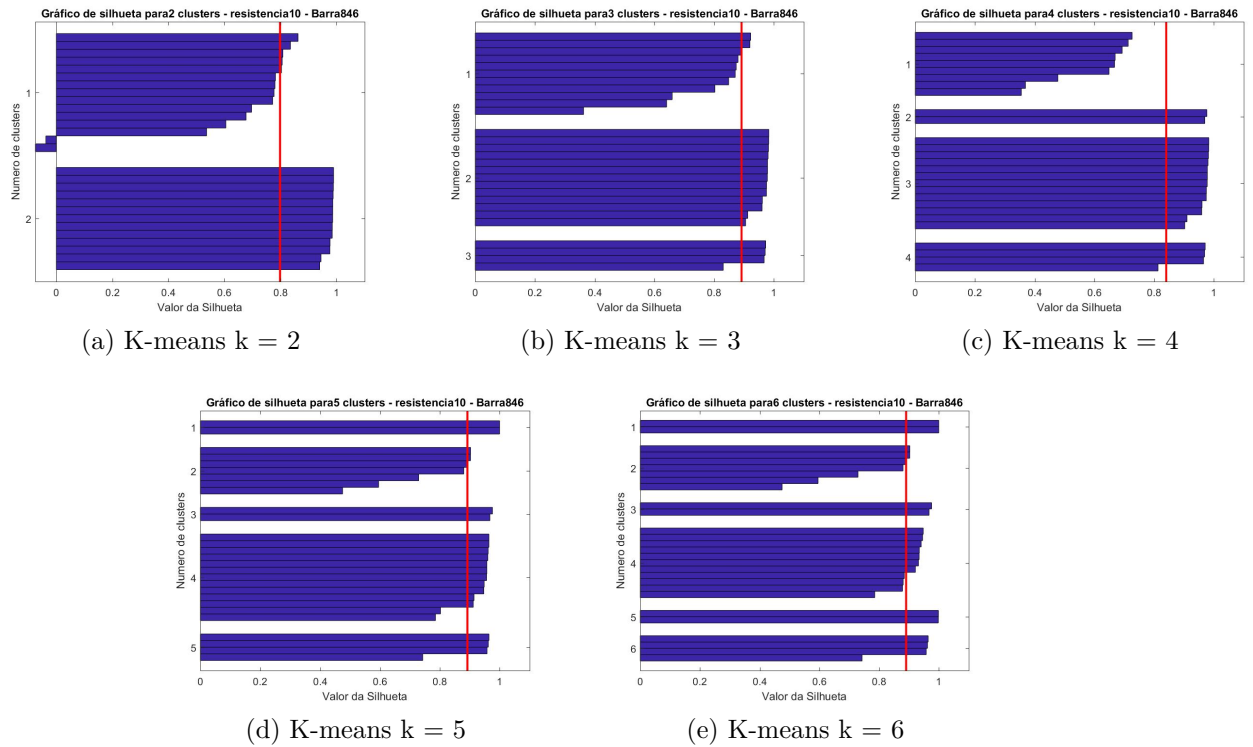


Figura 6 – Silhueta média do algoritmo k -means com $k = 2$ (a), $k = 3$ (b), $k = 4$ (c), $k = 5$ (d), $k = 6$ (e) de curto-circuito na barra 846

porém no grupo 1 existem barras com silhueta negativa o que indica um grupo mau formado, o que corrobora com a análise da silhueta média que indica o pior resultado (0.798 Tabela 3) entre as simulações do algoritmo k -means para valores de k entre 2 a 6.

O Gráfico 6c apesar de não ter silhuetas negativas o grupo 1 mostra uma certa distancia em relação a silhueta média indicando não ser o melhor agrupamento.

Com os Gráficos 6b, 6d e 6e é observado uma boa divisão dos grupos e uma boa aproximação da silhueta média o que mostra os melhores resultados do algoritmo k -means com o k igual a 6 com valor 0,890 e k igual a 3 e 5 com valor igual 0,892.

A Figura 7, mostra os resultados da Tabela 4 para barra 846 com os algoritmos k -Means (KM), Gaussian Mixture Models (GMM) e DBSCAN.

O resultado do algoritmo DBSCAN para barra 846 com o valor da silhueta média 0.798 igual aos algoritmos k -meas e GMM apresenta valor da silhueta de algumas barras do grupo 1 negativa, indicando má formação dos grupo. Essa situação é gerada pelo número de k diferente onde o k para o k -means e GMMM é 5 e o DBSCAN com k igual a 2.

A verificação de compactação intra-grupo e separação inter-grupos foi realizada para todas as 575 simulações dos 4 algoritmos que precisam indicar o valor de k e para as 115 simulações referente ao algoritmo DBSAN. As Tabelas 5 e 6 trazem um panorama geral da validação de todas as simulações.

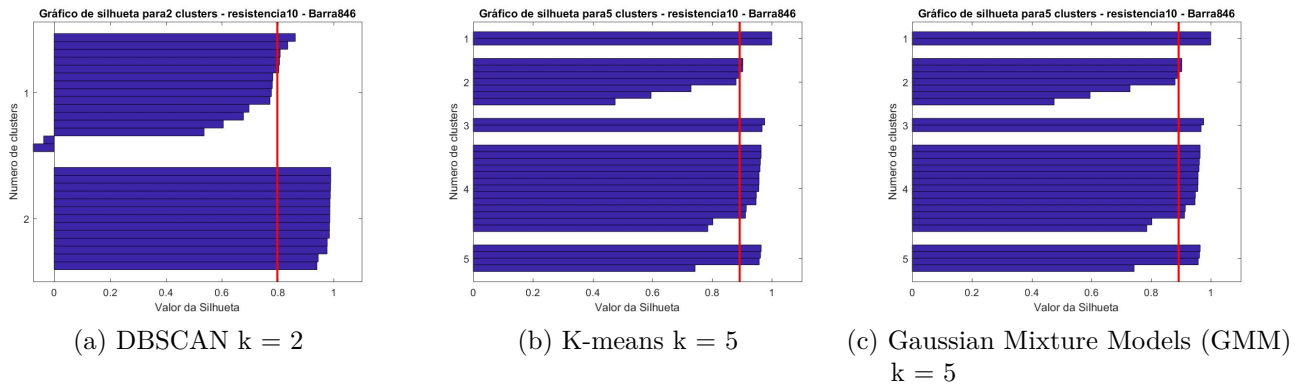


Figura 7 – Comparação da silhueta do algoritmo k-means com $k = 2$ (a), $k = 3$ (b), $k = 4$ (c), $k = 5$ (d), $k = 6$ (e) e fuzzy c-means $k = 2$ (f), $k = 3$ (g), $k = 4$ (h), $k = 5$ (i), $k = 6$ (j) de curto-circuito na barra 846

Tabela 5 – Quantidade de simulações com indicação do melhor k para cada algoritmo

K	HAC	KM	MMG	FCM
2	4	10	10	6
3	4	51	51	51
4	0	0	0	0
5	0	13	6	0
6	25	33	34	2
Total	33	107	101	59

A Tabela 5 mostra que para cada conjunto de Barra e Resistência qual é o melhor k por algoritmo (entre os quatro algoritmos validades que necessitam da indicação do valor de k) levando em consideração o maior valor de largura da Silhueta. Onde se destaca os algoritmos KM e o MMG com k igual a 3 com 51 ocorrências em ambos. Outra observação importante é que o k igual a 4 não teve ocorrência como melhor k em nenhum dos algoritmos. É necessário ressaltar que o melhor valor da largura da silhueta pode se repetir para o mesmo k em mais de um dos algoritmos avaliados, conseqüentemente será computado em cada coluna do algoritmo.

Tabela 6 – Quantidade de ocorrência do valor de k para o algoritmo de clusterização DBSCAN

k	2	3	4	5	6	Total
Quantidade	22	72	17	4	0	115

A Tabela 6, mostra a quantidade de escolhas do valor de k para o algoritmos de Clusterização DBSCAN, onde se destaca o o k igual a 3 com 72 indicações.

Para os resultados realizados para validação dos agrupamentos pelos algoritmos de clusterização Agglomerative Hierarchical Clustering (HAC), k -Means (KM), Gaussian

Mixture Models (GMM), Fuzzy c-means (FCM) e DBSCAN a partir dos dados originados por distúrbios na rede de energia elétrica por afundamento de tensão através da verificação de compactação intra-grupo e separação inter-grupos dos algoritmos de clusterização. Teve o intuito de aplicar a metodologia proposta neste trabalho apenas nos algoritmos que se destaquem desta análise. Porém com os resultados apresentados, apesar de ser possível identificar alguns algoritmos em destaque, também se observou valores consideráveis em outros algoritmos. O que encorajou a comparação de todos os algoritmos na metodologia como todo (algoritmo de clusterização combinado com regra de decisão) expostos na sessão a seguir.

3.2 Resultado do Método de Localização dos Distúrbios de Afundamento de Tensão

Nesta sessão serão apresentados os resultados obtidos utilizando toda a abordagem proposta (algoritmo de clusterização e Algoritmo PART) para localização dos distúrbios de afundamento de tensão no sistema test feeder IEEE 34-bus.

Os resultados computacionais foram obtidos da seguinte forma: 1) O agrupamento foi desenvolvido no ambiente MATLAB ([MATHWORKS, 1996](#)). 2) O treinamento da regra de decisão foi desenvolvido na Ferramenta WEKA, disponibilizada pelo grupo de aprendizado de máquina da Universidade de Waikato ([WITTEN et al., 2016](#)).

A base de dados para teste se constitui dos casos de simulação que ocasionam afundamento de tensão, sendo formada por simulação de curto-circuito trifásico, onde, para cada tipo, variou-se a resistência entre os valores 10, 20, 30, 40 e 50 Ω e o ângulo de incidência 0, para os algoritmos de Clusterização Agglomerative Hierarchical Clustering (HAC), k -Means (KM), Gaussian Mixture Models (GMM) e Fuzzy c-means (FCM) os agrupamentos foram formados para a indicação do k a partir de 2 até 6, contabilizando 575 simulações para cada algoritmo sendo 115 para cada valor de k . O algoritmo DBSCAN por apontar o valor de k automaticamente gerou 115 simulações como base para treinamento da regra de decisão. Esses dados são utilizados para realizar a indução do conjunto de regras e avaliar a metodologia.

Nesse contexto, é possível chamar a função responsável por induzir o conjunto de regras a partir de uma entrada que representa o conjunto de treinamento e os parâmetros de configurações do algoritmo PART, as regras geradas podem ser visualizadas na Tabela 7.

Tabela 7 – Base de Regras para localização dos distúrbios de Afundamento de Tensão induzida a partir dos parâmetros definidos no vetor das características apresentado na Tabela 2

<i>k</i>-means	
<i>k</i>	Regras
2	Curtose RMS > 0.4678: Cluster 1 : Cluster 2
3	Curtose Valor Médio > 2.1556 AND desvio padrão RMS grupo 1 <= 0.56319: Cluster 2 Assimetria Valor Médio > 2.1042: Cluster 1 : Cluster 3
4	Desvio Padrao RMS cluster 3 > 0.65479: Cluster 3 Distancia Média cluster 4 > 0.064191: Cluster 4 Curtose distância Cluster 2 <= -0.63327: Cluster 1 : Cluster 2
5	Desvio Padrao RMS Cluster 4 > 0.6315: Cluster 4 Curtose distância cluster 1 <= 0.4486 AND Curtose RMS Cluster 2 > 0.59532: Cluster 3 Curtose Distância Cluster 1 <= 0.4486: Cluster 2 : Cluster 5
6	Desvio Padrão RMS Cluster 3 <= 0.83733 AND Desvio Padrão RMS Cluster 2 <= 0.69544 AND Desvio Padrão RMS Cluster 4 > 0.86556 AND Curtose RMS Cluster 4 <= 0.96069: Cluster 4 Curtose RMS Cluster 3 <= 0.56078: Cluster 3 Desvio Padrão RMS Cluster 5 > 0.57156: Cluster 5 Curtose RMS Cluster 1 > 0.60426: Cluster 2 : Cluster 6
Fuzzy C-Means	
<i>k</i>	Regras
2	Desvio Padrão RMS Cluster 2 <= 0.43533: Cluster 1 : Cluster 2
3	Curtose RMS Cluster 1 <= 0.96069 AND Desvio Padrão RMS Cluster 1 <= 0.59759: Cluster 2 Curtose RMS Cluster 2 <= 0.96069: Cluster 1 : Cluster 3
	Amplitude RMS Cluster 4 > 0.84118 AND
Continua na próxima página	

	Desvio Padrão RMS Cluster 3 > 0.73206 AND Desvio Padrão PICO Cluster 3 > 0.01743: Cluster 3 Desvio Padrão RMS Cluster 2 > 0.77036: Cluster 2 Amplitude RMS Cluster 4 > 0.84118: Cluster 1 : Cluster 4
5	Amplitude RMS Cluster 5 > 0.84118 AND Desvio Padrao RMS Cluster 4 > 0.65479 AND Desvio Padrao RMS Cluster 3 <= 0.93534: Cluster 4 Amplitude RMS Cluster 5 > 0.84118 AND Desvio Padrão RMS Cluster 3 > 0.77378 AND Curtose Valor Médio Cluster 2 > 1.1872: Cluster 3 Desvio Padrão RMS Cluster 2 > 0.58135: Cluster 2 : Cluster 5
6	Desvio Padrão Distância <= 0.86726 AND Desvio Padrão RMS Cluster 2 <= 0.83878 AND Desvio Padrão RMS Cluster 5 > 0.73206 AND Sulhueta Média Cluster 1 <= 0.99968: Cluster 5 Sulhueta Média Cluster 6 > 0.64181 AND Desvio Padrão RMS Cluster 2 <= 0.83878 AND Desvio Padrão RMS Cluster 3 > 0.76326 AND Amplitude RMS Cluster 4 > 0.012147: Cluster 4 Desvio Padrão RMS Cluster 2 > 0.8345: Cluster 3 Silhueta média Cluster 6 <= 0.64181: Cluster 6 : Cluster 2
MMG	
k	Regras
2	Desvio Padrão RMS Cluster 2 > 0.50421: Cluster 2 : Cluster 1
3	Desvio Padrão RMS Cluster 1 <= 0.42092 AND Curtose RMS Cluster 2 <= 0.96069: Cluster 2 Curtose RMS Cluster 1 <= 0.61763: Cluster 1 : Cluster 3
4	Desvio ppadrão RMS cluster 3 > 0.50892: Cluster 3 Curtose RMS Cluster 4 <= 0.011791: Cluster 1 : Cluster 4
	Curtose Pico Cluster 3 > 0 AND Desvio Padrão RMS Cluster 4 > 0.67051: Cluster 4 Curtose Pico Cluster 3 > 0 AND

Continua na próxima página

	<p>Curtose RMS Cluster 1 > 0.56078 AND Curtose RMS Cluster 2 > 0.56078: Cluster 3 Curtose RMS Cluster 5 <= 0.011791 AND Média Pico Cluster 3 <= 0.00849: Cluster 1 Curtose RMS Cluster 1 <= 0.64139: Cluster 5 : Cluster 2</p>
6	<p>Desvio Padrão RMS Cluster 4 > 0.86164: Cluster 4 Desvio Padrão RMS Cluster 3 > 0.73506: Cluster 3 Desvio Padrão RMS Cluster 6 > 0.027685 AND Desvio Padrão RMS Cluster 6 <= 0.73714: Cluster 5 Curtose Distancia Cluster 2 > 0.49684: Cluster 2 : Cluster 6</p>
HAC	
k	Regras
2	<p>Assimetria RMS > 1.4286: Cluster 2 : Cluster 1 (55.0)</p>
3	<p>Desvio Padrão RMS Cluster 1 <= 0.5656 AND Curtose Pico Cluster 1 <= 0.11996: Cluster 2 Silhueta Média Cluster 2 <= 0.99571: Cluster 1 : Cluster 3</p>
4	<p>Desvio ppadrão RMS cluster 2 > 0.76343: Cluster 2 Desvio ppadrão RMS cluster 3 <= 0.65817 AND Curtose Valor Médio Cluster 1 > 1.5752: Cluster 1 (26.0) Desvio ppadrão RMS cluster 4 <= 0.14737: Cluster 3 : Cluster 4</p>
5	<p>Desvio Padrão RMS Cluster 3 > 0.65817: Cluster 3 Curtose Fator de Pico Cluster 5 <= 0.15263 AND Curtose RMS Cluster 1 <= 0.79612 AND Curtose Pico Cluster 4 > 0: Cluster 4 (11.0) Curtose RMS Cluster 5 <= 0.011791 AND Curtose RMS Cluster 4 > 0.1876: Cluster 2 (5.0) Curtose RMS Cluster 1 <= 0.56078: Cluster 5 : Cluster 1</p>
6	<p>Desvio Padrão RMS Cluster 4 > 0.80572: Cluster 4 Curtose RMS Cluster 3 <= 0.65485: Cluster 3 Curtose RMS Cluster 1 <= 0.79612 AND Assimetria Pico <= 1.467: Cluster 5 Curtose RMS Cluster 1 > 0.56078: Cluster 2</p>

Continua na próxima página

: Cluster 6
DBSCAN
Regras
Assimetria Valor Médio ≤ 2.1207 : Cluster 2 (56.0)
Desvio Padrão RMS Cluster 4 ≤ 0.19785 AND
Desvio Padrão RMS Cluster 3 ≤ 0.54474 : Cluster 1
Curtose Pico Cluster 3 > 0.10793 : Cluster 4 (9.0)
: Cluster 3 (8.0/1.0)

Com os algoritmo de clusterização os dados de entrada são representados de forma mais organizados. Dessa forma, o conjunto de regra é pequeno e as regras são simples e de fácil interpretação, como mostra a Tabela 7. Com isso, simplificando o trabalho do classificador.

A Tabela 8 apresenta o resultado do método de localização para o algoritmo PART aplicado sobre os clusters fornecidos pelos algoritmos Agglomerative Hierarchical Clustering (HAC), k -Means (KM), Gaussian Mixture Models (GMM) e Fuzzy c-means (FCM), para os diferentes valores de k , e para o algoritmo DBSCAN que aponta o valor de k . Destaca-se que a taxa de acerto apresenta o percentual de casos que o cluster determinado por meio das regras induzidas envolve a barra de origem do distúrbio.

Tabela 8 – Percentual das instâncias classificadas corretamente obtido por regra de decisão PART para cada conjunto de características gerados pelos algoritmos de clusterização avaliado

Número de k	Regra de decisão PART com os Algoritmos				
	k -Means	FCM	MMG	HAC	DBSCAN
2	100,00	99,13	99,13	100,00	
3	99,13	98,26	99,13	95,65	
4	98,26	93,04	98,26	96,52	98,26
5	95,65	90,43	97,39	97,39	
6	92,17	88,70	93,91	96,52	

Pela Tabela 8 observa-se que o melhor resultado foi corresponde a aplicação do algoritmo de clusterização com k igual a 2, porém essa divisão resulta em clusters que agregam uma quantidade expressiva de barras não atendendo a expectativa inicial de reduzir significativamente a área onde se encontra o nó que gerou a falha. Para k igual a 3 a utilização do algoritmo k -means e MMG apresenta uma taxa de acerto maior que a obtida utilizando os outros algoritmo, entretanto em alguns casos ainda fornece uma região de localização muito extensa. com k igual a 6 e 5 observa-se que se tem os piores resultados o que indica uma fragmentação muito grande dos agrupamentos dificultando a identificação do grupo que pertence a barra que originou a falta.

Dessa forma, a configuração que alcança o melhor equilíbrio equivale a utilização dos algoritmos k -means e MMG com k igual a 4 que apresenta uma taxa de acerto igual a 98.26%. O algoritmo DBSCAN também atinge a mesma taxa de acerto da melhor configuração, tem a vantagem de não precisar indicar o valor de k , porém tem contra si a necessidade de informar os parâmetros Eps (define o raio máximo da vizinhança) e MinPts (Densidade mínima, ou seja, número mínimo de objetos para formar um cluster). Com isso endossa a escolha que a melhor configuração corresponde a utilização dos algoritmos k -means e MMG com k igual a 4, pois apresentam taxa de acerto próxima e a metodologia aponta uma região de localização pequena, constituída com poucas barras.

4 Conclusões e Trabalhos Futuros

4.1 Conclusões

Este trabalho apresenta uma metodologia baseada em algoritmo de clusterização combinado com regra de decisão para localizar a região do distúrbio de afundamento de tensão em sistema de distribuição com smart meters. A abordagem é desenvolvida sobre duas etapas, onde o algoritmo de clusterização é responsável por agrupar as características advindas dos medidores em clusters distintos e concisos. Em seguida, o método baseado em regra de decisão é responsável por detectar qual cluster agrega a barra de origem do distúrbio.

Para a tarefa de clusterização os algoritmos de clusterização Agglomerative Hierarchical Clustering (HAC), K-Means (KM), Gaussian Mixture Models (GMM), Fuzzy c-means (FCM) e DBSCAN foram testados e comparados. O conjunto de regra de decisão foi induzido utilizando o algoritmo PART. De acordo com as figuras que apresentam os gráficos da silhueta e os resultados da regra de decisão pode-se verificar que quanto maior o número de k mais fragmentado será o conjunto de entrada o que dificulta a decisão da região. E quando o valor de k é igual a 2 a precisão da escolha do cluster que contem a falta é alta com 100% ou 99.13%, porem a área de cada cluster em geral é bem extensa.

Pelos resultados comprova-se que a metodologia apresentou resultados satisfatórios, apresentando, na maioria dos cenários uma taxa de acerto maior que 90%. No geral o algoritmo k-means apresentou resultados melhores que os outros algoritmos.

4.1.1 Limitações e Trabalhos Futuros

A seguir, são apontadas limitações desta pesquisa que poderão ser abordadas em trabalhos futuros:

- **Escalonamento da metodologia:** com o propósito de fortalecer a justificativa da abordagem proposta, seria interessante realizar os testes em um Sistema IEEE Node Test Feeder de 13, 37, 123 nós (barras). Inferindo sobre a manutenção do padrão de propagação do evento.
- **Implementar o processo iterativo:** implementar o processo completo de localização em um processo iterativo inserindo o resultado da região com a falta na metodologia proposta consecutivamente até encontrar a região mínima do distúrbio de afundamento de tensão. Analisando se a cada iteração o valor de k , se mantém ou se faz mais eficaz o decrementado do mesmo.

- **Ambiente real:** Testar a metodologia proposta em um sistema elétrico real.

Referências

- AHN, S.-J. et al. A new approach to determine the direction and cause of voltage sag. *Journal of Electrical Engineering and Technology*, The Korean Institute of Electrical Engineers, v. 3, n. 3, p. 300–307, 2008. Citado na página 11.
- ALI, S. et al. A machine learning approach to meter placement for power quality estimation in smart grid. *IEEE Transactions on Smart Grid*, IEEE, v. 7, n. 3, p. 1552–1561, 2016. Citado na página 3.
- APTÉ, C.; WEISS, S. Data mining with decision trees and decision rules. *Future generation computer systems*, Elsevier, v. 13, n. 2-3, p. 197–210, 1997. Citado na página 4.
- BALOUJI, E.; SALOR, O. Eigen-analysis based power quality event data clustering and classification. In: IEEE. *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*. [S.l.], 2014. p. 1–5. Citado na página 9.
- BHATTACHARYYA, S. Power quality requirements and responsibilities at the point of connection. *Technische Universiteit Eindhoven*, 2011. Citado na página 1.
- BISHOP, C. M. *Pattern recognition and machine learning*. [S.l.]: springer, 2006. Citado na página 20.
- BORGES, F. A. S. *Método híbrido baseado no algoritmo k-means e regras de decisão para localização das fontes de variações de tensões de curta duração no contexto de Smart Grid*. Tese (Doutorado) — Universidade de São Paulo, 2017. Citado 2 vezes nas páginas 15 e 23.
- CHANG, G. et al. A new procedure for tracking the source location. In: IEEE. *Power Engineering Society General Meeting, 2007. IEEE*. [S.l.], 2007. p. 1–4. Citado 2 vezes nas páginas 1 e 2.
- CHEESEMAN, P.; STUTZ, J. Bayesian classification (autoclass): Theory and results in advances in knowledge discovery and data mining eds. *Articles FALL.*, p. 51, 1996. Citado na página 3.
- CORREIA, M. S. B. B. Probabilidade e estatística. 2003. Citado na página 23.
- DAVIDSON, I.; RAVI, S. S. Using instance-level constraints in agglomerative hierarchical clustering: theoretical and empirical results. *Data mining and knowledge discovery*, Springer, v. 18, n. 2, p. 257–282, 2009. Citado na página 21.
- DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, n. 2, p. 224–227, 1979. Citado na página 25.
- DAY, W. H.; EDELSBRUNNER, H. Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of classification*, Springer, v. 1, n. 1, p. 7–24, 1984. Citado na página 21.

- DONI, M. V. Análise de cluster: métodos hierárquicos e de particionamento. *São Paulo: Universidade Presbiteriana Mackenzie, São Paulo*, 2004. Citado na página 18.
- DUAN, R. et al. Data mining & pattern recognition of voltage sag based on k-means clustering algorithm. In: IEEE. *Power & Energy Society General Meeting, 2015 IEEE*. [S.l.], 2015. p. 1–5. Citado 2 vezes nas páginas 1 e 10.
- DUGAN, R. C. et al. *Electrical power systems quality*. [S.l.]: mcgraw-Hill New York, 1996. v. 2. Citado 2 vezes nas páginas 1 e 7.
- DUNN, J. C. Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, Taylor & Francis, v. 4, n. 1, p. 95–104, 1974. Citado na página 25.
- ESTER, M. et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Kdd*. [S.l.: s.n.], 1996. v. 96, n. 34, p. 226–231. Citado na página 20.
- FAHAD, A. et al. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE transactions on emerging topics in computing*, IEEE, v. 2, n. 3, p. 267–279, 2014. Citado 2 vezes nas páginas 14 e 18.
- FAISAL, M. F.; MOHAMED, A. Identification of sources of voltage sags in the malaysian distribution networks using svm based s-transform. In: IEEE. *TENCON 2009-2009 IEEE Region 10 Conference*. [S.l.], 2009. p. 1–6. Citado na página 3.
- FANG, X. et al. Smart grid—the new and improved power grid: A survey. *IEEE communications surveys & tutorials*, IEEE, v. 14, n. 4, p. 944–980, 2012. Citado na página 3.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37, 1996. Citado na página 4.
- FEEDERS, D. T. Ieee pes distribution system analysis subcommittee’s, distribution test feeder working group. *ed*, 2013. Citado na página 26.
- FILHO, J. M. et al. A pca-based approach for substation clustering for voltage sag studies in the brazilian new energy context. *Electric Power Systems Research*, Elsevier, v. 136, p. 31–42, 2016. Citado 3 vezes nas páginas 1, 2 e 7.
- FRANK, E.; WITTEN, I. H. Generating accurate rule sets without global optimization. In: *Proc. of the Fifteenth International Conference on Machine Learning, 1998*. [S.l.: s.n.], 1998. p. 144–151. Citado 2 vezes nas páginas 22 e 23.
- GIL, V. de O.; FERRARI, F.; EMMENDORFER, L. Investigação da aplicação de algoritmos de agrupamento para o problema astrofísico de classificação de galáxias. *Revista Brasileira de Computação Aplicada*, v. 7, n. 2, p. 52–61, 2015. Citado na página 22.
- HAMZAH, N.; MOHAMED, A.; HUSSAIN, A. A new approach to locate the voltage sag source using real current component. *Electric Power Systems Research*, Elsevier, v. 72, n. 2, p. 113–123, 2004. Citado na página 11.
- HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011. Citado 2 vezes nas páginas 4 e 19.

- HARIYANTO, N.; ANGGORO, B.; NOEGROHO, R. New probabilistic approach for identification event severity index due to short circuit fault. In: IEEE. *Electrical Engineering and Computer Science (ICEECS), 2014 International Conference on*. [S.l.], 2014. p. 1–5. Citado na página 10.
- HARTIGAN, J. A. Clustering algorithms. Wiley, 1975. Citado na página 19.
- HE, X. et al. Laplacian regularized gaussian mixture model for data clustering. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 23, n. 9, p. 1406–1418, 2011. Citado na página 20.
- JAIN, A.; MURTY, M.; FLYNN, P. Data clustering: a review, acm computing survey. 31 (3), p. 264–323, 2011. Citado na página 25.
- JÚNIOR, W. M. d. S. *Mineração em dados do ENEM para a predição do desempenho acadêmico no âmbito da Rede Federal de Educação Tecnológica*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2018. Citado na página 22.
- LACERDA, G. S. Análise da influência de cargas não-lineares em dispositivos de medição de grandezas elétricas. 2017. Citado na página 16.
- LEONARDO, H.; OLIVEIRA, J. Sistema de medição de energia e monitoria da qualidade da rede elétrica. *Revista de Engenharia e Pesquisa Aplicada*, v. 2, n. 1, 2016. Citado na página 16.
- LUZ, G. L. Qualidade de energia: um estudo de caso. Fundação de Ensino e Pesquisa do Sul de Minas, 2018. Citado na página 16.
- MAHELA, O. P.; SHAIK, A. G. Power quality recognition in distribution system with solar energy penetration using s-transform and fuzzy c-means clustering. *Renewable energy*, Elsevier, v. 106, p. 37–51, 2017. Citado na página 11.
- MALANGE, F. C. V. Rede neuro-fuzzy-wavelet para detecção e classificação de anomalias de tensão em sistemas elétricos de potência. Universidade Estadual Paulista (UNESP), 2010. Citado na página 8.
- MATHWORKS, I. *MATLAB: Application program interface guide*. [S.l.]: MathWorks, 1996. v. 5. Citado na página 31.
- MEIRA, C. A. et al. Analysis of coffee leaf rust epidemics with decision tree [análise da epidemia da ferrugem do cafeeiro com árvore de decisão]. *Tropical Plant Pathology*, 2008. Citado na página 4.
- MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. [S.l.]: Editora UFMG, 2005. Citado na página 18.
- MOHAMMADI, Y.; MORADI, M. H.; LEBORGNE, R. C. Locating the source of voltage sags: Full review, introduction of generalized methods and numerical simulations. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 77, p. 821–844, 2017. Citado 3 vezes nas páginas 1, 2 e 3.
- MORADI, M. H.; MOHAMMADI, Y. Voltage sag source location: A review with introduction of a new method. *International Journal of Electrical Power & Energy Systems*, Elsevier, v. 43, n. 1, p. 29–39, 2012. Citado na página 1.

- PARSONS, A. C. et al. A direction finder for power quality disturbances based upon disturbance power and energy. In: IEEE. *Harmonics and Quality of Power Proceedings, 1998. Proceedings. 8th International Conference On*. [S.l.], 1998. v. 2, p. 693–699. Citado na página 2.
- PASSOS, F. O. et al. An alternative approach to locating voltage sag source side at the point of common coupling based on power-flow information. *Journal of Control, Automation and Electrical Systems*, Springer, v. 26, n. 5, p. 579–587, 2015. Citado na página 8.
- PERES, S. M. et al. Tutorial sobre fuzzy-c-means e fuzzy learning vector quantization: Abordagens híbridas para tarefas de agrupamento e classificação. *Revista de Informática Teórica e Aplicada*, v. 19, n. 1, p. 120–163, 2012. Citado na página 20.
- PESSANHA, J. et al. Construindo tipologias de curvas de carga com o programa r. *Podes-Revista Eletrônica Pesquisa Operacional para o Desenvolvimento*, v. 7, n. 1, p. 29–54, 2015. Citado na página 21.
- POLLARD, K. S.; LAAN, M. J. V. D. Cluster analysis of genomic data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. [S.l.]: Springer, 2005. p. 209–228. Citado na página 25.
- PRADHAN, A. K.; ROUTRAY, A. Applying distance relay for voltage sag source detection. *IEEE Transactions on Power Delivery*, IEEE, v. 20, n. 1, p. 529–531, 2005. Citado na página 2.
- REHMAN, M.; MEHDI, S. A. Comparison of density-based clustering algorithms. *Lahore College for Women University, Lahore, Pakistan, University of Management and Technology, Lahore, Pakistan*, Citeseer, 2005. Citado na página 20.
- ROMERO, M.; GALLEGO, L.; PAVAS, A. Fault zones location on distribution systems based on clustering of voltage sags patterns. In: IEEE. *Harmonics and Quality of Power (ICHQP), 2012 IEEE 15th International Conference on*. [S.l.], 2012. p. 486–493. Citado 3 vezes nas páginas 1, 7 e 9.
- ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, Elsevier, v. 20, p. 53–65, 1987. Citado na página 25.
- SHAREEF, H.; MOHAMED, A. An alternative voltage sag source identification method utilizing radial basis function network. IET, 2013. Citado 2 vezes nas páginas 3 e 11.
- SMITH, J. C.; HENSLEY, G.; RAY, L. Ieee recommended practice for monitoring electric power quality. *IEEE Std (Revision of IEEE Std 1159-1995)*, p. c1–81, 26, 2009. Citado 3 vezes nas páginas 2, 3 e 8.
- SOUZA, P. V. G. d. Formação de conjuntos de sistemas de distribuição quanto aos fundamentos de tensão. 2016. Citado na página 18.
- STARCZEWSKI, A.; KRZYŻAK, A. Performance evaluation of the silhouette index. In: SPRINGER. *International Conference on Artificial Intelligence and Soft Computing*. [S.l.], 2015. p. 49–58. Citado na página 25.

- TAYJASANANT, T.; LI, C.; XU, W. A resistance sign-based method for voltage sag source detection. *IEEE Transactions on Power Delivery*, IEEE, v. 20, n. 4, p. 2544–2551, 2005. Citado na página 12.
- VANDELEENE, N. et al. Master thesis: Individual differences across neural basis of pain. Université de Liège, Liège, Belgique, 2018. Citado na página 25.
- VARGAS, R. R. d. Uma nova forma de calcular os centros dos clusters em algoritmos de agrupamento tipo fuzzy c-means. Universidade Federal do Rio Grande do Norte, 2012. Citado na página 20.
- VENDRAMIN, L.; CAMPELLO, R. J.; HRUSCHKA, E. R. Relative clustering validity criteria: A comparative overview. *Statistical analysis and data mining: the ASA data science journal*, Wiley Online Library, v. 3, n. 4, p. 209–235, 2010. Citado na página 21.
- WANG, N.; WANG, S.; JIA, Q. The method to reduce identification feature of different voltage sag disturbance source based on principal component analysis. In: IEEE. *Transportation Electrification Asia-Pacific (ITEC Asia-Pacific), 2014 IEEE Conference and Expo*. [S.l.], 2014. p. 1–6. Citado na página 10.
- WITTEN, I. H. et al. *Data Mining: Practical machine learning tools and techniques*. [S.l.]: Morgan Kaufmann, 2016. Citado na página 31.
- WON, D.-J. et al. A new algorithm to locate power-quality event source with improved realization of distributed monitoring scheme. *IEEE Transactions on Power Delivery*, IEEE, v. 21, n. 3, p. 1641–1647, 2006. Citado na página 11.
- XU, R.; WUNSCH, D. Survey of clustering algorithms. *IEEE Transactions on neural networks*, Ieee, v. 16, n. 3, p. 645–678, 2005. Citado na página 18.
- YAN, Y. et al. A survey on smart grid communication infrastructures: Motivations, requirements and challenges. *IEEE communications surveys & tutorials*, IEEE, v. 15, n. 1, p. 5–20, 2013. Citado na página 2.
- ZADEH, L. A. Fuzzy sets. *Information and control*, Elsevier, v. 8, n. 3, p. 338–353, 1965. Citado na página 19.

Apêndices

APÊNDICE A – Resultado da análise dos Algoritmos de Clusterização

Tabela 9 – Comparação entre algoritmos de clusterização Agglomerative Hierarchical Clustering (HAC), *k*-Means (KM), Gaussian Mixture Models (GMM) e Fuzzy c-means (FCM) através do valor da Largura da silhueta (LS) para as todas as barras com as resistências 10, 20, 30, 40 e 50 Ω

Barra	Resistência	K	Silhueta				Melhor Algoritmo	Melhor k
			HIC	KM	MG	FCM		
802	10	2	0,706	0,706	0,706	0,706	Todos	6
802	10	3	0,829	0,829	0,829	0,829	Todos	
802	10	4	0,760	0,760	0,760	0,760	Todos	
802	10	5	0,811	0,811	0,811	0,782	HIC-KM-MG	
802	10	6	0,831	0,831	0,831	0,678	HIC-KM-MG	
806	10	2	0,648	0,648	0,648	0,648	Todos	
806	10	3	0,828	0,828	0,828	0,828	Todos	
806	10	4	0,753	0,753	0,753	0,753	Todos	
806	10	5	0,770	0,770	0,792	0,734	MG	
806	10	6	0,819	0,819	0,819	0,659	HIC-KM-MG	
808	10	2	0,913	0,913	0,913	0,487	HIC-KM-MG	2
808	10	3	0,749	0,808	0,808	0,808	KM-MG-FCM	
808	10	4	0,820	0,838	0,838	0,838	KM-MG-FCM	
808	10	5	0,860	0,860	0,860	0,860	Todos	
808	10	6	0,904	0,904	0,904	0,858	HIC-KM-MG	
812	10	2	0,888	0,888	0,888	0,813	HIC-KM-MG	
812	10	3	0,798	0,825	0,825	0,825	KM-MG-FCM	
812	10	4	0,827	0,859	0,859	0,851	KM-MG	
812	10	5	0,893	0,893	0,893	0,835	HIC-KM-MG	
812	10	6	0,935	0,935	0,935	0,813	HIC-KM-MG	
814	10	2	0,813	0,862	0,862	0,601	KM-MG	6
814	10	3	0,792	0,841	0,827	0,841	KM-FCM	
814	10	4	0,800	0,875	0,800	0,803	KM	
814	10	5	0,867	0,867	0,912	0,883	MG	
814	10	6	0,952	0,952	0,952	0,776	HIC-KM-MG	

Continua na próxima página

816	10	2	0,810	0,857	0,857	0,569	KM-MG	6
816	10	3	0,789	0,839	0,823	0,839	KM-FCM	
816	10	4	0,797	0,877	0,797	0,803	KM	
816	10	5	0,867	0,867	0,912	0,884	MG	
816	10	6	0,952	0,952	0,952	0,778	HIC-KM-MG	
824	10	2	0,818	0,844	0,818	0,818	KM	
824	10	3	0,800	0,836	0,822	0,836	KM-FCM	
824	10	4	0,809	0,826	0,809	0,826	KM-FCM	
824	10	5	0,835	0,835	0,835	0,835	Todos	
824	10	6	0,904	0,904	0,904	0,815	HIC-KM-MG	
828	10	2	0,817	0,842	0,817	0,817	KM	6
828	10	3	0,800	0,835	0,823	0,835	KM-FCM	
828	10	4	0,809	0,821	0,809	0,821	KM-FCM	
828	10	5	0,828	0,828	0,828	0,828	Todos	
828	10	6	0,897	0,897	0,897	0,826	HIC-KM-MG	
830	10	2	0,783	0,783	0,594	0,625	HIC-KM	
830	10	3	0,823	0,856	0,856	0,856	KM-MG-FCM	
830	10	4	0,809	0,868	0,868	0,749	KM-MG	
830	10	5	0,837	0,888	0,837	0,828	KM	
830	10	6	0,877	0,877	0,877	0,843	HIC-KM-MG	
832	10	2	0,798	0,798	0,798	0,798	Todos	3
832	10	3	0,855	0,898	0,898	0,898	KM-MG-FCM	
832	10	4	0,809	0,817	0,817	0,817	KM-MG-FCM	
832	10	5	0,786	0,890	0,890	0,662	KM-MG	
832	10	6	0,857	0,887	0,887	0,749	MG	
834	10	2	0,787	0,787	0,787	0,787	Todos	
834	10	3	0,857	0,890	0,890	0,890	KM-MG-FCM	
834	10	4	0,833	0,861	0,802	0,802	KM	
834	10	5	0,837	0,909	0,909	0,698	KM-MG	
834	10	6	0,872	0,872	0,909	0,600	MG	
836	10	2	0,814	0,814	0,814	0,814	Todos	5
836	10	3	0,864	0,903	0,903	0,903	KM-MG-FCM	
836	10	4	0,817	0,791	0,791	0,757	HIC	
836	10	5	0,817	0,905	0,905	0,742	KM-MG	
836	10	6	0,868	0,755	0,900	0,578	MG	
840	10	2	0,815	0,815	0,815	0,815	Todos	
840	10	3	0,863	0,904	0,904	0,904	KM-MG-FCM	

Continua na próxima página

840	10	4	0,818	0,810	0,788	0,754	HIC	
840	10	5	0,813	0,902	0,902	0,458	KM-MG	
840	10	6	0,866	0,793	0,897	0,483	MG	
842	10	2	0,786	0,786	0,786	0,786	Todos	
842	10	3	0,855	0,889	0,889	0,889	KM-MG-FCM	
842	10	4	0,831	0,859	0,799	0,799	KM	
842	10	5	0,835	0,905	0,905	0,756	KM-MG	5
842	10	6	0,868	0,868	0,905	0,618	MG	
844	10	2	0,776	0,776	0,776	0,776	Todos	
844	10	3	0,849	0,880	0,880	0,880	KM-MG-FCM	
844	10	4	0,830	0,858	0,799	0,799	KM	
844	10	5	0,832	0,894	0,894	0,672	KM-MG	6
844	10	6	0,858	0,852	0,895	0,630	MG	
846	10	2	0,798	0,798	0,798	0,798	Todos	
846	10	3	0,855	0,892	0,892	0,892	KM-MG-FCM	
846	10	4	0,806	0,840	0,793	0,793	KM	
846	10	5	0,819	0,892	0,892	0,714	KM-MG	5
846	10	6	0,856	0,890	0,890	0,704	KM-MG	
848	10	2	0,799	0,799	0,799	0,799	Todos	
848	10	3	0,855	0,892	0,892	0,892	KM-MG-FCM	
848	10	4	0,806	0,837	0,790	0,790	KM	
848	10	5	0,816	0,889	0,889	0,712	KM-MG	3
848	10	6	0,853	0,888	0,888	0,707	KM-MG	
850	10	2	0,810	0,858	0,858	0,570	KM-MG	
850	10	3	0,788	0,838	0,822	0,838	KM-FCM	
850	10	4	0,795	0,878	0,795	0,801	KM	
850	10	5	0,866	0,866	0,912	0,885	MG	6
850	10	6	0,952	0,952	0,952	0,779	HIC-KM-MG	
852	10	2	0,789	0,789	0,789	0,789	Todos	
852	10	3	0,855	0,898	0,898	0,898	KM-MG-FCM	
852	10	4	0,810	0,822	0,822	0,822	KM-MG-FCM	
852	10	5	0,790	0,892	0,892	0,761	KM-MG	3
852	10	6	0,858	0,891	0,890	0,760	KM	
854	10	2	0,782	0,782	0,597	0,629	HIC-KM	
854	10	3	0,823	0,856	0,856	0,856	KM-MG-FCM	
854	10	4	0,808	0,871	0,871	0,748	KM-MG	
854	10	5	0,839	0,890	0,839	0,829	KM	5
854	10	6	0,878	0,878	0,878	0,841	HIC-KM-MG	

Continua na próxima página

858	10	2	0,798	0,798	0,798	0,798	Todos	5
858	10	3	0,856	0,898	0,898	0,898	KM-MG-FCM	
858	10	4	0,809	0,812	0,812	0,812	KM-MG-FCM	
858	10	5	0,816	0,898	0,898	0,744	KM-MG	
858	10	6	0,863	0,887	0,894	0,733	MG	
860	10	2	0,801	0,801	0,801	0,801	Todos	5
860	10	3	0,864	0,900	0,900	0,900	KM-MG-FCM	
860	10	4	0,826	0,802	0,802	0,802	HIC	
860	10	5	0,832	0,915	0,915	0,727	KM-MG	
860	10	6	0,878	0,878	0,915	0,693	MG	
862	10	2	0,811	0,811	0,811	0,811	Todos	3
862	10	3	0,860	0,900	0,900	0,900	KM-MG-FCM	
862	10	4	0,814	0,802	0,786	0,765	HIC	
862	10	5	0,812	0,897	0,897	0,619	KM-MG	
862	10	6	0,860	0,801	0,892	0,681	MG	
802	20	2	0,734	0,734	0,734	0,734	Todos	6
802	20	3	0,804	0,821	0,821	0,821	KM-MG-FCM	
802	20	4	0,772	0,772	0,777	0,769	MG	
802	20	5	0,831	0,831	0,818	0,785	HIC-KM	
802	20	6	0,843	0,843	0,785	0,709	HIC-KM	
806	20	2	0,718	0,718	0,718	0,718	Todos	5
806	20	3	0,797	0,797	0,797	0,797	Todos	
806	20	4	0,761	0,767	0,761	0,761	KM	
806	20	5	0,791	0,808	0,791	0,768	KM	
806	20	6	0,800	0,800	0,800	0,583	HIC-KM-MG	
808	20	2	0,901	0,901	0,901	0,494	HIC-KM-MG	2
808	20	3	0,805	0,806	0,792	0,806	KM-FCM	
808	20	4	0,795	0,818	0,813	0,818	KM-FCM	
808	20	5	0,849	0,849	0,849	0,849	Todos	
808	20	6	0,869	0,869	0,869	0,799	HIC-KM-MG	
812	20	2	0,892	0,892	0,892	0,892	Todos	6
812	20	3	0,797	0,820	0,820	0,825	FCM	
812	20	4	0,831	0,857	0,857	0,857	KM-MG-FCM	
812	20	5	0,889	0,889	0,889	0,889	Todos	
812	20	6	0,919	0,919	0,919	0,809	HIC-KM-MG	
814	20	2	0,887	0,887	0,887	0,829	HIC-KM-MG	6
814	20	3	0,829	0,843	0,829	0,843	KM-FCM	

Continua na próxima página

814	20	4	0,860	0,860	0,829	0,845	HIC-KM	
814	20	5	0,874	0,874	0,839	0,874	HIC-KM-FCM	
814	20	6	0,948	0,948	0,948	0,844	HIC-KM-MG	
816	20	2	0,828	0,884	0,884	0,828	KM-MG	
816	20	3	0,793	0,843	0,827	0,843	KM-FCM	
816	20	4	0,800	0,862	0,825	0,843	KM	
816	20	5	0,860	0,860	0,909	0,875	MG	6
816	20	6	0,949	0,949	0,949	0,853	HIC-KM-MG	
824	20	2	0,840	0,878	0,878	0,878	KM-MG-FCM	
824	20	3	0,803	0,835	0,835	0,835	KM-MG-FCM	
824	20	4	0,814	0,815	0,815	0,815	KM-MG-FCM	
824	20	5	0,763	0,807	0,807	0,807	KM-MG-FCM	6
824	20	6	0,833	0,884	0,884	0,797	KM-MG	
828	20	2	0,841	0,878	0,878	0,841	KM-MG	
828	20	3	0,804	0,834	0,844	0,834	MG	
828	20	4	0,815	0,814	0,814	0,814	HIC	
828	20	5	0,769	0,800	0,800	0,800	KM-MG-FCM	2
828	20	6	0,839	0,877	0,877	0,804	KM-MG	
830	20	2	0,825	0,825	0,825	0,825	Todos	
830	20	3	0,799	0,829	0,829	0,829	KM-MG-FCM	
830	20	4	0,796	0,851	0,851	0,749	KM-MG	
830	20	5	0,824	0,881	0,824	0,833	KM	6
830	20	6	0,873	0,873	0,918	0,812	MG	
832	20	2	0,769	0,785	0,769	0,785	KM-FCM	
832	20	3	0,862	0,901	0,901	0,901	KM-MG-FCM	
832	20	4	0,826	0,835	0,835	0,835	KM-MG-FCM	
832	20	5	0,808	0,896	0,896	0,795	KM-MG	3
832	20	6	0,869	0,869	0,894	0,788	MG	
834	20	2	0,798	0,798	0,798	0,798	Todos	
834	20	3	0,873	0,913	0,913	0,913	KM-MG-FCM	
834	20	4	0,835	0,836	0,836	0,803	KM-MG	
834	20	5	0,812	0,909	0,909	0,791	KM-MG	3
834	20	6	0,880	0,880	0,906	0,713	MG	
836	20	2	0,801	0,801	0,801	0,801	Todos	
836	20	3	0,869	0,911	0,911	0,911	KM-MG-FCM	
836	20	4	0,831	0,835	0,835	0,765	KM-MG	
836	20	5	0,808	0,900	0,900	0,787	KM-MG	3
836	20	6	0,870	0,899	0,899	0,672	KM-MG	

Continúa na próxima página

840	20	2	0,801	0,801	0,801	0,801	Todos	3
840	20	3	0,869	0,911	0,911	0,911	KM-MG-FCM	
840	20	4	0,830	0,834	0,834	0,764	KM-MG	
840	20	5	0,807	0,899	0,899	0,786	KM-MG	
840	20	6	0,868	0,897	0,897	0,654	KM-MG	
842	20	2	0,798	0,798	0,798	0,798	Todos	3
842	20	3	0,873	0,913	0,913	0,913	KM-MG-FCM	
842	20	4	0,834	0,836	0,836	0,802	KM-MG	
842	20	5	0,812	0,909	0,909	0,791	KM-MG	
842	20	6	0,879	0,879	0,906	0,708	MG	
844	20	2	0,800	0,800	0,800	0,800	Todos	3
844	20	3	0,873	0,912	0,912	0,912	KM-MG-FCM	
844	20	4	0,834	0,836	0,836	0,805	KM-MG	
844	20	5	0,812	0,910	0,910	0,789	KM-MG	
844	20	6	0,880	0,880	0,907	0,725	MG	
846	20	2	0,802	0,802	0,802	0,802	Todos	3
846	20	3	0,871	0,912	0,912	0,912	KM-MG-FCM	
846	20	4	0,832	0,837	0,837	0,801	KM-MG	
846	20	5	0,811	0,904	0,904	0,787	KM-MG	
846	20	6	0,873	0,873	0,902	0,691	MG	
848	20	2	0,802	0,802	0,802	0,802	Todos	3
848	20	3	0,870	0,911	0,911	0,911	KM-MG-FCM	
848	20	4	0,831	0,836	0,836	0,800	KM-MG	
848	20	5	0,810	0,903	0,903	0,786	KM-MG	
848	20	6	0,872	0,872	0,901	0,675	MG	
850	20	2	0,884	0,884	0,884	0,828	HIC-KM-MG	6
850	20	3	0,827	0,843	0,827	0,843	KM-FCM	
850	20	4	0,862	0,862	0,825	0,844	HIC-KM	
850	20	5	0,875	0,875	0,859	0,875	HIC-KM-FCM	
850	20	6	0,948	0,948	0,948	0,851	HIC-KM-MG	
852	20	2	0,754	0,781	0,754	0,781	KM-FCM	6
852	20	3	0,860	0,898	0,898	0,898	KM-MG-FCM	
852	20	4	0,826	0,834	0,834	0,834	KM-MG-FCM	
852	20	5	0,807	0,893	0,893	0,796	KM-MG	
852	20	6	0,866	0,908	0,892	0,762	KM	
854	20	2	0,824	0,824	0,824	0,824	Todos	
854	20	3	0,800	0,830	0,830	0,830	KM-MG-FCM	

Continua na próxima página

854	20	4	0,796	0,855	0,855	0,748	KM-MG	
854	20	5	0,827	0,883	0,827	0,835	KM	
854	20	6	0,873	0,873	0,918	0,812	MG	
858	20	2	0,787	0,787	0,787	0,787	Todos	3
858	20	3	0,869	0,909	0,909	0,909	KM-MG-FCM	
858	20	4	0,832	0,833	0,833	0,782	KM-MG	
858	20	5	0,808	0,901	0,901	0,793	KM-MG	
858	20	6	0,873	0,873	0,898	0,776	MG	
860	20	2	0,802	0,802	0,802	0,802	Todos	
860	20	3	0,873	0,913	0,913	0,913	KM-MG-FCM	
860	20	4	0,834	0,838	0,838	0,770	KM-MG	
860	20	5	0,813	0,908	0,908	0,791	KM-MG	
860	20	6	0,879	0,879	0,906	0,715	MG	
862	20	2	0,801	0,801	0,801	0,801	Todos	3
862	20	3	0,868	0,910	0,910	0,910	KM-MG-FCM	
862	20	4	0,830	0,833	0,833	0,764	KM-MG	
862	20	5	0,807	0,899	0,899	0,785	KM-MG	
862	20	6	0,868	0,897	0,897	0,638	KM-MG	
802	30	2	0,690	0,690	0,690	0,690	Todos	6
802	30	3	0,858	0,858	0,858	0,858	Todos	
802	30	4	0,813	0,813	0,813	0,813	Todos	
802	30	5	0,847	0,847	0,847	0,847	Todos	
802	30	6	0,841	0,876	0,802	0,807	KM	
806	30	2	0,706	0,706	0,706	0,706	Todos	6
806	30	3	0,819	0,819	0,819	0,819	Todos	
806	30	4	0,816	0,816	0,816	0,816	Todos	
806	30	5	0,829	0,829	0,829	0,829	Todos	
806	30	6	0,828	0,828	0,856	0,793	MG	
808	30	2	0,886	0,886	0,886	0,439	HIC-KM-MG	2
808	30	3	0,772	0,796	0,758	0,796	KM-FCM	
808	30	4	0,820	0,820	0,820	0,793	HIC-KM-MG	
808	30	5	0,856	0,856	0,856	0,856	Todos	
808	30	6	0,863	0,863	0,863	0,861	HIC-KM-MG	
812	30	2	0,891	0,891	0,891	0,808	HIC-KM-MG	6
812	30	3	0,796	0,815	0,815	0,816	FCM	
812	30	4	0,832	0,853	0,853	0,853	KM-MG-FCM	
812	30	5	0,873	0,873	0,873	0,873	Todos	
812	30	6	0,891	0,891	0,891	0,721	HIC-KM-MG	

Continua na próxima página

814	30	2	0,897	0,897	0,897	0,897	Todos	6
814	30	3	0,826	0,842	0,826	0,842	KM-FCM	
814	30	4	0,843	0,843	0,829	0,847	FCM	
814	30	5	0,859	0,859	0,842	0,863	FCM	
814	30	6	0,934	0,934	0,934	0,827	HIC-KM-MG	
816	30	2	0,896	0,896	0,896	0,896	Todos	
816	30	3	0,825	0,842	0,825	0,842	KM-FCM	
816	30	4	0,849	0,849	0,827	0,845	HIC-KM	
816	30	5	0,864	0,864	0,844	0,864	HIC-KM-FCM	
816	30	6	0,940	0,940	0,940	0,837	HIC-KM-MG	
824	30	2	0,849	0,890	0,890	0,890	KM-MG-FCM	2
824	30	3	0,809	0,849	0,849	0,849	KM-MG-FCM	
824	30	4	0,820	0,811	0,820	0,811	HIC-MG	
824	30	5	0,759	0,787	0,874	0,787	MG	
824	30	6	0,832	0,867	0,867	0,768	KM-MG	
828	30	2	0,849	0,890	0,890	0,890	KM-MG-FCM	
828	30	3	0,809	0,848	0,848	0,848	KM-MG-FCM	
828	30	4	0,820	0,820	0,820	0,811	HIC-KM-MG	
828	30	5	0,765	0,779	0,872	0,779	MG	
828	30	6	0,837	0,858	0,858	0,779	KM-MG	
830	30	2	0,840	0,840	0,858	0,840	MG	5
830	30	3	0,782	0,811	0,811	0,811	KM-MG-FCM	
830	30	4	0,782	0,841	0,841	0,745	KM-MG	
830	30	5	0,814	0,878	0,832	0,832	KM	
830	30	6	0,868	0,868	0,868	0,710	HIC-KM-MG	
832	30	2	0,745	0,775	0,745	0,775	KM-FCM	
832	30	3	0,860	0,902	0,902	0,902	KM-MG-FCM	
832	30	4	0,827	0,837	0,837	0,743	KM-MG	
832	30	5	0,805	0,891	0,891	0,797	KM-MG	
832	30	6	0,861	0,893	0,893	0,868	KM-MG	
834	30	2	0,767	0,788	0,767	0,788	KM-FCM	6
834	30	3	0,869	0,909	0,909	0,909	KM-MG-FCM	
834	30	4	0,833	0,842	0,842	0,810	KM-MG	
834	30	5	0,812	0,905	0,905	0,803	KM-MG	
834	30	6	0,872	0,872	0,905	0,919	FCM	
836	30	2	0,774	0,787	0,774	0,787	KM-FCM	
836	30	3	0,870	0,911	0,911	0,911	KM-MG-FCM	

Continua na próxima página

836	30	4	0,836	0,838	0,838	0,807	KM-MG	
836	30	5	0,809	0,901	0,901	0,804	KM-MG	
836	30	6	0,869	0,916	0,901	0,699	KM	
840	30	2	0,774	0,787	0,774	0,787	KM-FCM	
840	30	3	0,870	0,911	0,911	0,911	KM-MG-FCM	
840	30	4	0,837	0,837	0,837	0,808	KM-MG	
840	30	5	0,808	0,901	0,901	0,804	KM-MG	6
840	30	6	0,869	0,916	0,901	0,687	KM	
842	30	2	0,767	0,788	0,767	0,788	KM-FCM	
842	30	3	0,869	0,909	0,909	0,909	KM-MG-FCM	
842	30	4	0,833	0,842	0,842	0,810	KM-MG	
842	30	5	0,812	0,905	0,905	0,803	KM-MG	6
842	30	6	0,872	0,872	0,905	0,918	FCM	
844	30	2	0,769	0,788	0,769	0,788	KM-FCM	
844	30	3	0,870	0,910	0,910	0,910	KM-MG-FCM	
844	30	4	0,834	0,811	0,841	0,811	MG	
844	30	5	0,812	0,905	0,905	0,803	KM-MG	3
844	30	6	0,873	0,873	0,905	0,747	MG	
846	30	2	0,774	0,788	0,774	0,788	KM-FCM	
846	30	3	0,871	0,912	0,912	0,912	KM-MG-FCM	
846	30	4	0,837	0,839	0,839	0,810	KM-MG	
846	30	5	0,811	0,904	0,904	0,806	KM-MG	3
846	30	6	0,872	0,872	0,904	0,707	MG	
848	30	2	0,775	0,787	0,775	0,787	KM-FCM	
848	30	3	0,871	0,912	0,912	0,912	KM-MG-FCM	
848	30	4	0,837	0,839	0,839	0,811	KM-MG	
848	30	5	0,810	0,903	0,903	0,806	KM-MG	3
848	30	6	0,872	0,872	0,903	0,698	MG	
850	30	2	0,896	0,896	0,896	0,896	Todos	
850	30	3	0,825	0,842	0,825	0,842	KM-FCM	
850	30	4	0,848	0,848	0,827	0,845	HIC-KM	
850	30	5	0,863	0,863	0,841	0,863	HIC-KM-FCM	6
850	30	6	0,939	0,939	0,939	0,834	HIC-KM-MG	
852	30	2	0,725	0,759	0,725	0,759	KM-FCM	
852	30	3	0,860	0,900	0,900	0,900	KM-MG-FCM	
852	30	4	0,832	0,831	0,831	0,796	HIC	
852	30	5	0,802	0,892	0,892	0,803	KM-MG	3
852	30	6	0,865	0,865	0,893	0,861	MG	

Continua na próxima página

854	30	2	0,839	0,839	0,857	0,839	MG	5
854	30	3	0,783	0,812	0,812	0,812	KM-MG-FCM	
854	30	4	0,783	0,844	0,844	0,743	KM-MG	
854	30	5	0,816	0,880	0,835	0,835	KM	
854	30	6	0,869	0,869	0,869	0,738	HIC-KM-MG	
858	30	2	0,755	0,783	0,755	0,783	KM-FCM	3
858	30	3	0,863	0,903	0,903	0,903	KM-MG-FCM	
858	30	4	0,827	0,837	0,837	0,797	KM-MG	
858	30	5	0,805	0,894	0,894	0,796	KM-MG	
858	30	6	0,862	0,895	0,895	0,840	KM-MG	
860	30	2	0,771	0,789	0,771	0,789	KM-FCM	3
860	30	3	0,870	0,911	0,911	0,911	KM-MG-FCM	
860	30	4	0,835	0,842	0,842	0,810	KM-MG	
860	30	5	0,812	0,905	0,905	0,805	KM-MG	
860	30	6	0,872	0,872	0,905	0,682	MG	
862	30	2	0,774	0,787	0,774	0,787	KM-FCM	6
862	30	3	0,870	0,911	0,911	0,911	KM-MG-FCM	
862	30	4	0,836	0,837	0,837	0,807	KM-MG	
862	30	5	0,808	0,900	0,900	0,803	KM-MG	
862	30	6	0,868	0,915	0,900	0,681	KM	
802	40	2	0,665	0,665	0,665	0,665	Todos	3
802	40	3	0,879	0,879	0,879	0,879	Todos	
802	40	4	0,762	0,762	0,762	0,762	Todos	
802	40	5	0,830	0,830	0,830	0,830	Todos	
802	40	6	0,819	0,872	0,872	0,819	KM-MG	
806	40	2	0,679	0,679	0,679	0,679	Todos	6
806	40	3	0,863	0,863	0,863	0,863	Todos	
806	40	4	0,751	0,759	0,751	0,759	KM-FCM	
806	40	5	0,829	0,829	0,829	0,829	Todos	
806	40	6	0,794	0,794	0,865	0,798	MG	
808	40	2	0,861	0,861	0,861	0,451	HIC-KM-MG	6
808	40	3	0,792	0,802	0,792	0,802	KM-FCM	
808	40	4	0,831	0,831	0,831	0,732	HIC-KM-MG	
808	40	5	0,810	0,846	0,846	0,846	KM-MG-FCM	
808	40	6	0,846	0,846	0,864	0,747	MG	
812	40	2	0,881	0,881	0,867	0,503	HIC-KM	
812	40	3	0,747	0,814	0,805	0,814	KM-FCM	

Continua na próxima página

812	40	4	0,829	0,852	0,841	0,852	KM-FCM	
812	40	5	0,873	0,873	0,860	0,873	HIC-KM-FCM	
812	40	6	0,897	0,897	0,897	0,717	HIC-KM-MG	
814	40	2	0,897	0,897	0,897	0,897	Todos	
814	40	3	0,832	0,842	0,832	0,842	KM-FCM	
814	40	4	0,834	0,834	0,836	0,848	FCM	
814	40	5	0,850	0,850	0,846	0,858	FCM	6
814	40	6	0,924	0,924	0,924	0,806	HIC-KM-MG	
816	40	2	0,897	0,897	0,897	0,897	Todos	
816	40	3	0,833	0,843	0,833	0,843	KM-FCM	
816	40	4	0,840	0,840	0,836	0,848	FCM	
816	40	5	0,857	0,857	0,849	0,861	FCM	6
816	40	6	0,932	0,932	0,932	0,817	HIC-KM-MG	
824	40	2	0,850	0,894	0,894	0,894	KM-MG-FCM	
824	40	3	0,811	0,852	0,852	0,852	KM-MG-FCM	
824	40	4	0,820	0,820	0,820	0,820	Todos	
824	40	5	0,877	0,763	0,877	0,763	HIC-MG	2
824	40	6	0,824	0,845	0,726	0,750	KM	
828	40	2	0,850	0,894	0,894	0,894	KM-MG-FCM	
828	40	3	0,810	0,850	0,850	0,850	KM-MG-FCM	
828	40	4	0,819	0,819	0,819	0,819	Todos	
828	40	5	0,874	0,757	0,874	0,701	HIC-MG	2
828	40	6	0,826	0,839	0,807	0,761	KM	
830	40	2	0,844	0,868	0,868	0,844	KM-MG	
830	40	3	0,768	0,799	0,799	0,799	KM-MG-FCM	
830	40	4	0,771	0,833	0,833	0,738	KM-MG	
830	40	5	0,804	0,876	0,804	0,827	KM	5
830	40	6	0,862	0,862	0,862	0,768	HIC-KM-MG	
832	40	2	0,717	0,753	0,717	0,753	KM-FCM	
832	40	3	0,859	0,899	0,899	0,899	KM-MG-FCM	
832	40	4	0,834	0,829	0,829	0,799	HIC	
832	40	5	0,803	0,892	0,817	0,805	KM	3
832	40	6	0,866	0,866	0,893	0,853	MG	
834	40	2	0,739	0,774	0,739	0,774	KM-FCM	
834	40	3	0,864	0,904	0,904	0,904	KM-MG-FCM	
834	40	4	0,833	0,812	0,838	0,812	MG	
834	40	5	0,809	0,899	0,826	0,803	KM	3
834	40	6	0,867	0,867	0,900	0,817	MG	

Continua na próxima página

836	40	2	0,747	0,778	0,747	0,778	KM-FCM	3
836	40	3	0,863	0,905	0,905	0,905	KM-MG-FCM	
836	40	4	0,830	0,835	0,835	0,775	KM-MG	
836	40	5	0,803	0,889	0,825	0,797	KM	
836	40	6	0,857	0,857	0,892	0,771	MG	
840	40	2	0,747	0,778	0,747	0,778	KM-FCM	3
840	40	3	0,862	0,905	0,905	0,905	KM-MG-FCM	
840	40	4	0,829	0,835	0,835	0,774	KM-MG	
840	40	5	0,803	0,887	0,825	0,796	KM	
840	40	6	0,855	0,891	0,891	0,751	KM-MG	
842	40	2	0,739	0,774	0,739	0,774	KM-FCM	3
842	40	3	0,864	0,904	0,904	0,904	KM-MG-FCM	
842	40	4	0,833	0,811	0,838	0,811	MG	
842	40	5	0,808	0,898	0,826	0,803	KM	
842	40	6	0,867	0,867	0,900	0,817	MG	
844	40	2	0,741	0,775	0,741	0,775	KM-FCM	3
844	40	3	0,864	0,904	0,904	0,904	KM-MG-FCM	
844	40	4	0,832	0,810	0,837	0,810	MG	
844	40	5	0,808	0,897	0,826	0,801	KM	
844	40	6	0,865	0,865	0,899	0,788	MG	
846	40	2	0,747	0,778	0,747	0,778	KM-FCM	3
846	40	3	0,864	0,906	0,906	0,906	KM-MG-FCM	
846	40	4	0,831	0,838	0,838	0,778	KM-MG	
846	40	5	0,806	0,891	0,827	0,798	KM	
846	40	6	0,859	0,859	0,894	0,784	MG	
848	40	2	0,748	0,779	0,748	0,779	KM-FCM	3
848	40	3	0,864	0,906	0,906	0,906	KM-MG-FCM	
848	40	4	0,830	0,838	0,838	0,777	KM-MG	
848	40	5	0,806	0,891	0,827	0,798	KM	
848	40	6	0,858	0,858	0,894	0,779	MG	
850	40	2	0,896	0,896	0,896	0,896	Todos	6
850	40	3	0,832	0,843	0,832	0,843	KM-FCM	
850	40	4	0,841	0,841	0,836	0,848	FCM	
850	40	5	0,857	0,857	0,846	0,786	HIC-KM	
850	40	6	0,931	0,931	0,931	0,815	HIC-KM-MG	
852	40	2	0,701	0,738	0,701	0,738	KM-FCM	
852	40	3	0,858	0,897	0,897	0,897	KM-MG-FCM	

Continua na próxima página

852	40	4	0,838	0,803	0,824	0,803	HIC	
852	40	5	0,803	0,891	0,812	0,807	KM	
852	40	6	0,869	0,869	0,893	0,832	MG	
854	40	2	0,843	0,843	0,867	0,843	MG	
854	40	3	0,769	0,800	0,800	0,800	KM-MG-FCM	
854	40	4	0,772	0,836	0,836	0,738	KM-MG	
854	40	5	0,806	0,877	0,806	0,829	KM	5
854	40	6	0,862	0,862	0,862	0,687	HIC-KM-MG	
858	40	2	0,728	0,764	0,728	0,764	KM-FCM	
858	40	3	0,863	0,902	0,902	0,902	KM-MG-FCM	
858	40	4	0,835	0,807	0,835	0,807	MG	
858	40	5	0,808	0,897	0,823	0,807	KM	3
858	40	6	0,869	0,869	0,899	0,813	MG	
860	40	2	0,744	0,776	0,744	0,776	KM-FCM	
860	40	3	0,864	0,905	0,905	0,905	KM-MG-FCM	
860	40	4	0,832	0,809	0,837	0,809	MG	
860	40	5	0,807	0,895	0,826	0,801	KM	3
860	40	6	0,863	0,863	0,897	0,774	MG	
862	40	2	0,747	0,778	0,747	0,778	KM-FCM	
862	40	3	0,862	0,905	0,905	0,905	KM-MG-FCM	
862	40	4	0,830	0,835	0,835	0,775	KM-MG	
862	40	5	0,803	0,888	0,825	0,796	KM	3
862	40	6	0,856	0,856	0,891	0,755	MG	
802	50	2	0,636	0,636	0,636	0,636	Todos	
802	50	3	0,881	0,881	0,881	0,881	Todos	
802	50	4	0,738	0,775	0,784	0,777	MG	
802	50	5	0,808	0,808	0,789	0,808	HIC-KM-FCM	3
802	50	6	0,797	0,860	0,778	0,765	KM	
806	50	2	0,658	0,658	0,658	0,650	HIC-KM-MG	
806	50	3	0,873	0,873	0,873	0,873	Todos	
806	50	4	0,733	0,759	0,765	0,759	MG	
806	50	5	0,799	0,763	0,776	0,799	HIC-FCM	3
806	50	6	0,759	0,759	0,745	0,756	HIC-KM	
808	50	2	0,843	0,843	0,843	0,481	HIC-KM-MG	
808	50	3	0,813	0,813	0,813	0,810	HIC-KM-MG	
808	50	4	0,838	0,838	0,838	0,838	Todos	
808	50	5	0,803	0,803	0,819	0,803	MG	6
808	50	6	0,869	0,869	0,869	0,773	HIC-KM-MG	

Continua na próxima página

812	50	2	0,874	0,874	0,862	0,503	HIC-KM	6
812	50	3	0,770	0,809	0,807	0,809	KM-FCM	
812	50	4	0,839	0,848	0,844	0,848	KM-FCM	
812	50	5	0,883	0,883	0,883	0,883	Todos	
812	50	6	0,898	0,898	0,898	0,711	HIC-KM-MG	
814	50	2	0,894	0,894	0,894	0,829	HIC-KM-MG	
814	50	3	0,840	0,840	0,840	0,843	FCM	
814	50	4	0,818	0,818	0,844	0,803	MG	
814	50	5	0,835	0,835	0,849	0,849	MG-FCM	
814	50	6	0,905	0,905	0,905	0,774	HIC-KM-MG	
816	50	2	0,897	0,897	0,897	0,897	Todos	6
816	50	3	0,833	0,842	0,833	0,842	KM-FCM	
816	50	4	0,826	0,826	0,836	0,796	MG	
816	50	5	0,843	0,843	0,849	0,859	FCM	
816	50	6	0,916	0,916	0,916	0,798	HIC-KM-MG	
824	50	2	0,843	0,891	0,891	0,891	KM-MG-FCM	
824	50	3	0,787	0,827	0,834	0,827	MG	
824	50	4	0,795	0,799	0,799	0,761	KM-MG	
824	50	5	0,721	0,764	0,839	0,764	MG	
824	50	6	0,795	0,848	0,848	0,771	KM-MG	
828	50	2	0,891	0,891	0,891	0,891	Todos	2
828	50	3	0,801	0,826	0,834	0,826	MG	
828	50	4	0,761	0,799	0,796	0,761	KM	
828	50	5	0,774	0,755	0,845	0,694	MG	
828	50	6	0,791	0,838	0,838	0,780	KM-MG	
830	50	2	0,845	0,873	0,873	0,845	KM-MG	
830	50	3	0,755	0,787	0,787	0,787	KM-MG-FCM	
830	50	4	0,758	0,819	0,819	0,737	KM-MG	
830	50	5	0,789	0,871	0,814	0,814	KM	
830	50	6	0,852	0,880	0,880	0,769	KM-MG	
832	50	2	0,696	0,736	0,696	0,736	KM-FCM	3
832	50	3	0,857	0,895	0,895	0,895	KM-MG-FCM	
832	50	4	0,837	0,805	0,823	0,805	HIC	
832	50	5	0,803	0,891	0,811	0,806	KM	
832	50	6	0,869	0,869	0,893	0,828	MG	
834	50	2	0,719	0,756	0,719	0,756	KM-FCM	
834	50	3	0,862	0,901	0,901	0,901	KM-MG-FCM	

Continua na próxima página

834	50	4	0,839	0,812	0,826	0,812	HIC	
834	50	5	0,805	0,892	0,814	0,803	KM	
834	50	6	0,868	0,868	0,894	0,805	MG	
836	50	2	0,725	0,761	0,725	0,761	KM-FCM	3
836	50	3	0,861	0,902	0,902	0,902	KM-MG-FCM	
836	50	4	0,836	0,808	0,827	0,808	HIC	
836	50	5	0,803	0,887	0,816	0,801	KM	
836	50	6	0,861	0,890	0,838	0,781	KM	
840	50	2	0,725	0,761	0,725	0,761	KM-FCM	
840	50	3	0,862	0,903	0,903	0,903	KM-MG-FCM	
840	50	4	0,836	0,809	0,828	0,809	HIC	
840	50	5	0,804	0,887	0,817	0,801	KM	
840	50	6	0,861	0,861	0,839	0,772	HIC-KM	
842	50	2	0,719	0,756	0,719	0,756	KM-FCM	3
842	50	3	0,862	0,901	0,901	0,901	KM-MG-FCM	
842	50	4	0,838	0,812	0,826	0,812	HIC	
842	50	5	0,805	0,892	0,815	0,803	KM	
842	50	6	0,867	0,867	0,894	0,804	MG	
844	50	2	0,720	0,757	0,720	0,757	KM-FCM	3
844	50	3	0,862	0,901	0,901	0,901	KM-MG-FCM	
844	50	4	0,838	0,812	0,826	0,812	HIC	
844	50	5	0,804	0,891	0,815	0,802	KM	
844	50	6	0,866	0,866	0,893	0,809	MG	
846	50	2	0,724	0,761	0,724	0,761	KM-FCM	3
846	50	3	0,862	0,903	0,903	0,903	KM-MG-FCM	
846	50	4	0,837	0,811	0,830	0,811	HIC	
846	50	5	0,806	0,890	0,818	0,806	KM	
846	50	6	0,864	0,864	0,892	0,782	MG	
848	50	2	0,724	0,761	0,724	0,761	KM-FCM	3
848	50	3	0,862	0,903	0,903	0,903	KM-MG-FCM	
848	50	4	0,837	0,811	0,830	0,811	HIC	
848	50	5	0,806	0,890	0,819	0,819	KM	
848	50	6	0,864	0,864	0,841	0,777	HIC-KM	
850	50	2	0,897	0,897	0,897	0,897	Todos	6
850	50	3	0,832	0,842	0,832	0,842	KM-FCM	
850	50	4	0,827	0,827	0,836	0,795	MG	
850	50	5	0,843	0,843	0,846	0,857	FCM	
850	50	6	0,916	0,916	0,916	0,797	HIC-KM-MG	

Continua na próxima página

852	50	2	0,677	0,720	0,677	0,720	KM-FCM	3
852	50	3	0,852	0,888	0,888	0,888	KM-MG-FCM	
852	50	4	0,836	0,809	0,813	0,809	HIC	
852	50	5	0,797	0,887	0,887	0,802	KM-MG	
852	50	6	0,867	0,867	0,841	0,808	HIC-KM	
854	50	2	0,844	0,872	0,872	0,844	KM-MG	
854	50	3	0,757	0,788	0,788	0,788	KM-MG-FCM	
854	50	4	0,759	0,822	0,822	0,737	KM-MG	
854	50	5	0,792	0,872	0,817	0,817	KM	
854	50	6	0,852	0,881	0,881	0,770	KM-MG	
858	50	2	0,707	0,747	0,707	0,747	KM-FCM	3
858	50	3	0,859	0,897	0,897	0,897	KM-MG-FCM	
858	50	4	0,837	0,809	0,822	0,809	HIC	
858	50	5	0,802	0,890	0,890	0,803	KM-MG	
858	50	6	0,867	0,867	0,892	0,823	MG	
860	50	2	0,722	0,759	0,722	0,759	KM-FCM	3
860	50	3	0,862	0,902	0,902	0,902	KM-MG-FCM	
860	50	4	0,838	0,812	0,827	0,812	HIC	
860	50	5	0,805	0,891	0,816	0,803	KM	
860	50	6	0,866	0,866	0,893	0,797	MG	
862	50	2	0,724	0,761	0,724	0,761	KM-FCM	3
862	50	3	0,861	0,902	0,902	0,902	KM-MG-FCM	
862	50	4	0,836	0,808	0,828	0,808	HIC	
862	50	5	0,803	0,887	0,817	0,801	KM	
862	50	6	0,861	0,861	0,838	0,772	HIC-KM	