

Clusterização de distribuidoras para análise de custo-benefício da implantação das Redes Elétricas Inteligentes no Brasil

Anésio de Leles Ferreira Filho, *Dr.*
Marcos Venicius Leite Vasconcelos
Departamento de Engenharia Elétrica
Universidade de Brasília
leles@ene.unb.br
marcosleitevasc@gmail.com

Davi Vidal Rôla Almeida
Superintendência de Regulação dos Serviços de Distribuição
Agência Nacional de Energia Elétrica
davividal@aneel.gov.br

Resumo—Este trabalho objetiva apresentar os resultados da avaliação de três algoritmos de clusterização de 20 distribuidoras, a saber, média das distâncias, distância dos centroides e *Ward*. Empregam-se três atributos para a análise de custo-benefício da implantação das Redes Elétricas Inteligentes (REI) no Brasil, redução das perdas não técnicas e da demanda de ponta e melhoria da qualidade do serviço. Inicialmente, faz-se uma exposição sobre as diferenças entre as empresas. Em seguida, discorre-se sobre o processo de clusterização, os algoritmos utilizados, os parâmetros de homogeneidade e heterogeneidade e a definição dos atributos adotados nesta investigação. Por meio de simulações computacionais, chega-se a um número de *clusters* que permite a execução da análise de custo-benefício da implantação das REI no Brasil, sem grande perda de homogeneidade nos grupos. Considerando que o setor elétrico conta com 63 distribuidoras, este trabalho mostra-se como uma importante ferramenta para a execução de estudos relacionados às REI.

Palavras-chave— Clusterização, Custo-benefício, Distribuição de energia elétrica, Redes Elétricas Inteligentes.

I. INTRODUÇÃO

Alguns dos inúmeros projetos pilotos e trabalhos acadêmicos executados no Brasil que versam sobre as Redes Elétricas Inteligentes (REI) avaliaram o custo-benefício da sua implantação em nível nacional. Todavia, não existe até o momento uma análise dessa natureza que considere as características específicas das concessionárias de distribuição do País [1].

Tal avaliação se justifica devido à grande heterogeneidade existente entre as áreas de concessão de distribuição no Brasil, expressa pelas diferenças em relação à extensão territorial, ao nível de receita, à tarifa média, ao mercado de fornecimento, aos indicadores de qualidade do serviço, à demanda de ponta, aos patamares de perdas de energia, dentre outros aspectos [2].

A execução de uma análise de custo-benefício para cada uma das 63 concessionárias de distribuição do Brasil caracteriza-se como um trabalho complexo, considerando-se o seu escopo e a carência de informações de algumas empresas. Nesse sentido, o agrupamento das distribuidoras com base na

seleção de atributos, assinala os resultados desta investigação como primordiais para a execução adequada de estudos vinculados às REI.

Com base nos aspectos ora mencionados, surgiu a ideia de desenvolvimento deste trabalho que objetiva apresentar os resultados da avaliação de três algoritmos de clusterização para classificação de 20 concessionárias de distribuição com base em três atributos sensíveis para a análise de custo-benefício da implantação das REI no Brasil. Para tanto, inicialmente são expostas as diferenças entre as áreas de concessão de distribuição. Em seguida, são exibidos os processos de clusterização existentes na literatura, os três algoritmos selecionados para avaliação, os parâmetros de homogeneidade e heterogeneidade e os atributos considerados. Na sequência, apresentam-se os resultados e as conclusões.

II. DESENVOLVIMENTO

A. Diferenças entre as áreas de concessão

Atuam hoje no Brasil 63 concessionárias e 38 permissionárias de distribuição, que atendiam a mais de 76 milhões de unidades consumidoras ao final de 2014 [3]. Dessas unidades consumidoras, 85% eram residenciais, com consumo médio mensal de 173 kWh [3]. As concessionárias são responsáveis pelo atendimento de mais de 99% do mercado e do número de consumidores [3], representando quase a totalidade da prestação do serviço de distribuição no País.

As concessionárias de distribuição tiveram uma receita de fornecimento da ordem de R\$ 128 bilhões em 2014¹ [3], porém de forma significativamente discrepante. A concessionária com maior receita arrecadou quase R\$ 12 bilhões em 2014 e a de menor receita arrecadou menos de R\$ 7 milhões no mesmo ano, menos de 0,06% da maior receita [3].

Essas diferenças também se estendem para o consumo mensal médio, que, ao final de 2014, foi de 1160 kWh para a concessionária com o maior consumo, enquanto a menor obteve média de 207 kWh [3]. Na mesma linha, verificando-

¹ Com impostos federais e estaduais.

se a qualidade do serviço, existem empresas com limites de continuidade que representam mais que o dobro de outras [4]. Quanto às compensações pela infringência desses limites, há uma empresa que compensou seus consumidores em R\$ 59 milhões em 2014, enquanto outra compensou em menos de R\$ 7 no mesmo ano [5].

No tocante às perdas de energia, verificaram-se percentuais calculados de perdas totais que oscilavam entre 3,11% e 39,88% ao final do 3º Ciclo de Revisões Tarifárias Periódicas, em 2014 [2]. Diferenças também são verificadas em aspectos relacionados à densidade de carga, extensão de rede, distribuição de consumidores por classe de consumo e nível de tensão, entre outros aspectos, o que justifica as discrepâncias entre as tarifas médias de fornecimento, que variavam de 258,65 R\$/MWh a 578,59 R\$/MWh, ao final de 2014 [3].

Neste trabalho, com vistas a avaliar a robustez dos algoritmos de clusterização, utilizaram-se as 10 maiores e as 10 menores concessionárias de distribuição em relação ao mercado de fornecimento. A Tabela 1 apresenta algumas informações referentes a essas empresas que elucidam suas diferenças.

TABELA 1. INFORMAÇÕES SOBRE AS 10 MAIORES E AS 10 MENORES CONCESSIONÁRIAS DE DISTRIBUIÇÃO - 2014 [3].

Distribuidora	Região	Receita de Fornecimento com Impostos (R\$)	Nº de Unidades Consumidoras ao Final do Ano	Tarifa Média com Impostos (R\$/MWh)	Consumo Médio ao Final do Ano (R\$/consumidor)	Compensação de Continuidade (R\$)
ELETROPAULO	SE	11.896.862.254,47	6.699.648,00	314,12	473,04	27.748.909,85
CEMIG	SE	11.517.567.993,42	7.936.965,00	427,62	270,77	20.122.581,83
COPEL	S	8.746.876.034,68	4.326.952,00	361,32	487,05	15.063.305,07
CPFL PAULISTA	SE	8.047.720.293,05	4.077.555,00	353,11	461,72	6.279.556,21
LIGHT	SE	9.090.414.983,98	3.738.727,00	422,80	498,44	29.730.170,98
COELBA	NE	5.743.512.367,28	5.565.729,00	351,33	247,23	30.688.507,40
CELESC	S	6.091.766.387,36	2.680.073,00	373,13	524,44	8.087.356,64
ELEKTRO	SE/CO	5.068.024.648,93	2.439.353,00	389,16	453,86	6.828.733,16
CELG	CO	4.415.528.007,93	2.716.002,00	377,16	339,85	58.996.899,11
CELPE	NE	4.144.271.346,17	3.428.012,00	369,53	291,34	20.323.235,14
ELETRORCAR	S	77.302.830,34	35.709,00	419,63	428,98	78.643,05
DEMEI	S	56.474.291,34	31.158,00	411,94	392,72	98.556,62
HIDROPAN	S	47.808.118,44	17.337,00	398,53	605,59	12.611,75
CHESP	CO	55.138.022,19	34.108,00	483,33	268,12	260.400,69
CERR	N	23.131.848,34	41.294,00	231,14	206,92	-
EFLUL	S	26.330.887,10	6.401,00	358,00	1.161,60	243,69
UHENPAL	S	27.567.047,31	15.165,00	407,61	386,87	86.348,10
MUX ENERGIA	S	23.813.691,48	10.577,00	383,72	489,29	543,92
FORCEL	S	20.076.337,96	7.113,00	362,03	673,66	6,42
EFLJC	S	6.974.011,64	3.556,00	425,78	408,63	22,98

Nota-se da Tabela 1 que todas as regiões do país foram representadas nas empresas analisadas, o que permite avaliar se a regionalidade interfere no resultado do processo de clusterização.

B. Clusterização

Processos de clusterização têm se tornado cada vez mais importantes nos últimos tempos, uma vez que se vive em um mundo repleto de dados e informações [6]. Uma das tarefas mais relevantes na análise de dados é a classificação, que busca identificar propriedades similares entre os objetos para agrupá-los [6].

Um *cluster*, ou grupo, pode ser descrito como um conjunto de entidades (ou objetos) que possuem similaridades. Da mesma forma, entidades de *clusters* diferentes são dissimilares [7].

Algoritmos de clusterização agrupam os objetos em análise em uma determinada quantidade de *clusters*.

Existem dois métodos principais de algoritmos de clusterização, o particional e o hierárquico. O primeiro método pressupõe a definição inicial da quantidade de *clusters* que se deseja no processo de agrupamento, ao passo que o segundo permite a escolha posterior da quantidade de *clusters*. Nesse último, constrói-se a árvore de agrupamentos (ou dendrograma), o que permite a visualização de cada etapa do processo.

O método hierárquico pode ser do tipo divisivo ou aglomerativo. No primeiro caso, parte-se de um único grupo com todos os n objetos analisados para a formação de n grupos com um objeto cada. Por outro lado, o método hierárquico aglomerativo parte de n grupos com um objeto cada para a formação de um grande grupo com todos os objetos [7].

Tendo em vista que a definição da quantidade de *clusters* é uma etapa deste trabalho e que o método hierárquico permite uma melhor visualização do processo de agrupamento, optou-se pela utilização do método hierárquico aglomerativo.

No método aglomerativo, após a representação de cada objeto como um único *cluster*, calcula-se a matriz de proximidade dos objetos com base em uma função de distância, usualmente a distância euclidiana, que pode ser representada como exibido na equação (1) [8]:

$$d_{(i,j)} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (1)$$

Onde i e j são dois objetos analisados e p a quantidade de atributos que representa cada objeto.

Neste trabalho, o número total de objetos é igual a 20, indicando o total de concessionárias, e o número de atributos é igual a 3. Portanto, p assume o valor 3 e i e j representam pares de distribuidoras.

Posteriormente, os objetos e *clusters* vão sendo agrupados a partir das menores distâncias calculadas com base em um algoritmo específico.

C. Algoritmos de Clusterização analisados

A ferramenta utilizada para clusterização das concessionárias foi o SAS® Enterprise Guide 6.1. Nessa ferramenta, estão disponíveis diferentes algoritmos de clusterização, tanto para o método particional quanto para o método hierárquico.

No método hierárquico o programa SAS® Enterprise Guide 6.1 disponibiliza três algoritmos de clusterização hierárquica aglomerativa, que se diferenciam pela forma de cálculo da distância entre os *clusters*. São eles: (1) média das distâncias entre os objetos, (2) distância dos centroides e (3) Ward. A seguir, têm-se uma breve descrição de cada um deles.

1) Algoritmo da média das distâncias entre os objetos

De acordo com [8], o algoritmo de média das distâncias calcula a distância entre dois *clusters* utilizando a média das similaridades em cada *cluster*, equação (2).

$$d_{(R)(S)} = \frac{\sum_r \sum_s d_{rs}}{n_R n_S} \quad (2)$$

Onde r pertence ao *cluster* R , s pertence ao *cluster* S , d_{rs} representa a medida de proximidade entre r e s calculada pela distância euclidiana, e n_R e n_S representam o número de objetos em cada *cluster*.

II) Algoritmo da distância dos centroides

Partindo da matriz de proximidades, o algoritmo da distância dos centroides executa a primeira combinação usando a média ponderada dos dois *clusters* mais similares, formando um novo *cluster* T [8].

Os centroides dos *clusters* r e s são apresentados na equação (3):

$$\bar{y}_r = \frac{\sum_r y_r}{n_r} \quad \text{e} \quad \bar{y}_s = \frac{\sum_s y_s}{n_s} \quad (3)$$

Onde y_r representa os objetos do *cluster* R e y_s representa os objetos do *cluster* S .

A distância entre dois *clusters* é dada pela equação (4):

$$d_{rs} = \|\bar{y}_r - \bar{y}_s\| \quad (4)$$

O centroide do novo *cluster* será dado pela equação (5):

$$\bar{y}_t = \frac{(n_R \bar{y}_r + n_S \bar{y}_s)}{(n_R + n_S)} \quad (5)$$

Onde y_s representa o centroide do *cluster* s .

Na sequência, a distância para um terceiro *cluster* será dada pelo quadrado da distância euclidiana de T para o centroide do terceiro *cluster*, U , por meio da equação (6):

$$d_{tu}^2 = \left(\frac{n_R}{n_R + n_S}\right) d_{rs}^2 + \left(\frac{n_S}{n_R + n_S}\right) d_{ru}^2 - \left(\frac{n_R n_S}{n_R + n_S}\right) d_{rs}^2 \quad (6)$$

III) Algoritmo Ward

O terceiro algoritmo, *Ward* ou soma incremental dos quadrados, baseia-se no menor incremento da soma dos quadrados dos erros, sendo esta definida para um único *cluster* com todos os objetos conforme equação (7) [8]:

$$SSE = \sum_{i=1}^n (y_i - \bar{y})'(y_i - \bar{y}) = \sum_{i=1}^n \|y_i - \bar{y}\|^2 = T \quad (7)$$

Sendo $\bar{y}$ o centroide de um dado *cluster* e y_i seus elementos.

Assim, na combinação dos *clusters* R e S formando o *cluster* T , o quadrado da soma dos erros de T será, equação (8):

$$SSE_t = \sum_t \|y_t - \bar{y}\|^2, t \in T \quad (8)$$

Busca-se, então, o menor valor do incremento do quadrado da soma dos erros, conforme representado pela equação (9):

$$SSE_t - (SSE_r + SSE_s) \quad (9)$$

Tendo apresentado os algoritmos utilizados neste trabalho, passa-se a considerar os conceitos e métricas dos parâmetros homogeneidade e heterogeneidade, que serão importantes para a avaliação dos algoritmos e para definição do número de *clusters*.

D. Homogeneidade e Heterogeneidade

O parâmetro homogeneidade representa a similaridade ou proximidade entre dois objetos, ao passo que a heterogeneidade indica a dissimilaridade ou distância entre os objetos. Esses conceitos podem ser estendidos para o estudo de *clusters*, indicando a similaridade dos elementos nos grupos. A seguir, apresenta-se uma métrica para cálculo da homogeneidade e heterogeneidade de *clusters*.

Tomando-se a soma dos quadrados das distâncias para n *clusters*, conforme apresentado anteriormente, repetida na equação (10) [8]:

$$T = \sum_{i=1}^n \|y_i - \bar{y}\|^2 \quad (10)$$

E a soma dos quadrados dentro de um *cluster* K , equação (11) [8]:

$$SSE_k = \sum_i \|y_i - \bar{y}_k\|^2 \quad (11)$$

Define-se R^2 para k *clusters*, equação (12) [8]:

$$R_k^2 = \frac{T - \sum_k SSE_k}{T} \quad (12)$$

Portanto, para $R^2 = 1$ tem-se o máximo de heterogeneidade entre os *clusters* e de homogeneidade dentro dos *clusters*. Da mesma forma, quanto mais R^2 se aproxima de 0, mais se perde em homogeneidade dentro dos *clusters* e heterogeneidade entre eles.

SR^2 é definido de acordo com a equação (13) [8]:

$$SR^2 = R_k^2 - R_{k-1}^2 \quad (13)$$

Assim, quanto maior SR^2 , maior a diferença entre R^2 de um agrupamento e R^2 do agrupamento seguinte. Portanto, maior a perda de homogeneidade dentro dos *clusters*.

E. Definição dos atributos

Em [9], no estudo do custo-benefício da implantação das REI no Brasil, realizou-se análise de sensibilidade a respeito de diferentes variáveis: custo de aquisição dos medidores inteligentes e do *in home display*, tempo de análise, taxa de desconto, percentual de redução de perdas não técnicas, percentual de melhoria de eficiência energética e da qualidade do serviço.

Das variáveis analisadas, o custo de aquisição de medidores e do *in home display* não sofrem grandes variações entre as concessionárias de distribuição. Em que pesem os diferentes poderes de compra dessas empresas, espera-se que o aumento

da produção desses dispositivos eleve a disponibilidade em todo o território nacional, o que contribuirá para o preço final de compra depender mais de negociação entre as partes do que das características específicas das áreas de concessão. O tempo de análise e a taxa de desconto, que têm como referência o período de vigência do contrato de concessão e a taxa de remuneração, respectivamente, também são iguais para todas as empresas.

Dessa forma, consideram-se neste trabalho como atributos sensíveis para a clusterização o percentual de redução das perdas não técnicas e as melhorias de eficiência energética e da qualidade do serviço. No tocante à melhoria de eficiência energética, adota-se a redução da demanda de pico; e com respeito à qualidade do serviço, considera-se a recuperação de receita decorrente da redução da duração total de interrupções no fornecimento de energia elétrica.

F. Levantamento de Dados

Os montantes de perdas não técnicas foram acessados nos resultados do 3º Ciclo de Revisões Tarifárias Periódicas das distribuidoras. Foram utilizadas as informações de perdas apuradas pelas empresas. Para se alcançarem os totais reduzidos com a implantação das Redes Inteligentes aplicou-se o percentual de redução utilizado em [9], de 33%.

As informações referentes à demanda máxima das empresas foram retiradas do Plano de Desenvolvimento da Distribuição (PDD), considerando o investimento necessário para atendimento da demanda anual. De posse dos valores de demanda de pico aplicou-se o percentual utilizado em [9], de 2,5% de crescimento da demanda evitado.

Com referência à recuperação de receita pela redução da duração média das interrupções no fornecimento, adotou-se o valor apurado do indicador Duração Equivalente de Interrupção por Unidade Consumidora (DEC) em 2014 por cada empresa. Essas informações são encaminhadas anualmente pelas empresas à ANEEL. Sobre esse valor foi aplicado o percentual de redução utilizado em [9], de 40%.

Considerando que os atributos referem-se a grandezas diferentes, as reduções decorrentes da implantação das novas funcionalidades tecnológicas foram monetizadas. O montante reduzido de perdas não técnicas foi precificado pelo custo médio de compra da energia de cada empresa, ao passo que a redução da demanda teve como referência o custo evitado com a expansão do sistema. Por fim, a redução devido à melhoria do DEC baseou-se na tarifa média de fornecimento de cada concessionária.

Posteriormente, com vistas a evitar distorções decorrentes do porte das empresas, os atributos monetizados foram normalizados pela receita anual de cada concessionária.

A Tabela 2 apresenta os atributos normalizados utilizados no agrupamento das 20 concessionárias de distribuição.

TABELA 2. DADOS NORMALIZADOS UTILIZADOS NA ANÁLISE DOS ALGORITMOS DE CLUSTERIZAÇÃO

Distribuidora	Nº Objeto	Dados Normalizados		
		Redução da Demanda de Ponta	Recuperação de Receita	Perdas Não Técnicas
ELETROPAULO	1	0,055763277	0,000403653	0,01476087
CEMIG	2	0,069593312	0,000491781	0,008542441
COPEL	3	0,043394722	0,000639726	0,006174064
CPFL PAULISTA	4	0,039182207	0,000316438	0,013962207
LIGHT	5	0,07043998	0,000559361	0,084386318
COELBA	6	0,083852492	0,001037443	0,018750418
CELESC	7	0,038570539	0,000692237	0,004117123
ELEKTRO	8	0,058109327	0,000378539	0,002907132
CELG	9	0,032023497	0,001844749	0,003019297
CELPE	10	0,119222317	0,001127854	0,031253102
ELETROCAR	11	0,183931274	0,000740183	0,001528176
DEMEI	12	0,008085575	0,000386758	0,013076802
HIDROPAN	13	0,083981503	0,000307306	0,002685734
CHESP	14	0,029728554	0,000644749	0,001604006
CERR	15	0	0,004753425	0
EFLUL	16	0,006475278	0,000263927	0
UHENPAL	17	0,038024456	0,000821918	0
MUX ENERGIA	18	0,010199143	0,000361187	0,001616725
FORCEL	19	0	3,69863E-05	0
EFLJC	20	0,008123786	0,000150228	0

Os dados de redução da demanda de ponta iguais a zero significam que não foram encontradas informações desse atributo para essas empresas. Por outro lado, dados de perdas não técnicas iguais a zero indicam que essas empresas não possuem patamar significativo apurado desse atributo.

Essas empresas foram mantidas propositalmente no presente trabalho com vistas a avaliar a influência dessas lacunas de informações nos algoritmos analisados.

III. RESULTADOS

A. Avaliação dos algoritmos de clusterização

As Tabelas 3, 4 e 5 apresentam um extrato dos resultados da clusterização das concessionárias de distribuição com base nos atributos anteriormente discutidos. Foram empregados os algoritmos de média das distâncias, da distância dos centroides e de Ward.

TABELA 3. RESULTADOS DO PROCESSO DE CLUSTERIZAÇÃO UTILIZANDO O ALGORITMO DE MÉDIA DAS DISTÂNCIAS

Nº de Clusters	Clusters Agrupados		SR²	R²
17	CL19	OB18	0.0002	1.00
16	OB7	OB17	0.0002	1.00
15	OB15	OB19	0.0002	.999
14	OB3	CL16	0.0006	.999
13	CL15	CL17	0.0019	.997
12	CL14	CL18	0.0022	.995
11	OB1	OB8	0.0016	.993
10	CL12	OB4	0.0023	.991
9	OB12	CL13	0.0031	.988
8	CL11	OB2	0.0023	.985
7	OB6	OB13	0.0028	.982
6	CL8	CL7	0.0136	.969
5	CL10	CL9	0.0643	.905
4	CL6	CL5	0.1873	.717
3	OB10	OB11	0.0550	.662
2	CL4	OB5	0.1528	.509
1	CL2	CL3	0.5095	.000

TABELA 4. RESULTADOS DO PROCESSO DE CLUSTERIZAÇÃO UTILIZANDO O ALGORITMO DA DISTÂNCIA DOS CENTROIDES

Nº de Clusters	Clusters Agrupados		SR ²	R ²
17	CL19	OB18	0.0002	1.00
16	OB7	OB17	0.0002	1.00
15	OB15	OB19	0.0002	.999
14	OB3	CL16	0.0006	.999
13	CL15	CL17	0.0019	.997
12	CL14	CL18	0.0022	.995
11	CL12	OB4	0.0023	.992
10	OB1	OB8	0.0016	.991
9	CL10	OB2	0.0023	.988
8	OB12	CL13	0.0031	.985
7	OB6	OB13	0.0028	.982
6	CL9	CL7	0.0136	.969
5	CL11	CL8	0.0643	.905
4	CL6	CL5	0.1873	.717
3	OB10	OB11	0.0550	.662
2	CL4	OB5	0.1528	.509
1	CL2	CL3	0.5095	.000

TABELA 5. RESULTADOS DO PROCESSO DE CLUSTERIZAÇÃO UTILIZANDO O ALGORITMO DE WARD

Nº de Clusters	Clusters Agrupados		SR ²	R ²
17	CL19	OB18	0.0002	1.00
16	OB7	OB17	0.0002	1.00
15	OB15	OB19	0.0002	.999
14	OB3	CL16	0.0006	.999
13	OB1	OB8	0.0016	.997
12	CL14	OB4	0.0018	.995
11	CL15	CL17	0.0019	.993
10	CL13	OB2	0.0023	.991
9	CL12	CL18	0.0027	.988
8	OB6	OB13	0.0028	.986
7	OB12	CL11	0.0031	.982
6	CL10	CL8	0.0136	.969
5	CL6	OB10	0.0519	.917
4	CL9	CL7	0.0643	.853
3	CL5	OB5	0.0955	.757
2	CL3	OB11	0.2249	.532
1	CL2	CL4	0.5323	.000

Com base no parâmetro R² das Tabelas 3, 4 e 5 é possível perceber que para os três algoritmos, a formação de até 16 *clusters* não implica na perda de homogeneidade nos grupos. Pode-se também observar que os algoritmos se assemelham no comportamento dos agrupamentos até a formação de 5 *clusters*.

Por outro lado, na formação de 4 *clusters*, a redução do parâmetro R² dos dois primeiros métodos foi acima de 20%, enquanto que para o método *Ward* essa redução não chegou a 7%, indicando maior homogeneidade nos grupos com esse algoritmo. Com isso, nota-se quando da formação de 4 grupos, que o algoritmo de distância das médias e distância dos centroides reduzem de forma significativa o parâmetro R², ao passo que o algoritmo de *Ward* apresenta uma menor redução desse parâmetro.

Na formação de 3 *clusters*, o parâmetro R² permaneceu menor no resultado com *Ward* do que nos outros dois algoritmos.

Analisando-se ainda o parâmetro R² das Tabelas 3, 4 e 5, percebe-se que o algoritmo de *Ward* resulta em maior homogeneidade até a formação de 2 *clusters*.

Tendo como referência o parâmetro SR² das Tabelas 3, 4 e 5, que indica a perda de homogeneidade dentro dos grupos, é possível identificar que a formação de 4 *clusters* nos dois

primeiros algoritmos resultou em um valor maior do que o do algoritmo de *Ward*.

Portanto, considerando que o processo de clusterização objetiva formar grupos internamente homogêneos, é possível concluir que a adoção do algoritmo de *Ward* exibe o melhor desempenho dentre os três em avaliação.

B. Definição da quantidade de clusters

A Figura 1 apresenta a árvore de agrupamento das 20 concessionárias em análise.

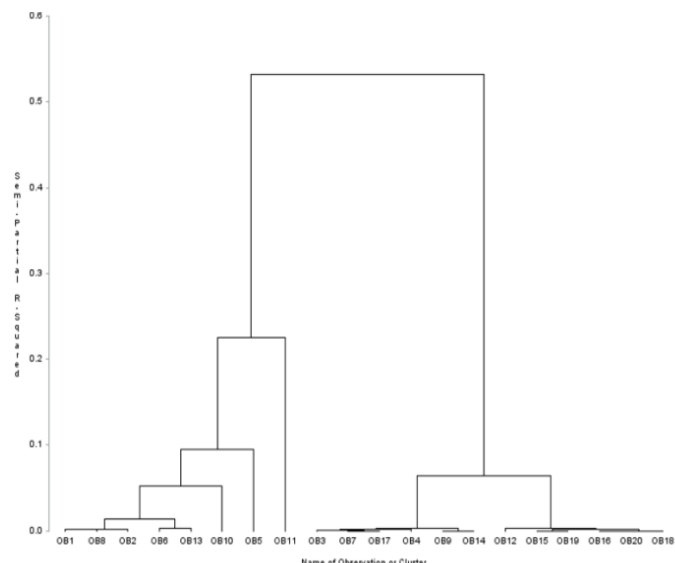


Figura 1. Dendrograma dos agrupamentos.

Percebe-se na Figura 1 que a formação de 1 *cluster*, representada pela linha horizontal mais alta do gráfico, que agrupa os objetos da esquerda até o número 11, aos da direita iniciados pelo número 3, apresenta o máximo de heterogeneidade dentro dos *clusters*, uma vez que o parâmetro SR² atinge seu maior valor, acima de 0,5.

Por outro lado, nos agrupamentos de 20 a 3 *clusters*, o valor de SR² permanece abaixo de 0,1, indicando uma pequena perda de homogeneidade.

A Figura 2 apresenta o gráfico do parâmetro SR² em relação aos agrupamentos dos objetos e diminuição da quantidade de *clusters*.

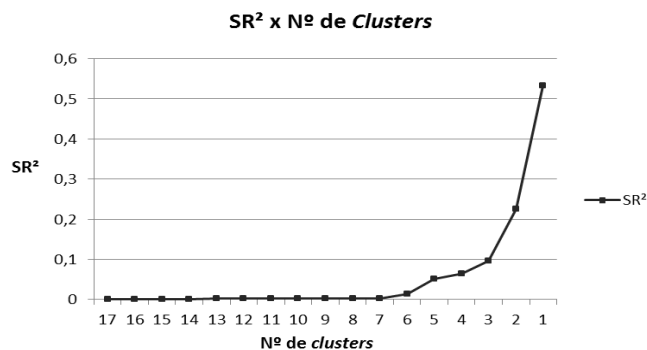


Figura 2. Gráfico da evolução do parâmetro SR² em relação às etapas de agrupamento.

Percebe-se da análise da Figura 2 que o primeiro crescimento significativo no parâmetro SR^2 ocorreu no agrupamento de 7 para 6 *clusters*. Contudo, esse parâmetro permaneceu abaixo de 0,1 até a formação de 3 *clusters*. Na composição de 2 grupos, esse parâmetro salta para o patamar acima de 0,1, indicando forte perda de homogeneidade em relação aos agrupamentos antecedentes.

Assim, conclui-se que na presente análise, considerando o algoritmo de Ward, o agrupamento em número não menor do que 3 *clusters* resulta em grande homogeneidade nos grupos. Objetivando-se diminuir a complexidade da análise de custo-benefício da implantação das REI, recomenda-se o menor número de grupos possível de grupos sem grande perda de homogeneidade. Por esta razão, será considerada neste trabalho a classificação das empresas em 3 *clusters*.

C. Clusters formados Formação de Clusters

A Tabela 6 apresenta os 3 grupos formados com base nos atributos analisados e no algoritmo de Ward.

TABELA 6. CLUSTERS FORMADOS.

CLUSTER 1	CLUSTER 2	CLUSTER 3
ELETROPAULO	ELETROCAR	COPEL
CEMIG		CPFL PAULISTA
LIGHT		CELESC
COELBA		CELG
ELEKTRO		DEMEI
CELPE		CHESP
HIDROPAN		CERR
		EFLUL
		UHENPAL
		MUX ENERGIA
		FORCEL
		EFLJC

Percebe-se na Tabela 6 que a concessionária Eletrocar ficou agrupada isoladamente, o que indica que, com base nos atributos analisados, ela apresenta a menor proximidade em relação às demais. As outras empresas foram classificadas em grupos com mais de 6 objetos, o que indica que há proximidade entre elas.

Analisando-se o *cluster* 1, percebe-se que nele há uma maior presença de empresas de grande porte, apesar da existência de uma concessionária pequena, a saber, a Hidropan. Por outro lado, percebe-se na diversidade de empresas do *cluster* 3, que o porte das distribuidoras não influenciou de maneira preponderante o processo de clusterização.

Assim, percebe-se que o tamanho do mercado das empresas não é atributo preponderante para classificação para a análise da implantação das REI. Considerando-se que os *clusters* 1 e 3 possuem distribuidoras de diferentes regiões do país, percebe-se que esse também não é um atributo suficiente para o agrupamento das concessionárias.

Adicionalmente, ao se focar a influência da falta de informações em algumas empresas, nota-se que as distribuidoras que apresentaram lacuna de informações ou dados iguais a zero foram agrupadas no mesmo conjunto, o que indica que este aspecto pode ter afetado o resultado.

IV. CONCLUSÕES

Este trabalho avaliou os algoritmos de clusterização de concessionárias de distribuição de energia elétrica, a saber, de média das distâncias, da distância dos centroides e de Ward com base em três atributos sensíveis para análise de custo-benefício da implantação das REI no Brasil.

Os resultados permitem concluir que para a clusterização das 10 maiores e das 10 menores distribuidoras, o algoritmo de Ward apresenta os grupos mais homogêneos dentre os três analisados.

Também é possível concluir que a classificação das empresas em quantidade de grupos maior ou igual a três não acarreta grandes perdas de homogeneidade.

Os resultados também apontam que a região de atuação e o porte das concessionárias não são fatores preponderantes para o agrupamento analisado, uma vez que dois *clusters* foram compostos por empresas de diferentes regiões e tamanhos.

Além disso, a falta de informações de algumas concessionárias pode ser fator importante na composição dos agrupamentos caso se conduza um processo de clusterização nacional.

Assim, ao se utilizar o algoritmo de Ward e a composição em 3 *clusters*, há uma redução da complexidade da análise de custo-benefício da implantação das REI, dado que pode-se conduzir 3 análises em substituição de 20 individuais. Esse resultado evidencia a importância deste trabalho para estudos da aplicação dessas redes no Brasil.

V. REFERÊNCIAS

- [1] Ministério de Ciência, Tecnologia e Inovação, “Redes Elétricas Inteligentes: Contexto Nacional”, Brasília, 2011.
- [2] Notas Técnicas do 3º Ciclo de Revisões Tarifárias Periódicas, ANEEL, 2014.
- [3] Sítio eletrônico da ANEEL, Sistema de Apoio à Decisão. Disponível em: < <http://www.aneel.gov.br/area.cfm?idArea=550>>. Acesso em 20 de dezembro de 2015.
- [4] Sítio eletrônico da ANEEL, Indicadores Coletivos de Continuidade. Disponível em: <http://www.aneel.gov.br/area.cfm?id_area=80>. Acesso em 20 de dezembro de 2015.
- [5] Sítio eletrônico da ANEEL, Compensação de Continuidade. Disponível em: <http://www.aneel.gov.br/aplicacoes/Compensacao_de_Continuidade_Conformidade_v2/>. Acesso em 20 de dezembro de 2015.
- [6] R. Xu and D. C. Wunsch II, “Clustering”, New Jersey: Wiley, 2009, pp. 11-46.
- [7] L. Kaufman and P. J. Rousseeuw, “Finding Groups in Data, an introduction to cluster analysis”, New Jersey: Wiley, 2005, pp. 3-49.
- [8] N. H. Timm, “Applied Multivariate Analysis”, New York: Springer, 2002, pp. 515-530.
- [9] H. Lamin, “Análise de Impacto Regulatório da implantação de redes inteligentes no Brasil”, Doutor, Tese de Doutorado, Faculdade de Tecnologia, Departamento de Engenharia Elétrica, Universidade de Brasília, 2013.
- [10] N. Kagan, M. Gouvea, F. Maia, D. Duarte, J. Labronici, D. S. Guimarães, A. B. Neto, J. F. R. da Silva, F. Particelli, “Análise de custos e benefícios da implantação de um plano nacional de implantação”, Rio de Janeiro: Synergia, 2013.
- [11] Ministério de Minas e Energia, “Relatório do Grupo de Trabalho de Redes Elétricas Inteligentes”, Brasília, 2010.