



Modulação de Comunidades Locais de Energia Renovável

JOÃO MANUEL LOPES DA COSTA

novembro de 2022

Modelação de Comunidades Locais de Energia Renovável

João Manuel Lopes da Costa

Departamento de Engenharia Eletrotécnica

Mestrado em Engenharia Eletrotécnica e de Computadores

Sistemas e Planeamento Industrial

Relatório elaborado para satisfação parcial dos requisitos da Unidade Curricular de
Tese/Dissertação do Mestrado em Engenharia Eletrotécnica e de Computadores

Candidato: João Manuel Lopes da Costa, Nº 1170665, 1170665@isep.ipp.pt

Orientação científica: Bruno Canizes, brm@isep.ipp.pt | Zita Vale, zav@isep.ipp.pt

GECAD – Grupo de Investigação em Engenharia e Computação Inteligente para a

Inovação e o Desenvolvimento

Empresa: All Fix, Lda.

Supervisão: Sérgio Rocha, sergio.rocha@globalunica.pt



Departamento de Engenharia Eletrotécnica

Mestrado em Engenharia Eletrotécnica e de Computadores

Sistemas e Planeamento Industrial

2022

Agradecimentos

Gostaria de agradecer a todas as pessoas que, de forma direta ou indireta, fizeram parte do meu percurso académico e que me apoiaram na realização da presente dissertação.

Em primeiro lugar, agradeço a toda a minha família. Aos meus pais, avós e irmã, pelas palavras de incentivo e todo o esforço que fizeram para que eu estivesse onde estou hoje. Agradeço à minha namorada, por toda a paciência e por todo o apoio, principalmente nos momentos mais difíceis.

Aos meus amigos e colegas de curso, por todos os momentos de amizade que passamos juntos e por tudo o que aprendi do vosso lado.

Ao meu supervisor Dr. Sérgio Rocha que me acolheu na sua empresa e me permitiu participar na implementação de uma Comunidade de Energia Renovável, para além de me facultar vários dados que foram indispensáveis para a elaboração desta dissertação.

Agradeço aos meus orientadores, Dr. Bruno Canizes e Dra. Zita Vale, pelo voto de confiança prestado, pela paciência, ajuda e disponibilidade e ao Grupo de Investigação em Engenharia e Computação Inteligente para a Inovação e o Desenvolvimento (GECAD).

Resumo

A transição do setor elétrico atual para um novo modelo é inevitável. O futuro que se aproxima promete novos projetos e desafios, para todos os que se encontram envolvidos nesta cadeia de valor. A evolução tecnológica veio permitir o surgimento de novos conceitos, como é o caso das comunidades de energia renovável, a produção energética renovável descentralizada e até o armazenamento da mesma, são alguns dos fatores que possibilitam ao consumidor ter um papel mais ativo no setor elétrico e a contribuir consideravelmente para o alcance dos objetivos ambientais.

Com a introdução das Comunidades de Energia Renovável, surgiu a necessidade de desenvolver ferramentas de gestão e otimização das mesmas, principalmente na gestão da energia produzida e da carga, levando sempre em consideração as várias regras de implementação. Desta forma, a presente dissertação apresenta inicialmente uma revisão bibliográfica sobre o conceito de comunidade de energia renovável e sobre as diferentes técnicas de agrupamento que poderão ser aplicadas neste contexto. Além disso, é desenvolvida uma metodologia capaz de agrupar os diferentes membros em comunidades, de acordo com as suas características energéticas e geográficas.

Com o objetivo de testar e validar o trabalho desenvolvido foram criados vários dados de entrada, e recolhidos uma série de dados provenientes de uma comunidade de energia renovável real. Este estudo permitiu verificar qual o comportamento e adaptação de cada algoritmo de agrupamento no processo de formação das várias Comunidades de Energia Renovável. A análise dos resultados, demonstrou a adequação, versatilidade e robustez da metodologia desenvolvida, evidenciando a utilidade da mesma na tomada de decisão ao nível do dimensionamento das Comunidades de Energia Renovável.

Palavras-Chave

Algoritmos de Agrupamento, Comunidades de Energia Renovável, Evolução Tecnológica, Otimização, Produção Descentralizada

Abstract

The transition of the current electricity sector to a new model is inevitable. The coming future promises new projects and challenges, for all those involved in this value chain. Technological developments have enabled new concepts to emerge, such as renewable energy communities, decentralized renewable energy production and even storage, are some of the factors that enable the consumer to play a more active role in the electricity sector and contribute considerably to achieving environmental objectives.

With the introduction of Renewable Energy Communities, the need arose to develop management and optimization tools for them, mainly in the management of the energy produced and the load, always taking into account the various implementation rules. In this way, this dissertation presents initially a literature review on the concept of renewable energy community and on the different clustering techniques that could be applied in this context. Moreover, a platform capable of grouping different members into communities according to their energy and geographical characteristics is developed.

In order to test and validate the work developed, several input data were created, and a series of data from a real renewable energy community was collected. This study allowed to verify the behavior and adaptation of each clustering algorithm in the process of formation of the various Renewable Energy Communities. The analysis of the results demonstrated the suitability, versatility and robustness of the methodology developed, highlighting its usefulness in decision making at the Renewable Energy Communities sizing level.

Keywords

Clustering Algorithms, Decentralized Renewable Energy Production, Optimization, Renewable Energy Communities, Technological Developments

Índice

AGRADECIMENTOS	I
RESUMO.....	III
ABSTRACT.....	V
ÍNDICE DE FIGURAS.....	IX
ÍNDICE DE TABELAS.....	XI
ACRÓNIMOS	XII
NOMENCLATURAS.....	XIII
1. INTRODUÇÃO	1
1.1.CONTEXTUALIZAÇÃO.....	1
1.2.MOTIVAÇÃO	3
1.3.OBJETIVOS.....	4
1.4.CALENDARIZAÇÃO	4
1.5.ORGANIZAÇÃO DO RELATÓRIO	5
2. COMUNIDADES DE ENERGIA RENOVÁVEL.....	7
2.1.LEGISLAÇÃO E REGULAMENTAÇÃO.....	9
2.1.1. <i>Autoconsumo</i>	10
2.1.1.1. Autoconsumo Individual.....	10
2.1.1.2. Autoconsumo Coletivo.....	11
2.1.1.3. Comunidades de energia renovável.....	12
2.2. <i>PROSUMERS</i>	15
2.3.MODELOS DE REPARTIÇÃO	16
2.3.1. <i>Repartição Baseada no Consumo</i>	16
2.3.2. <i>Repartição Baseada em Coeficientes</i>	17
2.3.3. <i>Repartição Híbrida</i>	18
2.4.TIPOS DE COMUNIDADES DE ENERGIA RENOVÁVEL	19
2.4.1. <i>Comunidades de Energia Homogénea</i>	20
2.4.2. <i>Comunidades de Energia Mista</i>	20
2.4.3. <i>Comunidade de Energia Autossuficiente</i>	20
2.5.ASPETOS CONCLUSIVOS	21
3. TÉCNICAS E ALGORITMOS DE AGRUPAMENTO	22
3.1.TÉCNICAS DE AGRUPAMENTO.....	23
3.1.1. <i>Agrupamento Particional</i>	23
3.1.2. <i>Agrupamento Hierárquico</i>	24
3.1.3. <i>Agrupamento Baseado em Densidade</i>	26

3.2. ALGORITMOS DE AGRUPAMENTO ESPACIAL	27
3.2.1. <i>K-MEANS</i>	27
3.2.2. <i>DBSCAN</i>	28
3.2.3. <i>Linkage Algorithms</i>	31
3.2.4. <i>Comparação entre os Algoritmos de Agrupamento</i>	33
3.3. ASPETOS CONCLUSIVOS.....	35
4. METODOLOGIA DE AGRUPAMENTO E CLASSIFICAÇÃO.....	36
4.1. IDENTIFICAÇÃO E CLASSIFICAÇÃO DAS COMUNIDADES DE ENERGIA RENOVÁVEL	36
4.1.1. <i>Comunidades de Energia Homogéneas</i>	37
4.1.2. <i>Comunidades de Energia Mista</i>	38
4.1.3. <i>Comunidades de Energia Autossuficientes</i>	38
4.2. MODELO DESENVOLVIDO	39
4.2.1. <i>Estrutura do Modelo</i>	40
4.3. ASPETOS CONCLUSIVOS.....	47
5. CASO DE ESTUDO	48
5.1. CARACTERIZAÇÃO E ANÁLISE DOS DADOS	48
5.2. COMUNIDADE DE ENERGIA RENOVÁVEL - NORTE DE PORTUGAL.....	52
5.3. SIMULAÇÕES E RESULTADOS	56
5.3.1. <i>DBSCAN</i>	58
5.3.2. <i>K-MEANS</i>	59
5.3.3. <i>Linkage Algorithms</i>	60
5.4. MÉTRICAS DE AVALIAÇÃO DE AGRUPAMENTO	62
5.5. ASPETOS CONCLUSIVOS.....	64
6. CONCLUSÕES E TRABALHOS FUTUROS	65
REFERÊNCIAS.....	69
ANEXOS.....	75
ANEXO A- FLUXOGRAMAS DAS FUNÇÕES UTILIZADAS	75
ANEXO B- BASE DE DADOS	78

Índice de Figuras

Figura 1 - Calendarização da Dissertação	5
Figura 2 - Exemplo ilustrativo do Autoconsumo Individual [20]	11
Figura 3 - Exemplos ilustrativos do Autoconsumo Coletivo [20]	12
Figura 4 - Exemplos ilustrativos de duas tipologias de CER [20]	13
Figura 5 – Conceito energético de Prosumer	15
Figura 6 - Esquema demonstrativo do Agrupamento Divisivo e Aglomerativo [36]	24
Figura 7 - Pseudocódigo do Algoritmo K-Means [39]	28
Figura 8 - Pseudocódigo do Algoritmo DBSCAN [44]	30
Figura 9 - Demonstração do cálculo da distância entre clusters nos Linkage Algorithms [52]	32
Figura 10 - Pseudocódigo dos Linkage Algorithms [46]	33
Figura 11 - Comparação do valor $E(t)$ nos diferentes tipos de Comunidades de Energia [53]	39
Figura 12 - Fluxograma do Funcionamento Geral do Modelo	41
Figura 13 - Pseudocódigo da Função Inicial	42
Figura 14 - Pseudocódigo da Função Importar Ficheiro	43
Figura 15 - Pseudocódigo da Função Conversão de dados	44
Figura 16 - Esquema de possíveis classificações para cada CER	45
Figura 17 - Pseudocódigo da Função Classificação	46
Figura 18 - Localização geográfica dos membros em estudo (Google Earth Pro)	50
Figura 19 - Gráfico da distribuição geográfica dos membros por região	50
Figura 20 - a) Valores de Produção Anual dos membros b) Valores de Consumo Anual dos membros	51
Figura 21 - Gráfico comparativo dos valores de $E(t)$	52
Figura 22 – Gráfico da Produção de Energia Estimada dos Centros Eletroprodutores	56
Figura 23 - Gráfico do Consumo e Produção de Energia Estimados da CER	56
Figura 24 - Gráfico comparativo do valor de energia líquida dos membros	57
Figura 25 - Gráfico e tabela da distribuição e classificação das CER (DBSCAN)	58
Figura 26 - Gráfico e tabela da distribuição e classificação das CER (K-MEANS)	59
Figura 27 - Gráfico e tabela da distribuição e classificação das CER (Ward's Method)	60
Figura 28 - Gráfico e tabela da distribuição e classificação das CER (Complete-Link)	61
Figura 29 - Gráfico e tabela da distribuição e classificação das CER (Average-Link)	61
Figura 30 - Gráfico e tabela da distribuição e classificação das CER (Single-Link)	62

Índice de Tabelas

<i>Tabela 1 - Condições para o exercício da atividade de produção de energia renovável [16]</i>	14
<i>Tabela 2 - Comparação entre os Modelos de Repartição baseados em coeficientes e no consumo.</i>	17
<i>Tabela 3 - Pontos Fortes e Pontos Fracos das técnicas de Agrupamento Particional</i>	24
<i>Tabela 4 - Pontos Fortes e Pontos Fracos das técnicas de Agrupamento Hierárquico.</i>	25
<i>Tabela 5 - Pontos Fortes e Pontos Fracos das técnicas de Agrupamento Baseado em Densidade.</i>	26
<i>Tabela 6 – Comparação entre os Algoritmos K-MEANS, DBSCAN e LINKAGE ALGORITHMS</i>	34
<i>Tabela 7 - Variáveis usadas no modelo desenvolvido</i>	43
<i>Tabela 8 - Características das instalações elétricas de cada consumidor</i>	54
<i>Tabela 9 - Características das instalações elétricas de cada produtor</i>	55
<i>Tabela 10 - Avaliação do desempenho dos algoritmos de agrupamento</i>	64

Acrónimos

AT - Alta Tensão

BT - Baixa Tensão

CER - Comunidade de Energia Renovável

DBSCAN - *Density-Based Spatial Clustering of Applications with Noise*

HEC - Comunidade de Energia Homogénea

IU – Instalação de Utilização

MAT - Muito Alta Tensão

MCP - Mera Comunicação Prévia

MEC - Comunidade de Energia Mista

MT - Média Tensão

OPTICS - *Ordering Points To Identify the Clustering Structure*

RESP - Rede Elétrica de Serviço Público

RND - Rede Nacional de Distribuição

RNT - Rede Nacional de Transporte

SEC - Comunidade de Energia Autossuficiente

SEE - Sistema Elétrico de Energia

UPAC - Unidade de Produção para Autoconsumo

Nomenclaturas

ER_{IU_i} - Energia repartida na IU_i

EC_{IU_i} - Energia consumida na IU_i

$\sum_j E_{UPACj}$ – Total de energia produzida pelas UPAC

F_{IU_i} - Fator de repartição na IU_i

i - Índice da IU

j - Índice da Unidade de Produção

$E(t)$ - Energia Líquida

1. INTRODUÇÃO

Neste capítulo, é realizada uma exposição introdutória do tema da dissertação, contextualizando-o, bem como as motivações que levam à necessidade da criação de novas ferramentas de otimização e gestão das energias renováveis. São igualmente traçados os objetivos desta dissertação, a sua calendarização e estrutura.

1.1. CONTEXTUALIZAÇÃO

O desenvolvimento das fontes de energia renováveis e distribuídas, como eólica e solar, dispersas ao longo da rede de distribuição estão a ganhar grande importância e a contribuir para os objetivos políticos e ambientais. No entanto, a variabilidade e a intermitência

dessas fontes de energia colocam novos desafios à gestão e planeamento da rede elétrica. Atualmente, a procura global por energia elétrica está a aumentar, a transição para os veículos elétricos está a crescer a cada ano, e como consequência, espera-se um aumento considerável do consumo energético na rede elétrica, embora a mesma seja alimentada maioritariamente por fontes de energia não renováveis, pelo que existe inevitavelmente a necessidade de minimizar a pegada de carbono, implementando sistemas de produção e armazenamento de energia renovável em larga escala [1]–[3].

As comunidades de energia renovável (CER) nos últimos anos têm sido alvo de bastante interesse tanto a nível industrial, como a nível investigacional, uma vez que são os pilares fundamentais para a construção de uma rede energética inteligente [4]. As expectativas são elevadas uma vez que poderá trazer soluções para problemas que enfrentamos atualmente como a elevada procura energética e a gestão das cargas [5], [6]. Além disso, alguns modelos cooperativos entre comunidades de energia renovável distribuídas foram alvo de investigação em várias aplicações, como por exemplo, na otimização de mecanismos contra a perda de energia, substituindo as unidades centrais de gestão de energia por dispositivos inteligentes [7] e na gestão de carga, possibilitada pela troca e partilha de energia produzida localmente [8]–[10].

O desenvolvimento científico com o intuito de estudar a implementação e integração das CER na rede elétrica é de extrema importância, uma vez que estas prometem ser um fator fundamental para o aumento da eficiência da rede elétrica [11]. *Hong et al.*, propõem abordagens para a identificação de comunidades de energia e interação entre as mesmas dentro de uma rede elétrica inteligente, onde a produção e partilha de energia predomina ao longo da rede elétrica [12]. No entanto, verificaram-se algumas limitações relativamente à variedade de algoritmos de agrupamento em estudo, assim como a inexistência de uma análise e comparação entre os mesmos, de modo a tentar identificar quais os mais eficientes na realização dos agrupamentos. Tendo em consideração a inexistência desses mesmos estudos, que analisem e compararem o comportamento e eficiência de cada tipo de algoritmo no agrupamento de CER, seria muito importante identificar quais os algoritmos que demonstram melhores resultados neste processo.

Os problemas de descoberta comunitária, tendem a agrupar dados que partilham características semelhantes ou se encontram próximos uns dos outros, como por exemplo, analisando conjuntos de dados geográficos para a identificação de comunidades locais [13], ou detetando grupos de indivíduos com interesses semelhantes dentro da rede [14]. No entanto, os problemas de descoberta de comunidades de energia são consideravelmente diferentes, pelo que não podem ser comparados com os anteriormente mencionados. A principal diferença reside no critério de agrupamento, pois na análise de dois ou mais membros dentro de uma comunidade de energia renovável deve-se ter em consideração não só as distâncias geográficas na rede elétrica, mas também a energia líquida de cada um no espaço temporal [5].

1.2. MOTIVAÇÃO

Novos desafios são gerados à medida que o setor elétrico evolui, o que conseqüentemente obriga ao desenvolvimento de novos projetos, não só na forma de como a energia é gerada, mas também como esta é distribuída. O principal objetivo passa por aumentar a eficiência e a qualidade de serviço da rede de distribuição, tirando o máximo partido da produção e consumo de energias renováveis localmente. As comunidades de energia renovável prometem trazer várias vantagens para este setor e por tal razão, necessitam de uma especial atenção, principalmente nesta fase ainda tão embrionária. Surge então a necessidade de desenvolver ferramentas eficientes, implementadas em softwares que reduzam o esforço na busca pela melhor solução, considerando sempre todas as alternativas possíveis.

1.3. OBJETIVOS

Nesta dissertação, propõe-se o desenvolvimento de uma metodologia de agrupamento e classificação, aplicável na formação de comunidades de energia renovável. Deste modo, pretende-se facilitar a fase de projeto de cada CER, tornando o processo replicável em ambientes com características distintas. Pretende-se, assim, atingir os seguintes objetivos:

- Utilizar, adaptar e aprimorar modelos para identificar os tipos de comunidades locais de energia renovável;
- Calcular a energia líquida das CER, resultante da diferença entre a geração de energia local e a carga;
- Selecionar e adaptar métodos adequados para *clustering*;
- Desenvolver um modelo onde serão apresentadas e classificadas as comunidades locais de energia renovável partindo da introdução de dados ou importação dos mesmos.

1.4. CALENDARIZAÇÃO

O tema abordado nesta dissertação relativo à “Modelação de Comunidade de Energia Renovável” teve uma duração aproximada de 9 meses, sendo o mesmo realizado no âmbito da Unidade Curricular de Tese/Dissertação do Mestrado em Engenharia Eletrotécnica e de Computadores. A calendarização da dissertação encontra-se representada na figura 1, ao longo deste período foi efetuado um estudo aprofundado sobre as comunidades de energia renovável e as várias técnicas de agrupamento, o que resultou no desenvolvimento de uma ferramenta capaz de agrupar vários membros em comunidades, de acordo com as suas características energéticas e geográficas.

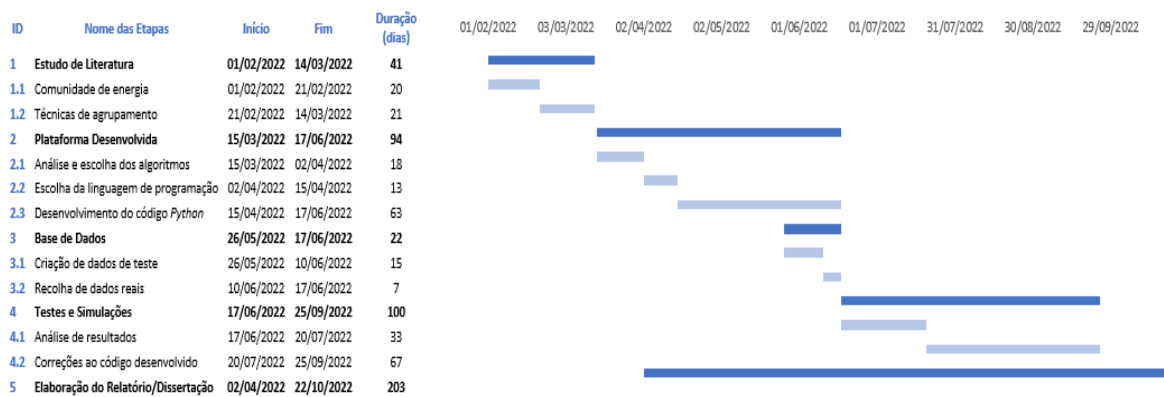


Figura 1 - Calendarização da Dissertação

1.5. ORGANIZAÇÃO DO RELATÓRIO

A dissertação está organizada em 7 capítulos, que serão abordados de seguida.

Capítulo 1: apresentação de uma contextualização ao tema abordado, motivação, objetivos principais a atingir, a calendarização do trabalho desenvolvido e a organização da dissertação.

Capítulo 2: apresentação do tema “Comunidade de Energia Renovável”, onde é abordada a legislação e regulamentação das diferentes modalidades de autoconsumo, os *Prosumers*¹, os diferentes modelos de repartição e os tipos de comunidade de energia renovável.

Capítulo 3: apresentação do tema “Técnicas e Algoritmos de Agrupamento”, onde são abordadas as principais técnicas de agrupamento e os algoritmos que serão utilizados para o desenvolvimento do modelo de agrupamento.

Capítulo 4: elaboração da metodologia utilizada para a identificação e classificação das comunidades de energia renovável. Abordagem ao modelo desenvolvido, assim como ao funcionamento e estrutura do mesmo.

¹ *Prosumer* na sua terminologia anglo-saxónica, representa um consumidor energético que produz e partilha a sua própria energia, gerada localmente, com os outros utilizadores da rede elétrica.

Capítulo 5: análise e caracterização dos dados constituintes do caso de estudo, assim como demonstração das simulações e resultados obtidos através do modelo desenvolvido.

Capítulo 6: principais conclusões retiradas desta dissertação e propostas para trabalhos futuros.

2. COMUNIDADES DE ENERGIA RENOVÁVEL

O modelo de produção centralizado de energia elétrica tem vindo a predominar até à atualidade, no entanto a introdução e o desenvolvimento de fontes de energia renovável estão a permitir a transição para um sistema mais sustentável e distribuído. O principal objetivo passa por atribuir aos consumidores um papel mais ativo, no que diz respeito à produção de energia renovável, contribuindo assim para o cumprimento dos objetivos ambientais.

De acordo com o Plano Nacional de Energia-Clima 2021-2030, Portugal tem como objetivo entre 2021 e 2030 alcançar uma quota de 47% de energia consumida resultante de fontes renováveis. Para que este objetivo seja possível de alcançar, serão necessários fortes investimentos no setor, uma vez que pelo menos 80% da produção total terá de ser fornecida por fontes renováveis [15].

O regime jurídico resultante do Decreto-Lei Nº 162/2019 de 25 de outubro, vem desenvolver o conceito já existente de autoconsumo individual, assim como introduzir novos conceitos, nomeadamente as comunidades de energia renovável e o autoconsumo coletivo [16]. Já o Decreto-lei Nº 15/2022 de 14 de janeiro foi desenvolvido de forma a consolidar o conceito de CER eliminando assim algumas incertezas que ainda persistiam [17].

O autoconsumo individual de energia tem aumentado bastante nos últimos anos, e esta nova legislação veio permitir a criação de contratos bilaterais entre consumidores e o desenvolvimento de comunidades de energia renovável, com o principal objetivo de possibilitar a partilha dos excedentes de energia renovável produzida em cada instalação, fazendo assim um melhor aproveitamento da energia produzida em excesso. Devido a estas alterações, tudo indica que a descentralização da produção energética irá aumentar consideravelmente, uma vez que o consumidor poderá produzir, partilhar e até armazenar a sua própria energia elétrica.

O desejo de intensificar a utilização de fontes renováveis e de alcançar o objetivo da neutralidade carbónica deu origem ao desenvolvimento das comunidades de energia renovável. Uma CER pode ser constituída por membros de vários tipos: empresas públicas ou privadas, consumidores domésticos, comércio, autarquias, indústrias entre outros. Esta deve ser gerida pelos seus participantes, que se devem localizar nas proximidades dos projetos de energia renovável desenvolvidos pela mesma [16].

O estudo da viabilidade das CER é de elevada importância principalmente quando se trata da sua implementação em núcleos rurais mais remotos, onde muitas das vezes, os consumidores têm bastante dificuldade em adquirir energia elétrica. Todavia, a implementação destes projetos é condicionada por vários aspetos, sejam eles técnicos ou até mesmo ambientais e geográficos, como por exemplo a radiação solar e as características do terreno, pelo que se justifica o estudo dos mesmos de modo a criar processos que possam ser utilizados no seu dimensionamento [18].

2.1. Legislação e Regulamentação

As comunidades de energia renovável surgiram a partir da Diretiva Europeia 2018/2001 [19], sendo que esta foi introduzida no regime jurídico português através do Decreto-Lei 162/2019 aprovado a 25 de outubro de 2019 e posteriormente consolidada pelo Decreto-lei Nº 15/2022 de 14 de janeiro. A diretiva teve como propósito a promoção da produção de energia renovável pelos estados europeus através do autoconsumo individual e coletivo, que já existiam até ao momento, e com a introdução de um novo conceito denominado de comunidades de energia renovável. A partir deste momento começou a ser possível a produção, consumo, armazenamento, partilha e venda de energia renovável.

As CER foram desenvolvidas com o objetivo de cativar os próprios consumidores de energia a serem parte ativa na produção da mesma, no que diz respeito à energia renovável. De acordo com o Decreto-Lei 162/2019, uma CER é “uma pessoa coletiva” constituída nos termos do presente Decreto-Lei, com ou sem fins lucrativos, com base numa adesão aberta e voluntária dos seus membros, sócios ou acionistas, os quais podem ser pessoas singulares ou coletivas, de natureza pública ou privada, incluindo, nomeadamente, pequenas e médias empresas ou autarquias locais, que seja autónoma dos seus membros ou sócios, mas por eles efetivamente controlada, desde que e cumulativamente: os membros ou participantes estejam localizados na proximidade dos projetos de energia renovável ou desenvolvam atividades relacionadas com os projetos de energia renovável da respetiva comunidade de energia; os referidos projetos sejam detidos e desenvolvidos pela referida pessoa coletiva; e a pessoa coletiva tenha por objetivo principal propiciar aos membros ou às localidades onde opera a comunidade benefícios ambientais, económicos e sociais em vez de lucros financeiros [16].

No entanto existiam ainda algumas incertezas no Decreto-Lei Nº 162/2019 relativamente ao tema, principalmente no que diz respeito à inexistência de critérios geográficos, elétricos e sociais que selecionem os membros de cada CER. O Decreto-lei Nº 15/2022 permitiu o esclarecimento de tais incertezas, pois foram definidas distâncias máximas permitidas entre as Unidades de Produção para Autoconsumo (UPAC) e as Instalações de

Utilização (IU), que constituem um requisito para o exercício da atividade de produção para autoconsumo. Segundo o presente decreto-lei, entende-se abrangidas pelo conceito de proximidade as UPAC e a(s) IU ligadas por linha direta ou rede interna ou, quando operem através da Rede Elétrica de Serviço Público (RESP) nos diferentes níveis de tensão, desde que cumpram uma das seguintes condições:

- Quando, no caso de UPAC ligadas às redes de distribuição de energia elétrica em Baixa Tensão (BT), a IU e a UPAC não distem entre si mais de 2 km de distância geográfica ou, em alternativa, estejam ligadas ao mesmo posto de transformação;
- Estejam ligadas na mesma subestação, no caso de UPAC ligadas à Rede Nacional de Distribuição (RND) e à Rede Nacional de Transporte (RNT), desde que não seja ultrapassada a distância geográfica entre as UPAC e as IU de 4 km no caso de ligação em Média Tensão (MT), de 10 km nas ligações em Alta Tensão (AT) e de 20 km nas ligações em Muito Alta Tensão (MAT) [17].

2.1.1. AUTOCONSUMO

Dentro do autoconsumo existem três categorias diferentes: o autoconsumo individual, o autoconsumo coletivo e as comunidades de energia renovável. Esta última categoria é a novidade inserida na legislação portuguesa, que por sua vez será o maior alvo de estudo nesta dissertação. Todas as categorias mencionadas têm como base a produção de energia proveniente de fontes renováveis, distinguindo-se entre elas principalmente pela maneira de como a energia gerada é consumida [16].

2.1.1.1. AUTOCONSUMO INDIVIDUAL

Quando a produção de energia se efetua no interior da instalação onde é consumida, denomina-se autoconsumo individual, ou seja, quando a UPAC está conectada apenas a uma única IU, podendo esta ser uma habitação, uma empresa, um espaço comercial entre

outros, como pode demonstrar a figura 2. O autoconsumo individual permite ao consumidor final não só produzir e consumir a energia produzida, mas também a possibilidade de a armazenar em baterias, ou vender através da injeção da mesma na RESP [16].

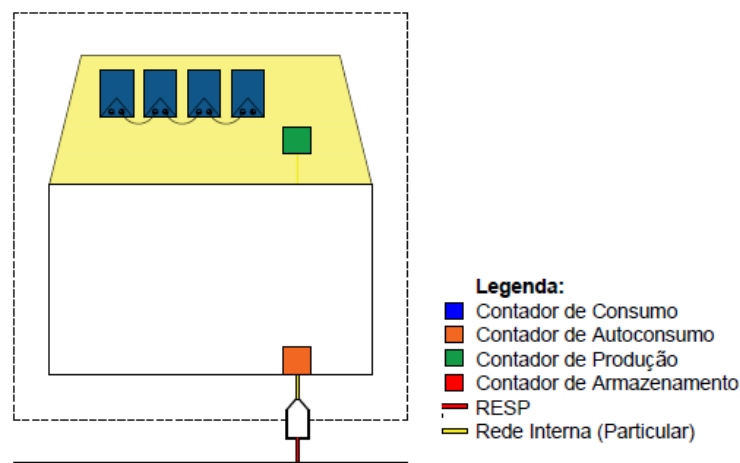


Figura 2 - Exemplo ilustrativo do Autoconsumo Individual [20]

2.1.1.2. AUTOCONSUMO COLETIVO

O autoconsumo coletivo é em parte semelhante ao autoconsumo individual, uma vez que ambos partem da produção de energia através de fontes renováveis, no entanto, este possibilita que várias IU possam se conectar a uma ou mais UPAC, quando se verifica uma proximidade geográfica e elétrica. Possíveis exemplos deste modelo são autoconsumidores que se encontram no mesmo edifício, zona de moradias ou apartamentos, unidades industriais, agrícolas entre outras, que se localizam geograficamente próximas entre si, e que têm acesso à energia produzida pela(s) UPAC(s), como pode demonstrar a figura 3.

A distribuição da energia pelos diversos consumidores poderá ser através da rede interna do edifício, ou através da RESP, sendo neste caso necessário o pagamento da respetiva taxa de utilização da rede.

Uma das grandes vantagens deste modelo comparativamente ao autoconsumo individual, é o facto de todos os consumidores envolvidos no projeto contribuírem para o pagamento do respetivo investimento e das manutenções necessárias ao longo do período de vida do projeto [19], para além de consumirem a energia renovável produzida localmente ao

máximo e conseqüentemente diminuirão a injeção na rede, uma vez que existirão vários de tipos de perfis de consumo associados.

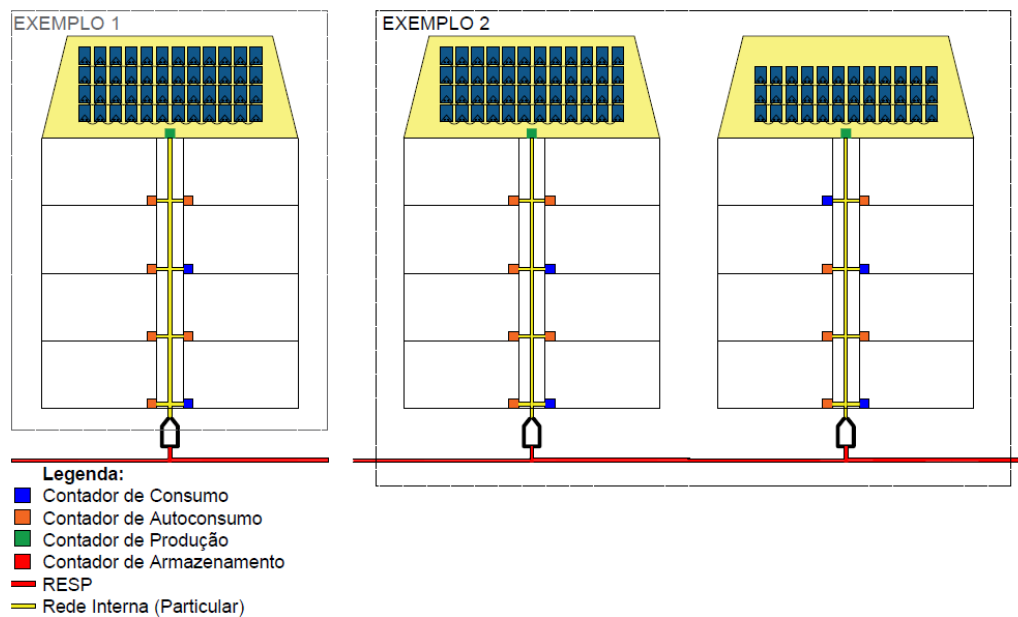


Figura 3 - Exemplos ilustrativos do Autoconsumo Coletivo [20]

2.1.1.3. COMUNIDADES DE ENERGIA RENOVÁVEL

Pode fazer parte deste projeto qualquer consumidor seja ele singular ou coletivo, de natureza privada ou pública. A possível integração de diversos tipos de consumidores nestas comunidades, permite aos intervenientes passarem a pertencer à cadeia do setor energético, tornando-se assim mais ativos na produção de energia elétrica proveniente de fontes renováveis. Para além disso, toda esta atividade resulta numa redução da utilização dos recursos do Sistema Elétrico de Energia (SEE), nomeadamente no uso das redes de transporte e distribuição de energia, que conseqüentemente contribui para a redução de perdas e congestionamentos de energia [19].

Existem algumas limitações relativamente à implementação de uma comunidade de energia renovável, uma delas é a necessidade de proximidade elétrica e geográfica entre os consumidores e a(s) UPAC(s) [16].

A introdução do conceito das CER teve como principal objetivo a produção e respetivo consumo de energia renovável, com custos comparativamente inferiores aos do mercado de energia, para além da contribuição para o cumprimento das metas ambientais. Este conceito comparativamente com o autoconsumo coletivo, diferencia-se na medida de que a gestão é efetuada por uma pessoa coletiva formada pelos diversos membros da comunidade, assim como outros participantes no autoconsumo. Já no autoconsumo coletivo a gestão é limitada ao agrupamento de autoconsumidores, administrado por um regulamento interno estipulado [21].

Como se pode verificar na figura 4, a CER representada pelo exemplo 1 apenas utiliza a rede de baixa tensão para efetuar a partilha da energia, o que faz com que o preço total da energia diminua, pois os consumidores apenas terão de pagar a tarifa de acesso à rede de baixa tensão, economizando nas taxas de acesso às redes superiores.

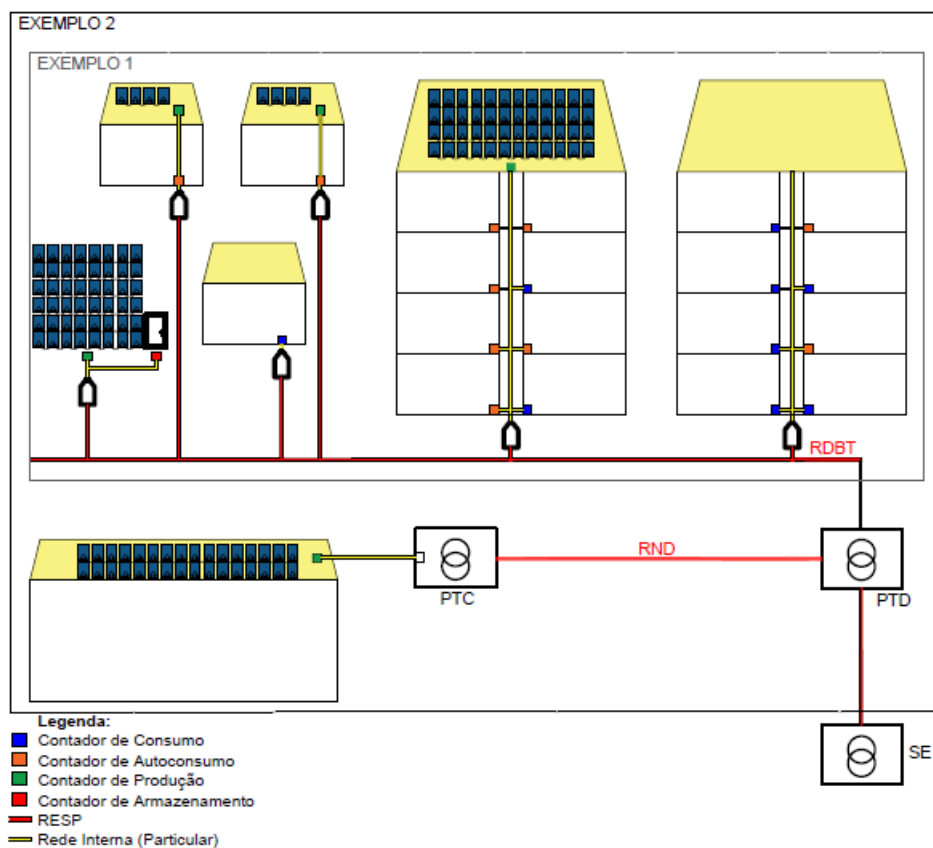


Figura 4 - Exemplos ilustrativos de duas tipologias de CER [20]

A gestão de toda a energia consumida e produzida por uma CER torna-se um grande desafio para o operador de rede, pelo que foram criados sistemas de contagem inteligentes. Desta forma é obrigatória a instalação de contadores inteligentes aptos a realizarem a telecontagem dos respetivos dados de consumo e produção de energia a cada 15 minutos [16].

A potência a instalar varia de acordo com os consumos energéticos dos vários elementos que constituem a CER, para além disso, está sujeita a diversas condições para o exercício da atividade de produção de energia renovável. Estas mesmas condições variam mediante a potência instalada e dividem-se em 4 categorias, como se pode verificar na tabela 1.

Potência Instalada	Requerimentos
<u>$P \leq 350W$</u>	<ul style="list-style-type: none"> • Sem controlo prévio.
<u>$350W < P \leq 30KW$</u>	<ul style="list-style-type: none"> • Mera comunicação prévia (MCP).
<u>$30Kw < P \leq 1MW$</u>	<ul style="list-style-type: none"> • Registo prévio para a instalação da UPAC. • Parecer do Operador de rede em caso de injeção na RESP. • Certificado de exploração.
<u>$P \geq 1MW$</u>	<ul style="list-style-type: none"> • Licença de produção e de exploração. • Prévia atribuição de capacidade de reserva em caso de injeção na RESP > 1MW.

Tabela 1 - Condições para o exercício da atividade de produção de energia renovável [16]

Para além destas condições jurídicas é importante salientar que existem outras condições a nível técnico de necessitam de ser consideradas, nomeadamente a instalação de proteções de interligação/homopolar para efeitos de ligação à rede para UPAC com uma potência instalada superior a 250kW [16].

2.2. PROSUMERS

Atualmente, a procura global por energia está a aumentar, a transição para os veículos elétricos está a crescer a cada ano, e como consequência, espera-se um aumento considerável do consumo energético na rede elétrica, embora a mesma seja alimentada maioritariamente por fontes de energia não renováveis [1]. A produção e distribuição de energia descentralizada tem demonstrado importantes avanços nos últimos anos, abrindo novos horizontes e oportunidades para além dos clássicos sistemas centralizados. A evolução para sistemas energéticos descentralizados apenas é possível com o desenvolvimento tecnológico do sistema de comunicação e das energias renováveis [22].

Existe uma enorme expectativa na chamada “rede elétrica inteligente” para o fornecimento de energia sustentável utilizando como base fluxos bidirecionais de dados, proporcionados por infraestruturas avançadas de controlo, informação e comunicação [23]. Os consumidores são um dos elementos mais importantes nesta rede inteligente, uma vez que eles irão produzir e partilhar energia com outros utilizadores e com a rede, tornando-se assim *prosumers* [24]. O *prosumer* irá ajudar a colmatar as preocupações ambientais, sociais e económicas relacionadas com o aumento da procura de energia, tendo estes também um papel vital na gestão dos picos de consumo energético. A rede elétrica inteligente irá permitir a conectividade e gestão dos vários *prosumers* formando assim comunidades energéticas, sendo estas classificadas de acordo com critérios, tais como o consumo e a produção de energia de cada *prosumer* [25].

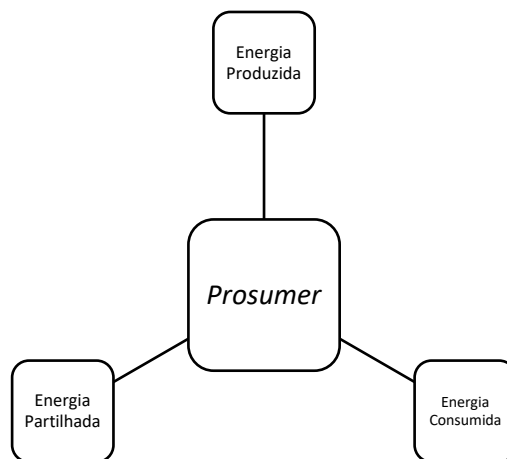


Figura 5 – Conceito energético de *Prosumer*

2.3. MODELOS DE REPARTIÇÃO

A possibilidade de partilha de energia entre os vários elementos de uma CER é sem dúvida o ponto forte da sua implementação, pelo que a escolha do modelo de repartição dessa mesma energia tem de ser a mais adequada possível às características da CER, garantindo altos níveis de eficiência.

2.3.1. REPARTIÇÃO BASEADA NO CONSUMO

Este modelo de repartição tem como base o consumo energético de cada IU face ao consumo de todas as IU pertencentes à CER, garantindo a distribuição conforme as necessidades. Este modelo torna-se vantajoso na gestão da repartição da energia, tendo como principal objetivo a maximização do consumo da energia produzida localmente pela CER, minimizando a injeção de excedente na rede. É de salientar que esta repartição é feita com base em dados provenientes do ORD e é definida pela expressão matemática (1) [19], [26], [27]:

$$ER_{IU_i} = \frac{EC_{IU_i}}{\sum_i EC_{IU_i}} \times \sum_j E_{UPAC_j} \quad (1)$$

Em que:

- ER_{IU_i} - Energia repartida na IU_i ;
- EC_{IU_i} - Energia consumida na IU_i ;
- $\sum_j E_{UPAC_j}$ - Total de energia produzida pelas UPAC;
- i - Índice da IU;
- j - Índice da Unidade de Produção.

2.3.2. REPARTIÇÃO BASEADA EM COEFICIENTES

O modelo de repartição baseado em coeficientes efetua a repartição da energia proporcionalmente, de acordo com os coeficientes previamente estipulados. É caracterizado por ser um modelo muito simples, no entanto por vezes torna-se ineficiente pois despreza o consumo energético, pelo que quando se verifica um baixo consumo, o excedente é injetado na rede em vez de ser repartido pelas IU adjacentes. Este modelo é definido pela expressão matemática (2) [11], [18], [19].

$$ER_{IU_i} = \frac{F_{IU_i}}{\sum_i F_{IU_i}} \times \sum_j E_{UPAC_j} \quad (2)$$

Em que:

- ER_{IU_i} - Energia repartida na IU_i ;
- F_{IU_i} - Fator de repartição na IU_i ;
- $\sum_j E_{UPAC_j}$ - Total de energia produzida pelas UPAC;
- i - Índice da IU;
- j - Índice da Unidade de Produção.

<u>Modelo de Repartição</u>	Vantagens	Desvantagens
Baseada em Coeficientes	Repartição energética proporcional.	Injeção de excedente na rede quando o consumo é baixo.
Baseada no Consumo	Maximiza o autoconsumo, minimizando a injeção de excedente na rede.	Todo o controlo é detido exclusivamente pelo ORD.

Tabela 2 - Comparação entre os Modelos de Repartição baseados em coeficientes e no consumo.

2.3.3. REPARTIÇÃO HÍBRIDA

Actualmente este modelo ainda não se encontra em execução, no entanto é de extrema relevância o estudo do mesmo para possibilitar uma futura implementação. Este tira partido dos benefícios dos modelos apresentados anteriormente, permitindo um controlo inteligente da partilha de energia, aumentando a eficiência e otimizando o autoconsumo. Este modelo é definido pela expressão matemática (3) [19], [26], [27]

$$ER_{IU_i} = \frac{F_{IU_i} \times EC_{IU_i}}{\sum_i (F_{IU_i} \times EC_{IU_i})} \times \sum_j E_{UPAC_j} \quad (3)$$

Em que:

- ER_{IU_i} - Energia repartida na IU_i ;
- F_{IU_i} - Fator de repartição na IU_i ;
- $\sum_j E_{UPAC_j}$ - Total de energia produzida pelas UPAC;
- EC_{IU_i} - Energia consumida na IU_i ;
- i - Índice da IU;
- j - Índice da Unidade de Produção.

2.4. TIPOS DE COMUNIDADES DE ENERGIA RENOVÁVEL

O avanço tecnológico permitiu nos últimos anos tornar a rede de energia elétrica mais inteligente, o que levou conseqüentemente ao surgimento de novos sistemas de recolha de dados não só ao nível do consumo de energia elétrica, mas também de produção, com a integração dos contadores bidirecionais [28]. A partir deste momento começou a ser possível produzir energia e consumir ou injetar a mesma na rede elétrica, dando origem ao novo conceito de microrede [29].

Uma microrede pode operar em rede ao conectar-se com outras microredes, quer seja por meio da RESP ou através de ligações internas pertencentes às próprias microredes, ou pode operar em modo independente. Normalmente a energia produzida por estas microredes é de origem renovável (ex: painéis solares, turbinas eólicas), no entanto também poderão ter origem não renovável (ex: células de combustível). O principal objetivo destas microredes é alcançar a independência energética ou o mais próximo possível, diminuindo assim a dependência relativamente à rede elétrica, pois embora que de uma forma limitada, estas podem continuar a operar mesmo que a rede elétrica não esteja em funcionamento [30]. Para além disso, existem várias outras vantagens que tornam interessante o estudo das microredes, como a gestão da energia consumida/produzida quando estas se agrupam e formam uma comunidade de energia.

As comunidades de energia irão permitir uma maior flexibilidade na utilização da energia produzida localmente, tornando possível o controlo da partilha de energia e do balanço da carga. A energia líquida resultante destas comunidades, nada mais é que a diferença entre a energia produzida e a energia consumida (carga), durante um período de tempo [31].

De seguida serão abordados os vários tipos de comunidades de energia tendo como referência a energia líquida resultante dos seus membros que as constituem, assim como as respetivas distâncias espaciais entre os mesmos.

2.4.1. COMUNIDADES DE ENERGIA HOMOGÉNEA

Uma comunidade de energia homogénea pode ser definida como um conjunto de membros cujas energias líquidas são exclusivamente positivas ou negativas em qualquer momento, ou seja, todas os membros da comunidade conseguem se alimentar com a energia produzida localmente, ou todos os membros da comunidade necessitam de abastecimento energético externo [12], [32].

2.4.2. COMUNIDADES DE ENERGIA MISTA

Entre os vários membros existentes, alguns produzem energia em excesso, enquanto que outros necessitam de energia externa para superarem a energia consumida. Portanto, os membros adjacentes com excesso de energia podem comercializar ou compartilhar a sua energia produzida localmente, sem ser necessário recorrer à rede eléctrica, garantindo assim uma melhor resiliência e confiabilidade do fornecimento de energia.

Este tipo de comunidades de energia são constituídas por membros cujas energias líquidas podem ser positivas ou negativas. Idealmente uma comunidade de energia mista deverá conter membros geograficamente próximos e que permitam atingir um equilíbrio energético [12], [32].

2.4.3. COMUNIDADES DE ENERGIA AUTOSSUFICIENTES

Correspondem a um conjunto de membros com energia líquida total positiva, ou seja, neste tipo de comunidade o valor de energia líquida individual de cada membro não tem relevância, mas sim o valor da energia líquida total da comunidade, sendo que para ser autosuficiente este valor terá de ser obrigatoriamente positivo [12], [32].

2.5. ASPETOS CONCLUSIVOS

As CER apresenta-se como uma solução inovadora para os problemas atuais do setor energético e para a sociedade. Esta veio permitir a participação de novos atores e a diversificação dos mesmos no mercado elétrico, contribuindo significativamente para a redução dos preços da energia. Para além disso, contribuem para a satisfação das necessidades energéticas da forma mais sustentável possível, priorizando soluções renováveis e sem emissões de gases nocivos.

O aumento constante do consumo energético global e conseqüentemente o congestionamento das redes, deixará de ser um problema, uma vez que as CER permitirão diminuir a distância elétrica entre o produtor e o consumidor, atrasando a necessidade de reforço das linhas.

No entanto, ainda existem certas questões, tanto práticas como jurídicas, que se encontram sem resposta concreta, levando assim à necessidade de constantes atualizações dos regimes jurídicos e regulamentos que regem a prática da exploração das CER.

3. TÉCNICAS E ALGORITMOS DE AGRUPAMENTO

O desenvolvimento das tecnologias de armazenamento de dados resultou num crescimento exponencial das bases de dados, pelo que surgiu a necessidade de criar ferramentas que permitissem analisar e agrupar esta enorme quantidade de informação. Uma das técnicas capaz de analisar e agrupar uma grande quantidade de dados em pequenos grupos, de acordo com as suas características é o *clustering*². Esta técnica é definida através da função semelhança, que procede à aglomeração dos dados com semelhanças entre si, dividindo-os em grupos ou *clusters* [33].

² *Clustering* na sua terminologia anglo-saxónica, representa agrupamento e consiste num conjunto de técnicas de prospeção de dados que visa fazer agrupamentos automáticos de dados segundo o seu grau de semelhança.

3.1. TÉCNICAS DE AGRUPAMENTO

O agrupamento pode ser realizado de várias formas, uma vez que existem n categorias de algoritmos de agrupamento. É importante salientar que cada categoria tem os seus próprios pontos fortes e pontos fracos, o que significa que certos algoritmos poderão se adaptar melhor a um conjunto de dados de entrada do que outros. A escolha de um algoritmo de agrupamento adequado a um conjunto de dados, geralmente é uma tarefa difícil, devido ao número de opções disponíveis. Existem vários fatores importantes que afetam essa decisão, tais como as características dos *clusters*, o número de *outliers*, os recursos do conjunto de dados, etc.

As três maiores categorias de algoritmos de agrupamento são: Agrupamento Particional, Agrupamento Hierárquico, Agrupamento Baseado em Densidade [34].

3.1.1. AGRUPAMENTO PARTICIONAL

O agrupamento particional divide os objetos de dados em grupos não sobrepostos, ou seja, nenhum objeto poderá pertencer a mais de um *cluster* e cada *cluster* deve ter pelo menos um objeto. Estes algoritmos de agrupamento funcionam por meio de um processo iterativo que distribui subconjuntos de objetos de dados em k agrupamentos, no entanto, estes necessitam que o número de *clusters* (k) seja definido a priori. O algoritmo *K-MEANS* e *K-MEDOIDS* são exemplos de algoritmos de agrupamento particional, estes são classificados como não determinísticos, ou seja, eles podem gerar resultados diferentes em duas execuções separadas, mesmo que as execuções tenham por base os mesmos dados de entrada [34].

A tabela 3 apresenta os pontos fortes e pontos fracos na utilização de algoritmos que utilizam a técnica de agrupamento particional:

Pontos Fortes	Pontos Fracos
<ul style="list-style-type: none"> ✓ Eficiência elevada quando os aglomerados têm uma forma esférica. ✓ Escaláveis em relação à complexidade do algoritmo. 	<ul style="list-style-type: none"> ✓ Ineficiência ao tratar grupos de dados com diferentes densidades. ✓ Baixa performance no tratamento de <i>clusters</i> com formas complexas e tamanhos diferentes.

Tabela 3 - Pontos Fortes e Pontos Fracos das técnicas de Agrupamento Particional

3.1.2. AGRUPAMENTO HIERÁRQUICO

O agrupamento hierárquico é um método comumente usado para determinar *clusters* constituídos por pontos de dados semelhantes em espaços multidimensionais [35]. Este caracteriza-se por determinar as atribuições dos *clusters* estabelecendo uma hierarquia. Existem duas abordagens distintas para a implementação deste método como demonstra o esquema da figura 6.

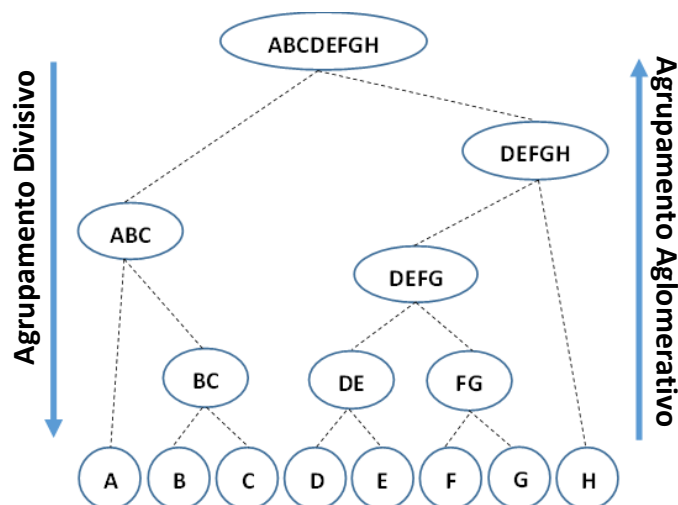


Figura 6 - Esquema demonstrativo do Agrupamento Divisivo e Aglomerativo [36]

- **Agrupamento Aglomerativo** – Inicialmente os pontos começam separados e de seguida estes são agrupados, a cada iteração, até se atingir um único *cluster* constituído por todos os pontos.
- **Agrupamento Divisivo** – Inicia com um único *cluster* constituído por todos os pontos e, de seguida em cada iteração, são separados os pontos com menor nível de similaridade, até se formarem n *clusters*.

Estes dois métodos originam uma hierarquia baseada em árvores de pontos chamados de dendrograma. Eles são comumente usados no estudo de *clusters* hierárquicos antes de decidir o número de *clusters* a atribuir ao conjunto de dados. À semelhança do agrupamento particional, no agrupamento hierárquico o número de clusters pode ser pré-determinado pelo usuário.

O agrupamento hierárquico é um processo determinístico, pelo que as atribuições dos *clusters* não sofrerão qualquer alteração ao executar o algoritmo mais do que uma vez com os mesmos dados de entrada, ao contrário do agrupamento particional [34]. Um exemplo deste tipo de algoritmos são os *Linkage Algorithms*.

A tabela 4 apresenta os pontos fortes e pontos fracos na utilização de algoritmos que utilizam a técnica de agrupamento hierárquico.

Pontos Fortes	Pontos Fracos
<ul style="list-style-type: none"> ✓ Revelam um maior número de detalhes sobre o relacionamento entre os objetos de dados. ✓ Elabora um dendrograma interpretável. 	<ul style="list-style-type: none"> ✓ Complexidade elevada, exigindo um poder computacional elevado. ✓ Sensível a ruídos.

Tabela 4 - Pontos Fortes e Pontos Fracos das técnicas de Agrupamento Hierárquico.

3.1.3. AGRUPAMENTO BASEADO EM DENSIDADE

O agrupamento baseado em densidade determina as atribuições dos *clusters* com base na densidade de pontos de dados numa região. Dentro do conjunto de dados de entrada existem regiões onde a densidade de pontos é elevada separadas por regiões onde a densidade é menor, sendo que *clusters* são atribuídos às regiões onde a densidade de pontos é mais elevada. Os algoritmos que utilizam esta técnica de agrupamento, destacam-se em relação aos anteriormente mencionados, uma vez que não exigem que o usuário especifique o número de *clusters*. No entanto, existem parâmetros adicionais que necessitam de ser fornecidos, como por exemplo, a distância máxima entre pontos para que estes pertençam ao mesmo *clusters*. Como exemplos de algoritmos de agrupamento baseados em densidade, existe o *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*) e o *OPTICS* (*Ordering Points To Identify the Clustering Structure*) [34].

A tabela 5 apresenta os pontos fortes e pontos fracos na utilização de algoritmos que utilizam a técnica de agrupamento baseado em densidade:

Pontos Fortes	Pontos Fracos
<ul style="list-style-type: none">✓ Maior eficiência no tratamento de grupos de formas não esféricas.✓ Resistência ao ruído/<i>outliers</i>.	<ul style="list-style-type: none">✓ Baixa eficiência de agrupamento em regiões de alta densidade.✓ Dificuldade na identificação de aglomerados de diversas densidades.

Tabela 5 - Pontos Fortes e Pontos Fracos das técnicas de Agrupamento Baseado em Densidade.

3.2. ALGORITMOS DE AGRUPAMENTO ESPACIAL

No decorrer deste subcapítulo serão abordados três algoritmos: *K-MEANS*, *DBSCAN* e *Linkage Algorithms*, uma vez que serão o ponto de partida para o estudo das diferentes comunidades de energia que se irá desenvolver.

3.2.1. K-MEANS

O algoritmo *K-MEANS* é um algoritmo de aprendizado não supervisionado iterativo que tenta particionar um conjunto de dados em K subgrupos ou *clusters* distintos e não sobrepostos, onde cada ponto apenas pertence a um grupo. Este tenta tornar os dados pertencentes a um *cluster* os mais semelhantes possíveis, enquanto mantém os *clusters* distintos entre si. Quanto menor for a variação dentro de um *cluster*, mais homogêneos serão os dados dentro do mesmo.

O *K-MEANS* é considerado um dos algoritmos de agrupamento mais utilizados devido à sua simplicidade. É caracterizado por ter uma abordagem ambiciosa ao descobrir o agrupamento que minimiza o SSE (soma do erro ao quadrado), convergindo para uma solução local em vez de uma solução ótima global. Inicialmente o algoritmo distribui aleatoriamente os pontos do conjunto D em k *clusters* e faz o cálculo dos centroides através da média dos pontos do *cluster* C_i . De seguida realiza a atribuição dos *clusters*, onde cada ponto $X_j \in D$ é agregado ao *cluster* que tem o centroide Z_i mais próximo do ponto. Por fim, os centroides são actualizados através da média de todos os pontos pertencentes ao *cluster* C_i . É importante salientar que estas últimas duas fases são realizadas repetidamente até alcançarem um mínimo local, ou seja, o algoritmo converge se os centroides não se alterarem após uma iteração [37], [38]. Todo este processo se pode resumir ao pseudocódigo representado na figura 7.

ALGORITHM: K-MEANS

Result: $C = \{C_1, \dots, C_k\}$, $C_j, j \in 1, \dots, k$

$C_j = \text{random } x_i \text{ in } \chi, j = 1, \dots, k, C_j \neq C_i \forall i \neq j;$

do

$C_j = \emptyset, j = 1, \dots, k;$

foreach $x_i \in \chi$ **do**

$j = \text{argminD}(C_j, x_i);$

$C_j = C_j \cup x_i$

end

foreach $C_i \in C$ **do**

$C_i = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i;$

end

while convergence;

Figura 7 - Pseudocódigo do Algoritmo *K-Means* [39]

3.2.2. DBSCAN

O *DBSCAN* é um algoritmo clássico e um dos mais importantes algoritmos de agrupamento baseado em densidade espacial. Pode ser aplicado a elevados conjuntos de dados que possuem *outliers* e em simultâneo descobre *clusters* com variados formatos com uma eficiência aceitável, mesmo com a presença de ruído. Este encontra agrupamentos baseando-se na vizinhança dos objetos, onde a densidade agregada a um ponto é calculada por meio da contagem do número de pontos vizinhos numa determinada região em volta desse ponto. O *DBSCAN* detém a capacidade de encontrar *clusters* tendo em consideração as características dos dados, não sendo necessário definir o número de *clusters* antecipadamente, pelo que permite a formação de grupos com formatos arbitrários. No entanto, embora seja uma boa solução, ele possui o mesmo problema relacionado com a escalabilidade em comparação com outros algoritmos de agrupamento. No processo de agrupamento é necessário calcular a densidade, ou seja, a distância de um ponto a todos os outros, implicando assim um elevado poder computacional ao trabalhar com conjuntos de dados de elevada dimensão.

Existem métodos que visam reduzir o alto custo computacional do *DBSCAN* no tratamento de grandes conjuntos de dados. Estes métodos resumem-se a duas abordagens, a primeira passa por realizar a consulta do vizinho mais próximo de um modo mais eficiente, a segunda abordagem utiliza técnicas de amostragem para encontrar uma solução aproximada, utilizando apenas um conjunto de instâncias para representar a totalidade da base de dados, o que reduz significativamente o tempo de execução do *DBSCAN* [40].

Uma importante característica deste algoritmo é o facto de ser capaz de identificar *outliers*, além disso este também consegue integrar diversas funções de distância em simultâneo, em contrapartida, necessita dos seguintes parâmetros de entrada:

- **Raio de ϵ -vizinhança de um ponto:** define o raio de vizinhança ϵ para cada ponto da base de dados. O algoritmo *DBSCAN* verifica o número de pontos contidos no raio ϵ para cada ponto existente na base de dados, se esse número exceder o limite definido, então é formado um *cluster*;
- **Número mínimo de pontos (η):** define o limite mínimo de pontos num dado raio de ϵ -vizinhança, que um ponto necessita de possuir para ser considerado um ponto central e conseqüentemente iniciar a formação de um *cluster* [41], [42].

A partir do momento que o algoritmo forma um dado *cluster* C_i , todos os pontos contidos no mesmo tem recursividade, permitindo assim que novos pontos possam ser adicionados. O processo de consulta à vizinhança mais próxima é repetida até que todos os pontos que estejam dentro do raio de ϵ -vizinhança sejam verificados. Após esta etapa, o algoritmo encerra o crescimento do C_i e segue para o próximo ponto não visitado até ao momento, da base de dados. Quando todos os pontos da base de dados forem visitados e classificados, o processo é finalizado. Todo este processo se pode resumir ao pseudocódigo representado na figura 8. O *DBSCAN* distingue três tipos de pontos dentro do conjunto de dados:

- **Pontos centrais:** Localizam-se no interior de uma região densa, onde existem pelo menos η pontos no raio de ϵ -vizinhança desse ponto;

- **Pontos de borda:** Localizam-se na fronteira de uma região densa, ou seja, são pontos que estão na ϵ -vizinhança de algum ponto central, embora não sejam considerados pontos centrais;
- **Outliers:** Não são considerados pontos centrais nem de borda, pelo que não são conectados por densidade a nenhum outro ponto, e conseqüentemente não pertencem a nenhum cluster [42], [43].

ALGORITHM: DBSCAN

Input: DB: Database

Input: ϵ : Radius

Input: minPts: Density threshold

Input: dist: Distance function

Data: label: Point labels, initially undefined

```

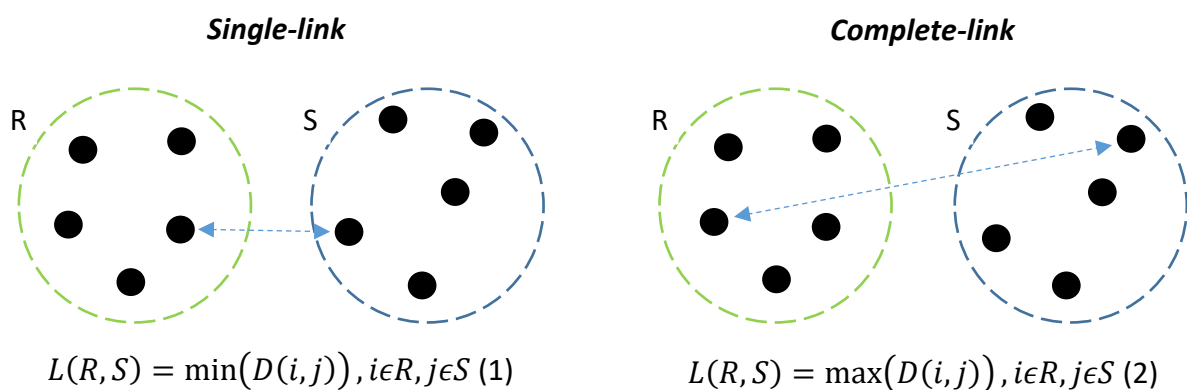
foreach point p in database DB do
    if label(p)  $\neq$  undefined then continue
    Neighbors N  $\leftarrow$  RangeQuery(DB,dist,p, $\epsilon$ )
    if |N| < minPts then
        |   label(p)  $\leftarrow$  Noise
        |   continue
    c  $\leftarrow$  next cluster label
    label(p)  $\leftarrow$  c
    Seed set S  $\leftarrow$  N \ {p}
    foreach q in S do
        |   if label(q) = Noise then label(q)  $\leftarrow$  c
        |   if label(q)  $\neq$  undefined then continue
        |   Neighbors N  $\leftarrow$  RangeQuery(DB,dist,q, $\epsilon$ )
        |   label(q)  $\leftarrow$  c
        |   if |N| < minPts then continue
        |   S  $\leftarrow$  S  $\cup$  N
  
```

Figura 8 - Pseudocódigo do Algoritmo *DBSCAN* [44]

3.2.3. LINKAGE ALGORITHMS

Os *Linkage Algorithms* utilizam o método de agrupamento hierárquico aglomerativo, ou seja, consideram que a fusão de *clusters* é baseada na distância entre os mesmos [45]. Existem vários tipos de algoritmos de ligação, e estes dividem-se em dois grandes grupos, os algoritmos baseados em métodos gráficos e os algoritmos baseados em métodos geométricos [46]. Dentro dos algoritmos baseados em métodos gráficos, os mais importantes e conseqüentemente os mais utilizados são o *Single-link*, *Average-link* e *Complete-link* [45]. Relativamente aos algoritmos baseados em métodos geométricos o mais comumente usado é o *Ward's Method* [47]. Todos eles se baseiam na similaridade e na distância euclidiana entre os vários pontos de dados, mas a principal diferença entre eles é o modo de como essa mesma distância é calculada, como pode-se verificar nas ilustrações e expressões matemáticas representadas na figura 9 [48].

- **Single-link:** A distância entre dois *clusters* é dada pela margem de distância mínima entre os pontos dos dois *clusters* (1).
- **Complete-link:** A distância entre dois *clusters* é dada pela margem de distância máxima entre os pontos dos dois *clusters* (2).
- **Average-link:** A distância entre os dois *clusters* é dada pela média de todas as distâncias entre os pontos dos dois *clusters* (3).
- **Ward's Method:** A distância entre os dois *clusters* é dada pela soma dos desvios ao quadrado dos pontos dos *clusters* aos centroides (4) [35].



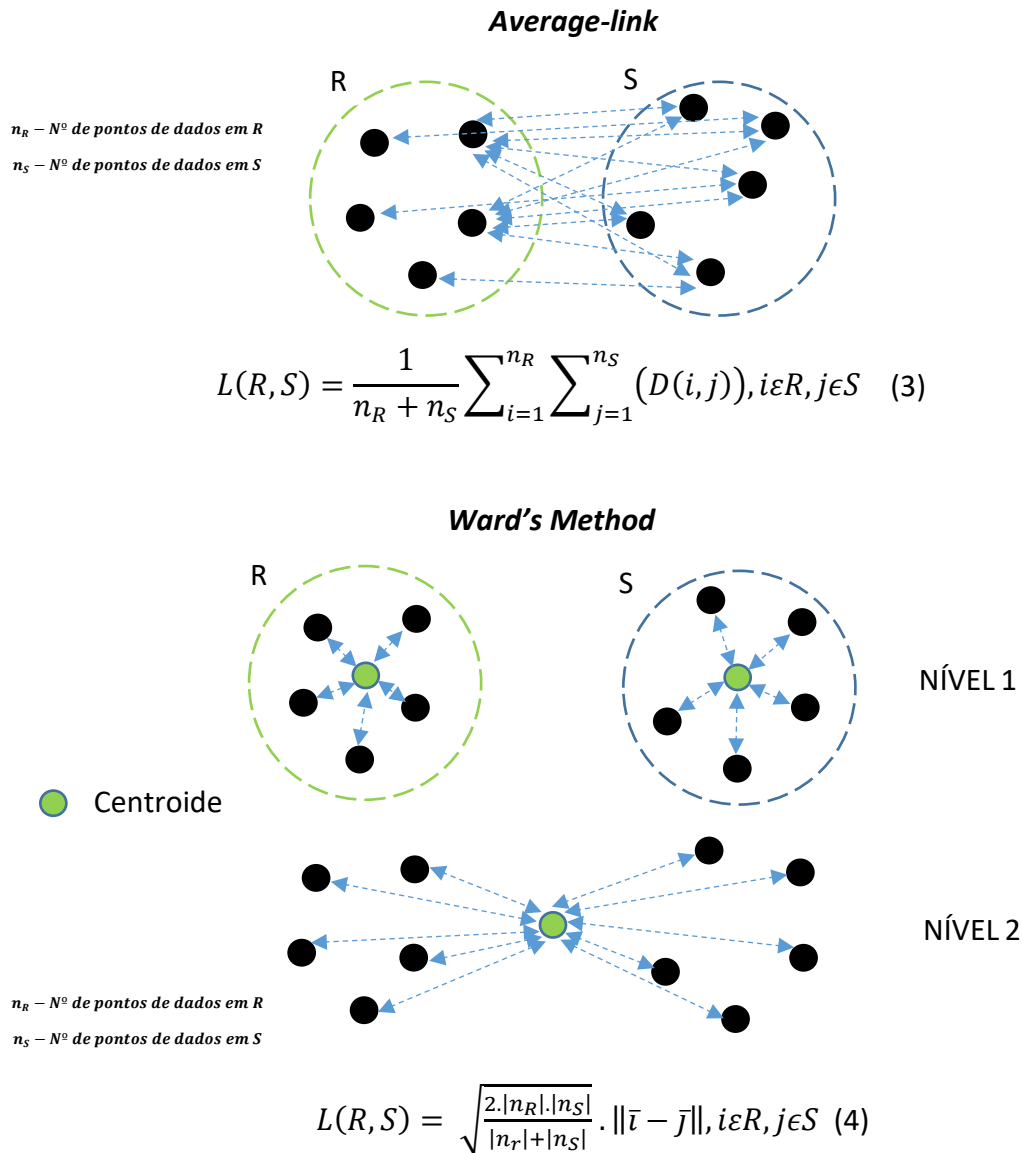


Figura 9 - Demonstração do cálculo da distância entre clusters nos *Linkage Algorithms* [52]

O *Single-link* é sensível à presença de ruído e às mudanças significativas na densidade dos aglomerados, no entanto apresenta uma insensibilidade total relativamente à forma e tamanho dos aglomerados. O *Average-link* tem alguma dificuldade em lidar com aglomerados de forma e tamanhos distintos, pelo que pode facilmente falhar nestes casos. O *Complete-link* não é afetado pelo ruído, no entanto, tem dificuldade em tratar aglomerados com formas convexas, para além de poder dividir grandes *clusters* [45]. Por fim, o *Ward's Method* de entre todos os tipos de algoritmos mencionados é o mais completo e consequentemente o que gera melhores resultados, pois produz melhores hierarquias de *clusters*, para além de ser o menos suscetível a ruídos e discrepâncias [36].

O funcionamento do algoritmo pode resumir-se ao pseudocódigo apresentado na figura 10. O algoritmo inicia com a importação da base de dados e o número de clusters pretendidos. De seguida, calcula a matriz proximidade a partir da função da distância que varia de acordo com o tipo de algoritmo escolhido, como visto anteriormente. Por fim realiza continuamente a pesquisa da solução ótima, fazendo a atualização da matriz proximidade a cada iteração, até que reste apenas um cluster.

ALGORITHM: Linkage Algorithms

Input: Data set X , n - number of samples, k - number of clusters

for $i = 1$ to n **do**

 Compute the proximity matrix (usually based on the distance function) for the k clusters;

end for

for $j = 1$ to k **do**

 Compute/Search the minimal distance $d(C_i, C_j) = \min_{1 \leq m, l \leq k, m \neq l} d(C_m, C_l)$ where $d(., .)$ is the distance function

end for

for $p = 1$ to k **do**

 Update the proximity matrix by computing the distances between the cluster C_{ij} and the other clusters;

end for

REPEAT steps 2 and 3 until only one cluster remains.

Figura 10 - Pseudocódigo dos *Linkage Algorithms* [46]

3.2.4. COMPARAÇÃO ENTRE OS ALGORITMOS DE AGRUPAMENTO

Embora todos os algoritmos abordados tenham como principal objetivo criar os melhores clusters possíveis, existem várias diferenças entre eles que poderão ser discutidas de forma a conseguir prever possíveis resultados, ou até concluir à priori quais serão os mais apropriados para certos tipos de base de dados. Na tabela 6 pode-se verificar as principais diferenças entre os algoritmos em estudo.

<u><i>K-Means</i></u>	<u><i>DBSCAN</i></u>	<u><i>LINKAGE ALGORITHMS</i></u>
Eficiente a tratar grandes conjuntos de dados.	Baixa eficiência ao manipular conjuntos de dados de elevada dimensão.	Eficiente a tratar grandes conjuntos de dados.
Sensível ao ruído (<i>outliers</i>).	Elevada imunidade a ruído e a dados com valores discrepantes.	<i>Complete-link</i> e <i>Ward's Method</i> não são afetado pela presença de ruído, ao contrário do <i>Single-link</i> .
Os <i>clusters</i> formados têm forma aproximadamente esférica e o mesmo tamanho que o recurso.	O <i>clusters</i> formados têm forma arbitrária e podem ter ou não o mesmo tamanho que o recurso.	<i>Single-link</i> apresenta insensibilidade total relativamente à forma e tamanho dos aglomerados ao contrário do <i>Average-link</i> .
Necessário definir o número de clusters previamente.	Número de clusters não necessita de ser especificado.	Necessário definir o número de <i>clusters</i> previamente.
Imune a conjuntos de dados com densidades variáveis.	Sensível a conjuntos de dados dispersos e com densidades variáveis.	<i>Single-link</i> é sensível a conjuntos de dados com densidades variáveis.
Necessita de 1 parâmetro de entrada: Número de <i>clusters</i> .	Necessita de dois parâmetros de entrada: Raio de ϵ -vizinhança e o Número mínimo de pontos .	Necessita de 1 parâmetro de entrada: Número de <i>clusters</i> .
Difícil detecção de anomalias, pois atribui todos os pontos a <i>clusters</i> , mesmo que estes tenham valores anômalos.	Diferencia regiões de alta densidade que se encontram separadas por regiões de baixa densidade.	<i>Complete Link</i> tem dificuldade em tratar aglomerados com formas convexas.

Tabela 6 – Comparação entre os Algoritmos *K-MEANS*, *DBSCAN* e *LINKAGE ALGORITHMS*

3.3. ASPETOS CONCLUSIVOS

Embora os algoritmos de agrupamento discutidos ao longo deste capítulo tentem resolver questões como escalabilidade e complexidade, pode-se verificar que um algoritmo de agrupamento perfeito que compreende todos os problemas e se adapta a qualquer conjunto de dados é uma noção idealista.

No contexto real consegue-se identificar uma diversidade enorme de algoritmos, sendo que os mesmo podem-se basear em várias técnicas de agrupamento distintas. No entanto, o maior desafio permanece em encontrar qual o melhor algoritmo no tratamento de um conjunto de dados específico. Para tal, é necessário a realização de um estudo prévio sobre os respetivos algoritmos alvo, concluido quais os pontos fortes e fracos de cada um, de forma possibilitar uma melhor escolha. Esta escolha passa pela procura de algoritmos que utilizam métodos mais dinâmicos, adaptativos e inovadores que decifram padrões significativos que efetivamente satisfaçam os requisitos de lidar com elevados volumes de dados, insensíveis a ruídos, não afetados pela ordem de entrada e sem conhecimento prévio do domínio .

4. METODOLOGIA DE AGRUPAMENTO E CLASSIFICAÇÃO

Neste capítulo é apresentada inicialmente a metodologia para a identificação e classificação das CER. É abordado também o funcionamento e a estrutura do modelo desenvolvido ao longo desta dissertação, capaz de importar um conjunto de dados referentes a possíveis membros, e de acordo com as características de cada membro, realizar o agrupamento em várias CER.

4.1. IDENTIFICAÇÃO E CLASSIFICAÇÃO DAS COMUNIDADES DE ENERGIA RENOVÁVEL

Definindo N como o número de membros existentes numa comunidade, $E(t)$ a energia líquida por unidade de tempo t e D a distância espacial entre dois membros numa rede elétrica. Pode-se observar que se $E(t) < 0$, então o membro necessita de solicitar energia externa para se sustentar energeticamente, se $E(t) > 0$ então o membro detém uma quantidade excessiva de energia que pode ser partilhada. Com base nestas variáveis realizar-se-á a distinção entre os diferentes tipos de comunidades de energia.

4.1.1. COMUNIDADES DE ENERGIA HOMOGÉNEAS

Quando se trata de uma comunidade de energia homogénea (HEC) a sua classificação é relativamente fácil em relação às restantes, uma vez que estas apenas poderão ser constituídas por um conjunto de membros cujas energias líquidas são unicamente positivas ($E(t) > 0$) ou unicamente negativas ($E(t) < 0$).

Aplicando este conceito ao nível real, poderá se classificar uma comunidade como homogénea, por exemplo, à medida que a construção de novos edifícios aumenta, também aumenta o número de consumidores naquela área, caso nenhum opte pela instalação de fontes de energia renováveis, este conjunto de consumidores no ponto de vista energético terão sempre uma energia líquida negativa e como tal poderão ser classificados como uma comunidade de energia homogénea negativa. Outro exemplo será a implementação de um conjunto de UPAC sobredimensionadas numa certa área com o intuito de armazenar a energia excedente para ser posteriormente utilizada, este conjunto de membros também poderá ser classificado como uma comunidade de energia homogénea, neste caso positiva.

Para que seja possível a identificação de diferentes HEC na rede elétrica são essências duas variáveis, a energia líquida ($E(t)$) e as distâncias geográficas entre os membros (D). Este problema de identificação da comunidade energética pode ser considerado como um problema de agrupamento baseado nas distâncias geográficas entre os membros, onde o número de grupos é dado por K e a energia líquida da HEC pode ser agregada em momentos diferentes. No entanto, existem outras restrições que terão de se ter em consideração na implementação deste conceito, uma vez que a $E(t)$ de cada HEC não é ilimitada, pois o fornecimento de energia depende da capacidade das UPAC instaladas. Nestes casos, o

número de comunidades é desconhecido e devem-se aplicar as seguintes restrições adicionais: se $\Sigma E(t)$ da HEC for positiva, esta não poderá exceder um limite superior positivo LS ; caso $\Sigma E(t)$ da HEC seja negativa, esta não poderá ser inferior a um limite inferior LI .

4.1.2. COMUNIDADES DE ENERGIA MISTA

Como se acabou de verificar anteriormente, de entre os possíveis milhares de membros existentes na rede elétrica, alguns possuem $E(t)$ positiva, ou seja, excesso de energia que poderá ser partilhada ou armazenada, enquanto outros possuem $E(t)$ negativa, ou seja, necessitam de solicitar energia externamente. Estes membros podem conectar-se e formar uma comunidade de energia mista (MEC), uma vez que os membros podem partilhar o excedente resultante da produção energética local, de modo a reduzir a perda de energia na distribuição a partir da RESP e a melhorar a fiabilidade do fornecimento de energia da mesma. Neste processo de partilha, não só os membros com $E(t)$ negativa saem favorecidos, por se fornecerem de energia mais barata, mas também os membros com $E(t)$ positiva, uma vez que estes podem aumentar a rentabilização da energia produzida localmente.

Uma MEC ideal pode ser definida como um conjunto de membros geograficamente próximos entre si, onde existe um equilíbrio entre a produção de energia e o consumo, de modo que toda a energia que é produzida localmente seja também consumida localmente, evitando assim o desperdício de energia ou a injeção na rede. Por exemplo, um membro com $E(t)$ positivo e outro membro com $E(t)$ negativo podem suprir mutuamente as suas necessidades energéticas utilizando a energia produzida localmente.

4.1.3. COMUNIDADES DE ENERGIA AUTOSSUFICIENTES

As comunidade de energia autossuficientes (SEC) podem ser classificadas como um caso especial dentro do grupo das MEC, uma vez que também são constituídas por microredes com $E(t)$ positiva e negativa, no entanto existe uma importante diferença a salientar, que

é o facto dos membros que constituem uma SEC conseguirem suprir totalmente o consumo energético com a energia produzida localmente através das UPAC. Neste caso, pode-se considerar que o $\sum E(t) > 0$, o que torna bastante interessante o estudo destas comunidades, pois estas não dependem em parte da rede elétrica, criando assim várias vantagens para quem pertence às mesmas, como por exemplo o facto de não serem afetadas quando surgem apagões em larga escala.

De forma a consolidar os conceitos relativos à identificação dos tipos de CER abordados anteriormente, a figura 11 demonstra a forma de como essa classificação se processa, tendo em consideração a energia líquida de cada CER.

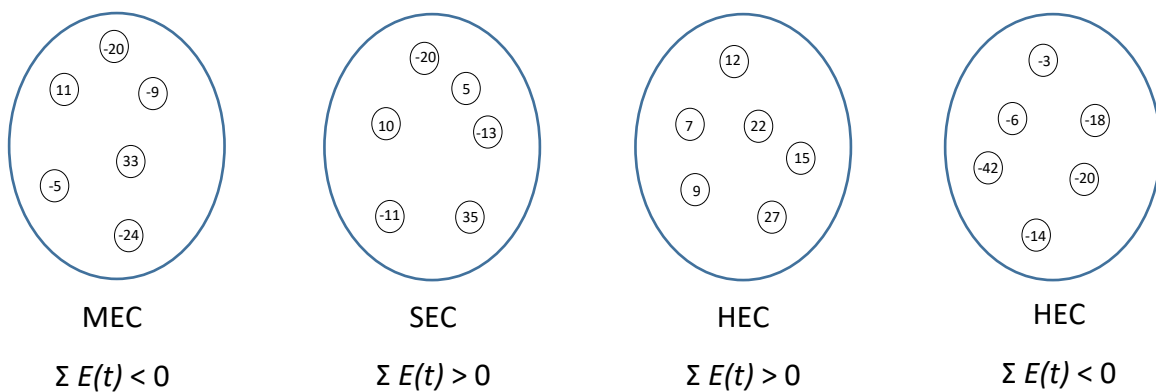


Figura 11 - Comparação do valor $E(t)$ nos diferentes tipos de Comunidades de Energia [53]

4.2. MODELO DESENVOLVIDO

O modelo desenvolvido ao longo desta dissertação tem como principal objetivo fazer o agrupamento e a classificação dos tipos de comunidades locais de energia partindo da importação de dados de um conjunto de membros. Para tal, foi necessário criar uma base de dados, esta foi desenvolvida em Excel e é composta pelos dados de georreferenciação e energia líquida referentes a cada membro. A linguagem de programação escolhida para

o desenvolvimento desta aplicação foi a linguagem Python³, não só pela popularidade que esta tem vindo a ganhar entre a comunidade científica, mas também por possuir diversas bibliotecas e estruturas para a análise de dados, baseadas em algoritmos, que é o grande foco desta dissertação. Os softwares utilizados no desenvolvimento deste estudo foram o Visual Studio Code⁴, onde se desenvolveu todo o código de programação, o Excel⁵ que serviu de base de dados para armazenar toda a informação relativa aos membros, e por fim o Google Earth Pro⁶ que foi usado para obter as coordenadas geográficas de cada membro. De seguida será abordada a estrutura do código desenvolvido, assim como todas as funcionalidades e variantes do mesmo.

4.2.1. ESTRUTURA DO MODELO

O modelo desenvolvido ao longo deste estudo caracteriza-se pelo desenvolver de três processos, como se pode verificar no fluxograma representado na figura 12, onde inicialmente existe a conversão das coordenadas geográficas dos diferentes membros, possibilitando a georreferenciação gráfica dos mesmos, de seguida são executados os vários algoritmos em estudo, nomeadamente os *Linkage Algorithms*, *K-MEANS* e *DBSCAN*, onde todos os dados são agrupados. No final, dá-se a classificação dos agrupamentos formados no processo anterior, ou seja, das comunidades de energia renovável.

³ <https://www.python.org/>

⁴ <https://code.visualstudio.com/>

⁵ <https://www.microsoft.com/pt-pt/microsoft-365/excel>

⁶ <https://www.google.com/intl/pt-PT/earth/>

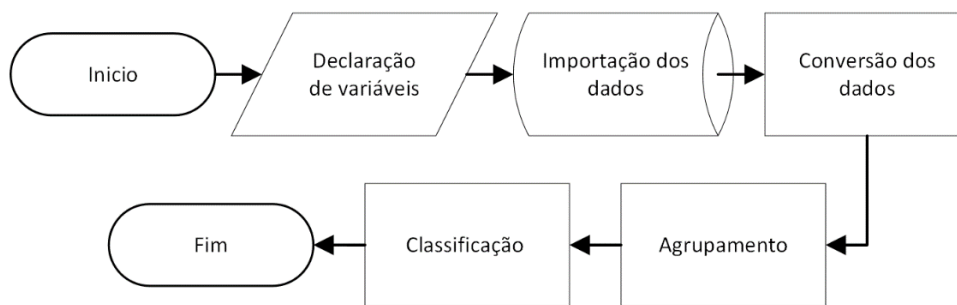


Figura 12 - Fluxograma do Funcionamento Geral do Modelo

A função inicial, representada no fluxograma da figura 13, caracteriza-se por ser a função principal do código onde se declaram as variáveis necessárias para a correta execução do mesmo e onde são chamadas as restantes funções integrantes. Esta função pode variar ligeiramente de acordo com o algoritmo abordado, uma vez que alguns deles necessitam de dados de entrada distintos. No entanto existem algumas variáveis que se mantêm de algoritmo para algoritmo como é caso do “fator_correlação_energético”, esta variável representa o fator de correlação que tem como objetivo facilitar o agrupamento de dados por parte de cada algoritmo, pois como os dados de entrada (distâncias e energia) tem ordens de grandeza relativamente distintas, torna-se muito difícil ou até impossível o agrupamento dos dados de forma eficiente sem a utilização deste fator.

As variáveis “maxenergia” e “minenergia” representam respetivamente o limite superior e inferior da energia líquida da CER. O objetivo destas variáveis é limitar o valor da energia líquida da CER, seja ela positiva ou negativa, pois existem certos casos onde esta limitação poderá ser uma mais valia, por exemplo, devido ao fornecimento limitado da energia a cada uma das comunidades descobertas, a carga global de procura poderá ser obrigada a ter um limite superior, ou em algumas comunidades, a diferença entre a procura e a oferta globais em cada comunidade pode ter de ser limitada a um pequeno número, com o objetivo de otimizar o desempenho da rede.

Com a necessidade de representar geograficamente cada membro da comunidade, foi indispensável a recolha dos valores de latitude e longitude de cada membro e o armazenamento dos mesmos na base de dados criada em *Excel*.

As variáveis anteriormente abordadas são estritamente necessárias para o funcionamento do modelo, já as seguintes variáveis a abordar são apenas necessárias para o funcionamento de certos algoritmos, como é caso da “max_distância” e da “min_membros” que representam respetivamente, a distância máxima entre membros para que estes possam pertencer à mesma comunidade e o número mínimo de membros por comunidade, estes parâmetros apenas são necessários para o funcionamento do algoritmo *DBSCAN*. A variável “n_clusters” representa o número de clusters que o algoritmo terá de formar ao realizar o agrupamento, sendo que esta é apenas necessária declarar nos *Linkage Algorithms* e no *K-MEANS*.

Função Inicial

Saída: --

Entrada: --

- Inicializa** max_distância
- Inicializa** fator_correlação_energético
- Inicializa** min_membros
- Inicializa** max_energia
- Inicializa** min_energia
- Inicializa** base_dados (Excel)
- Inicializa** n_clusters
- Chama** Importar Ficheiro
- Chama** Conversão de dados
- Chama** DBSCAN / K-means / Agglomerative Clustering
- Chama** Classificação
- Chama** Gráficos

Figura 13 - Pseudocódigo da Função Inicial

A função “Importar Dados” é a função responsável por importar as informações referentes a cada membro da base de dados criada em Excel. Esta importa os dados que se encontram nas colunas com o nome “id”, “energia”, “lat” e “long”, como se pode verificar no pseudocódigo da figura 14. Cada variável irá armazenar um tipo de dados, sendo que este se encontram detalhados na tabela 7.

Variáveis	Descrição dos Dados
id	Identificação de cada membro.
energia	Energia Líquida Anual de cada membro (energia produzida - energia consumida) em MWh/Ano.
lat	Latitude referente à localização de cada membro.
lon	Longitude referente à localização de cada membro.

Tabela 7 - Variáveis usadas no modelo desenvolvido

Após a importação destes dados, eles são guardados de forma a possibilitar o uso dos mesmos nas funções seguintes.

Função Importar Dados

Saída: -

Entrada: localização do ficheiro

Atribui dados, lista = ler excel(localização do ficheiro)

Define precisão de casas decimais = 14

Atribui pd = dados, colunas("id", "energia", "lat", "lon")

Retorna linhas de dados

Figura 14 - Pseudocódigo da Função Importar Ficheiro

A função "Conversão de Dados" utiliza os dados guardados anteriormente pela função "Importar Ficheiro", mais concretamente a latitude e longitude relativa à localização de cada membro, e faz a conversão dos dados que se encontram no sistema de referência WGS84 para o sistema UTM. Esta conversão permite posteriormente a representação de cada membro num referencial cartesiano.

A biblioteca utilizada no processo de conversão de coordenadas foi o "Proj", esta faz a conversão de coordenadas geoespaciais entre diversos sistemas de referência. Para que esta conversão ocorra com sucesso, para além das coordenadas geográficas, são necessários alguns parâmetros de entrada adicionais, nomeadamente o hemisfério e a

zona onde se encontram as coordenadas a converter, o sistema de referência inicial e o sistema de referência pretendido no final do processo.

Após efetuada a conversão de todos os dados, estes são armazenados num vetor, juntamente com a energia líquida de cada membro multiplicada pelo fator de correlação. Após este processo, os dados encontram-se devidamente tratados, possibilitando a importação dos mesmos nos vários algoritmos de agrupamento em estudo e a representação dos mesmos num gráfico cartesiano de 3 dimensões, o seguinte pseudocódigo representado na figura 15, demonstra sequencialmente o funcionamento desta função.

Função Conversão de Dados

Saída: x

Entrada: base_dados[id,latitude,longitude,energia], fator_correlação_energético

Inicializa myProj = [sistema_referência inicial, zona, sistema_ referência final, unid.]

PARA cada membro da base_dados

distância_x, distância_y = proj (base_dados["longitude"], base_dados["latitude"])

x = (distância_x, distância_y, base_dados ["energia"] * fator_correlação_energético)

Figura 15 - Pseudocódigo da Função Conversão de dados

A função "Classificação" faz a classificação dos *clusters* previamente criados pelos algoritmos. As comunidades de energia poderão ser classificadas como Autossuficientes ou Não Autossuficientes e simultaneamente como Mistas ou Homogéneas, como demonstra o esquema da figura 16.

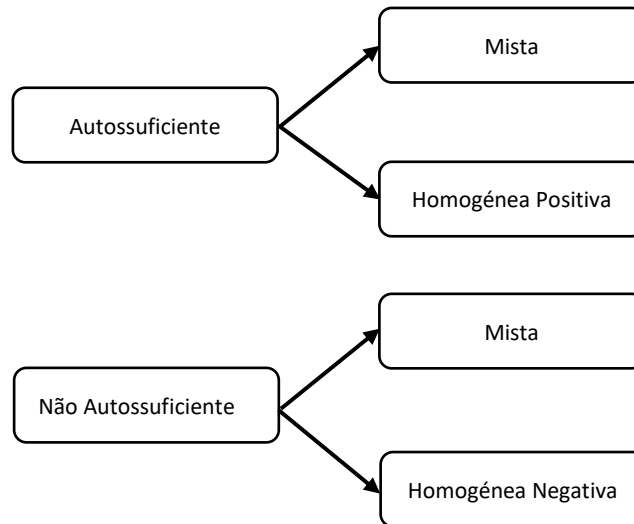


Figura 16 - Esquema de possíveis classificações para cada CER

Como se pode verificar no pseudocódigo representado na figura 17, a função inicialmente repõe a ordem de grandeza relativa aos valores de energia líquida de cada comunidade, uma vez que o agrupamento já ocorreu e são necessários os valores reais de modo a possibilitar a correta classificação. De seguida inicia-se a classificação de cada comunidade, se todos os membros pertencentes à comunidade possuírem apenas energia líquida positiva, então esta é classificada como “Homogénea Positiva”, caso se verifique o contrário, ou seja, todos os membros pertencentes à comunidade possuírem apenas energia líquida negativa, então a mesma é classificada como “Homogénea Negativa”. A comunidade é classificada como mista se os membros que a compõem possuírem energias líquidas negativas e positivas.

A classificação da comunidade de acordo com o soma de todas as energias líquidas de cada membro que a constitui, divide-se em dois tipos: “Autossuficiente” e “Não Autossuficiente”. Teoricamente, para que uma comunidade seja considerada “Autossuficiente” esta terá de possuir um valor de $E(t) \geq 0$, no entanto, optou-se por inserir um intervalo de valores maior de forma a agilizar o processo de classificação, pelo que se o valor de $E(t)$ se situar entre o intervalo de valores $[-20, \text{max_energia}]$, então esta é classificada como “Autossuficiente”. Caso o valor de $E(t)$ da comunidade se situar entre o intervalo de valores $[\text{min_energia}, -20[$, esta é classificada como “Não Autossuficiente”.

Todas estas funções poderão ser analisadas de uma forma mais simples e sequencial nos fluxogramas que se encontram representados no Anexo A.

Função Classificação

Saída: classificação, número de membros, energia, id

Entrada: *labels*, *fator_correlação_energético*, *max_energia* e *min_energia*

```
PARA elementos pertencentes a labels
|
| SE elementos pertencerem a clusters
|   | Inicializa o cluster + 1
|   | Inicializa o cluster com energia / fator_correlação_energético
|
| SENÃO
|   | Inicializa o cluster = 1
|   | Inicializa o cluster com energia / fator_correlação_energético
|
PARA j pertencente ao número de clusters
|
| SE número de pontos de ruído = 0
|   | Inicializa j + 1
|
| SE valor da energia pertencer ao intervalo [-20, max_energia]
|   | Inicializa a classificação como "Autossuficiente"
|
| SE valor da energia pertencer ao intervalo [min_energia, -21]
|   | Inicializa a classificação como "Não Autossuficiente"
|
| SE energia de todos os membros pertencentes ao cluster > 0
|   | Adiciona à classificação " – Homogénea Positiva"
|
| SE energia de todos os membros pertencentes ao cluster < 0
|   | Adiciona à classificação " - Homogénea Negativa"
|
| SE classificação = 'Não Autossuficiente' ou 'Autossuficiente'
|   | Adiciona à classificação " - Mista"
|
Classifica o cluster
```

Figura 17 - Pseudocódigo da Função Classificação

4.3. ASPETOS CONCLUSIVOS

A metodologia proposta para a realização do agrupamento dos membros em comunidades de energia renovável baseia-se em três algoritmos de agrupamento, o *K-MEANS*, *DBSCAN* e *Linkage Algorithms*, sendo que estes utilizam técnicas de agrupamento distintas.

O modelo desenvolvido agrupa os diferentes membros em CER dependendo das suas características, nomeadamente o valor de energia líquida anual e a localização geográfica, através da latitude e longitude de cada membro.

O estudo para a obtenção de CER que se aproximem do ideal não é o único objetivo desta dissertação, é igualmente importante saber identificar e classificar uma CER de acordo com as suas características energéticas. Como se pôde verificar nos capítulos anteriores, as CER dividem-se em três tipos, HEC, MEC e SEC. Estas podem ser facilmente descobertas se o valor de energia líquida de cada membro pertencente à CER for conhecido.

Para que seja possível comparar os vários resultados obtidos é essencial garantir que todos os algoritmos se encontram nas mesmas condições de igualdade, garantindo assim uma maior credibilidade ao estudo.

5. CASO DE ESTUDO

Neste capítulo, é apresentada a visão geral dos dados utilizados para o estudo, assim como a caracterização e análise dos mesmos. Os aspetos focados serão fundamentais para o processo de teste e validação do modelo que irá distribuir os diversos membros em grupos, formando assim comunidades de energia, para além de classificar as mesmas de acordo com o valor das respetivas energias líquidas. Também serão analisadas as simulações e resultados obtidos, provenientes do modelo desenvolvido.

5.1. CARACTERIZAÇÃO E ANÁLISE DOS DADOS

Sendo o tema abordado nesta dissertação bastante desconhecido e inovador, conseqüentemente apenas foi possível obter parte dos dados de entrada de forma real,

sendo que a maior parte dos dados utilizados no estudo resultaram da estimativa de valores produção e consumo energético de diversas tipologias de instalações elétricas. Isto deve-se ao facto de na atualidade não existir nenhuma base de dados disponível que possa fornecer livremente dados de consumo e produção das instalações elétricas. No entanto, mesmo que fosse possível, não iria trazer benefícios significativos ao desenvolvimento do estudo, face aos dados criados artificialmente.

Foram criados artificialmente dados de 170 membros localizados no distrito do Porto. Para cada membro foram registados os valores da produção e consumo de energia, e as respetivas localizações geográficas, recorrendo aos valores de latitude e longitude.

Relativamente aos dados reais, estes foram recolhidos do projeto de uma comunidade de energia renovável, que se encontra em desenvolvimento na zona norte de Portugal. Estes dados foram obtidos através das faturas de consumo energético de cada membro da comunidade, e do cálculo da energia produzida pelas UPAC que se encontram associadas à mesma.

É importante salientar que nem todos os membros terão valores de produção de energia e consumo, uma vez que se pode tratar de um membro que apenas produz energia, sendo este apenas constituído por uma UPAC, ou então um membro apenas com consumo de energia, não havendo neste caso nenhuma unidade de produção energética associada.

As localizações geográficas dos membros criados artificialmente foram definidas de forma arbitrária, no entanto as mesmas encontram-se todas localizadas dentro do distrito do Porto, e distribuídas pelas principais cidades/regiões. Como se pode verificar na figura 18, a cidade onde se encontra o maior número de membros é a cidade do Porto, uma vez que também é a cidade onde se encontra uma maior densidade populacional e industrial, o que leva conseqüentemente um maior consumo/produção de energia elétrica.

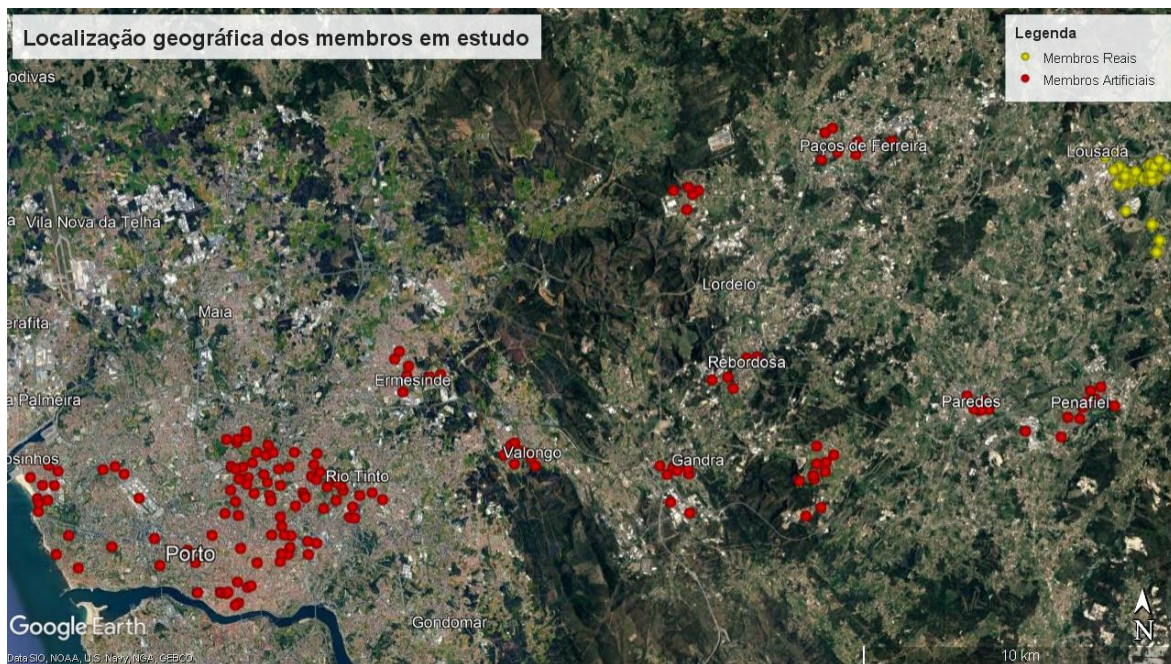


Figura 18 - Localização geográfica dos membros em estudo (Google Earth Pro)

Como se pode observar na figura 19, cerca de 43% dos membros localizam-se na cidade do Porto, sendo que as restantes 57% encontram-se distribuídas nas regiões vizinhas, nomeadamente em Valongo, Lousada, Ermesinde, Paredes, Paços de Ferreira, Penafiel, Rebordosa e Gandra.

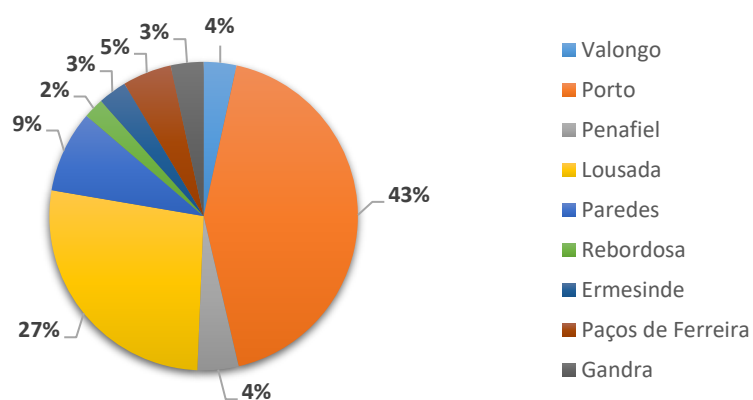


Figura 19 - Gráfico da distribuição geográfica dos membros por região

Relativamente aos dados de produção e consumo de energia, estes têm como unidade de medida o MWh/Ano. A atribuição dos valores de consumo e produção de energia dos membros criados tiveram como base vários projetos realizados durante o meu percurso profissional como projetista de instalações fotovoltaicas. Como é de prever, o dimensionamento de uma central fotovoltaica começa inicialmente pela análise das necessidades energéticas do cliente, pelo que é essencial ter uma noção de qual o consumo anual médio da instalação em estudo. Após a realização de todo o dimensionamento, ficará definida a potência ideal a instalar, e como tal, facilmente se consegue prever qual a produção energética anual da UPAC, tendo sempre em consideração fatores condicionantes como a radiação solar, a eficiência dos equipamentos utilizados, etc...

Os valores de consumo e produção de energia criados variam entre os [0 , 400] MW/h Ano, como se pode observar na figura 20, sendo estes referentes a instalações de várias tipologias, representando desde os consumidores particulares até às grandes indústrias.

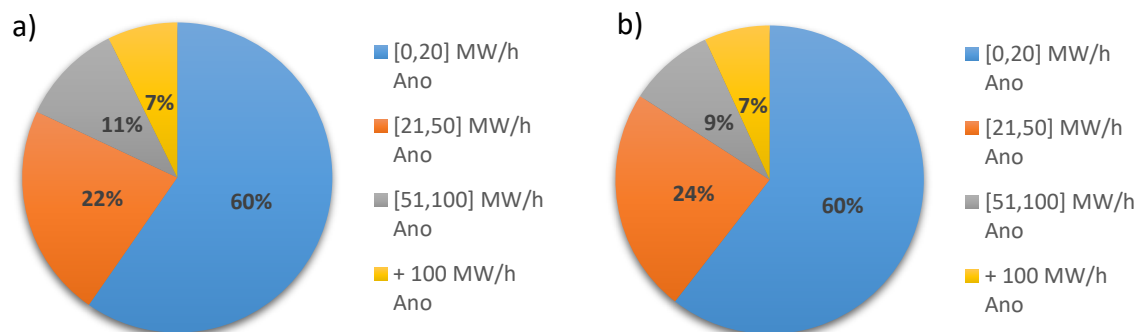


Figura 20 - a) Valores de Produção Anual dos membros b) Valores de Consumo Anual dos membros

O valor $E(t)$ nada mais é que o valor da energia líquida de um membro, ou seja, a diferença entre a produção e o consumo energético. Este valor poderá ser positivo, isto quando a produção excede o consumo, ou então negativo, quando se verifica o contrário. No conjunto de dados criados para o desenvolvimento deste estudo, verifica-se que existe no global um maior consumo de energia comparativamente à produção, sendo esses valores de 64% e 36% respetivamente, como demonstra o gráfico da figura 21.

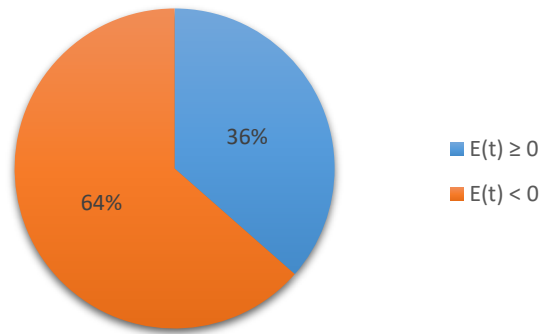


Figura 21 - Gráfico comparativo dos valores de $E(t)$

Todos os dados relativos aos membros criados podem ser consultados no Anexo B.

5.2. COMUNIDADE DE ENERGIA RENOVÁVEL - NORTE DE PORTUGAL

A comunidade de energia renovável caracteriza-se por ser uma comunidade mista, onde existem cerca de 14 centros electroprodutores (energia fotovoltaica) e cerca de 63 membros consumidores. Dentro deste conjunto de membros, estão associadas várias tipologias de instalações, desde residenciais até comerciais e industriais. Por questões de confidencialidade não foi possível identificar os associados individualmente, assim a entidade responsável pela CER e a sua localização exata.

Como já referido anteriormente a proximidade entre as UPAC e as IU constitui um requisito para o exercício da atividade de produção do autoconsumo, pelo que em particular nesta CER, todas as UPAC se encontram ligadas às redes de distribuição de energia elétrica em Baixa Tensão, e a distância geográfica entre as UPAC e as IU não é superior a 2km [17].

A potência de ligação encontra-se limitada a 50 % da potência contratada das IU com perfil de consumo em BTN e 50 % da potência requisitada das IU para outros perfis de consumo, pelo que a potência total de ligação referente ao conjunto de instalações fotovoltaicas da

CER é cerca de 309 kWn, o que apresenta menos de metade do total das potências certificadas das IU que neste caso é 660,06 KVA, como se pode verificar na tabela 7 e 8, onde estão descritas as características das instalações elétricas de cada membro.

Nº Consumidor	Tipologia da Instalação	Potência Contratada (kVA)	Consumo Anual Estimado (MWh/Ano)
1	BTN	3,45	6,144
2	BTN	6,90	8,244
3	BTN	6,90	0,768
4	BTE	41,40	0,000
5	BTN	6,90	7,902
6	BTN	20,70	16,852
7	BTN	3,45	3,552
8	BTN	6,90	2,400
9	BTN	4,60	3,252
10	BTN	6,90	0,348
11	BTE	41,41	68,196
12	BTN	27,60	19,848
13	BTN	13,80	5,296
14	BTN	3,45	0,365
15	BTN	3,45	9,012
16	BTN	20,70	3,960
17	BTN	20,70	17,412
18	BTN	20,70	5,700
19	BTN	10,35	0,540
20	BTN	3,45	0,060
21	BTN	3,45	1,188
22	BTN	10,35	1,023
23	BTN	10,35	2,304
24	BTN	6,90	5,280
25	BTN	10,35	4,140
26	BTN	3,45	2,148
27	BTN	3,45	1,572
28	BTN	10,35	3,528
29	BTN	6,90	3,672
30	BTN	10,35	3,072
31	BTN	1,15	0,108
32	BTN	1,15	0,084
33	BTN	1,15	0,144

34	BTN	3,45	1,548
35	BTN	17,25	1,560
36	BTN	3,45	1,536
37	BTN	3,45	0,680
38	BTN	6,90	0,240
39	BTN	6,90	0,024
40	BTN	3,45	0,480
41	BTE	77,00	118,416
42	BTN	6,90	6,096
43	BTN	3,45	2,772
44	BTN	41,4	11,148
45	BTN	4,6	9,572
46	BTN	4,6	6,960
47	BTN	13,8	4,005
48	BTN	3,45	0,180
49	BTN	3,45	0,048
50	BTN	3,45	1,536
51	BTN	3,45	0,084
52	BTN	3,45	0,804
53	BTN	3,45	0,408
54	BTN	3,45	0,540
55	BTN	17,25	2,088
56	BTN	17,25	8,016
57	BTN	17,25	3,084
58	BTN	3,45	0,744
59	BTN	3,45	0,780
60	BTN	6,9	3,451
61	BTN	6,9	3,789
62	BTN	10,35	3,511
63	BTN	3,45	3,108
Total	-	660,06	401,811

Tabela 8 - Características das instalações elétricas de cada consumidor

O modelo de repartição energético implementado na CER é a repartição baseada nos consumos onde toda a energia produzida e consumida pelos associados é monitorizada a cada 15 minutos pela ORD, sendo posteriormente os excedentes de cada centro electroprodutor distribuídos pelos restantes membros da CER de acordo com os respetivos consumos.

Os dimensionamentos das instalações fotovoltaicas tiveram em consideração o perfil estimado de consumo de energia elétrica relativo ao conjunto dos membros da CER, garantindo que quase a totalidade ou parte da energia produzida, seja consumida dentro da mesma, visando o autoconsumo e a partilha de energia. Foi necessário calcular também a energia produzida anualmente, pelo que foi necessário estimar o número médio de horas de radiação solar por dia na região considerada, o que permitiu calcular a energia média produzida anualmente por cada centro electroprodutor e consequentemente pela CER, totalizando um valor de 512,9 MWh/Ano, como se pode verificar na tabela 8 e no gráfico da figura 22. Para o cálculo do consumo médio anual da CER, foi necessário ter acesso ao consumo de cada membro, sendo que este foi disponibilizado pelos mesmos através das respetivas faturas de energia e totalizou um valor de 401,8 MWh/Ano, como se pode verificar na tabela 8 e no gráfico da figura 23.

Nº Produtor	Tipologia da Instalação	Potência Contratada (kVA)	Potência Instalada (kWp)	Potência de Ligação (kWn)	Produção Anual Estimada (MWh/Ano)
1	BTN	41,40	40	40	62,2
2	BTN	6,90	6	5	9,3
3	BTE	41,41	80	80	124,4
4	BTN	27,60	25	20	38,9
5	BTN	13,80	13	10	20,2
6	BTN	20,70	20	20	31,1
7	BTE	77,00	90	80	139,9
8	BTN	4,6	4	5	6,2
9	BTN	20,70	20	20	31,1
10	BTN	6,90	6	5	9,3
11	BTN	6,90	6	5	9,3
12	BTN	6,90	6	5	9,3
13	BTN	10,35	4	4	6,2
14	BTN	10,35	10	10	15,5
Total	-	295,51	330	309	512,9

Tabela 9 - Características das instalações elétricas de cada produtor

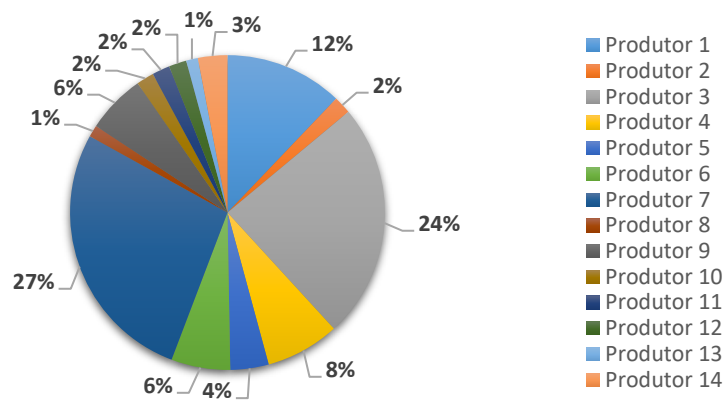


Figura 22 – Gráfico da Produção de Energia Estimada dos Centros Eletroprodutores

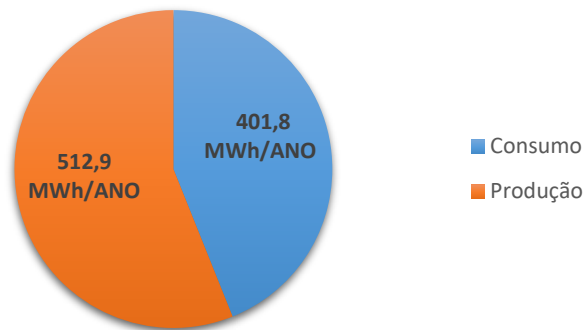


Figura 23 - Gráfico do Consumo e Produção de Energia Estimados da CER

Esta breve caracterização e análise da base de dados, irá servir para testar a metodologia desenvolvida, ou seja, o modelo de agrupamento e classificação de CER. Esta análise é fundamental, uma vez que irá possibilitar uma interpretação mais fácil dos resultados obtidos nas simulações, para além de permitir prever e até confirmar os mesmos.

5.3. SIMULAÇÕES E RESULTADOS

Para ter uma base de comparação igualitária entre todos os algoritmos foi necessário pré-definir algumas variáveis antes de executar os vários algoritmos, nomeadamente a distância máxima entre membros para que este possam pertencer ao mesmo cluster para

o algoritmo *DBSCAN* e pré-definir número de clusters para o algoritmo *K-MEANS* e os *Linkage Algorithms*.

O gráfico representado pela figura 24, contém os vários membros da base de dados que foi utilizada para a realização deste estudo. Esta base de dados é constituída por membros criados artificialmente e membros reais, relativos à Comunidade de Energia Renovável do Norte de Portugal. O respetivo gráfico permite verificar como estão distribuídos os diferentes membros a nível espacial e a nível energético, uma vez que todos os membros se encontram representados num referencial cartesiano de 3 dimensões, onde “X” representa a distância em latitude, “Y” a distância em longitude e “Z” a energia correlacionada. Pode-se também verificar que no global não existe nenhum aglomerado de pontos que se destaque pelo valor de energia líquida, uma vez que o gradiente de cores se encontra uniformemente distribuído.

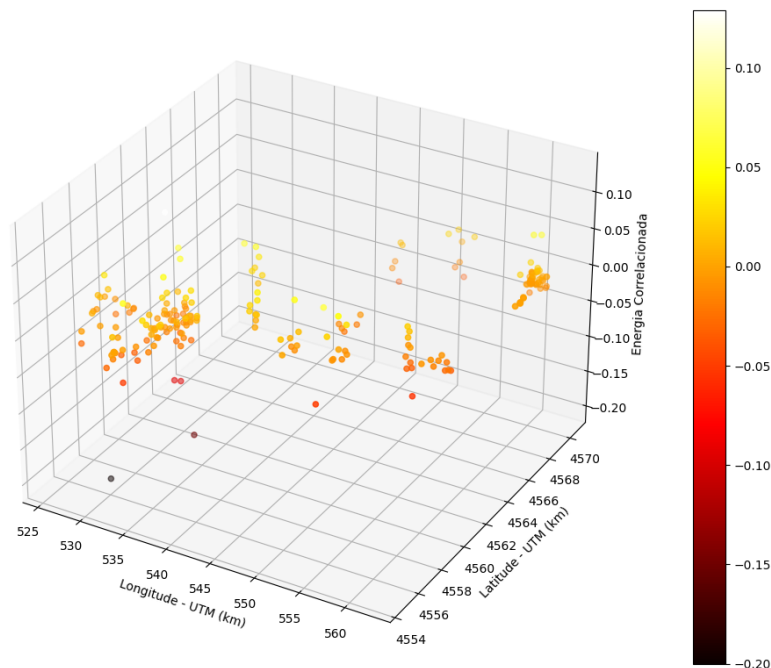


Figura 24 - Gráfico comparativo do valor de energia líquida dos membros

5.3.1. DBSCAN

O estudo foi desenvolvido tendo em consideração que todas as UPAC se encontram ligadas às redes de distribuição de energia elétrica em Baixa Tensão, pelo que segundo o Decreto-Lei n.º 15/2022, a distância geográfica entre as UPAC e as IU não pode ser superior a 2 km [17]. Para além disso, para que uma CER se possa formar, apenas são necessários dois membros, sendo assim definiu-se como parâmetros de entrada para o algoritmo DBSCAN: raio de ϵ -vizinhança de um ponto = 2 e número mínimo de pontos (η) = 2. Como se pode observar no gráfico e tabela representados na figura 25, o algoritmo DBSCAN agrupou o conjunto de membros em 10 comunidades, na qual existem 4 comunidades classificadas como não autossuficientes e 6 comunidades classificadas como autossuficientes, sendo que entre as mesmas existem 7 classificadas como mistas, 2 classificadas como homogéneas positivas e 1 classificada como homogénea negativa. Também se pode verificar que todos os membros da base de dados foram agrupados numa comunidade, não existindo assim pontos de ruído. E que o valor da energia líquida das CER varia entre os -497 MWh/Ano e os 249 MWh/Ano.

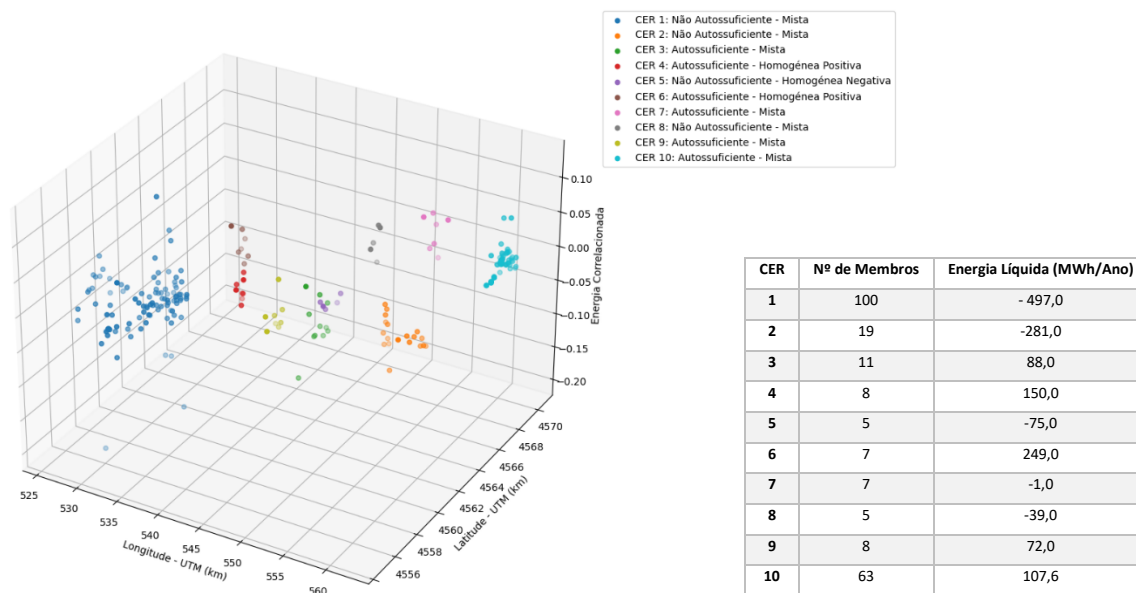


Figura 25 - Gráfico e tabela da distribuição e classificação das CER (DBSCAN)

5.3.2. K-MEANS

O único dado de entrada do algoritmo *K-MEANS* para além dos dados referentes aos membros, é o número de *clusters*, e como já referido anteriormente este número será igual em todos os algoritmos, de modo a conseguir comparar os diferentes dados obtidos de forma coerente. Sendo assim, definiu-se o número de clusters igual a 10 como o obtido no algoritmo *DBSCAN*. Os resultados obtidos pelo algoritmo *K-MEANS* foram bastante interessantes, pelo que se verificou uma distribuição dos membros pelas CER mais equilibrada a nível quantitativo, comparativamente com o algoritmo *DBSCAN*. Como já referido anteriormente, o algoritmo *K-MEANS* utiliza a técnica de agrupamento particional, ou seja, é um algoritmo não determinístico, e como tal pode gerar resultados diferentes a cada execução, tendo por base os mesmos dados de entrada. Grande parte dos resultados obtidos por este algoritmo foram muito semelhantes aos obtidos pelo método *Ward*, no entanto obtiveram-se alguns resultados distintos de todos os outros algoritmos em estudo, uma vez que em algumas simulações este não foi capaz de identificar a CER real, como se pode verificar no exemplo demonstrado no gráfico e tabela da figura 26. Verificou-se também que todos membros foram inseridos numa CER, uma vez que o algoritmo *K-MEANS* não deteta pontos de ruído.

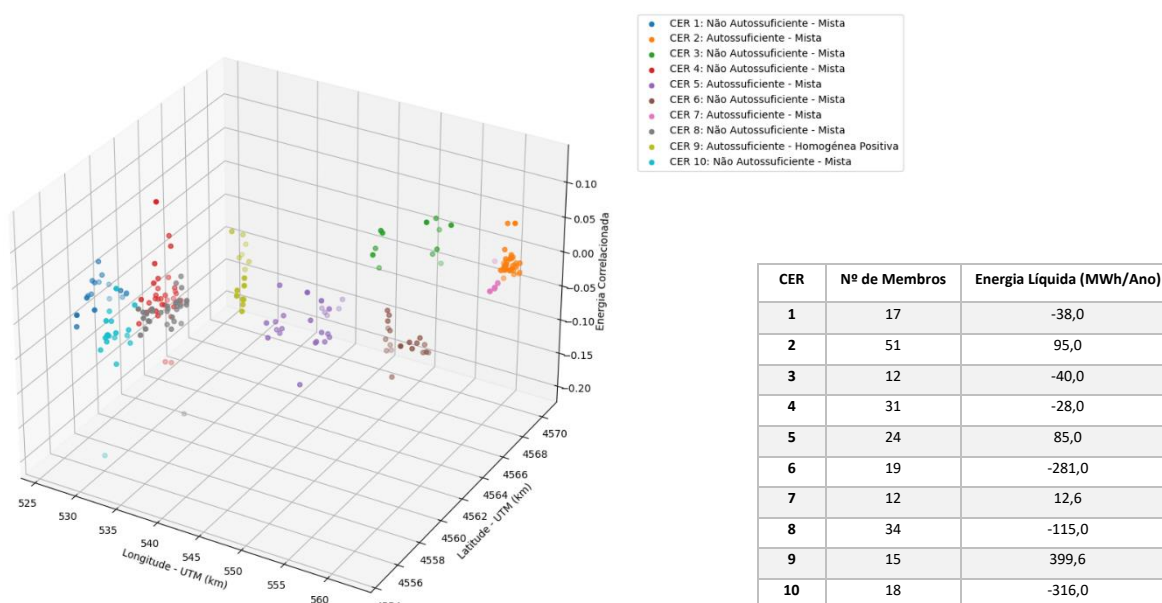


Figura 26 - Gráfico e tabela da distribuição e classificação das CER (*K-MEANS*)

5.3.3. LINKAGE ALGORITHMS

À semelhança do algoritmo *K-MEANS*, os *linkage algorithms* apenas necessitam do número de *clusters* para executarem o agrupamento dos dados, e como tal este número também será igual a 10. O método *Ward* formou 4 CER autossuficientes e 6 CER não autossuficientes, de entre as quais apenas uma foi classificada como homogénea positiva e as restantes como mistas. Como se pode verificar no gráfico da figura 28, este método foi o que formou comunidades mais similares, comparativamente com os outros métodos, tendo em consideração o número de membros, não existindo a formação de grandes aglomerados. No entanto a nível energético, foi o método que formou o maior número de comunidades não autossuficientes. Os respetivos valores de energia líquida de cada CER variaram entre os -347 MWh/no e os 399 MWh/Ano.

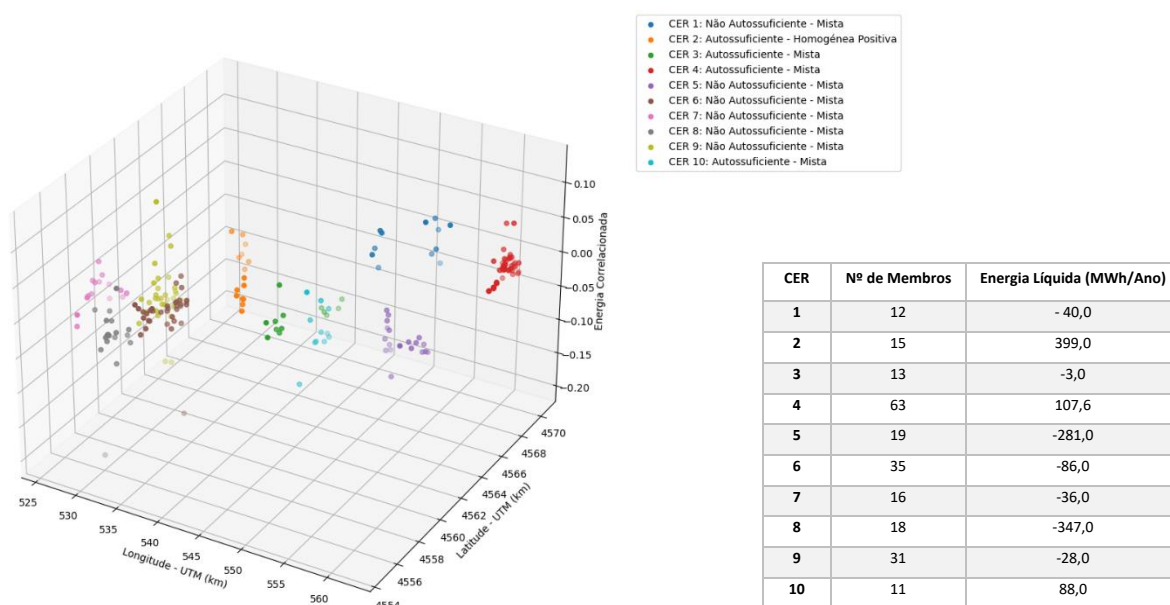


Figura 27 - Gráfico e tabela da distribuição e classificação das CER (Ward's Method)

O método *complete-link* formou 5 CER autossuficientes e 5 CER não autossuficientes, de entre as quais duas foram classificadas como homogéneas positivas e as restantes como mistas. Pode-se verificar na tabela da figura 29 que os valores das energias líquidas das CER formadas por este método são os que menos variam, pelo que pertencem ao menor intervalo de valores, nomeadamente [-281 , 249] MWh/Ano.

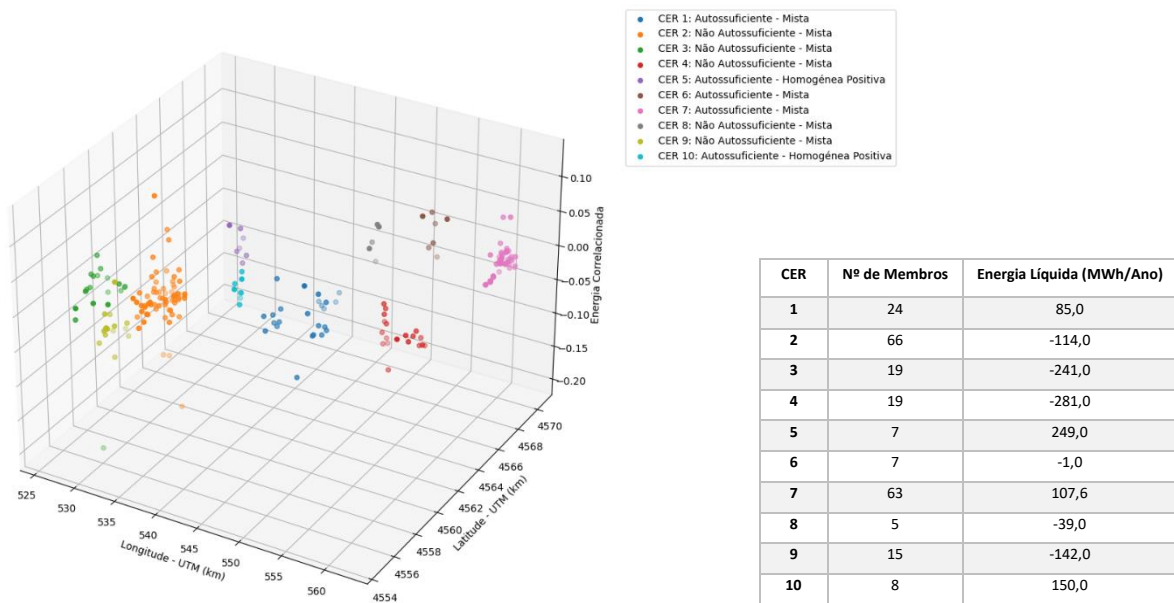


Figura 28 - Gráfico e tabela da distribuição e classificação das CER (Complete-Link)

O método *average-link* formou 6 CER autossuficientes e 4 CER não autossuficientes, de entre as quais duas foram classificadas como homogêneas positivas e as restantes como mistas. Este método foi o que conseguiu formar mais comunidade autossuficientes juntamente com o método *single-link*. Para além disso foi o método que obteve as CER com o valor de energia líquida mais próxima de 0, como se pode verificar na tabela da figura 30.

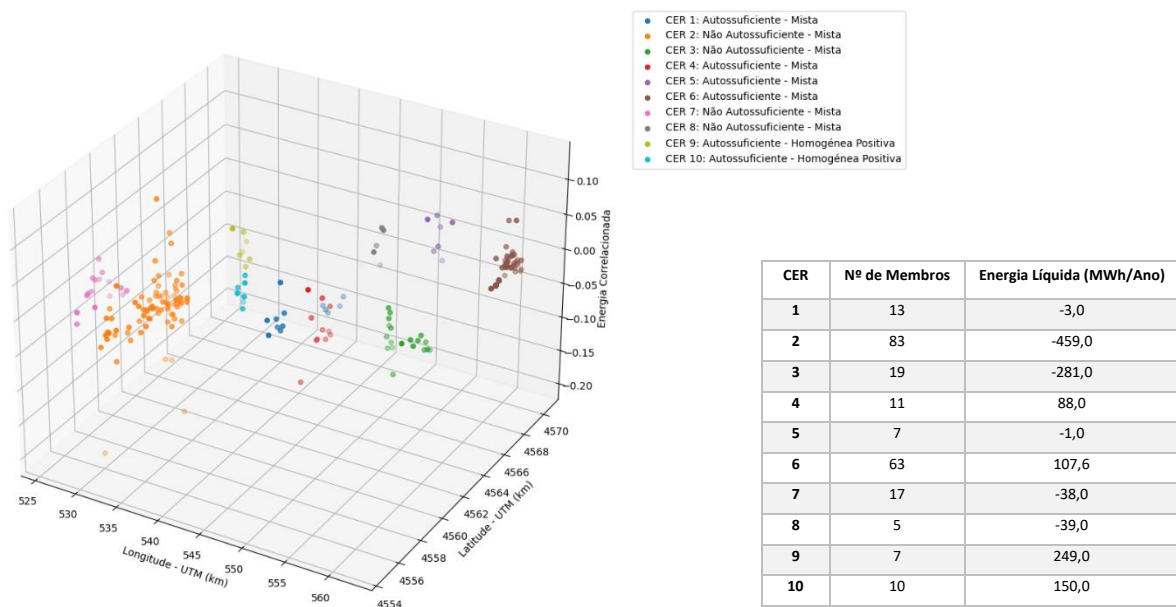


Figura 29 - Gráfico e tabela da distribuição e classificação das CER (Average-Link)

O método *single-link* formou exatamente o mesmo número de comunidades autossuficientes e não autossuficientes do que o método *average-link*, no entanto com uma pequena diferença na classificação, uma vez que formou 2 CER homogêneas positivas, 1 CER homogênea negativa e as restantes mistas, como se pode verificar no gráfico da figura 30. Este método tem várias semelhanças como algoritmo *DBSCAN*, uma vez que ambos foram os que conseguiram encontrar mais CER homogêneas, para além de formarem a CER com o maior número de membros e com a menor energia líquida.

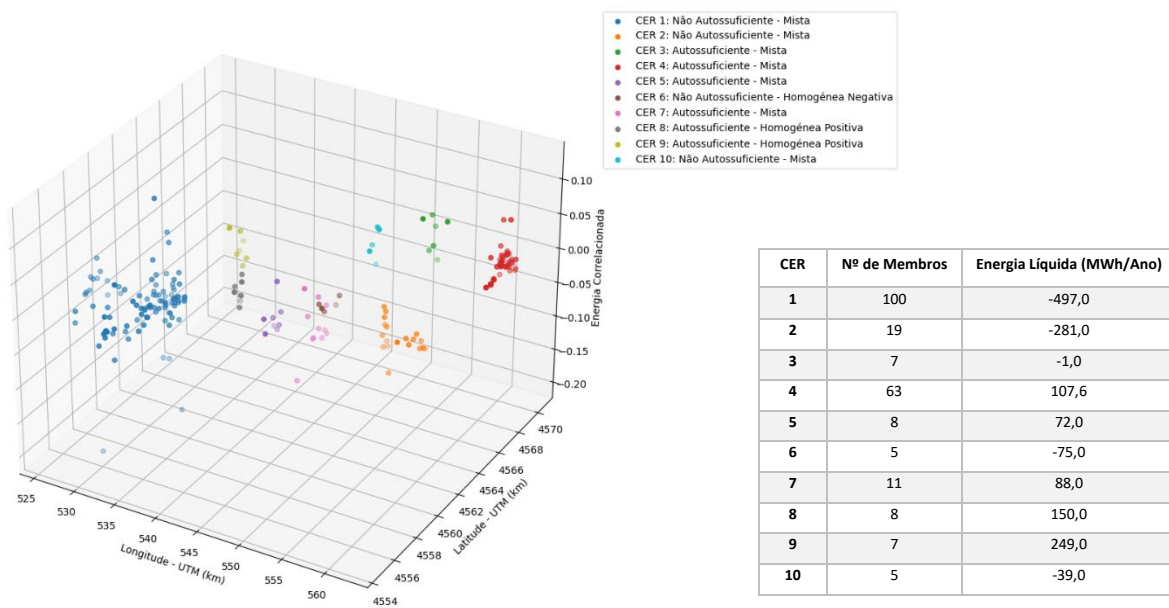


Figura 30 - Gráfico e tabela da distribuição e classificação das CER (*Single-Link*)

5.4. MÉTRICAS DE AVALIAÇÃO DE AGRUPAMENTO

A avaliação correta do desempenho dos algoritmos de agrupamento é fundamental. Em muitos casos, uma avaliação manual e qualitativa torna-se insuficiente, pelo que se recorre geralmente a métricas de avaliação do desempenho do agrupamento de forma a obter-se resultados mais precisos e facilmente comparáveis. Com o objetivo de identificar quais os algoritmos que geram melhores resultados de agrupamento, os vários algoritmos em estudo foram submetidos a três métricas de avaliação:

Pontuação de silhueta: A pontuação de silhueta é utilizada para medir a distância de separação entre clusters. Apresenta uma medida da proximidade de cada ponto de um aglomerado a pontos nos aglomerados vizinhos. A pontuação de silhueta retorna o coeficiente médio de silhueta aplicado em todas as amostras. O Coeficiente de Silhueta é calculado utilizando a média da distância do intra-agrupamento e do aglomerado mais próximo para todas as amostras, este coeficiente varia entre $[-1,1]$. Quanto maior for o Coeficiente de Silhueta (quanto mais próximo de 1), maior é a separação entre os aglomerados. Se o valor for 0, indica que a amostra está dentro ou muito perto do limite de decisão entre dois aglomerados vizinhos, enquanto um valor negativo indica que essas amostras podem ter sido atribuídas ao aglomerado errado [49].

Índice Calinski Harabaz: O índice de Calinski Harabaz baseia-se no princípio da razão da variância. Esta razão é calculada entre a soma da dispersão inter-cluster e a soma da dispersão intra-cluster para todos os clusters (onde a dispersão é a soma das distâncias ao quadrado). Quanto maior for o índice, melhor é o agrupamento [50].

Índice de Davies Bouldin: O Índice de Davies-Bouldin é definido como a medida de similaridade média de cada cluster com seu cluster mais semelhante, onde a similaridade é a razão entre as distâncias dentro do cluster e as distâncias entre os clusters. Desta forma, os aglomerados que estão mais afastados e menos dispersos levarão a uma melhor pontuação. A pontuação mínima é zero, e diferentemente da maioria das métricas de desempenho, quanto menor for o valor, melhor será o desempenho do aglomerado [51].

Ao analisar as classificações obtidas pelos vários algoritmos em estudo, verificaram-se diferenças significativas no desempenho. O algoritmo que conseguiu obter os Coeficientes de Silhueta mais próximos de 1 foi o método Average-Link, bem como o menor Índice Davies Bouldin, pelo que foi o que obteve a melhor classificação nestas duas métricas. O Índice Calinski Harabaz mais elevado foi obtido pelo Método Ward's e DBSCAN, pelo que foram, conseqüentemente, os que obtiveram a melhor classificação nesta métrica. O algoritmo que teve o pior desempenho global foi o K-MEANS. As classificações obtidas pelos algoritmos de agrupamento em estudo podem ser analisadas na Tabela 7.

	Silhueta	Calinski Harabaz	Davies Bouldin
K-MEANS	0.568	1951.153	0.578
DBSCAN	0.653	897.255	0.329
Ward's Method	0.619	2124.174	0.590
Complete-Link	0.631	1980.555	0.451
Average-Link	0.671	1760.073	0.429
Single-Link	0.653	897.255	0.329

Tabela 10 - Avaliação do desempenho dos algoritmos de agrupamento

5.5. ASPETOS CONCLUSIVOS

Ao analisar os resultados obtidos a partir das simulações realizadas no modelo desenvolvido, é possível apontar uma série de conclusões referentes aos agrupamentos realizados pelos diferentes algoritmos, nomeadamente no que diz respeito à quantidade de membros pertencentes a cada CER e à quantidade de HEC, MEC e SEC formadas por cada algoritmo. Também foi possível por à prova a metodologia de classificação desenvolvida, após a contabilização da energia líquida de cada CER. Após uma análise comparativa, evidenciou-se quais os algoritmos que identificaram corretamente a CER real, e para além disso confirmou-se a pré-análise realizada a esta CER em concreto. Por fim, foi posta à prova a qualidade dos agrupamentos formados pelos algoritmos em estudo, com o auxílio de várias métricas de avaliação.

6. CONCLUSÕES E TRABALHOS FUTUROS

Neste último capítulo, são abordadas as conclusões obtidas ao longo da realização desta dissertação, incluindo propostas de melhoria ao trabalho e futuros desafios.

O estudo das comunidades de energia renovável formadas por recursos energéticos distribuídos irá facilitar seguramente a gestão energética da rede elétrica, devido à possibilidade de cooperação entre membros e até entre as comunidades. Nesta dissertação, propôs-se uma série de metodologias para o agrupamento e identificação dos diferentes tipos de CER, nomeadamente as HEC, MEC e SEC. Também foi possível validar a eficácia e eficiência das abordagens utilizando um conjunto de dados formado por dados reais e dados criados.

O *DBSCAN* e o *single-link* revelaram ser os algoritmos que conseguiram identificar o maior número de HEC entre o conjunto de membros. Estes algoritmos poderão fornecer uma importante ajuda na gestão energética da rede, uma vez que poderão prever por exemplo qual a melhor localização para a instalação de novos recursos com grande capacidade de fornecimento energético para balancear HEC com grandes consumos de energia, fazendo com que não seja necessário solicitar grandes quantidades de energia à rede elétrica.

No entanto, o algoritmo *DBSCAN* mostrou-se ineficiente ao agrupar grandes aglomerados de elevada densidade, uma vez que o algoritmo apenas considera a distância entre cada membro, e não a distância entre os membros mais distantes no *cluster*. Por esta mesma razão, verificou-se que nem todas as CER formadas por este algoritmo satisfizeram a condição da distância máxima entre as UPAC e as IU.

Relativamente à comunidade de energia renovável real, não houve problema para a maior parte dos algoritmos em estudo na identificação da mesma, com exceção do algoritmo *K-MEANS*, uma vez que este apresenta uma elevada sensibilidade ao ruído. O modelo desenvolvido classificou a CER real como autossuficiente mista, o que confirmou a pré-análise realizada em concreto para esta CER.

Pôde-se concluir que os *linkage-algorithms* foram os que conseguiram obter resultados mais satisfatórios. Estes conseguiram formar o maior número de comunidades autossuficientes, cerca de 6 SEC. Também foram os algoritmos que obtiveram CER com os valores de energias líquidas mais próximas de 0, para além disso, identificaram o maior número de HEC.

Fazendo uma análise mais pormenorizada e qualitativa dos agrupamentos formados por cada algoritmo, os *linkage-algorithms* também se destacaram pela positiva, uma vez que obtiveram as melhores classificações nas métricas de avaliação a que foram submetidos: 0.631 na Pontuação de Silhueta, 2124.174 no índice de Calinski Harabaz e 0.329 no Índice de Davies-Bouldin.

O método de classificação das CER revelou-se eficiente, conseguindo classificar todas as CER formadas de forma correta. A classificação de cada CER, baseada na soma das energias líquidas de todos os membros constituintes da mesma, será no futuro uma informação

fundamental, quando novos membros se quiserem associar a CER já existentes ou mesmo na formação de novas CER adjacentes, uma vez que irá facilitar toda a gestão energética dentro das próprias CER e mesmo na RESP, permitindo assim um melhor balanceamento das cargas.

Ao longo desta dissertação foi abordada uma metodologia de auxílio ao dimensionamento de comunidades de energia renovável. O estudo desenvolvido poderá ser alvo de melhorias e novas abordagens. O modelo desenvolvido, capaz de formar e classificar comunidades de energia renovável, poderá ser melhorado com a criação de novas funcionalidades de forma a diminuir ou até eliminar as limitações verificadas. Sugere-se:

- O desenvolvimento de uma metodologia de planeamento, que, face às adversidades observadas na CER, sugira determinadas alterações que possam vir a ser efetuadas para melhoria da mesma (ex: instalação de novas UPAC);
- A integração de novos algoritmos de agrupamento ao modelo, assim como o aumento da base de dados e das características energéticas de cada membro;
- A realização de uma análise mais pormenorizada a cada membro, de forma a identificar o perfil de consumo/produção diário, com o objetivo de formar CER mais equilibradas energeticamente;
- O desenvolvimento de uma metodologia de dimensionamento de grupos de baterias a integrar nas UPAC existentes, de forma a colmatar o défice energético verificado em certos momentos do dia.

Por fim, vale destacar que do trabalho desta dissertação resultaram dois artigos científicos, os dois encontram-se em fase de submissão, um na revista científica *Energies* e outro para a conferência *IEEE PES GM 2023*.

REFERÊNCIAS

- [1] B. Canizes, J. Soares, Z. Vale, and J. M. Corchado, "Optimal distribution grid operation using DLMP-based pricing for electric vehicle charging infrastructure in a smart city," *Energies (Basel)*, vol. 12, no. 4, Feb. 2019, doi: 10.3390/en12040686.
- [2] P. Faria and Z. Vale, "Demand response in electrical energy supply: An optimal real time pricing approach," *Energy*, vol. 36, no. 8, pp. 5374–5384, 2011, doi: 10.1016/j.energy.2011.06.049.
- [3] C. Silva, P. Faria, A. Fernandes, and Z. Vale, "Clustering distributed Energy Storage units for the aggregation of optimized local solar energy," *Energy Reports*, vol. 8, pp. 405–410, Jun. 2022, doi: 10.1016/j.egyr.2022.01.043.
- [4] D. de São José, P. Faria, and Z. Vale, "Smart energy community: A systematic review with metanalysis," *Energy Strategy Reviews*, vol. 36, Jul. 2021, doi: 10.1016/j.esr.2021.100678.
- [5] J. Kennedy, P. Ciufo, and A. Agalgaonkar, "Intelligent load management in microgrids," 2012. doi: 10.1109/PESGM.2012.6345729.
- [6] B. Canizes, V. Silveira, and Z. Vale, "Demand response and dispatchable generation as ancillary services to support the low voltage distribution network operation," *Energy Reports*, vol. 8, pp. 7–15, Jun. 2022, doi: 10.1016/j.egyr.2022.01.040.
- [7] N. Rahbari-Asr, U. Ojha, Z. Zhang, and M. Y. Chow, "Incremental welfare consensus algorithm for cooperative distributed generation/demand response in smart grid," *IEEE Trans Smart Grid*, vol. 5, no. 6, 2014, doi: 10.1109/TSG.2014.2346511.
- [8] Y. Hong, J. Vaidya, H. Lu, P. Karras, and S. Goel, "Collaborative Search Log Sanitization: Toward Differential Privacy and Boosted Utility," *IEEE Trans Dependable Secure Comput*, vol. 12, no. 5, 2015, doi: 10.1109/TDSC.2014.2369034.

- [9] W. Saad, Z. Han, H. V. Poor, and T. Başar, “Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications,” *IEEE Signal Processing Magazine*, vol. 29, no. 5. 2012. doi: 10.1109/MSP.2012.2186410.
- [10] B. Canizes, J. Soares, F. Lezama, C. Silva, Z. Vale, and J. M. Corchado, “Optimal expansion planning considering storage investment and seasonal effect of demand and renewable generation,” *Renew Energy*, vol. 138, pp. 937–954, Aug. 2019, doi: 10.1016/j.renene.2019.02.006.
- [11] R. Barreto, C. Gonçalves, L. Gomes, P. Faria, and Z. Vale, “Evaluation Metrics to Assess the Most Suitable Energy Community End-Users to Participate in Demand Response,” *Energies (Basel)*, vol. 15, no. 7, 2022, doi: 10.3390/en15072380.
- [12] Y. Hong, S. Goel, H. Lu, and S. Wang, “Discovering energy communities for microgrids on the power grid,” in *2017 IEEE International Conference on Smart Grid Communications, SmartGridComm 2017*, 2018, vol. 2018-January. doi: 10.1109/SmartGridComm.2017.8340661.
- [13] M. Coscia, F. Giannotti, and D. Pedreschi, “A classification for community discovery methods in complex networks,” *Statistical Analysis and Data Mining*, vol. 4, no. 5. 2011. doi: 10.1002/sam.10133.
- [14] J. Yang, J. McAuley, and J. Leskovec, “Community detection in networks with node attributes,” 2013. doi: 10.1109/ICDM.2013.167.
- [15] DGEG, “Plano Nacional Energia e Clima 2021-2030 (PNEC 2030),” *Relatório Técnico*. 2019.
- [16] *Decreto-Lei n.º 162/2019 de 25 de outubro*. 2019. Accessed: May 15, 2022. [Online]. Available: <https://files.dre.pt/1s/2019/10/20600/0004500062.pdf>
- [17] Presidência do Conselho de Ministros, “Decreto-Lei n.º 15/2022 de 14 de janeiro,” *Diário da República*, 1.ª série, 2022, [Online]. Available: <https://files.dre.pt/1s/2022/01/01000/0000300185.pdf>

- [18] Guilherme Machado Figueiredo, "Gestão de Autoconsumo de Comunidades de Energia Renovável em Núcleos Rurais," FEUP, Porto, 2021.
- [19] Parlamento Europeu e do Conselho, "Diretiva (UE) 2018/2001 do Parlamento Europeu e do Conselho de 11 de dezembro de 2018," *Jornal Oficial da União Europeia*, vol. 2001, 2018.
- [20] DGEG, "Regulamento Técnico e de Qualidade." 2019.
- [21] M. Ramires, "Conferência Comunidades de Energia," 2020.
- [22] A. J. D. Rathnayaka, V. M. Potdar, T. Dillon, O. Hussain, and S. Kuruppu, "Goal-oriented prosumer community groups for the smart grid," *IEEE Technology and Society Magazine*, vol. 33, no. 1, 2014, doi: 10.1109/MTS.2014.2301859.
- [23] R. Zafar, A. Mahmood, S. Razzaq, W. Ali, U. Naeem, and K. Shehzad, "Prosumer based energy management and sharing in smart grid," *Renewable and Sustainable Energy Reviews*, vol. 82. 2018. doi: 10.1016/j.rser.2017.07.018.
- [24] I. G. Ciuciu, R. Meersman, and T. Dillon, "Social network of smart-metered homes and SMEs for grid-based renewable energy exchange," *IEEE International Conference on Digital Ecosystems and Technologies*, 2012, doi: 10.1109/DEST.2012.6227922.
- [25] F. Skopik and C. Wagner, "Novel energy saving opportunities in smart grids using a secure social networking layer," 2012. doi: 10.1109/COMPSAC.2012.75.
- [26] J. Duarte, "Gestão de Repartição e Transação de Valor em Comunidades de Energia Renovável," 2021.
- [27] G. J. A. P. M. Figueiredo, "Gestão de Autoconsumo de Comunidades de Energia Renovável em Núcleos Rurais," 2021, [Online]. Available: <https://repositorio-aberto.up.pt/handle/10216/133041>
- [28] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid - The new and improved power grid: A survey," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 4. 2012. doi: 10.1109/SURV.2011.101911.00087.

- [29] S. F. Bush, S. Goel, and G. Simard, "IEEE Vision for Smart Grid Communications: 2030 and Beyond Roadmap," *standard-download.org*, 2013.
- [30] "2002 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.02CH37309)," 2011. doi: 10.1109/pesw.2002.984925.
- [31] N. Carlisle, O. van Geet, and S. Pless, "Definition of a " Zero Net Energy " Community," *National Renewable Energy Laboratory*, no. November, 2009.
- [32] S. Xie *et al.*, "Discovering communities for microgrids with spatial-temporal net energy," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 6, 2019, doi: 10.1007/s40565-019-0543-4.
- [33] B. M. Varghese, a Unnikrishnan, and P. J. K, "Spatial Clustering Algorithms - an Overview," *Asian Journal of Computer Science And Information Technology*, vol. 1, 2013.
- [34] K. Arvai, "K-Means Clustering in Python: A Practical Guide," 2020. <https://realpython.com/k-means-clustering-python/> (accessed Jun. 03, 2022).
- [35] C. F. Olson, "Parallel algorithms for hierarchical clustering," *Parallel Comput*, vol. 21, no. 8, 1995, doi: 10.1016/0167-8191(95)00017-I.
- [36] S. I. Sultana, "HOW THE HIERARCHICAL CLUSTERING ALGORITHM WORKS," 2020. <https://dataaspirant.com/hierarchical-clustering-algorithm/> (accessed Jun. 11, 2022).
- [37] I. Dabbura, "K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks," 2018. <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a> (accessed Jun. 04, 2022).
- [38] D. H. F. Nunes, "Um breve estudo sobre o algoritmo K-means," Universidade de Coimbra, 2016.

- [39] S. Cuomo, V. de Angelis, G. Farina, L. Marcellino, and G. Toraldo, "A GPU-accelerated parallel K-means algorithm," *Computers and Electrical Engineering*, vol. 75, 2019, doi: 10.1016/j.compeleceng.2017.12.002.
- [40] I. D. M. Ventrone and D. Luchi, "A biased sampling method for applying DBSCAN," 1973.
- [41] N. S. Chauhan, "DBSCAN Clustering Algorithm in Machine Learning," 2022. <https://www.kdnuggets.com/2020/04/dbscan-clustering-algorithm-machine-learning.html> (accessed Jun. 04, 2022).
- [42] É. M. Gava *et al.*, "O Algoritmo Density-Based Spatial Clustering of Applications With Noise (DBSCAN) na Clusterização dos Indicadores de Dados Ambientais," vol. d, 2013.
- [43] "Difference between K-Means and DBScan Clustering," 2020. <https://www.geeksforgeeks.org/difference-between-k-means-and-dbscan-clustering/> (accessed Jun. 04, 2022).
- [44] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN," *ACM Transactions on Database Systems*, vol. 42, no. 3, 2017, doi: 10.1145/3068335.
- [45] S. Patel, S. Sihmar, and A. Jatain, "A study of hierarchical clustering algorithms," *2015 International Conference on Computing for Sustainable Global Development, INDIACom 2015*, vol. 3, no. 10, pp. 537–541, 2015.
- [46] N. Grozavu, "Thèse Classification Topologique pondérée : approches modulaires , hybrides et collaboratives," 2009.
- [47] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview, II," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 7, no. 6, 2017. doi: 10.1002/widm.1219.

- [48] A. M. Jarman, "Hierarchical cluster analysis: Comparison of single linkage, complete linkage, average linkage and centroid linkage method," *Georgia Southern University*, 2020.
- [49] J.-O. Palacio-Niño and F. Berzal, "Evaluation Metrics for Unsupervised Learning Algorithms," May 2019, [Online]. Available: <http://arxiv.org/abs/1905.05667>
- [50] A. Karna and K. Gibert, "Automatic identification of the number of clusters in hierarchical clustering," *Neural Comput Appl*, vol. 34, no. 1, pp. 119–134, Jan. 2022, doi: 10.1007/s00521-021-05873-3.
- [51] J. Xiao, J. Lu, and X. Li, "Davies Bouldin Index based hierarchical initialization K-means," *Intelligent Data Analysis*, vol. 21, no. 6, pp. 1327–1338, 2017, doi: 10.3233/IDA-163129.
- [52] AlindGupta, "ML | Types of Linkages in Clustering," 2021. <https://www.geeksforgeeks.org/ml-types-of-linkages-in-clustering/> (accessed Jun. 11, 2022).
- [53] S. Xie *et al.*, "Homogeneous and Mixed Energy Communities Discovery with Spatial-Temporal Net Energy," pp. 1–14, 2019, [Online]. Available: <http://arxiv.org/abs/1902.03916>

Anexo A - FLUXOGRAMAS DAS FUNÇÕES UTILIZADAS

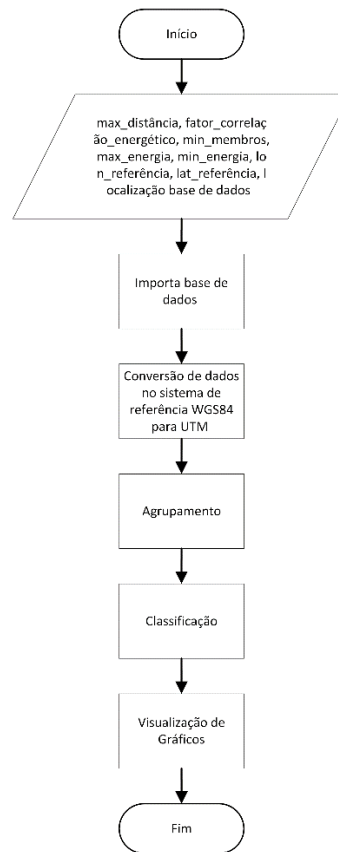


Figura A 1 - Fluxograma da Função Inicial

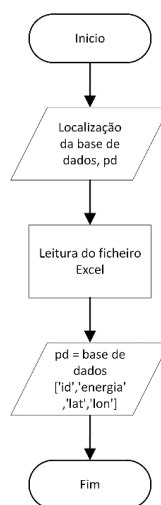


Figura A 2 - Fluxograma da Função Importar Dados

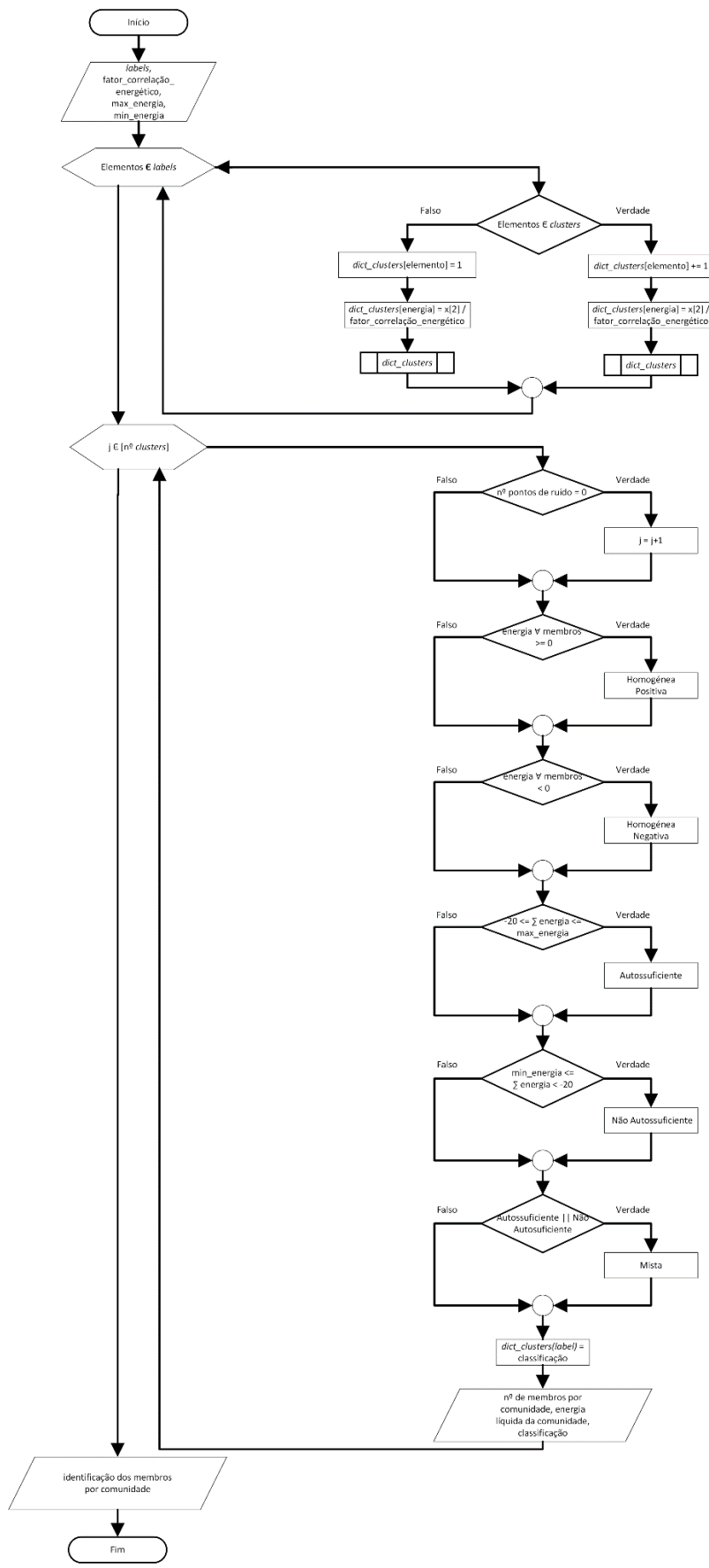


Figura A 3 - Fluxograma da Função Classificação

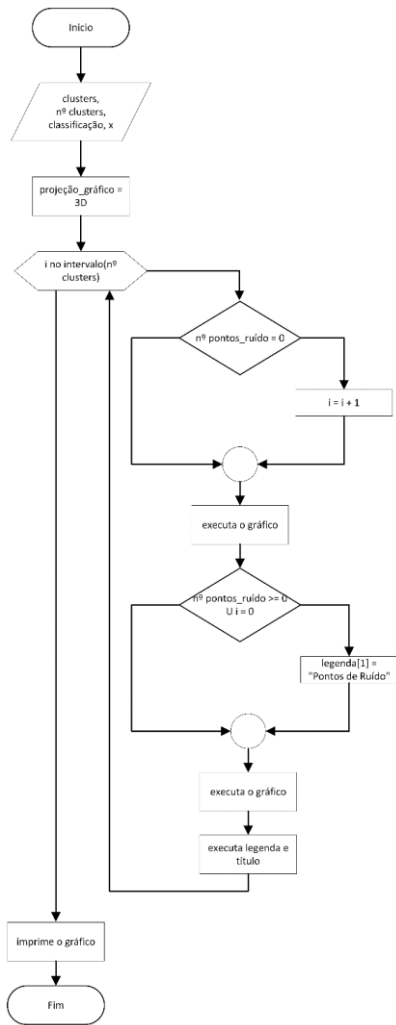


Figura A 4 - Fluxograma da Função Gráficos

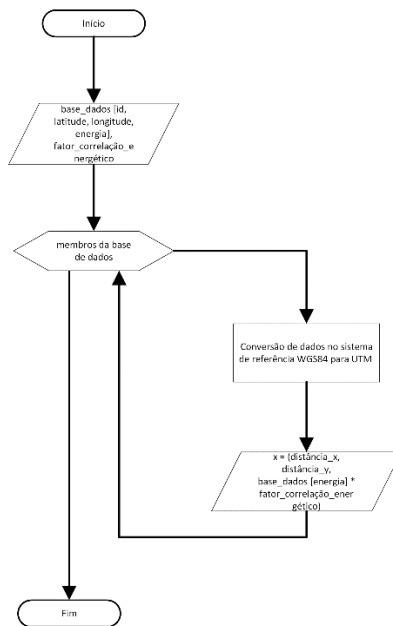


Figura A 5 - Fluxograma da Função Conversão de Dados

Anexo B - BASE DE DADOS

id	Produção Anual Estimada (MWh/Ano)	Consumo Anual Estimado (MWh/Ano)	Energia Líquida (MWh/Ano)	id	Produção Anual Estimada (MWh/Ano)	Consumo Anual Estimado (MWh/Ano)	Energia Líquida (MWh/Ano)
1	6	9	-3	34	20	35	-15
2	10	8	2	35	3	15	-12
3	40	50	-10	36	35	6	29
4	206	300	-94	37	46	20	26
5	200	150	50	38	12	6	6
6	365	350	15	39	15	10	5
7	234	330	-96	40	20	30	-10
8	64	0	64	41	25	25	0
9	70	50	20	42	14	20	-6
10	15	10	5	43	33	68	-35
11	20	25	-5	44	20	27	-7
12	40	45	-5	45	26	28	-2
13	60	33	27	46	41	11	30
14	5	14	-9	47	60	72	-12
15	0	27	-27	48	14	10	4
16	30	22	8	49	5	18	-13
17	25	30	-5	50	31	9	22
18	55	37	18	51	41	39	2
19	9	9	0	52	36	26	10
20	0	18	-18	53	13	21	-8
21	0	25	-25	54	29	45	-16
22	20	15	5	55	10	2	8
23	10	34	-24	56	35	20	15
24	50	39	11	57	17	34	-17
25	15	31	-16	58	42	15	27
26	39	43	-4	59	24	12	12
27	0	5	-5	60	7	32	-25
28	30	35	-5	61	60	99	-39
29	10	12	-2	62	130	1	129
30	12	20	-8	63	14	21	-7

31	5	13	-8	64	18	39	-21
32	10	14	-4	65	56	69	-13
33	8	16	-8	66	88	19	69
67	163	144	19	104	0	12	-12
68	0	3	-3	105	49	64	-15
69	0	200	-200	106	0	22	-22
70	50	82	-32	107	5	10	-5
71	0	33	-33	108	92	147	-55
72	10	25	-15	109	0	8	-8
73	20	22	-2	110	10	17	-7
74	5	4	1	111	70	97	-27
75	3	7	-4	112	25	10	15
76	11	13	-2	113	3	20	-17
77	25	19	6	114	57	89	-32
78	300	150	-150	115	43	33	10
79	5	33	-28	116	37	15	22
80	0	50	-50	117	0	35	-35
81	25	28	-3	118	7	70	-3
82	60	0	60	119	9	17	-8
83	33	0	33	120	350	400	-50
84	81	66	15	121	77	0	77
85	70	99	-29	122	59	35	24
86	44	33	11	123	8	15	-7
87	24	25	-1	124	3	10	-7
88	36	55	-19	125	15	14	1
89	5	30	-25	126	33	45	-12
90	4	12	-8	127	50	0	50
91	13	15	-2	128	55	26	29
92	10	20	-10	129	6	14	-8
93	11	16	-5	130	0	9	-9
94	2	8	-6	131	50	17	33
95	3	9	-6	132	26	10	16
96	7	0	7	133	35	31	4
97	10	0	10	134	19	10	9
98	0	8	-8	135	25	0	25
99	0	13	-13	136	189	151	38
100	15	0	15	137	77	55	22

101	50	78	-28	138	20	17	3
102	26	47	-21	139	12	25	-13
103	0	33	-33	140	30	53	-23
141	0	13	-13				
142	3	9	-6				
143	36	56	-20				
144	99	72	27				
145	188	117	71				
146	40	15	35				
147	271	215	56				
148	140	117	23				
149	179	155	24				
150	20	7	13				
151	150	188	-38				
152	125	99	26				
153	78	100	-22				
154	71	50	21				
155	50	47	3				
156	10	22	-12				
157	43	22	21				
158	20	33	-13				
159	60	73	-13				
160	35	12	13				
161	10	48	-38				
162	56	44	12				
163	20	14	6				
164	5	16	-11				
165	0	10	-10				
166	5	7	-2				
167	22	8	14				
168	55	0	55				
169	25	8	17				
170	33	30	3				

Tabela B 1 - Dados dos membros criados