Network Partitioning Algorithms for Electricity Consumer Clustering

by

Olamide Oladeji

B.Sc., Electrical & Electronic Engineering, Univ. of Ibadan (2015)

Submitted to the Institute for Data, Systems, and Society and the
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of

Master of Science in Technology and Policy

and

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2018

© Massachusetts Institute of Technology 2018. All rights reserved.

Author . . . .
Institute for Data, Systems, and Society
Department of Electrical Engineering and Computer Science
August 10, 2018

Certified by. . . . .
Professor Ignacio Perez-Arriaga
Visiting Professor, Sloan School of Management
Professor, Electrical Engineering, Comillas University
Thesis Supervisor

Certified by. . . . .
Professor Stefanie Jegelka
Department of Electrical Engineering and Computer Science
Thesis Reader

Accepted by . . . .
Professor Munther A. Dahleh
Director, Institute for Data, Systems, and Society

Accepted by . . . .
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Network Partitioning Algorithms for Electricity Consumer Clustering

by

Olamide Oladeji

Abstract

In many developing countries, access to electricity remains a significant challenge. Electrification planners in these countries often have to make important decisions on the mode of electrification and the planning of electrical networks for those without access, while under resource constraints. To facilitate the achievement of universal energy access, the Reference Electrification Model (REM), a computational model capable of providing techno-economic analysis and data-driven decision support for these planning efforts, has been developed.

Primary among REM's capabilities is the recommendation of the least-cost mode of electrification - i.e by electric grid extension or off-grid systems - for non-electrified consumers in a region under analysis, while considering technical, economic and environmental constraints. This is achieved by the identification of consumer clusters (either as clusters of off-grid microgrids, stand-alone systems or grid-extension projects) using underlying clustering methods in the model.

This thesis focuses on the development and implementation of partitioning algorithms to achieve this purpose. Building on previously implemented efforts on the clustering and recommendation capabilities of REM, this work presents the development, analysis and performance evaluation of alternative approaches to the consumer clustering process, in comparison with REM's previously incorporated clustering methodology. Results show that the alternative methodology proposed can compare favorably with the hitherto implemented method in REM. Consequently, the integration of the proposed network partitioning procedures within REM, as well as some potential future research directions, is discussed. Finally, this thesis concludes with a discourse on the social and regulatory aspects of energy access and electricity planning in developing countries, providing some perspectives on the development policies and business models that complement the technological contributions of this work.

Thesis Supervisor: Professor Ignacio Perez-Arriaga
Title: Visiting Professor, Sloan School of Management

Thesis Reader: Professor Stefanie Jegelka
Title: Assistant Professor of Electrical Engineering and Computer Science

Acknowledgments

The past two years at MIT have been some of my most memorable ever. Leaving one's comfort zone to move thousands of miles away for the pursuit of knowledge certainly isn't something one does without support and counsel. I would therefore like to use this medium to thank all those who have helped me through this journey of intellectual development and self-discovery. I list below a few of them.

- First I would like to thank my thesis supervisor; Professor Ignacio Perez-Arriaga, thank you for accepting me to work with you and for the immense professional mentorship and academic training which you have given me. Among many other learnings, it has been wonderful learning about the technical and regulatory aspects of the electric power sector from you.

- I am quite grateful to Professor Fernando de Cuadra of IIT-Comillas, who has also been an advisor to me, guiding me at every stage of my research over the past two years and whose ideas really helped shape a lot of this thesis. I cannot thank you enough!

- Dr. Robert Stoner; thank you for accepting me into the Universal Access Lab as a Tata Fellow. I have learned a great deal from you and it has been a pleasure.

- Professor Stefanie Jegelka; thank you for being available and very helpful with additional technical perspectives on this thesis.

- Thank you to the members of the UA team; Pedro Ciller (IIT-Comillas) who provided a lot of technical support with the code, Stephen, Cailinn and Turner; amazing as colleagues and as friends. Roxanne and Matt; it's been amazing navigating grad school together and I will cherish the memories. Andres Gonzalez Garcia & Dr Carlos Mateo Domingo at IIT-Comillas and Dr Reja Amatya, Dr Pablo Duenas, Gregoire, and Sika at the MIT end; thank you!

- The staff and students of MITEI, Tata Center, TPP and Ashdown - places that have been my home at MIT - thank you. Also, thank you to Barb for the many

interesting conversations and your open door and to my TPP colleagues and friends with whom I've shared such great memories.

- Nwike, Chukwuka, Mrs Uwadileke, Mrs Anyigbo, and Dr Olakanmi; thank you for being there especially at the very early stages of this journey, when studying at MIT was merely a dream for me.

- MIT *Naija* people who have impacted me at various times; from Kesiena to Tunde, Abiodun, Chibueze, and Uyiosa, *una* deserve accolades.

- Grace, Otega, Olaoluwa, Opemipo *make una chop special knuckle*. CJ, Mayowa, *atigbogbo awon ore mi*, thanks for the support and friendship.

- Joanne!

- My family. Thank you to the Oladejis and the Rajis for providing the warmth of family and for always being there. 'Bolanle, 'Lola, 'Dayo, Michelle, Michael, John, Bros T. *Ese e ku ife*. My parents: *Baba mi* who's always got my back, *ati iya mi - E ba mi kira fun mama mi*, *Orisa bi* Titilayo *ko si* - mama I did it! :D. Thank you all for believing!!!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Context

In many parts of the world, access to basic electricity services remains a significant challenge. In fact, in 2015, the International Energy Agency (IEA) estimated that about 1.1 billion lacked access to electricity [5]. The scale and intractability of the energy access challenge has drawn a lot of attention in recent years. For instance, in addition to traditional electrification strategy by grid-extension, off-grid micro-grids and stand-alone home systems have recently gained momentum as effective, alternative ways of providing access to energy. The proliferation of these technologies has also resulted in the rise of novel business models for facilitating access. For example, the Pay-As-You-Go model, which involves credit repayment via instalments, has allowed customers of solar home system companies to bypass large upfront costs in East Africa and gain access to electricity services [65]. Off-grid offerings such as these complement the extension of the existing distribution grid, to areas without access, by the utility or rural electrification agency.

Critical to the success of these efforts is planning. In developing countries where energy access remains a problem, the planner is typically faced with complex decision making while under several technical, economic, regulatory and political constraints. Developed countries typically have electrification rates of 100 percent with the utility planning process focusing on infrastructure reinforcement to meet demand growth

or to fortify the grid with advanced ICT. In developing countries, the situation is typically different. Many developing economies, such as India and many sub-Saharan countries, have significant amount of their populations who are unserved or underserved by electricity; in India, this population was estimated at about 230 million people in 2016 [5]. Many of these unserved parts of the population largely reside in rural areas which may be far from load centers and where the cost of electrification from the grid may be very significant. In addition, electric utilities and governments typically have a universal service obligation to provide electricity to consumers on demand as electricity is considered a human right. This means that the planner in these countries has to make decisions on what mode of electrification, how to electrify or design any networks, for millions of people while aligning with national electricity policies.

On the technological end, there have been many efforts at building software tools to aid electrification planning for energy access. Computational techniques can provide enabling data-driven platforms to analyze these planning efforts and to support data-driven decisions for large-scale electrification projects. For instance, the planner can utilize software tools to understand the most economical way to electrify a large amount of consumers in unserved or underserved regions, to simulate network growth and understand reinforcements required, to design optimal networks, to understand the trade-offs between energy sources and undertake optimal generation system design among others. Recent advancements in computational processing capabilities and techniques mean that many of these decisions can now be made more data-driven and undertaken for large-scale regions, often involving millions of consumers and within relatively short time periods.

The Reference Electrification Model (REM)[1] is one such computational model, which undertakes network designs and provides recommendations for mode of electrification - i.e. whether as electrical grid extension projects or off-grid microgrid clusters

---

[1]REM is a software developed at the Universal Energy Access Lab at MIT and IIT-Comillas University see: universalaccess.mit.edu/REM. REM is referred to multiple times in this thesis, especially with respect to the research work done by previous/other graduate researchers in the group on its development.

- based on technical and economic constraints. The identification of consumer clusters (either as clusters of off-grid microgrids, stand-alone systems or grid-extension projects) is central to these recommendation and network design capabilities of REM.

This thesis builds on the body of work in this area, in particular the multiple contributions incorporated in the Reference Electrification Model, providing additional tools and perspectives for computation-aided electricity planning. It focuses specifically on two important decisions made by planners; the decision of the least-cost electrification mode for consumers and the clustering of consumers, showing how computational techniques can aid these decisions.

It is important to note that even in the adoption of the technical tools developed and discussed in this thesis, significant regulatory and socio-political barriers to energy access and mobility remain, which must be overcome. This thesis therefore addresses some of those barriers as well and provides recommendations on how they may be overcome or mitigated.

1.2 Motivation

It has been mentioned that REM is able to provide recommendations on the least-cost electrification mode - i.e. whether as electrical grid extension projects or off-grid microgrid clusters - based on technical and economic constraints. In the hitherto implemented version of REM, this clustering and recommendation capability is achieved via a "bottom-up" agglomerative procedure whose development has been extensively discussed in [28] and [30]. This "bottom-up" method, as incorporated in REM, has also been tested on several electrification planning application cases, such as in Rwanda, Uganda and India [24] [28] and [30]. We refer to this approach as "bottom-up" since we start by assuming that each consumer is in its own cluster before systematically merging consumers based on least cost electrification modes. Ultimately, this "bottom-up" agglomeration procedure allows us to identify and designate consumers as either off-grid stand-alone, microgrid clusters, or grid-extension candidates. Such an approach contrasts with the "top-down", network partitioning

15

approaches which this thesis proposes. Rather than starting at the individual consumer level, a "top-down" approach involves starting at the network level with all consumers connected to the grid. With all consumers connected to the grid, the objective then becomes identifying which consumers are better disconnected from the grid into off-grid clusters based on electrification cost. This leads to the important issue of how to systematically partition the grid to achieve this, a key question the research presented herein sought to address.

There are multiple motivations for this work. One motivation for the examination of alternative approaches such as network partitioning or "top-down" methods to electricity consumer clustering is the nature of rural electrification planning policies in reality. National rural electrification policies in many developing countries typically address the provision of off-grid micro-grids in unserved communities as a 'stop-gap' till it becomes economically (or politically) favorable to ultimately connect each off-grid consumer to the grids. Thus, many off-grid microgrids will likely become grid-compatible or replaced by electric grids, as the grid eventually extends to areas with no previous grid network existence.

Since grid-extension planning by distribution utilities typically involves solving a substation facility location clustering problem, a potentially better or more realistic approach to the analysis of off-grid clusters (and grid-extension clusters) may be to analyze the clustering problem under the assumption that a future planner will eventually run a substation facility location clustering algorithm (via Reference Network Models or network planning tools) to connect consumers in any present off-grid clusters to the grid. Thus, it may be possible with such an approach that we save on future network investment or upstream reinforcement costs when the grid eventually extends and connects to any consumers which we designate for off-grid microgrid clusters in the present time, if we already clustered these off-grid microgrid candidates considering the entire grid network and possibility of future grid-connectivity.

Another motivation for the work is that by generating clusters from partitioning a reference network designed by a robust network planner, we can better account for features and constraints such as topography or forbidden zones in the resulting

clusters and designs. This is because the original reference network would have been designed to account for these features and a network-partitioning based clustering process would inherit these properties.

Finally, there is also a need to benchmark any developed methods by comparing results obtained from the implementation with those from other more established methods.

1.3 Thesis Questions

The issues raised in the previous section lead to multiple research questions on the development of alternative clustering algorithm approaches to the consumer clustering and electrification mode recommendation problem. Specifically:

- Can robust alternative approaches to the electricity consumer clustering functionality in REM be developed?

- How can distribution networks be partitioned to identify clusters for off-grid micro-grids and for grid-extension projects which will minimize future investment costs when future brownfield planning occurs?

- How do any such alternative computational methods developed compare to currently implemented approaches in REM or in the literature?

- What kinds of regulations, policies or business models are needed to complement technological solutions such as the Reference Electrification Model in the drive towards universal energy access?

1.4 Preview

The rest of this thesis is organized as follows. The second chapter presents a review of existing computational approaches to the electricity consumer clustering problem, providing an overview of many models and tools for distribution planning such as

those in the Reference Electrification Model (REM). In Chapter 3, the development of network-based methods for partitioning non-electrified consumers into off-grid and grid-connected consumers is discussed. Specifically, the chapter presents some greedy algorithms which overcome the limitations of previously implemented methods in REM and in literature by partitioning the distribution network of the future fully electrified region to identify consumer clusters for present planning. Chapter 4 addresses different methods for clustering consumers who have been designated as off-grid candidates into microgrid clusters. Finally, the complementary regulatory environment for technologically-enhanced universal access is discussed in chapter 5.

Chapter 2

Existing Computational Tools for Electrification Planning

As a financially-intensive process subject to regulation and government policies, electrification planning can be a multi-faceted one, involving the collection and analysis of data, the application and development of models to analyze the expansion of the grid, the quantification of investment costs and benefits, among other steps. In developing economies, distribution planning is particularly important and complex because decisions have to be made on how to electrify unserved and underserved communities while catering to increasing demand. To this end, there are a number of tools which have been developed and applied for distribution planning. This chapter provides an overview of some of these tools and reviews a number of computational methods related to those explored in this thesis.

2.1 Urban Electrification Planning

2.1.1 Commercial Urban Distribution Planning Tools

Simaris, developed by Siemens AG, is one example of one commercial software utilized for planning distribution [1]. In addition to technical simulations, Simaris provides information on space requirements and estimates the budget required for an urban

distribution planning project, taking inputs of the network and intended switchgear equipment. ETAP, under its ETAP Grid platform, also offers a similar software for the modeling, expansion planning and analysis of smart urban distribution networks. Unlike Simaris, ETAP Grid incorporates Geo-Informatics System (GIS) capabilities which allow the user to visualize and analyze the network geospatially. It however does not incorporate financial analyses [33].

2.1.2   Academic Research on Urban Distribution Planning

There are also many examples in the body of literature on the development of methods or tools for urban distribution planning. In general, these research works focus on the optimization associated with greenfield distribution network planning, with a number of objective functions seeking to minimize investment or system cost, subject to technical constraints such as connectivity, reliability and voltage and thermal limits. [8] [11] [50] [51] [58] [71] present some of those approaches. In [50], we see the distribution planning problem formulated as a two-stage mixed-integer optimization problem to minimize investment and operational cost with nonlinear constraints. [69] considers network reinforcement costs while [8] considers deployment risks of expansion plans as part of the objective function. [29] and [63] analyze the expansion planning problem under the incorporation of Energy Storage Systems.

Many other literature works on distribution planning optimization are reviewed extensively in [37]. [75] proposes a weighted Voronoi diagram approach to determine new location and capacities of substations while providing annual cost output. Some other works of research incorporate distributed generation (DGs) or urban microgrids into their models for urban distribution planning. For instance, [13] [41] [51] [58] and [71] formulate the network planning optimization algorithms to factor technical issues associated with grid-compatible microgrids such as those associated with large penetration of DERs in the grid. One of the most interesting applied cases of academic research on urban distribution planning can be seen in [49] where the network planning tool developed is integrated with Geographic Information System (GIS) and Supervisory Control and Data Acquisition (SCADA) platforms, and is applied in the

actual expansion planning of Shanghai, China. Another important, and perhaps the most relevant to the Reference Electrification Model presented in this thesis, is the Reference Network Model (RNM) whose development is discussed in [26]. The RNM generates street-level designs for distribution planning and has been extensively applied for actual distribution planning in Europe. The Reference Electrification Model which has been introduced and is the focus of this research builds on RNM in its approach to rural electrification planning.

2.2   Rural Electrification Planning

The body of research dedicated to rural electrification planning is significantly smaller than that for urban distribution planning. Rural electrification planning introduces additional complexities of how to electrify unserved consumers. It is also important to note that the countries which often undertake large-scale rural electrification planning are usually developing countries, which operate under more stringent funding constraints. There are few rural electrification planning tools, one of which is the Reference Electrification Model on which this research is based and which is presented in [30] [17] and [47].

The Network Planner discussed in [44] is one of the most documented computational tools for geospatial rural electrification planning. It determines if grid extension is favorable in comparison to microgrids using a modified version of Kruskal's shortest path algorithm, finding the shortest paths between potential grid-connected locations. One of its limitations however is the lack of topography. It also does not provide its recommendation at the household level, but at a community cluster level. Another limitation is the lack of network design. [44] shows an application of the Network Planner to rural electrification planning in Ghana.

[35] describes a GIS-based rural electrification planning tool called LAPER developed by the EDF. LAPER requires the user to input data on network, geographical limits and an initial state design for a given community of villages before then proceeding in a step-by-step series of replacement stages - using alternative energy sources

and configurations - in which it tries to connect as many villages as possible under the given geographical constraints. Its optimization algorithm seeks to optimize the global cost of electrification over the entire community of villages i.e. the sum of the investment and operation and maintenance costs. An interesting feature of LAPER is the incorporation of the following criteria:

- Political

- Financial resources

- Development

- Financial

- Inter-region balance.

LAPER outputs a GIS representation of the electrification mode. Developed in 2001, not much is seen in literature of the applications of LAPER besides its original description in [35].

[27] introduces IntiGIS which takes a different approach to GIS-based decision-support for electrification planning. IntiGIS calculates the 'Levelized Electric Cost' (LEC) of competing technologies (grid, PV, wind, etc.) for each community location provided and then outputs the most competitive technology for that location based on LEC calculations. IntiGIS provides this information on a GIS visualization output and does not provide any network designs [9] [10].

GEOSIM is a commercial rural electrification decision support software developed by Innovation Energie Development [40]. Like Network Planner, GEOSIM also incorporates GIS into its approach to least-cost electrification mode recommendation. Rather than consumer-level designs and analyses, GEOSIM finds 'development centers' within communities using a gravity probability model and determines the least-cost electrification mode to supply each of these centers. This gravity probability model is based on the HUFF model, an established spatial analysis model, and is represented by the equation below:

$$P_{ij} = \frac{\frac{\lambda_i}{d_{ij}^2}}{\sum_k \frac{\lambda_k}{d_{kj}^2}}$$

where λ is a measure of the 'attractiveness' or 'gravitation pull' of one site to another and d_{ij} represents the distance between sites i and j.

GEOSIM takes in a number of technical and economic inputs and also allows for users to receive outputs on details such as "the percentage of people living in an electrified settlement". In addition, it also indicates the location of isolated settlements - settlements that are too far from electrified development centers and it estimates investment plans to provide power for basic social amenities (school, hospital) in such isolated settlements [40].

In [18], we see another computational approach to rural electrification. The authors of [18] first review of three other tools for rural electrification; HOMER, Network Planner and GEOSIM, identifying the various limitations across all three. A multi-stage procedure for rural electrification planning, which has many similarities to the Reference Electrification Model in approach, is then presented. For instance, the authors of [18] consider the following inputs;

- Electrification Status: If not available, [18] proposes inferring through night lights satellite imagery, existing grid location, census data, availability of decentralized power generators, social infrastructure (schools, health facilities) data.

- Existing grid network.

- Socio-economic data and local resource data.

In addition, [18] presents an analysis of demand alongside above data for ultimate recommendation of least-cost electrification mode as either grid extension, mini-grids or stand-alone projects for consumer clusters. Villages are used as consumer clusters for the analysis and, again, as in [61] and [44], it can be observed that electrification mode recommendation is not at household level (as in REM) but at village or community levels. Furthermore, while the authors of [18] do not delve deep into the

underlying algorithms behind the various stages of the model/tool presented, the authors provide another paper on the application of the developed software for rural electrification planning in Nigeria. In [14], the analysis region - the entire country - was divided into consumer clusters. Consumer clusters were first identified based on applying a buffer zone of 500m on population raster data sets. The electrification status of consumers was then inferred based on night lights and data on the grid-connectivity status of schools in each cluster. Three thresholds are applied for the determination of least-cost electricity supply option. These are:

- All clusters within a 20-km buffer zone of electrified clusters are to be electrified by grid-extension.

- All clusters outside the grid extension below a population of 1,000 received recommendation of Standalone systems.

- All remaining clusters were then recommended to be electrified by PV-based microgrids.

[14] then concludes by estimating PV capacity required for microgrids based on general assumptions about average size per household and load profiles. Overall, 47,489 cluster regions (corresponding to 171 million people) were analyzed and 3,800 clusters of these were recommended off-grid microgrids.

In [55], another GIS-based rural electrification model is presented and applied to electrification planning in Ethiopia. The model takes in GIS inputs such as proximity to grid as well as resources data (solar, wind potential and mining reserves) and evaluates locational Levelized Cost of Electricity (LCOE) values in making visual recommendations of least-cost mode of electrification.

The developers of HOMER, a microgrid design, modelling and optimization tool, also address the issue of determining the least-cost electrification mode by introducing a metric called the 'Breakeven Grid Extension Distance' [31]. This variable is obtained from other costs and variable values such as the capital recovery rate and the costs of the grid-extension and the alternative off-grid project. When analyzing a region

with a few consumers, this distance metric may be compared to actual distance of the consumers to the grid as a quick way to determine which should be designated as grid-extension or off-grid [18].

Overall, we see that all the rural electrification planning tools reviewed are limited in the following:

- The lack of analyses of electrification mode on a building-by-building or consumer-level basis.

- The consumer groups or 'clusters' for analysis for electrification are not computationally-determined least-cost clusters of consumers. The tools above largely examined consumer groups for electrification recommendation based on pre-defined natural clusters such as villages or "development centers". The limitation of this is that there may be other possible groupings of consumers for service by microgrids or grid-extension which may have lower costs per off-grid generation or grid-extension project.

- The lack of incorporation of technical and geospatial network designs alongside recommendation mode outputs. REM overcomes the above limitations in its approach to electrification planning. This work particularly focus on the second point above i.e. REM's approach to the identification of clusters of consumers for least-cost electrification mode recommendation. The next sections of this review introduce clustering algorithms and more formally define the clustering problem in the context of rural electrification planning as well as the currently-implemented approach to clustering in REM.

2.3   The Regional Reference Electrification Model

The development of the Reference Electrification Model has sought to overcome the aforementioned limitations of existing planning tools for electrification planning for energy access. REM provides more granularity in its approach to electrification mode recommendation, by determining the least-cost electrification strategy at individual

building level. Its clustering algorithms also allow the identification of off-grid microgrid clusters as least-cost electrification option while incorporating both techno-economic and geospatial information. Finally, by leveraging the robustly tested Reference Network Model described in [26], REM is able to both provide final network designs for all recommended systems and better quantify network costs in its underlying recommendation and clustering processes.

REM can be utilized in two modes: as the Local Reference Electrification Model (LREM) and as the Regional Reference Electrification Model (RREM). The Local Reference Electrification Model allows users to design individual microgrid projects and quantify the cost of investment. Given multiple types of input such as economic data, geospatial information (on location of consumers), LREM is able to provide both the generation and network design complemented with all cost estimates. It also has a geospatial output capability and can be made to account for features such as street/road terrains in the design of the microgrid distribution network.

The Regional Reference Electrification Model is used for large scale electrification planning where decisions have to made on the mode of electrification for a given region. In addition to providing electrification mode recommendation to consumers in a given region i.e. as grid extension candidates, off-grid consumers or stand-alone systems candidates, RREM also designs the required network and quantifies the cost of electrification. These utilizations modes have been described in [47] [30] [24] and [28] . In providing electrification mode recommendations, REM utilizes clustering algorithms to cluster consumers together and takes advantage of the Reference Network Model (RNM) to design networks and evaluate network costs. A description of REM's traditional approach to electrification mode recommendation is presented later on in this chapter.

The development of LREM and discussions on its applications for off-grid microgrid system design are presented in [47]. [30] [24] and [28] focus on RREM. An overview of Regional REM and the underlying methods used in the model development are presented in [30]. Some preliminary results of applying RREM for large scale electrification planning are also presented. In [24], improvements on the underlying

methods are described with additional applications of REM for regions in Rwanda and Uganda. [24] also addresses methods for the estimation of electrification status and for the quantification of upstream reinforcement. [28] focuses on work done to adapt REM to terrains with significant topography challenges. Inspired by the topographic feature of the RNM and using Rwanda as a case study, the author of [28] describes a methodology to incorporate topography to the model. Rwanda serves as a great case-study for evaluating topographic-handling capabilities since its location at the East African Plateau means a lot of the landmass lies on challenging, hilly terrains. It should be noted that the methods presented in this thesis complement other REM development efforts such as the afore-described. In particular, this thesis focuses on alternative clustering and recommendation methods to that presented in [30] [24] and [28]. Before presenting these methods, several applications of clustering within the context of electricity network design and planning are reviewed and discussed.

2.4 Clustering Applications in Electricity Network Planning

2.4.1 Overview: Clustering Algorithms

Clustering algorithms seek to group data points together based on some measure of similarity. As an unsupervised machine learning method, clustering algorithms explore the structure of data points (such as proximity of points) based on a number of approaches and have been applied to a number of problems across various domains from healthcare to network design and transportation analysis [7] [73] [36]. Generally, from literature, the following classes of clustering algorithms can be observed:

- Connectivity-based/Hierarchical Clustering Algorithms: Hierarchical clustering involves the creation of clusters from the recursive merging or division of clusters in either a 'bottom-up' or "top-down" manner [7]. The clustering process for this type of algorithms can be represented by a dendrogram. There are two approaches to connectivity-based clustering:

– Agglomerative Clustering: This is a 'bottom-up' approach in which every point is assumed to be in its own cluster and clusters are successively merged together based on distances. The order of complexity of this algorithm is at least O(n3).

– Divisive Clustering: This is a disconnection, top-down based clustering approach in which all nodes are initially assumed to be connected into one large cluster and then splitting occurs based on some measure of distance dissimilarity.

- Partitional Clustering Algorithms: These are largely centroid-based algorithms which often require the user to input the number of clusters. They work by initializing a partition and moving through various partitions to find the clustering configuration that minimizes the overall dissimilarities. It can be seen that this is an NP-hard (non-deterministic polynomial time hard) optimization problem and so many partition algorithms utilize approximate solutions:

– K-means Clustering: This is an error-minimization partitional clustering algorithm which seeks to minimize sum of Sum of Squared Error as follows:

$$J = argmin \sum_{n=1}^{N} \sum_{k=1}^{K} ||x_n - \mu_k||^2 \tag{2.1}$$

– Other partition methods also include the K-medoids algorithm as described by [15] which uses datapoints as centers and can work with any distance metric. These partition methods in general are limited in their ability to obtain concave clusters and do not scale well on differently-sized clusters [43].

- Density Based Methods: Density based Methods determine clusters based on the definition that clusters are areas with higher probability densities than others. The Density Based Spatial Clustering of Applications with Noise (DB-SCAN) Algorithm is a primary example of this approach to clustering and has been widely used in geospatial clustering [43]. [43] however identifies some lim-

itations to DBSCAN as they show it cannot effectively find clusters of different density.

- Model-based clustering with neural networks (self-organizing maps in particular) and decisions trees which are good at characterizing each cluster but may not scale effectively to large data sets [62].

- We find some grid-based and soft computing based methods such as Fuzzy clustering, Genetic Algorithm clustering discussed in [62].

2.4.2 Clustering in Electrical Networks: State of the Art

Clustering algorithms have been applied in a number of ways in electricity networks. One major application area of clustering algorithms has been in network planning. The area of network expansion planning has attracted a number of computational methods in the attempt to minimize network costs while determining locations of substations [38] [57]. In [38], k-means clustering algorithm is parsed recursively alongside a network planner too as part of a heuristic approach for the optimal planning of greenfield MV/LV substations and network. The authors of [53] apply fuzzy C-means clustering (however, with predefined number of clusters) for a distribution system expansion planning optimization problem in contrast to the evolutionary algorithmic based approach seen in [52]. Gray theory is also another approach which has been applied to this problem [74]. The authors used it to determine energy saving potential measures, which then feed into a distribution network planning optimization formulation. In [23], the network partitioning problem in which graph-theoretic network clustering algorithms are used to partition a power network under 'electrical distance' constraints, is presented. The authors of [23] apply a hybrid-k-means to solve the multi-objective optimization clustering problem proposed. In addition, an application of a heuristic algorithm - 'imperialist competitive algorithm' - for optimal boundary clustering of urban microgrids can be seen in [56]. These examples show that partitioning or clustering algorithms have been extensively applied to electricity networks in general in literature albeit not much is seen in the specific context of

29

rural electrification planning as in REM.

Finally, this subsection is concluded with a discussion on a class of combinatorial optimization problems, from the operations research or optimization community, which closely approximate the off-grid clustering objective of rural electrification; the Facility Location Clustering Problem with unknown number of facilities. Facility location optimization has been extensively studied in computer science and operations research. This refers to the problem of locating facilities (for our case - off-grid microgrids) to serve a number of consumers while minimizing costs. For an off-grid microgrid, the costs to be minimized would be the total investment costs i.e. the sum of all off-grid generation and O & M cost ('facilities generation cost') and the network costs ('facilities transportation cost'). Many studies on this class of problem model situations in which the number of facilities is defined beforehand and the possible locations are discrete. These problems have been shown to be NP-hard and so different approximate solutions have been investigated in literature [25] [34] [48] [54]. Thus, computational methods from other research communities may be applicable to electricity planning problems. For instance, for the case of off-grid electrification, the modified facilities location problem would be that of a continuous facilities location problem with unknown numbers of facilities.

2.4.3 Consumer Clustering for Rural Electrification Planning: The Regional Reference Electrification Model

Within the context of the Reference Electrification Model, the clustering problem as defined in [30] as "the grouping of customers into candidate off-grid systems and grid extension projects". While there are a number of ways to approach this, one that is discussed in [24] and [30] is a Delaunay Triangulation-based agglomerative clustering algorithm.

Implemented Delaunay Triangulation Agglomerative Clustering Algorithm

The implemented Delaunay Triangulation agglomerative consumer clustering algorithm in REM involves dividing the clustering process into two steps:

1. Off-grid clustering

2. On-grid clustering

Under this approach, a Delaunay Triangulation (DT) connecting every consumer of an analysis region will be built. The preliminary clustering algorithm documented in [30] utilized a Minimum Spanning Tree instead of a DT. Afterwards, arcs of the Delaunay are sorted in increasing order of length and evaluated to determine if the two clusters located at its ends should be joined in one cluster. The assumption here is that initially all edges(connections) are not activated and so every customer node is in its own cluster of only that node. Agglomeration occurs when edges are activated - based on defined parameter comparisons - such that customers at both node ends of an edge are connected into same cluster. The algorithm loops several times until no new connection is activated. At the end of this process, the off-grid clusters would have been calculated. The idea of merging is such that the savings of being together compensate the extra connection costs. The parameters compared in merging two clusters may include:

1. The costs of off-grid generation for each isolated cluster.

2. The operation maintenance costs for each isolated cluster.

3. The costs of off-grid generation for the combined clusters.

4. The extra connection costs of an electrical line (equivalent distance) between two clusters.

5. The operation maintenance costs for the combined clusters.

For the on-grid clustering procedure; the inactive arcs of the DT are reevaluated, now comparing the cost/savings balance of being connected to the grid together

against being electrified separately (with at least one of them connected to the grid). Depending on the most inexpensive configuration for merging, the clusters are joined. The algorithm loops several times until no new connection is activated. At the end of this process, the on-grid clusters have been calculated. The costs parameters compared for off-grid clustering may include:

1. Generation cost of each isolated cluster node.

2. MV/LV transformers costs.

3. Cost of LV (or MV) line that connects two clusters.

4. Costs LV (or MV) lines that goes from the existing grid network to each cluster.

5. Energy cost and non-served energy cost.

6. Operation Maintenance costs of clusters

More comprehensive descriptions of this clustering method, including its implementation and applications, can be found in [24] and [30]. As raised in the introductory chapter of this thesis, this approach is limited in its incorporation of the full distribution network topology which may better capture useful information such as topography that can be relevant to clustering recommendations.

2.4.4 Conclusions

The review of the body of literature presented above highlighted the different ways that computational methods have been used for electricity infrastructure planning. We reviewed several tools which have been developed to facilitate rural electrification planning. From the review, we observe that the majority of tools in this space do not analyze electrification mode recommendation on a building-by-building level. Also, most approaches do not analyze other ways of clustering existing consumers into microgrids besides the pre-defined natural village boundaries. It is possible that considering artificial clusters and analyzing at the building level could lead to lower

costs in the recommendation results from these tools. The incorporation of other costs and savings, such as the microgrid network costs, management costs and savings upstream, when undertaking electrification mode recommendation is also another area of weakness observed for most of the tools. Simple heuristics, such as comparing the offgrid generation cost with the grid LCOE values or using distance to grid, may not lead to the true least-cost electrification planning solution as they fail to capture other important cost drivers. The latter part of this review chapter introduced the bottom-up clustering methodology in REM which attempts to overcome some of the limitations highlighted above. As mentioned, REM has been used for several large scale analysis and can provide consumer-by-consumer electrification mode recommendations. However, it may be possible to improve on REM's bottom-up clustering methodology or develop better performing alternatives.

As discussed, there are several opportunities for further work in the area of developing robust, scalable techniques which building on the computational methods and tools reviewed, overcome their limitations. For instance, it was highlighted that REM utilizes a reference network model, the RNM, which can design networks incorporating features such as topography using well tested underlying clustering and optimization algorithms. A top-down method to partition the RNM's output network helps us not only to better incorporate future grid connectivity but also to take advantage of RNM's abilities to design electrical infrastructure under geospatial constraints.

The next chapter therefore presents the development of a top-down computational methodology for electrification mode recommendation; one involving the partitioning of the distribution grid connecting all consumers to identify those who should be designated as off-grid or grid-connected. Subsequently, additional complementary perspectives to the off-grid clustering problem are also addressed. Considering the objectives laid out therein, it is pertinent that the developed methods be applied on datasets from real-world cases and the ensuing results compared with those obtained from the reviewed methods. This allows us to benchmark the algorithms and to establish if any of the hypotheses which motivated this work hold. To this end, results from implementation these methods on test cases are also presented and discussed.

Chapter 3

Partitioning the Distribution Network: Greedy Approaches

This chapter presents a number of algorithmic approaches for the partitioning of electricity consumers in order to identify those to be electrified as grid-extension consumers and those for off-grid consumers based on cost. Unlike the computational approaches to this problem which were reviewed in the previous chapter, the methods presented herein take a top-down approach by partitioning the distribution network in order to designate customers to be electrified by either off-grid or on-grid (grid-extension) projects.

This network or graph-based approach involves first designing a reference distribution network for the region under consideration such that all consumers are connected to the existing grid. The designed network represents the ideal future network if all consumers were able to be connected to the grid and costs were not a barrier. Since this distribution network is radial, all elements of the distribution network and their associated properties (such as costs) can then be represented in a tree data structure for computation purposes. This makes it possible to approach the consumer bi-partitioning problem (into off-grid and grid-extension partitions) as a tree pruning problem and explore computational strategies to that effect.

The rest of this chapter describes the development and implementation of algorithms which build on this idea of partitioning the network.

3.1 Designing the Distribution Network

As previously mentioned, designing the distribution network connecting all consumers to the grid is the first step in our network-based partitioning procedure. Thus, we assume that a network planner has a network design routine capable of designing an optimal (or quasi-optimal)[1] reference network for ALL new consumers in a given region and quantifying the costs of every element in this distribution network. Any reference network so designed from such a routine can then be assumed to have the same topological characteristics of the future grid when all consumers become fully connected to the grid.

There are many examples of such network designing tools[2] in literature and a number of them have been addressed in chapter 2. There are many properties of the distribution network that can be extracted from a distribution network design routine. From the topological properties of the network, it is possible to define pointers which show the hierarchy of power flow, from the grid all the way to the consumers and through every element in between. In addition to pointers of what is fed by what, other properties of elements in the network such as the length of line segments, the geo-location of equipment, the cost of all elements in the network, the voltage levels and other electrical, spatial and economic properties can also be adequately represented.

Being able to design the network and quantify properties facilitates the representation of the network by an equivalent tree; a capability upon which the ideas presented in this thesis rests. Thus, we can invoke (repeatedly) a reference network designing subroutine in order to design the network and subsequently construct an equivalent tree data structure for our partitioning purpose. A complete equivalent tree should have all the elements of the network properly represented, providing an important data structure for implementing many computational methods on the network.

---

[1]Network Optimality here refers the fact that the designed network from such a routine is expected to be the least-cost network.

[2]In the rest of this thesis, such a network design tool is referred to as a Reference Network Model or RNM.

3.2 Tree Data Structure for Radial Distribution Networks

It has been previously mentioned that since (rural) electricity distribution is hierarchical and typically radial, the synthesized optimal network can be represented by an equivalent tree for computation purposes. Using a tree data structure provides computational advantage as they are easily implementable and have had many well-established computation methods developed to exploit their structure. Constructing the tree involves the provision of pointers from any node indicating key identity of parent as well as children of given node. For the equivalent trees of radial distribution networks, three types of nodes exist:

1. Line Segment: Line segments refer to the electrical distribution lines which distribute power to consumers or transmit from one element in the tree to another.

2. Consumers: Consumers are the load consumption nodes which ultimately must be fed and have their demands met. It is easy to see that for the equivalent tree of a radial distribution network, a node is a leaf node if and only if it is a consumer node.[3]

3. Transformers: Transformer nodes are nodes which represent power transformer equipment for voltage transformation within the network. They link node elements at different voltage levels together.

3.2.1 Tree Data Structure Terminology

Before presenting the developed algorithms for partitioning the equivalent tree of a distribution network, it is important to describe standard computation terms used in describing tree data structures. These include:

---

[3]Consumer nodes have no children nodes by definition. Other types of nodes must have children by definition

- Root: The root of a tree is the highest node which has no ancestor or parent. The root node is an ancestor node to every other node downstream.

- Node: As with networks, nodes (or vertices) are one of two types of elements of trees. Edges are the other type of element and link two or more nodes together. It is important to note that the definition of 'node' here is distinct from the concept of electrical nodes in circuitry. In this thesis, a node refers to an object within a tree data structure which represents information corresponding to an element of the designed distribution network. Thus, the nodes in the equivalent tree of a distribution network could correspond to information representation for either a particular consumer, or a line segment or a distribution transformer. The edges between nodes of the tree then show the information dependencies between them.

- Leaf: A leaf node is a node which is terminal i.e. has no children nodes.

- Parent: A parent node, with respect to a child node, is a node which has an outgoing edge to other nodes called its children nodes.

- Child node: A child node (or children nodes) is used in reference to a parent node as a node which receives an incoming link or edge from a parent node.

- Ancestors: The ancestors of a node are all nodes upstream a node (all the way to the root) from which incoming edges proceed. Thus the root node of a given tree is an ancestor node to all nodes in the tree.

- Subtree/Downstream Offspring: The offspring of a node are all nodes downstream to the node (all the way to the leaf nodes). The subtree encompasses the node, children nodes of a node and the children nodes of those children nodes all the way to leaf nodes. For example, the subtree of the root node is the entire tree.

Some of these definitions are illustrated in Figure 3.1.
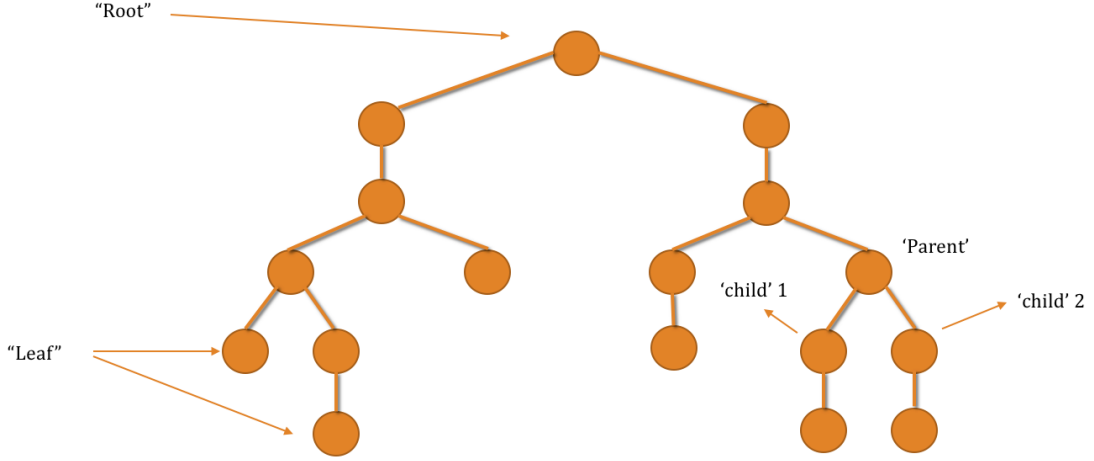
Figure 3-1: Sample Tree

3.3 Greedy Tree Pruning

Given our objective of partitioning or pruning the network tree to determine the least-cost electrification mode for every consumer node, a greedy approach to this problem may be considered. A greedy strategy means that the decision made locally at a node is the locally optimal one out of the two possible decisions i.e. to prune or otherwise. At any node, the optimal pruning decision is one determined from a local evaluation of cost and benefit of retaining the downstream subtree. In deciding that a node (and the consumer nodes in its subtree) be pruned away and designated as off-grid consumers, or retained in the tree as part of the grid extension, we compare the costs with the benefits of the decision. Specifically, the cost of introducing a new offgrid generation facility if the node and its downstream nodes are pruned must be less than the savings in cost in the grid from having it removed. For every node $i \in N = \{1, 2, ...n\}$ in an n-node tree, we can define and compute a decision value δ_i which tracks this decision as given by (3.1);

$$\delta_i = offGen_i + offCnse_i + netAdm_i - upStrm_i - onCnse_i - gridEc_i - selfC_i \quad (3.1)$$

where $offGen_i$, $offCnse_i$ and $netAdm_i$ are the costs incurred if the node's sub-

39

tree is pruned. $offGen_i$ is the downstream offgrid generation cost incurred if the node's subtree is pruned to off-grid, $netAdm_i$ captures the net management cost in providing off-grid generation rather than the grid while $offCnse_i$ is the cost of non-served energy associated with the resulting offgrid consumers. $offGen_i$ is also implemented such that it can capture the costs associated with the inclusion of an additional MV/LV transformer should the microgrid associated with the pruned subtree be a large (MV) microgrid.

$upStrm_i$, $onCnse_i$, $gridEc_i$ and $selfC_i$ are the savings or benefits incurred if pruning occurs. $upStrm_i$ is the total cost savings upstream if the downstream subtree is pruned, $onCnse_i$ is the cost of non-served grid energy saved if the node's subtree is pruned $gridEc_i$ is the grid energy cost saved after pruning, and $selfC_i$ is the self cost of node i which is saved if the node is pruned. $selfC_i$ also captures costs associated with the electrical losses of the node's equivalent element. A key assumption in our definition of δ_i is that there is negligible difference in network cost before and after pruning for the downstream subtree consumers. Without this assumption, we would have to evaluate any gains in the overall network cost as a result of pruning and factor it in the computation of δ_i.

If $\delta_i < 0$, then pruning is the locally optimal decision and is undertaken. An exhaustive search on all nodes in the tree, in a bottom-up fashion, can be undertaken until no further pruning is possible. Consumer nodes that are pruned can then be designated as off-grid customers while those still remaining in the tree are designated as grid-extension consumers.

In addition, the following criteria should also be met:

- Bottom-up Traversal: The tree must be traversed in a bottom-up fashion. That is, the exploration begins from a leaf node and all nodes must be examined before their ancestors.

- Pruning decisions are irreversible.

It is possible to traverse the tree in different ways such that the bottom-up criterion is satisfied. This leads to another question; how does the order of traversal affect

solutions? This is a salient question since some node elements of the tree have discrete-sized values as they were based on a standard catalog. For example, the order of removal of a 15kVA standard-sized transformer from a tree may significantly affect subsequent pruning decisions, since the node was discretely sized. Further discussion on how discrete catalog data is handled in the context of this work is discussed in section 3.3.2 of this chapter. For the traversal application at hand, one might expect that exploring nodes that are farthest from the root, or have the highest downstream demand, or both should make good candidates for early pruning. To address this, two traversal strategies - traversing the tree in order of the distance of path to the root node & in order of a defined and computed node variable called 'Moments' - were explored. These traversal methods were then bench-marked with a third standard tree traversal method - the post-order traversal strategy. These traversal strategies are discussed subsequently.

3.3.1 Tree Traversal

The order in which we visit nodes for pruning may be critical to the partition solutions obtained from our proposed pruning procedure. Since properties such as distance to the root and node power capacity, which are considered for pruning, are hierarchical and subtrees are not visited after a node is visited, it is possible that a node is prunable but its children are not.

One classic tree traversal technique from literature which satisfies our bottom up criterion is the post-order algorithm [42]. This is a depth first search method in which the children of a node are recursively traversed before the node. By recursively visiting the subtrees of every node before a node, the post-order algorithm guarantees that the root node is examined for pruning last. This process should generate a list of nodes to visit based on the post-order depth first search tree algorithm. The structure of the algorithm to obtain this is as follows:

If the root of the tree is passed as the node argument to the above procedure, the result *List* gives us a bottom up array of nodes.

Thus, for a tree such as in Figure 3.2, the post-order traversal strategy yields a

41

Algorithm 1 Generic Post-Order Algorithm

procedure MODIFIEDPOSTORDER($tree, node, List = []$)
    $childrenCount \leftarrow$ number of children of $node$
    if $childrenCount > 0$ **then**
        for i = 1 to childrenCount **do**
            MODIFIEDPOSTORDER(tree,node.child(i),List)
            $List$.append($node$)   ▷ append node to List
        end for
    else
        $List$.append($node$)
    end if
    return $List$ ▷ Final bottom up list is in List
end procedure
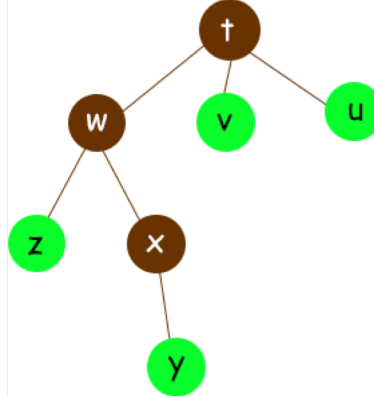
Figure 3-2: Sample Tree for Post Order Traversal: Traversal Order is $[Y, X, Z, W, V, U, T]$

list which agrees with our bottom up criterion. This traversal therefore allows us to benchmark the expected advantages of any heuristics we develop for traversing the tree.

Distance of Path To the Root Node

The distance of path to the root node or $PathToRootLength$ traversal heuristic is most intuitive considering the problem at hand. Other factors being equal, it should be expected that grid-extension costs increases with distances to the grid and consequently farther-off isolated consumers may likely be the best candidates for off-grid. The distance can therefore be imposed alongside the bottom-up criterion to determine order of exploring the tree. From a computational perspective, it is easy to define a

PathToRootLength property for every node which can be updated with each pruning decision. This is because each node in the tree already represents information on a corresponding distribution network element and its properties such as its length. The *PathToRootLength* property is then the sum of the length properties of all the ancestors to the root node, for a given node. A subroutine that computes and updates this property for nodes in the tree can thus be defined. This subroutine returns the nodes in order of their *PathToRootLength* property. To understand how this might work, consider again Figure 3.2. If such a subroutine is called and it is assumed that the spatial distribution of nodes in Figure 3.2 are similarly scaled to the real network, the *PathToRootLength* would return a list $[Y, X, Z, W, V, U, T]$.

'Moments'

While distance to the root seems a good criterion for tree traversal for our electrification problem, there are other important parameters, such as the magnitude of downstream demand, not captured by such a strategy. Our expectation of any good traversal strategy is that nodes which are farther from the grid and have higher downstream capacity are pruned earlier because they imply higher grid costs, and thus higher savings when pruned. Factoring both the distance to the grid with the power delivered in the traversal strategy also helps to capture the voltage drop, which is a very relevant cost driver in real-life distribution network planning.

A node property, *Moments*, S, which captures this combination of downstream power and distance to the grid can thus be defined. For a tree of nodes n, the *Moments* S_i at a node as:

$$S_i = S_i * L_i + max\{S_{v1}, S_{v2}, S_{v3} \ldots\}$$

where $S_{v1}, S_{v2}, S_{v3}...$ are the moments of node i's children and L_i is the length of node i.

These values can be initially pre-computed as the equivalent tree is constructed from the designed network. Based on the above, a procedure based on the combination of power-distance or *Moments* can be described for identifying the next best candidate node to pruning evaluation as follows:

- pre-sort nodes from highest to lowest moments.

- Check sorted array if nodes have had all their downstream nodes explored or they have no node, if yes terminate tree traversal else go to next node in sort, proceed.

- Return next best candidate node for pruning as highest value in sorted nodes array.

The two traversal strategies described above were implemented and are described in section 3.4 of this chapter. Also presented are results of implementation of the post-order traversal method as a benchmark for performance evaluation of the distance to root and post order method.

3.3.2 Partitioning Algorithm I.

Given an order to traverse the tree for pruning and the local decision criterion as in equation 3.1, an overall partitioning procedure for pruning the tree can be described as in Algorithm 2.

In Algorithm 2, all consumers are initially connected to the grid-extension partition S_{ongrid} while the offgrid partition $S_{offgrid}$ is empty. For every node i visited in the tree, the local decision variable δ_i is computed and, if negative, the downstream consumer indices are appended to the offgrid partition $S_{offgrid}$. The tree properties are then updated and the process is repeated until all nodes have been evaluated. Recall that the downstream offgrid generation cost incurred is $offGen_i$ while $offCnse_i$ is the Cost of Non-Served energy (CNSE) of node i's downstream consumers if offgrid. In addition, $upStrm_i$, $onCnse_i$, $gridEc_i$ and $selfC_i$ are the upstream savings, the grid cost of non-served energy, the grid energy cost and the node self cost respectively.

The downstream generation cost $offGen_i$ is computed by simply calling a generation design function or look-up table using node i's subtree consumer nodes properties as arguments. The look-up table function has already been previously incorporated in REM as described in [30]. The grid cost of non served energy $onCnse_i$ and grid energy cost, $gridEc_i$ are similarly determined by invoking these downstream consumer

44

Algorithm 2 Network Partitioning Algorithm I

procedure GREEDYPRUNDELTA
    $tree \leftarrow$ `construct network tree procedure`   ▷ tree has n nodes
    $S_{offgrid} \leftarrow \emptyset$
    $S_{total} \leftarrow \{1, 2, ...k\}$ ▷ all k consumer initially grid-connected
    while `all nodes in` $tree$ `have not been evaluated` **do**
        $i \leftarrow$ `get Next Terminal Node to Evaluate`
        $\delta_i = offGen_i + offCnse_i + netAdm_i - upStrm_i - onCnse_i - gridEc_i - selfC_i$
        `set node.evaluated as True for node` i
        if $\delta_i \leq 0$ **then**
            $List_i \leftarrow$ `obtain Set of consumers' indices in subtree of` i
            $S_{offgrid} \leftarrow S_{offgrid} \cup List_i$
            `prune subtree of node` i
            `update` $tree$
        end if
    end while
    $S_{ongrid} \leftarrow S_{total} \setminus S_{offgrid}$
    return $S_{offgrid}, S_{ongrid}$
end procedure

---

node data to functions which then compute them as documented in the agglomerative 'bottom-up' clustering based version of REM [30]. The upstream savings $upStrm_i$ penalizes pruning that lead to little savings in cost of upstream grid infrastructure. To compute this, we evaluate the cost differences in all the ancestors to node i before versus after pruning.

Note that it is important to update upstream tree properties of the tree after pruning to account for the expected changes in a distribution network. Some properties of other nodes in the tree may change due to the removal of downstream nodes. Specifically, the following two properties need to be updated:

- Power capacity: When nodes downstream are pruned, capacity of upstream ancestor nodes should reduce correspondingly.

- Self Cost: This is typically a function of the power capacity.

These two properties may also affect other node properties such as the *moments* and $path-to-root$ which are used to identify next best pruning candidate. They are also highly discrete properties, which means that the use of discrete versus continuous

values when they are being updated from catalog must be considered. This is because cost and capacity in the distribution equipment catalogs, are stored in discrete values. For instance, transformer capacities may be cataloged as 5, 7.5, 10, 15, 25 kVA. However, as an example, the updating process may lead to a capacity value of 6.8kVA and its associated cost. We handle this discrete-continuous distinction with a number of heuristics. For instance, the continuous capacity is used to obtain discrete values for the self costs (during the update process and the computation of δ_i). Continuous cost values are used when computing upstream savings from cost differences.

Thus, updating the power capacity and self cost may lead to the existence of a different, lower-cost tree topology that connects the remaining nodes. Consequently, an ideal update of the tree structure would be a re-design of the network/tree (with the RNM) using the remaining tree consumer nodes after every pruning. This, however, may not be time-feasible in practice, under large scale planning conditions. Section IV of this chapter further addresses the tree updating process and the algorithm implementation heuristics.

Executing Algorithm 2 returns the partitions of consumers designated as off-grid and grid-extension consumers. Passing the grid-extension consumer partition to a network design module (RNM) allows us to obtain final grid-extension network designs, one of the objectives of REM. For the off-grid consumers list, we can then run an additional off-grid clustering procedure to cluster these off-grid consumers - based on total cost - into different microgrid clusters before then designing their associated off-grid networks. Chapter 4 of this thesis addresses the off-grid clustering process as specific to REM and in general.

.

3.3.3 Partitioning Algorithm II

Another greedy method which has been explored involves defining a proxy for the decision value, pre-computing this decision value proxy for all nodes and then greedily pruning the tree in order of the decision value.

Recall that $\delta_i = offGen_i + offCnse_i + netAdm_i - upStrm_i - onCnse_i - gridEc_i -$

$selfC_i$ where δ_i can be interpreted as the decision value or 'net benefit' of assigning the downstream consumers of a node i to offgrid. We can therefore define another greedy strategy as one which we pre-compute δ_i for all nodes $i \in N = \{1, 2, ...n\}$ in our n-node tree and then prune accordingly at each node. Using this method, the best candidate for pruning at each step would be the node with the most negative decision value while nodes with positive decision values are not prunable. An advantage of this strategy is that a predefinition of the order of pruning such as using $Moments$ or $PathtoRootLength$ is unneeded unlike the previously described algorithm.

It should however be noted that since after each pruning the upstream nodes are updated, we must re-compute delta for all nodes upstream and re-sort before checking for the next best candidate.

The overall approach can be summarized in Algorithm 3.

Algorithm 3 Network Partitioning Algorithm II

procedure GREEDYPRUNDELTACARLOS
    $tree \leftarrow$ construct network tree procedure $\triangleright$ tree has n nodes
    $S_{offgrid} \leftarrow \emptyset$
    $S_{total} \leftarrow \{1, 2, ...k\}$ $\triangleright$ all k consumer initially grid-connected
    $n \leftarrow$ count of nodes in tree
    for i = 1 to n **do** $\triangleright$ could be in any order
        $\delta_i = offGen_i + offCnse_i + netAdm_i - upStrm_i - onCnse_i - gridEc_i - selfC_i$
    end for
    $sortedIDs \leftarrow$ sort nodes from -ve to +ve δ
    while there are still -ve δs & they have not been evaluated **do**
        $j \leftarrow$ get next most negative Terminal node index from $sortedIDs$
        set node.evaluated as True for node j
        if $\delta_j \leq 0$ & node j has not been evaluated **then**
            $List_j \leftarrow$ obtain Set of consumers' indices in subtree of j
            $S_{offgrid} \leftarrow S_{offgrid} \cup List_j$
            prune subtree of node j
            update $tree$
        end if
    end while
    $S_{ongrid} \leftarrow S_{total} \setminus S_{offgrid}$
    return $S_{offgrid}, S_{ongrid}$
end procedure

3.4 Implementation: Heuristics and Storage

The algorithms previously described were implemented and tested on a number of cases in order to understand and comparatively evaluate their performance. Considering that the methods described in this thesis are expected to be used for real-world computation enhanced planning via the Reference Electrification Model (REM), it is important to describe how they were implemented and additional heuristics introduced for scalability.

As is the case with other modules in REM, MATLAB was used as the programming platform for implementing these methods. The following graphic (Figure 3.3) represents the flow of input to output for partitioning the customers.
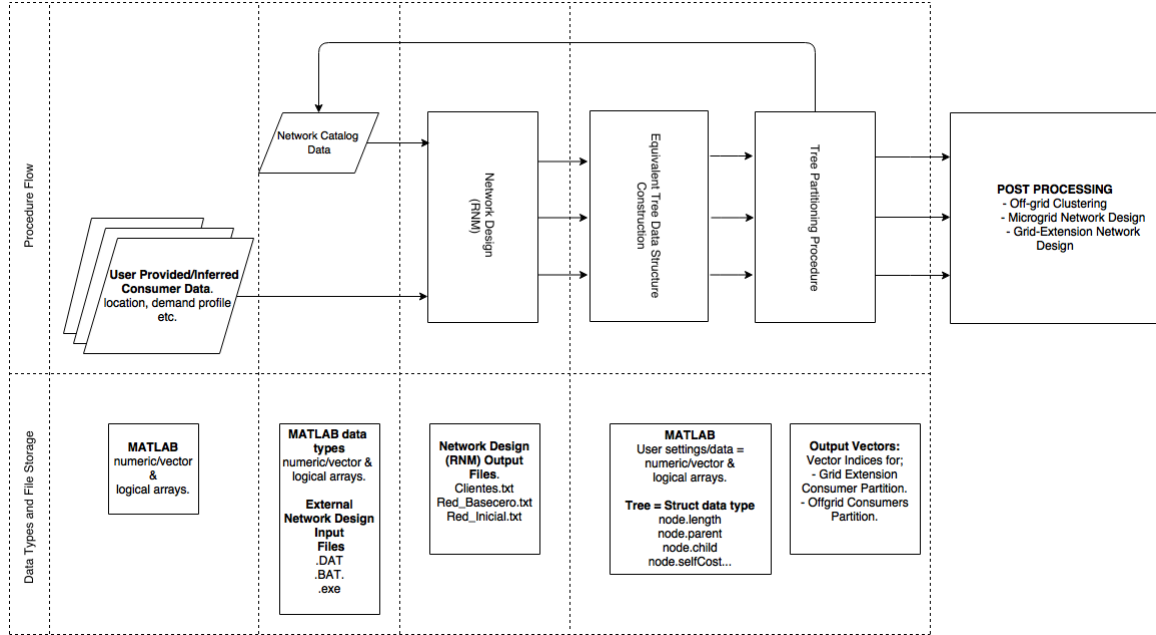
Figure 3-3: Implementation within REM

The $top - down$ procedure(s) first receives input data for both consumers and network, in addition to configuration parameters, storing them as internal variables in MATLAB. A call to the RNM for network design is then made using these inputs and other external RNM-specific files. The network design produces network files such as *.shp* and *.txt* files which are then parsed to build the tree data structure. The tree is implemented via MATLAB's *struct* data type with fields of the *struct*

corresponding node properties or pointers to parent and children. Redesigning the tree after pruning will involve invoking the network design subroutine and the external RNM-specific files alongside the remaining consumers to re-design the network.

In addition, the implementation of Algorithm 2 in MATLAB is achieved with several heuristics. Most of these heuristics are related to the *tree* updating procedure after pruning. Recall that after pruning, upstream effects are updated with the self cost, power capacity and the self losses of upstream nodes updated. The approach to updating the self cost and power capacity values involve piece-wise linear interpolation using the user-provided network equipment catalog data. This is because actual elements are discrete, and we need to pass on smooth cost signals, as order-independent as possible, when traversing and pruning the tree. Losses can also be similarly linearly interpolated.

The frequency of re-design of the network/consequent reconstruction of the tree after pruning is another parameter that can be varied. Rather than designing a new network after every subtree is pruned, a user-defined frequency parameter can be used to control how often the network tree is reconstructed. The user can provide a value between 0 and 1 corresponding to the fraction of total grid power that can be pruned downstream before the network is redesigned. The most accurate scenario - which however has the worst run-time complexity - is thus when the tree is redesigned after every subtree pruning occurs. On the other hand, the best run-time complexity scenario - at the expense of accuracy - occurs when there is only a single initial tree design without any updates after pruning. Real world planning cases will involve design for millions of consumers, and this user-defined variable can be used to speed up computation as a trade-off against accuracy.

3.5   Results

To conclude this chapter, we present results of the implementations of the algorithms described in this chapter and compare results of these implementations to those of the previously implemented consumer clustering method in REM. First presented

are results of comparisons of the tree traversal strategies for evaluating pruning node candidates.

3.5.1 Comparative Analysis of Tree Traversal Strategy

As discussed in section 3.3.1, it may be possible that the order in which the tree is traversed for evaluating nodes may be significant in the final results obtained. To determine if any such effects exist and the nature of these effects, the two primary methods of traversal discussed; $Moments$ and $PathToRootLength$ were compared alongside a third method $PostOrder$ using a small test case of 520 consumers. Using the same test case to evaluate all three methods, a sensitivity analysis using grid reliability as the varying parameter was undertaken to evaluate the performance differences (as measured by global costs and spatial characteristics of the partitions) across the methods. Grid reliability is an important parameter to explore because it is critical to grid connection as it affects the cost of non-served energy.
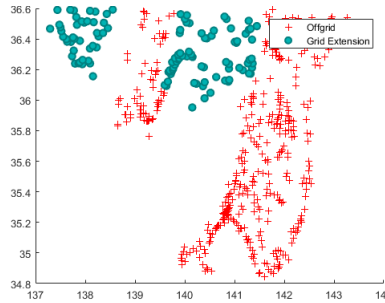

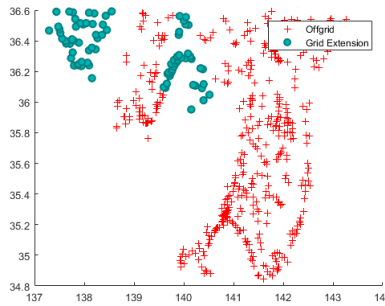Figure 3-4: Test Case 1 (520 Consumers) Result using $Moments$: 90% grid reliability

Figure 3-5: Test Case 1 Result using $PathToRootLength$: 90% grid reliability
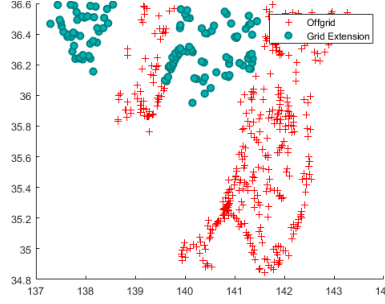
Figure 3-6: Test Case 1 Result using *PostOrder*: 90% grid reliability

Table 3.1:   Traversal Method Comparison at 90% Reliability: Test Case 1

Traversal Strategy	Offgrid Cost ($)	Grid Extension Cost ($)	Total Cost ($)
PathToRoot	190,293	100,606	290,899
PostOrder	217,214	70,660	287,874
Moments	217,214	70,570	287,783

Figures 3.4, 3.5 & 3.6 show the results of implementing the three algorithms on the same region using the *GreedyPrunDelta* procedure. At the same grid reliability of 90%, the three methods show a large amount of overlap in results, appearing visually similar. Reducing the grid reliability to 85%, the number of offgrid candidate nodes increases as seen in the following figures:

Figure 3-7: Test Case 1 (520 Consumers) Result using *Moments*: 85% grid reliability

At 70%, as observable in Figure 3-10, all three traversal methods lead to fully off-grid consumers, as the very low grid reliability becomes heavily penalized, favoring off-grid systems.

Quantifying and comparing the total costs (generation and network) across all
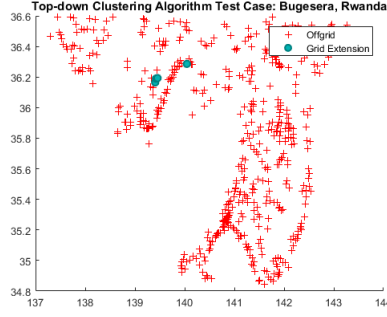
Figure 3-8: Test Case 1 Result using *PathToRootLength*: 85% grid reliability
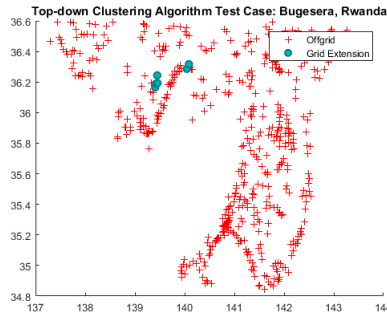
Figure 3-9: Test Case 1 Result using *PostOrder*: 85% grid reliability

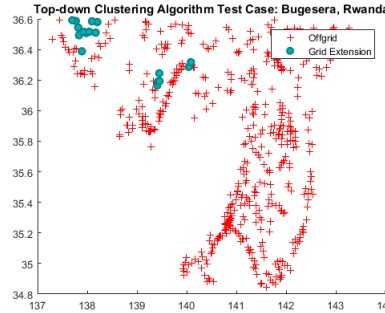methods provides more insight into similarities or difference in their performance. As seen in Tables 3.1 & 3.2, for the given test case 1 the *Moments* approach leads to almost the same total cost as the *PostOrder* method for this particular test case, indicating that there exist at least some geospatial distribution of consumers for which these two methods produce same traversal order.

Table 3.2: Traversal Method Comparison at 85% Reliability: Test Case 1

Traversal Strategy	Offgrid Cost ($)	Grid Extension Cost ($)	Total Cost ($)
PathToRoot	273,139	13,571	286,710
PostOrder	271,910	15,168	287,079
Moments	239,833	45,294	285,127
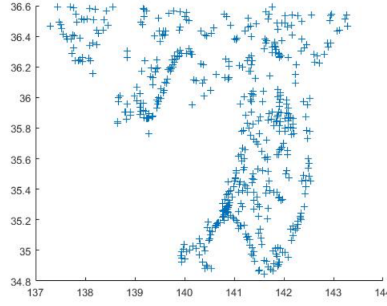
Figure 3-10: Test Case 1 Result for all methods at 70% grid reliability (All Consumers Offgrid).

3.5.2 'Bottom Up' - 'Top Down' REM Comparison

One of the objectives of this thesis is to explore how top-down based approaches to the consumer electrification mode recommendation and clustering problem compare to bottom-up methods such as that previously implemented in REM [30]. To this end, a larger test case (with 6688 consumers) was selected to study the performances of these two approaches. Figure 3.11 shows the geographic distribution of these 6688 consumers in the test case region as well as the surrounding existing grid.
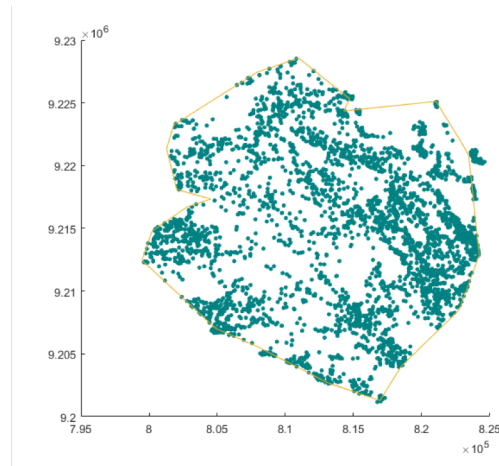
Figure 3-11: Consumers Distribution for Test Case 2

For the top-down method, $greedyPrunDelta$ was utilized as the tree partitioning procedure with the overall REM procedure implemented as presented in Figure 3.2. The DT-agglomerative method is the same discussed in chapter 2. Figures 3.12 & 3.13 show the final designs using the bottom-up and top-down methods respectively. To

interprete the visualized results, note that offgrid microgrid networks and standalone consumers are represented with green lines and purple dots respectively, while grid extension project candidates are delineated with red and blue lines for the MV & LV grid network, respectively.

Base Case Results

Using input data provided and user-defined parameters, both approaches were applied on test case 2. Before undertaking sensitivity analysis (the variation of input parameters to observe changes in input) on the test case, a reference base case has to be defined. The sensitive parameters under consideration in the application presented in this thesis are the user-defined diesel cost and reliability of the existing grid network. For the reference case, the diesel cost is set at $0.8/L$ and the grid reliability at 90%. The grid energy cost value is also set at $0.08 per kWh$. Running the test case using these values lead to the results presented in Figures 3.12 and 3.13.
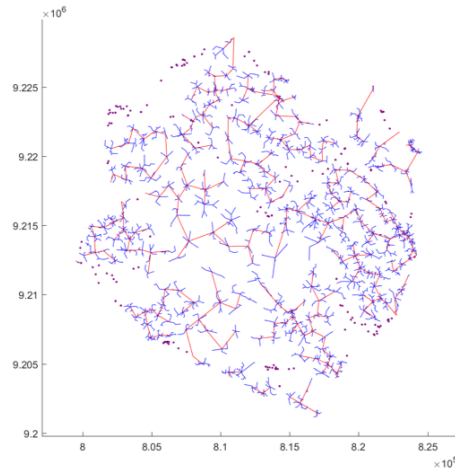
Figure 3-12: Base Case: Test Case 2 Result using Agglomerative '$Bottom-up$' based method

As can be observed, the results for this particular scenario and its associated input data show the top-down partitioning method leading to lower costs. The fraction

54

Table 3.3:   Base Case Bottom-up Results Summary

System Type	Microgrids	Isolated	Grid	All
Number of Customers	0	145	6543	6688
Annual System Cost ($)	0	139,706	3,378,329	3,518,035
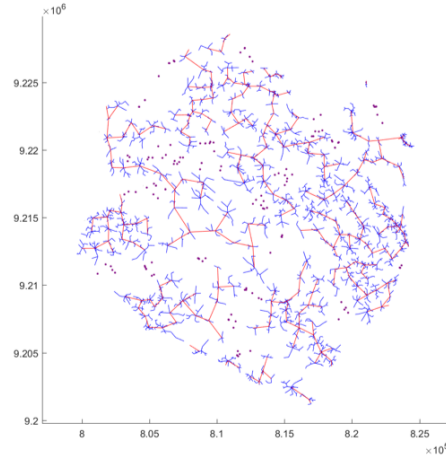
Figure 3-13: Base Case: Test Case 2 Result using 'Top-down' method

Table 3.4:   Base Case Top-down Results Summary

System Type	Microgrids	Isolated	Grid	All
Number of Customers	0	175	6513	6688
Annual System Cost ($)	0	115,756	3,330,305	3,446,061

of consumers assigned to off-grid versus grid-extension are also similar. The next subsections show how this test case responds to variations in certain user-defined parameters.

Sensitivity Analysis: Diesel Cost

Keeping all other data and parameters constant, the diesel cost was varied and results examined to see differences and sensitivities of both approaches to this parameter. Since the diesel cost parameter affects primarily the off-grid generation cost, intuitively, higher diesel cost should lead to less microgrid clusters than otherwise. Figures 3-14 to 3-19 show the diesel cost sensitivity analysis results for both approaches. For Diesel Cost = **$0.7/L**, the results are as follows:

Figure 3-14: Diesel Cost = $0.7/L: Test Case 2 Result using Agglomerative '$Bottom-up$' based method

Table 3.5: Diesel Cost = $0.7/L: Bottom-up Results Summary

System Type	Microgrids	Isolated	Grid	All
Number of Customers	0	157	6333	6688
Annual System Cost ($)	0	125,336	3,304,488	3,520,583

Table 3.6: Diesel Cost = $0.7/L: Top-down Results Summary

System Type	Microgrids	Isolated	Grid	All
Number of Customers	0	161	6527	6688
Annual System Cost ($)	0	128,529	3,327,160	3,455,689
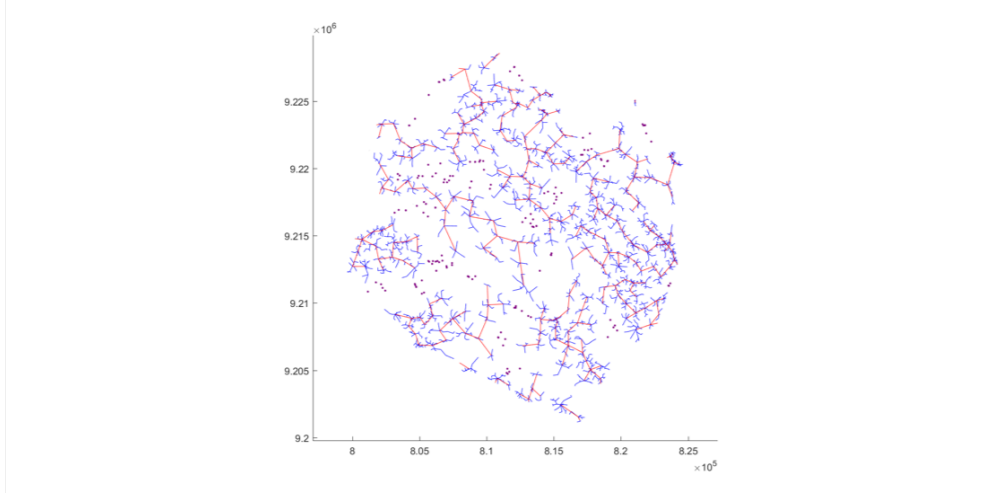
Figure 3-15: Diesel Cost = $0.7/L: Test Case 2 Result using using 'Top-down' method

For Diesel Cost = **$0.6/L**, the results are as follows:
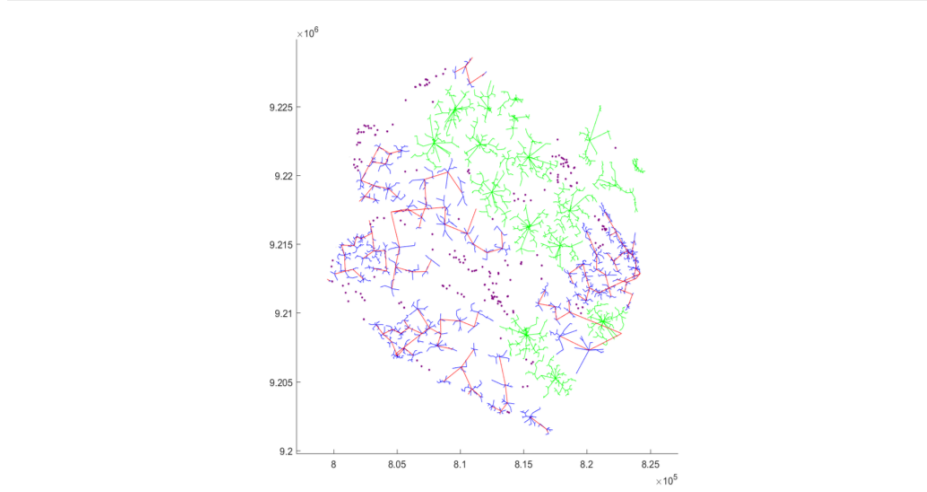
Figure 3-16: Diesel Cost = $0.6/L: Test Case 2 Result using Agglomerative '*Bottom−up*' based method

Table 3.7: Diesel Cost = $0.6/L: Bottom-up Results Summary

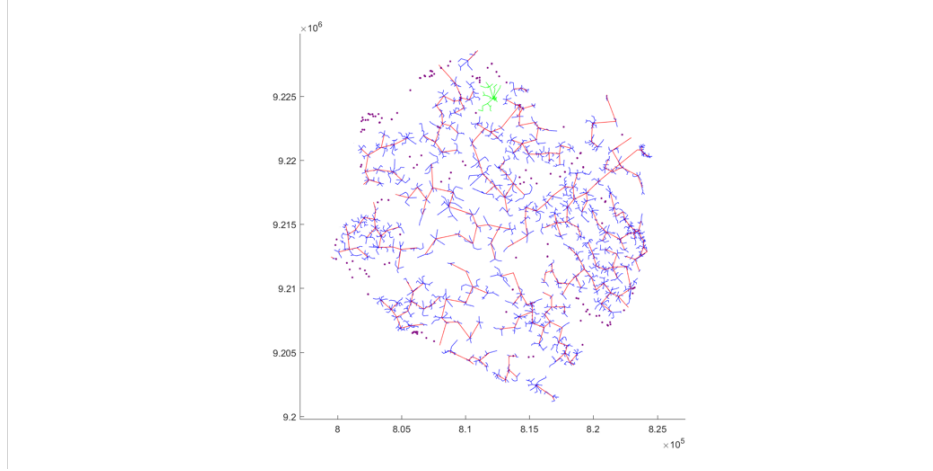| System Type	Microgrids	Isolated	Grid	All
Number of Customers	3602	234	2852	6688
Annual System Cost ($)	1,707,699	186,807	1,574,481	3,468,986

57
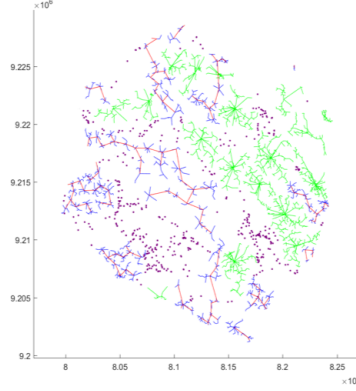
Figure 3-17: Diesel Cost = $0.6/L:Test Case 2 Result using using 'Top-down' method

Table 3.8: Diesel Cost = $0.6/L: Top-down Results Summary

System Type	Microgrids	Isolated	Grid	All
Number of Customers	3960	529	3199	6688
Annual System Cost ($)	1,872,419	422,311	1,295,432	3,500,162

For Diesel Cost = **$0.5/L**, the results are as follows:

Table 3.9: Diesel Cost = $0.5/L: Bottom-up Results Summary

System Type	Microgrids	Isolated	Grid	All
Number of Customers	6233	158	297	6688
Annual System Cost ($)	2,930,615	126,134	159,990	3,216,740

Table 3.10: Diesel Cost = $0.5/L: Top-down Results Summary

System Type	Microgrids	Isolated	Grid	All
Number of Customers	5685	170	833	6688
Annual System Cost ($)	2,656,906	135,714	490,731	3,283,352
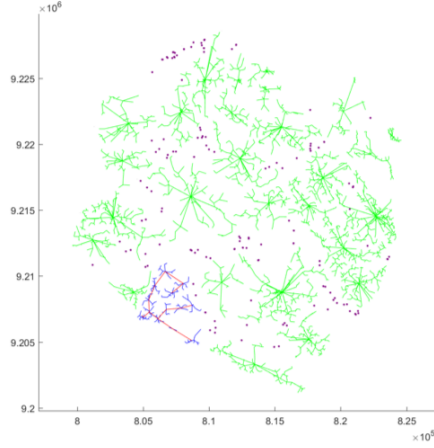
Figure 3-18: Diesel Cost = $0.5/L: Test Case 2 Result using Agglomerative '$Bottom-up$' based method
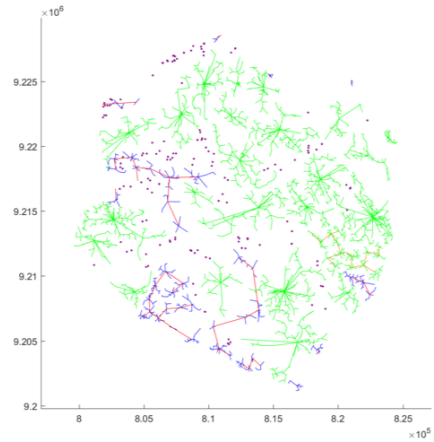
Figure 3-19: Diesel Cost = $0.5/L: Test Case 2 Result using using 'Top-down' method

The results from the diesel cost sensitivity analysis follow the expected intuitive trends; decreasing the diesel cost leads to more off-grid consumer candidates. Visually, both results have significant, though not exact, overlap in recommendations. In addition, in terms of cost, both approaches also lead to similar values; with the top-down leading to slightly lower system costs in all but one of the scenarios presented above.

Sensitivity Analysis: Grid Reliability

The sensitivity analysis was also repeated using the grid reliability level as the varying parameter. Intuitively, the expectation would be that at higher grid reliability, there would be more consumers assigned to the grid extension and vice-versa. Figures 3-20 to 3-23 show the grid reliability sensitivity analysis results for both approaches. The results for grid reliability at the reference value of 90 percent has already been presented.

At **100**% grid reliability, the results are as follows:
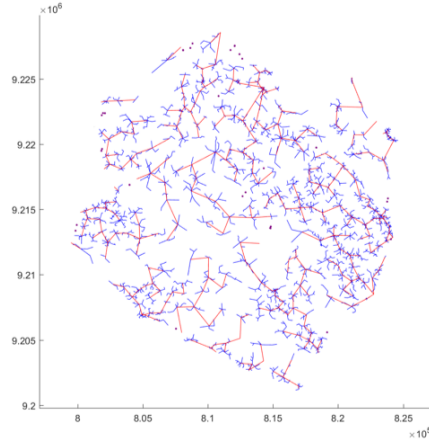
Figure 3-20: Grid Reliability at 100 percent: Test Case 2 Result using Agglomerative '$Bottom-up$' based method

Table 3.11:   100 percent reliability: Bottom-up Results Summary

System Type	Microgrids	Isolated	Grid	All
Number of Customers	0	34	6654	6688
Annual System Cost ($)	0	27,143	1,909,549	1,936,691
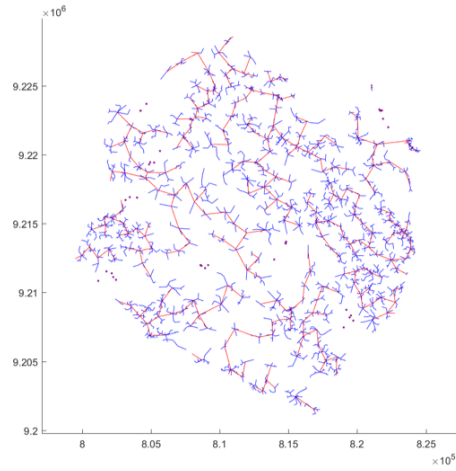
Figure 3-21: 100 percent reliability: Test Case 2 Result using using 'Top-down' method

Table 3.12: 100 percent reliability: Top-down Results Summary

System Type	Microgrids	Isolated	Grid	All
Number of Customers	0	37	6651	6688
Annual System Cost ($)	0	29,538	1,829,370	1,858,908

For grid reliability = **70**%, the results are as follows:
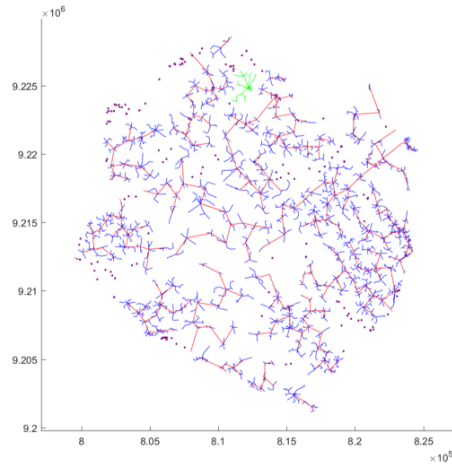
Figure 3-22: 70 percent reliability: Test Case 2 Result using Agglomerative '*Bottom−up*' based method

Table 3.13: 70 percent reliability: Bottom-up Results Summary

System Type	Microgrids	Isolated	Grid	All
Number of Customers	0	161	6527	6688
Annual System Cost ($)	0	128,529	3,327,160	3,455,689

Table 3.14:   70 percent reliability: Top-down Results Summary

System Type	Microgrids	Isolated	Grid	All
Number of Customers	198	157	6333	6688
Annual System Cost ($)	90,759	125,336	3,304,488	3,520,583
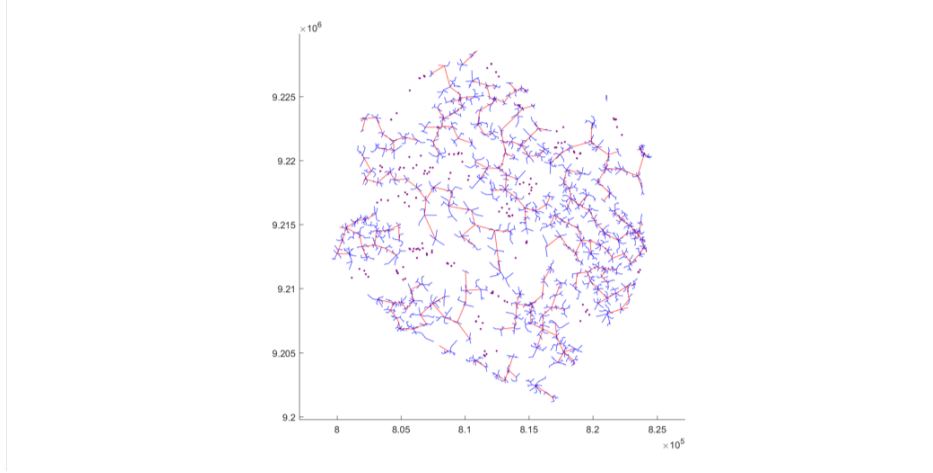
Figure 3-23: 70 percent reliability:Test Case 2 Result using using 'Top-down' method

As with the diesel cost sensitivity results, the results from the grid reliability sensitivity analysis follow the expected intuitive trends: decreasing the grid reliability led to more off-grid consumer candidates. Results from both approaches also have significant overlap in recommendations with similar cost values. At high reliability, the top-down led to lower system costs with the bottom-up dominating at lower reliability for the test case.

3.6   Network Partitioning: Insights and Future Work

Thus far, we have presented computational methods to address the technical aspects of planning electricity infrastructure for those without access. The review of the existing computational approaches to this problem highlighted the limitations that motivated the proposal of network partitioning methodology presented in this chapter. Implementing the proposed methodology, as we have shown, has led to promising results when benchmarked with the 'bottom-up' method incorporated within REM. This chapter concludes by discussing important insights garnered from the development and implementation of the network partitioning algorithms discussed herein, as well as suggestions on future work in this area.

- Network partitioning methods depend on the initial layout of the network designed by the network planning software. Sometimes, certain initial network layouts may lead to less desirable partition results. For example, the combination of low demand (as in rural areas) with large-sized components (lines and transformers) and a high number of candidate connection points, may lead to inadequate layouts. For future work, it may be worthwhile to refine the partitioning procedure such that it is capable of handling scenarios involving these.

- Partitioning the network allows factors which may be otherwise difficult to model, such as savings upstream of the grid due to planning decisions, to be captured. Upstream Network Reinforcement Costs - including upstream impact all the way to transmission and generation levels - may also be better incorporated using the approach presented and this can serve as an area of future research work.

- For very large-sized cases, it may be more computationally efficient to parse the input data to a pre-processing clustering module before running the network partitioning procedures discussed. This intermediate pre-processing step can be used to cluster or agglomerate customers into 'super-customers' before running the partitioning procedure. This can lead to significant time savings while obtaining reasonable, approximate results for these large cases.

- The sensitivity analyses showed that the 'top-down' approach led to lower system cost values than the 'bottom-up' method at relatively higher grid reliability and diesel cost values and the opposite was also true. Further analysis can be undertaken to understand the reason for these sensitivity results. Such analysis can help provide knowledge of what scenarios favor the use of the different methods, allowing a REM user to select the most suitable method on a case-by-case basis.

- It was observed that the piece-wise linear interpolation of catalog components,

used for incremental savings estimation, does not work properly when the smallest components are used. Thus, in low demand cases, the logic may have to be refined.

Chapter 4

The (Off-grid) Electricity Consumer Clustering Problem

The previous chapter addressed algorithms to partition a group of unelectrified consumers into partitions of grid-extension consumers and off-grid candidate consumers. This chapter addresses the clustering problem involved after such electrification mode recommendation has been done; the grouping of the off-grid consumers into off-grid microgrid clusters. This involves clustering the designated off-grid consumers into multiple microgrid clusters each served by an offgrid generation facility such that the overall cost of off-grid electrification is minimized. The following sections describe several clustering methods investigated to achieve this objective.

4.1 Problem Definition

Given a group of un-electrified candidate houses which have been designated for off-grid rural electrification, the objective is to obtain clusters of off-grid customers to be served with microgrid facilities. For such off-grid clusters, the spatial characteristics of houses and the economies of scale of a power system network mean that the goal is to connect as many people as possible together to any identified micro-grid generation center, while minimizing the entire network and investment costs i.e. network cost of lines and generation investment costs.

Thus if we identify k off-grid micro-grid clusters of consumers, the objective function for clustering if interpreted as an optimization problem can be defined as follows:

Minimize:

$$\sum_k C_{inv}^k \tag{4.1}$$

where: C_{inv}^k is the sum of C_{net}^k, the total network cost associated with the kth off-grid microgrid cluster and C_{gen}^k, the generation cost associated with the kth off-grid microgrid cluster.

It should be noted that the determination of the generation cost for a given off-grid cluster of consumer points i.e. the computation of C_{gen}^k for the kth cluster of consumers is in itself an optimization problem of optimizing different generation technology mixes subject to demand and supply constraints of the consumers in the given cluster. The determination of the network cost of a given off-grid cluster of k consumer points i.e. the C_{net}^k is a objective optimization problem in itself which can be approximately solved/generated by passing the coordinates of the cluster of points and their associated generation facility into an electrical network planning function (such as the Reference Network Model previously described) which then takes in a number of electrical parameters and builds a quasi-optimal network which has the necessary electrical wires, transformers, and other grid equipment for electricity supply. From the resulting network, the network cost, C_{net}^k which is the sum total of cost of edges (corresponding to cables) and required distribution equipment of the electrical network graph can then be estimated. This network design functionality as undertaken by a reference network model requires the following inputs:

- location (such as GPS coordinates) and characteristics of customers, and generation facilities.

- Technical and economic parameters, including: the discount rate, cost of ohmic losses, simultaneity factors at each voltage level, maximum voltage drop at each voltage level, load factors and loss factors in each voltage level.

- Technical characteristics of network equipment. This includes the following types of equipment:

  - LV, MV, and HV cables,

  - MV/LV transformer substations,

  - HV/MV substations,

  - Capacitors,

  - Voltage regulators and other reliability equipment.

Having defined the clustering objective above, described subsequently are some clustering methods and their implementation in the context of electricity consumer clustering.

4.2 REM's Traditional Approach to Clustering

Recall that the greedy 'bottom-up' procedure implemented in REM and described in chapter 2 of this thesis seeks to doubly partition consumers into either grid-extension or off-grid algorithms in addition to agglomerating similar consumers into clusters.

In [30], an MST based implementation of this approach is presented. The MST agglomerative algorithm involves first building a minimum spanning tree connecting all consumers, before then greedily agglomerating consumer nodes at the end of edges based on a connectivity measure until convergence. In the Reference Electrification Model (REM), the software whose development informs the work described in this thesis, the currently implemented off-grid clustering strategy involves building a Delauney Triangulation instead of an MST to connect all consumers in a procedure described subsequently.

4.2.1 A Delaunay Triangulation based Agglomerative Clustering Approach

As previously mentioned, the bottom-up clustering strategy currently implemented in the Reference Electrification Model involves clustering based on the Delaunay Triangulation (DT). A Delaunay Triangulation is a triangulation planar graph such that no node is inside the circumcircle of any triangle in the triangulation. Since this planar graph ensures that every node is connected to at least two edges (the degree > 1), clustering on the Delaunay triangulation rather than the minimum spanning tree means more edges or linkages are considered in the agglomeration process. Under this method, all consumers are initially in separate clusters but cluster agglomeration follows the arcs of the Delaunay. A consumer (or cluster of consumers) can thus be only be agglomerated with the others only if they at least some member of the other cluster is linkable to them by a Delaunay arc.

In order to apply the Delaunay Triangulation-based agglomerative strategy for clustering consumers as either microgrids clusters or grid-extension clusters, a formal definition of a connectivity or edge-activation measure between any two edges and their associated clusters, examined for agglomeration must be provided.

Given a Delaunay triangulation planar graph $G = (V, E)$ and $i, j \in$ V where nodes i and j are any pair of vertices of the graph connected by an edge, we can define the following cost parameters;

- The cost of off-grid generation and investment for each isolated cluster (or node) C_{gen}^i and C_{gen}^j on the sides of an edge.

- The total generation and investment costs for the combined clusters as a single node C_{gen}^{com}.

- The estimated incremental network costs for connection (equivalent line, including losses) between the clusters. Cx_l.

A binary edge-activation measure S_{ij}^{off} can be defined if two separate consumer nodes or clusters can be agglomerated for electrical connection into a single cluster served

by the same micro-grid generation facility as follows:

$$S_{ij}^{off} \in \{0, 1\}$$

$$S_{ij}^{off} = 1,$$

$$if \ C_{gen}^{com} + Cx_l < (C_{gen}^i) + C_{gen}^j$$

$$S_{ij}^{off} = 0 \ \text{otherwise}$$

Under this DT based clustering method, a cluster can be defined as a group of consumer nodes which are part of a connected component subgraph linked by activated edges.

With this defined, Algorithm 4 below describes the procedure to obtain the off-grid clusters. It should be noted that the assumption here is that initially all edges (connections) are not activated and so every customer node is in its own cluster of only that node. Agglomeration occurs when edges are activated - based on defined parameter comparisons - such that customers at both node ends of an edge are connected into same cluster. The edges and nodes of the graph are then updated appropriately to reflect this.

Algorithm 4 DT Bottom Up Clustering
___

  procedure OFFGRIDBOTTOMUP

    $DT \leftarrow$ construct Delaunay Triangulation connecting all consumers

    sort edges of DT in increasing length order

    initialize every consumer into separate clusters

    while there are still agglomerable nodes **do**

      evaluate shortest unactivated edge k of DT

      if $S_{ij}^{off} = 1$ **then**

        activate edge k

        merge clusters on either end of k associated with nodes $i \& j$

      end if

    end while

    return $clusters$

  end procedure

4.2.2 DT Agglomerative Algorithm Implementation

Herein, the results of applying the modified agglomerative clustering algorithm previously introduced to a small-scale offgrid electrification problem are discussed and

presented. Input data for the method include geolocation data for 1954 consumers, as well as information associated with the electrical demand profile for each consumer and hourly solar insolation data for input to the optimal generation sizing function. Figure 4.1 shows the geospatial distribution of the different consumers used for Test Case 2.

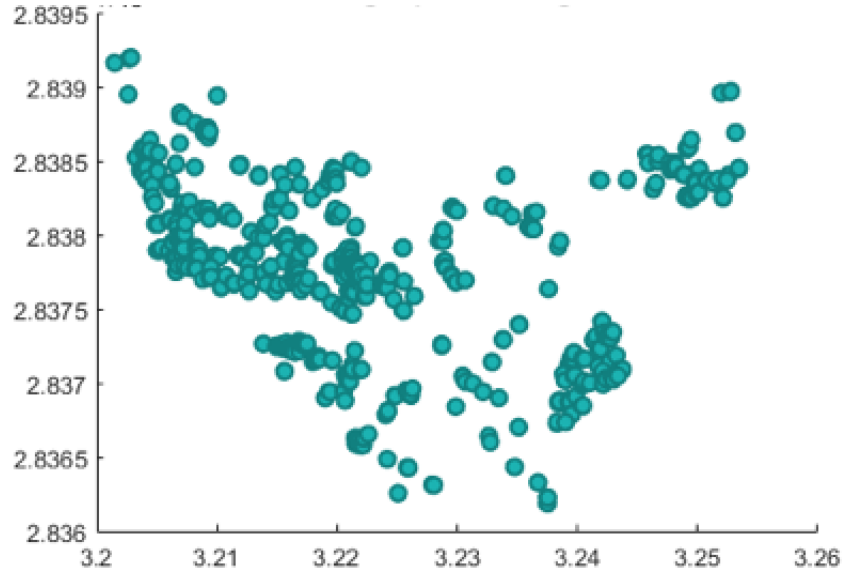Figure 4-1: Test Case 2: Consumers' Geospatial Distribution (1953 Consumers)

Running Algorithm 4 on these consumers leads to 9 microgrid clusters (visualized in Figure 4-2) with cost summary presented in Table 4.1.
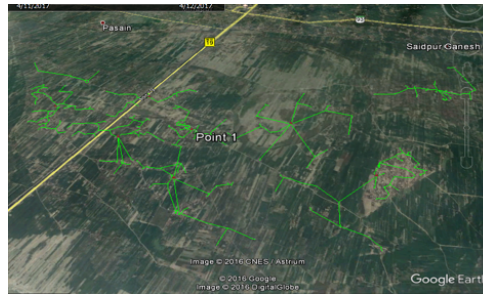
Figure 4-2: Off-grid Microgrid Clusters Using DT method

The cost summary shows that the generation cost, at almost 90% of the total cost, dominates over the network cost for this analysis.

Table 4.1: Cost Results of Algorithm 4 on Test Case: 2

Cost Parameter	Cost ($/year)
$TotalNetworkCost$	117650
$TotalGenInv.Cost$	975730
$TotalSystemAnnuity$	1093380

4.3 Other Explored Clustering Methods

Although small, the Test Case provided above has different spatial distribution of consumer locations allowing for comparison and benchmarking of Algorithm 4 alongside other clustering methods.

4.3.1 K-Means & K-Medoids Clustering

First presented, of the other methods explored, is a generic clustering algorithm - the k-means algorithm - which minimizes an objective cost function that is different from the offgrid microgrid cost function earlier defined. The K-means algorithm is an error-minimization partitional clustering algorithm which seeks to minimize sum of squared error while also requiring the user to input the number of clusters. As with many partitional clustering algorithms, it clusters data points by initializing a partition and moving through various partitions to find the clustering configuration that minimizes the overall dissimilarities. The k-means algorithm was applied to Test Case 2 and its outline is briefly described below.

Algorithm 5 K-Means Clustering

  procedure K-Means Algorithm
    Input data set and number of clusters k
    Initialize k cluster centers
    while till convergence **do**
       assign data points to the closest cluster center
       update cluster center based on assignment
    end while
    return $clusters$
  end procedure

To apply the k-means to these clusters of off-grid consumers, the techno-economic

network parameters and the associated generation sizing function for the population were retained. By setting the k value to equal the number of clusters or generation facilities from the agglomerative clustering case (i.e. k = 9), it is possible to examine if agglomeration using the defined edge-activation measure leads to lower cost than purely clustering based on Euclidean distance via k-means; such a comparison would answer the question of whether the domain-specific approach to clustering lead to a generic clustering method based on a different metric. It should be noted that the k-means algorithm leads to a locally optimal solution with respect to an objective function based on within group sum of squares. It might also be insightful to analyze and compare differences in cluster visualizations from both approaches.

Figure 4.3 shows results obtained from applying the k-means to the problem. Using the cluster result from the k-means, the generation and network design functions can be called to determine the cost of the generation facility C_{gen}^k and network associated with each cluster.
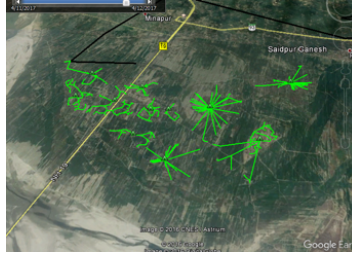

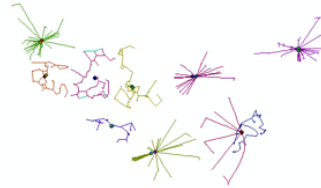Figure 4-3: K-means cluster of microgrids visualization

Figure 4-4: Test-Case 2: Designed Off-grid Microgrid Clusters from K-means

The electrical network per cluster of consumer nodes and their microgrid generation facility can then be designed thus providing the network cost associated with the cluster, C_{net}^k. It is thus easy to compute the overall cost by summing the generation

and network costs over all clusters and comparing with that obtained from using the previous algorithm.

We see the associated network designs of the microgrid clusters using k-means in Figure 4.4. The costs associated with this clustering approach are presented in Table 4.2. Visually, the clusters follow what one would intuitively expect if the clusters were based on proximity. However, comparing the cost results to the DT agglomerative method shows that the k-means method led to a higher overall system cost. The result also indicates that most of this cost difference in this case-study comes from the network cost component. This is not surprising since no direct network input data fed into the k-means clustering algorithm.

Table 4.2:   Costs of Microgrid Clusters Using Generic K-Means Method

Microgrid	Microgrid Network Cost($/yr)	Mg Gen Inv Cost ($/yr)	Microgrid Total Cost ($/yr)
mg1	11028	86720	97748
$mg2$	30592	118900	149492
$mg3$	32934	140460	173394
$mg4$	14020	124100	138120
$mg5$	7901	85030	92931
$mg6$	12870	77800	90670
$mg7$	12331	103280	115611
$mg8$	29551	185070	214621
$mg9$	11635	76110	87745
$Total$	162862	997470	1160332

The k-medoids is a centroid based algorithm similar to k-means but which unlike the k-means involves a data point being chosen as cluster centroid in the course of the algorithm. The k-medoids algorithm was also applied to the given data-points and compared, as in the K-means algorithm case, with the agglomerative clustering approach. Figure 4.5 and Figure 4.6 show the clustering and network design results of the K-medoids algorithm.

Again, after designing the microgrid cluster network based on this clustering result,
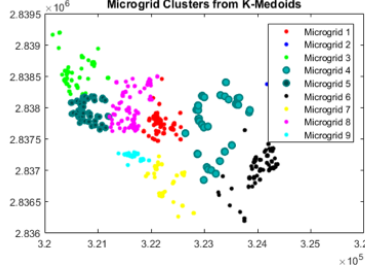
Figure 4-5: K-medoids cluster of microgrids visualization

the following network design is obtained:
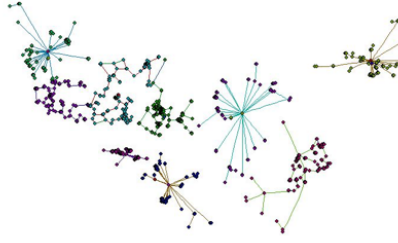
Figure 4-6: K-medoids cluster of microgrids visualization

Table 4.3: Costs of Microgrid Clusters Using Generic K-Medoids Method

Microgrid	Microgrid Network Cost($/yr)	Mg Gen Inv. Cost ($/yr)	Microgrid Total Cost ($/yr)
Total	174996	959785	1134781

Similar to the k-means results, the k-medoids clustering results also follow what one would intuitively expect if the clusters were based on proximity. There is also significant overlap with that of the k-means result although the overall cost is lower than the k-means but still higher than that resulting from the DT agglomerative clustering approach.

4.3.2   Density Based Clustering Approach

Examining C_{inv}^{k} which is based on generation and network costs suggests that a density-based clustering approach may lead to lower clustering costs results. This is because the network costs are computed from the summation of the costs ('weights)

of the designed radial electrical distribution network tree as well as any required distribution equipment such as transformers designed as part of the network. Thus, it can be reasonable to expect that clusters with higher density would have shorter edges and equivalently shorter network costs. While the generation cost is dependent on the number of consumers i.e. it increases, there are economies of scale associated with having more consumers per cluster. As discussed previously, the generation cost is determined by solving another optimization problem to find the minimum cost combination and value of generation resources (battery or diesel or solar PV) based on the demand profile of the consumer that would meet the pre-defined reliability constraints. For consumer nodes of the same type, the generation function (when called as a look-up table) generates a monotone function with economies of scale. For example, figure 4.7 below provides a plot of the computed (from the generation resource oracle function call) per customer generation costs for a microgrid cluster as the number of consumers in the cluster increases.
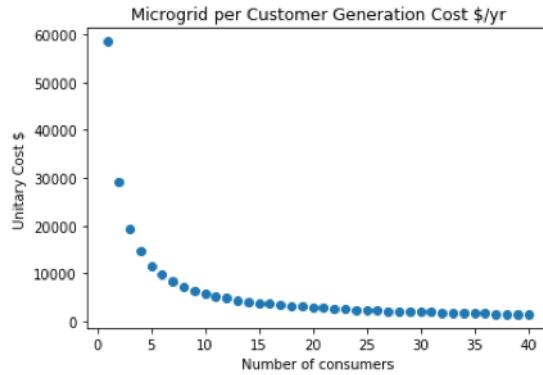
Figure 4-7: Microgrid Unitary Generation Cost

Thus, it can be seen that density-based algorithms which favor 'denser (more consumers over small area) micro-grid clusters may lead to lower cost solutions than has been obtained using the previously discussed algorithms. One such density-based clustering algorithm, the DBSCAN (Density Based Spatial Clustering of Applications with Noise) Algorithm is applied to the dataset of Test Case 2 to test this hypothesis and the results obtained are discussed subsequently.

The DBSCAN algorithm, as proposed by Ester et al. [32], designates some

data points as noise. For the purpose of our clustering problem, we require that no points are designated as noise and so by varying the two parameters of epsilon and min_points, we can eliminate noise. The epsilon parameter determines how close points are to be part of a cluster while the min_points value controls the minimum number of neighbours to a point in a cluster. Running the DBSCAN and varying both parameters, final clustering results for design involved a choice of parameter values epsilon = 200, and min_points =5 and 7 clusters with no noise. As with all clustering results, the electrical network for the microgrid clusters was designed producing the following network visualizations:
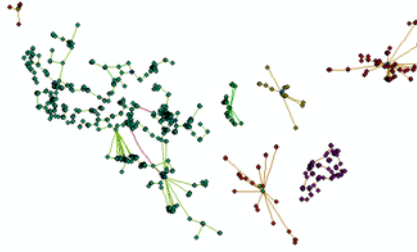
Figure 4-8: Test Case 2: Off Grid Microgrid Clusters Designed Using DBSCAN clustering

The costs are also tabulated as shown in Table 4.3.

Table 4.4: Costs of Microgrid Clusters Using DBSCAN

Microgrid	Microgrid Network Cost($/yr)	Mg Gen Inv. Cost ($/yr)	Microgrid Total Cost ($/yr)
mg1	530	58534	59064
mg2	105880	556030	661910
mg3	7900	85032	92932
mg4	2790	65977	68767
mg5	2060	60693	62753
mg6	6940	61635	68575
mg7	17410	105140	122550
Total	143510	993041	1136551

It can be observed that the total network cost using the density-based DBSCAN is less than that from the k-means and k-medoids algorithm, in line with the hypothesis.

78

Its network costs are second only to that of the agglomerative clustering presented. The generation costs are however only higher than that of k-means. The total cost of the DBSCAN approach is very close to the value as obtained for the k-medoids which yielded a lower aggregate generation cost value. For a better understanding of possible approaches using this generic approach to clustering electricity consumers, different Test Cases with varying geospatial and demand characteristics may need to be evaluated with the algorithm.

4.3.3   Spectral Clustering

Another method investigated is the application of nearest neighbor based spectral clustering, a graph-theoretic approach which involves the creation of a similarity/affinity matrix and a low-dimension embedding on this matrix before clustering in low-dimensional space. For the results presented below, a similarity-matrix based on 10 nearest neighbors is constructed and the clustering results as applied on the same data and using the same technical and economic parameters can be seen in Figure 4.9.
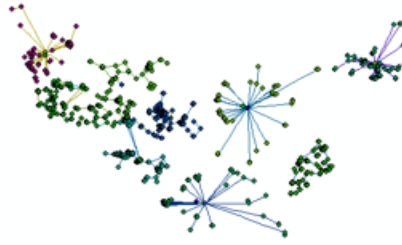
Figure 4-9: Test Case 2: Designed Off Grid Microgrid Clusters (10-NN Spectral Clustering Approach)

The network cost obtained is higher than with the DBSCAN algorithm, although there is a significantly lower total generation cost and the total cost with the spectral clustering approach is lower.

These results show that one possible direction for future work would be to explore spectral/graph cut-based clustering algorithms in which the computation of adjacency

Table 4.5: Costs of Microgrid Clusters Using 10-NN Spectral Clustering Method

Microgrid	Microgrid Network Cost($/yr)	Mg Gen Inv Cost ($/yr)	Microgrid Total Cost ($/yr)
Total	157192	965914	1123106

matrices are based on the parameters unique to the problem such as costs of electrical lines between two points, or other related parameters.

4.3.4   A Submodularity Clustering Approach

Considering the economies of scale property exhibited by the dominant generation cost, it may be possible to explore other scalable optimization approaches to this clustering problem which can take advantage of this property. To this end, we define below a class of set functions said to be submodular.

Definitions

A set function $f : 2^V \to \mathbb{R}$ is said to be submodular if for any $A, B \subseteq \mathrm{V}$,

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B) \tag{4.2}$$

The above is the classic mathematical definition for submodularity in which V represents the ground set on which the function f operates. Submodular functions can also be defined using incremental diminishing benefits. Under this equivalent definition, a set function $: 2^V \to \mathbb{R}$ is said to be submodular if, for any $A \subseteq \mathrm{B} \subseteq \mathrm{V}$ and $a \in V$,

$$f(A \cup \{a\}) - f(A) \geq f(B \cup \{a\}) - f(B) \tag{4.3}$$

These definitions express the property of diminishing marginal returns of a utility function (or economies of scale for a cost function). That is, when a new element is added to a set of elements, the marginal change in function value diminishes.

Submodularity in the Electricity Consumer Clustering problem

Submodularity clustering methods have been applied to coverage and facility location problems in literature [21] [67]. For the electricity consumer clustering problem, if we define the ground set V as the set of all consumers nodes which are to be clustered, then we can re-represent the problem as an optimization problem amenable to submodular clustering techniques.

Recall that the objective function of this clustering problem encompasses the generation costs and network costs. These two cost elements can be examined separately for submodular properties. This is because, an important property that is useful for submodular function optimization problems is the fact that the sum of submodular functions is also submodular [16]. Using definition (4.3) and the results from the oracle calls of the generation function, it is more straightforwrd to see how the microgrid generation function is submodular. Examining figure 4.10 shows this.
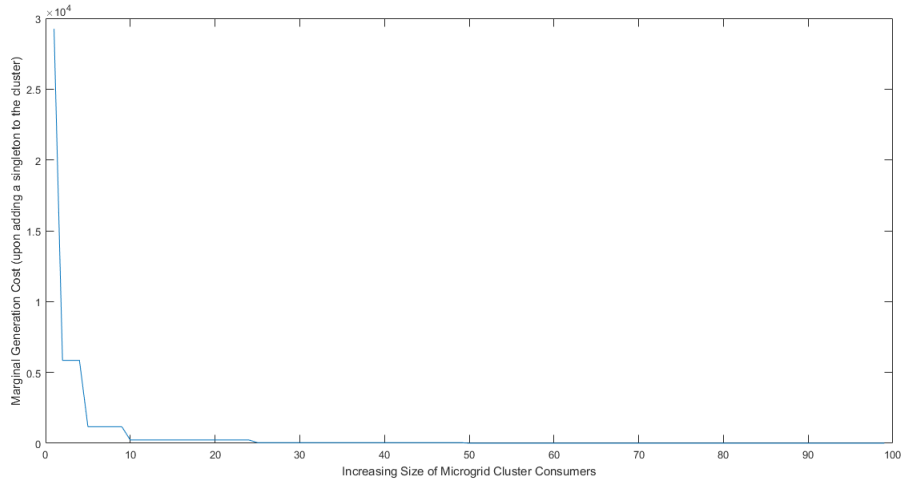
Figure 4-10: Marginal Generation Cost as Cluster Set Size Increases

Figure 4.10 is a plot showing how the marginal cost of adding an additional microgrid customer to a microgrid sets is non-increasing. In Figure 4.10, as we add a marginal customer node $\{a\}$ to the set A of nodes in a microgrid, the marginal cost $F(A \cup \{a\}) - F(A)$ can be seen to either decrease or stay the same. Thus for any two sets of microgrid consumer elements, A and B such that $A \subseteq B$, the following holds:

$f(A \cup \{a\}) - f(A) \geq f(B \cup \{a\}) - f(B)$, as in definition (4.3).

Note that for an empty set the cost function value is 0.

Also, since the sum of submodular functions is also submodular, the sum of generation costs of multiple microgrids is still submodular and can lend itself to submodular minimization methods.

To minimize over a clustering objective function of the sum of both the generation and network functions like has been done for the other methods described, a submodular network cost function would be necessary. If it can be shown that the associated network cost for a set of nodes is also submodular, then the overall cost becomes a submodular function. The computation of the network cost is however done by calling a black box network design function (the RNM) and does not readily lend itself to analysis.

This does not mean that the network aspects of the costs cannot be captured with a submodularity based optimization approach. Using what is known about how the network costs are computed by the network function, a proxy function with submodular properties can be designed on the set. Since the DBSCAN clustering method explored previously showed that network cost reduced with denser clusters, a submodular function which leads to denser clusters may capture network costs. If the location of the generation facility within a cluster of a set of nodes is fixed, a node with a high marginal network cost in this micro-grid cluster would be one farthest from the generation facility. Thus, we can define a function on the ground set as the maximum euclidean distance of the generation facility to nodes.

Again, using definition (4.3), it can be shown that this is submodular as follows. If the ground set is V, and subsets A and B are chosen such that $A \subseteq B \subseteq V$, then under this function definition $f(A)$ represents the maximum distance of the generator in A to any node in A and is monotone non-decreasing. The resultant cost function from adding a singleton node $\{e\}$ to A is $f(A \cup \{e\})$.

Note that $f(B \cup \{e\}) - f(B) \leq f(B \cup \{e\}) - f(A)$

since;

$$f(B \cup \{e\}) \geq f(B) \geq f(A) \geq 0$$

In addition, if the addition of the same singleton node $\{e\}$ to both A and B is considered, it can be seen based on the function definition that:

$f(B \cup \{e\}) \geq f(A \cup \{e\})$

since:

$A \cup \{e\} \subseteq B \cup \{e\}$ and the function is monotone non-decreasing.

In the case where $f(B \cup \{e\}) = f(A \cup \{e\})$, then from (4.2) and (4.5),
$f(B + \{e\}) \cup f(B) \leq f(A \cup \{e\}) - f(A)$ or,
$f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B)$

In the case where $f(B \cup \{e\})$ is strictly greater than $f(A \cup \{e\})$ i.e. when
$f(B \cup \{e\}) > f(A \cup \{e\})$

then $f(B) = f(B \cup \{e\})$. This is because, the defined set function operates as a max function on set elements. Since $f(B \cup \{e\}) > f(A \cup \{e\})$, then the argmax element can not be $\{e\}$ and is from set B. And since $B \subset (B \cup \{e\})$, then $f(B) = f(B \cup \{e\})$ Since $f(A \cup \{e\}) - f(A) \geq 0$,
$f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B) = 0$
Thus completing the proof.

Thus, a submodularity clustering method can be applied to an objective function which is the sum of the submodular generation cost and the submodular network proxy cost. The next section addresses the application of one such method to the clustering example described in previous sections of this chapter.

Application to Case-Study

Herein, the Test Case 2 offgrid consumer clustering problem is revisited using a submodular clustering approach with the objective function as previously formulated. Many algorithms exist for getting exact or approximate submodular optimization solutions and they have been extensively analyzed in [19] [20] [66] [67] [72] . In the SFO MATLAB toolbox described in [45], MATLAB implementations of multiple submodular function optimization methods are also provided for researchers to test some

common submodular function optimization algorithms. The application described in this subsection focuses on a greedy method presented in [72] known as the $GreedMin$. In [72], $GreedMin$ is proposed as an approximation algorithm to the submodular load balancing problem with the additional effect of obtaining resulting clusters with similar sizes. This load-balancing objective is desirable in a real world off-grid clustering setting, allowing the planner to have similarly-sized microgrid clusters per generation facility. A proper review of submodular load balancing in general is provided in [72].

The $GreedMin$ algorithm, as presented in [72], is described below.

Algorithm 6 GreedMin

    procedure GREEDMIN(f, m, V)
        Let $A_1, ..., = A_m = \emptyset$
        Let $R = V$
        while $R \neq \emptyset$ **do**
            $j^* \in \text{argmin}_j f(A_j)$
            $a^* \in \min_{a \in R} f(a|A_{j*})$
            $A_{j*} \leftarrow A_{j*} \cup a*$
            $R \leftarrow R \setminus a*$
        end while
        return $\{A_i\}_{i=1}^m$
    end procedure

Applying $GreedMin$ to the Test Case 2 and setting m to be equal to 9 as in the k based clustering algorithms previously implemented, clustering results were obtained which can be seen in Figure 4.11. As expected, the resulting clusters are similarly sized or 'balanced'. In addition, some of the clusters visually match those obtained with other algorithms discussed previously.

The overall costs figures associated with the $GreedMin$ algorithm are shown in table 4.6. At just $\$947,636$, it can be observed that the overall generation cost obtained with this method is lower than those obtained with other clustering methods. The overall clustering cost obtained is also relatively low. These low costs are complementary to the additional observation that the cluster sizes are also more uniform, or 'balanced', a consequence of the load balancing approach. It should be noted that the submodular clustering approach applied need not necessarily be a 'load-balancing' one and an area for future work on this may involve applying or developing other sub-

modular clustering approaches to this problem.
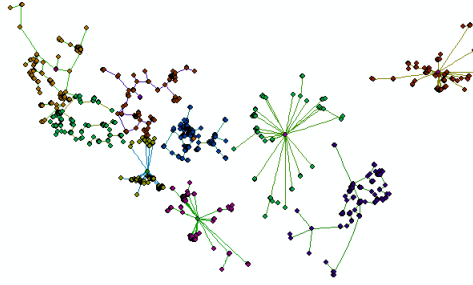
Figure 4-11: Test Case 2: Off Grid Microgrid Clusters based on GreedMin

Table 4.6: Costs of Microgrid Clusters Using 'GreedMin' Method

Microgrid	Microgrid Network Cost($/yr)	Mg Gen Inv. Cost ($/yr)	Microgrid Total Cost ($/yr)
mg1	16322	84550	100872
$mg2$	24215	99565	123780
$mg3$	31132	102167	133299
$mg4$	19370	163506	182876
$mg5$	21097	102910	124007
$mg6$	8044	102167	110211
$mg7$	7901	85033	92934
$mg8$	30070	112576	142646
$mg9$	11178	95163	106341
$Total$	169330	947636	1116966

4.4   Chapter Conclusion

In this chapter, the off-grid electricity consumer clustering problem has been extensively examined. Several computational methods from optimization and machine learning literature were implemented and applied on a test case for comparative analysis. In particular, clustering spectral and k-center clustering. A greedy heuristic formulation and a submodular load balancing approach to the clustering problem were also implemented and presented. The results showed many visual similarities

and differences as well as in overall costs across methods. For the cases analyzed, the DT agglomerative method led to lower costs in comparison to the generic clustering methods such as the k-centroid clustering methods. Results also show that submodularity formulations which exploit the economic of scale properties of the cost functions associated with this clustering problem may yield desirable results and may be a promising method for addressing a real-world electricity clustering problem. Finally, with respect to the overall Reference Electrification Model, the methods compared in this chapter can be integrated with the top-down network partition procedures previously presented in the following ways:

- As a post-processing step required after bi-partitioning consumers into grid-extension and off-grid project candidates. While the current implementation of the 'top-down' method uses the DT Agglomerative algorithm to cluster the off-grid candidates after partitioning, other methods such as the submodular clustering approach can be similarly integrated.

- As a pre-processing step before top-down network partitioning is undertaken for very large case studies. The ideas discussed can be extended to pre-cluster a data set with a large amount of input consumers into smaller super-consumers, thereby speeding up computation for these large-sized cases.

Chapter 5

Discussion and Conclusion

5.1 Beyond Computation-Aided Planning: The Socio-Regulatory Context of Electricity Access

As mentioned in the introductory chapter of this thesis, the problem of energy access is a multi-faceted one spanning economic, social, regulatory and technical barriers. Backed by a powerful computation aided planning tool such as the Reference Electrification Model (REM), a decision maker may still be unable to make requisite electrification decisions such as investment in the absence of suitable policy and an enabling business environment. This section examines some of the non-technical dimensions of the energy access challenge. Subsequently, some recommendations are provided which complement the techno-economic analytic capabilities of REM. It is hoped that these recommendations, if implemented, may ultimately facilitate universal energy access. Before delving into the recommendations, an overview of the different stakeholders in rural electrification decision-making is presented.

5.1.1 Stakeholders

The complexity of decision making for energy access is accentuated by the number of stakeholders in the electrification landscape. For example, there is usually an incumbent distribution utility company which ideally is responsible for ensuring universal

energy access for consumers in the territory of its operation. Utilities in developing countries are however subject to unique challenges which limit their ability to extend the grid to those without access. In addition, the consumers to be electrified can also be regarded as stakeholders since their demand requirements and ability to pay may affect optimal electrification mode recommendations and can choose to accept or reject electrification technological offerings. As members of the electorate, consumers may also inspire the politicization of government electrification decisions, by ensuring that only governments who plan to make electrification decisions favorable to them are elected or remain in office [46]. Thus, electrification decision-making in developing countries may be described as a complex interplay of stakeholders with the consumers at the center. Some of these decision-makers and stakeholders are discussed below.

Government Energy Departments

The government departments of power/energy in developing countries are in charge of defining the overall energy policy which will either spur or impede the pace of energy access. Such government departments may also have designated rural electrification agencies or parastatals which help execute different rural electrification projects in the country by providing finance or promoting rural electrification projects. As was highlighted previously, sometimes the policy decisions made by the government agency may be politically motivated or subject to regulatory capture since their decisions may affect their sustenance in authority and re-election. Government energy departments, as policy-implementation experts, usually oversee the high-level electrification decisions of the country which may affect all other stakeholders.

Electricity Regulatory Commissions

As promulgators of regulations which govern the various electricity provision activities from generation to distribution and retailing, regulators can help accelerate or mar energy access. For example, the absence of regulations such as those on off-grid microgrids - a situation in many countries - may affect both intending providers of energy services as well as potential consumers. Under such situations of regulatory

absence and uncertainty, energy service providers such as microgrid developers may be unsure of important factors affecting their businesses such as microgrid standards, where to locate their microgrids and how to structure remuneration. On the other hand, the consumers may also not be appropriately protected in the absence of regulation, and may be exposed to unregulated tarriffs, poor service standards or left without access as energy service providers go for the least risky and most economic projects or "the low hanging fruits".

The presence of regulation does not guarantee successes and regulatory measures may sometime have unintended consequences. "Bad" regulation can be a deterrent to proper investment and may lead to a failure in achieving policy goals and indeed, this is the situation in many developing countries in sub-Saharan Africa [12].

Electricity regulatory commissions whether at the state level or at the national level must take care to account for multiple scenarios and stakeholders, making them one of the most crucial players in the drive towards universal energy access.

Electricity Distribution Companies

Distribution utilities are also major stakeholders in electrification planning and decision making. Typically, a distribution utility by law is assigned a territory, for which they are in charge of electricity distribution and the provision of access to electricity to all consumers within. In line with regulatory standards, the utility's planners may then coordinate this distribution activity and obtain remuneration through agreed tariffs.

Structurally, the distribution utility could be public, as part of vertically-integrated government utility monopolies, or private or established under a public-private partnership. The grid extension activity for energy access is under the purview of distribution companies and would ideally be a solution for all consumers. However, as REM demonstrates, solutions which combine grid-extension with off-grid microgrids and systems may be more economical given the constraints such as tariff ceilings faced by these utilities and the cost of grid power distribution. Consequently, some of the off-grid microgrid sites, though under the territories of utilities, may then be served

by private microgrid developers, raising important questions such as what to do when the grid eventually arrives and the need for grid compatible microgrids.

In theory, with sound regulation, a planning tool such as REM, and an enabling business environment, a single utility may also be able to electrify all consumers using both grid extension and off-grid electrification strategies. However, many utilities in developing countries face multiple challenges which hamper both the distribution of electricity to already electrified consumers, and the extension of the grid to those without access. Some of these are discussed below.

- High Losses: With values up to as high as 60% for some utilities, the Aggregate Technical and Commercial Losses faced by many distribution companies in developing countries are typically very significant [70]. This is due to multiple factors ranging from the prevalence of electricity theft through illegal wiring, to the continued subsidization of tariffs even in the face of financial challenges. Thus, collection rates are low for many of these utilities, affecting their bottom-lines.

- The losses (and associated debt) by these distribution companies mean that they are unable to meet the power purchase agreement payments from generation. The distribution companies are reluctant to procure power (due to associated losses), making the generators run on low capacities. This in turn reduces the bankability of the generation activity as prospective generation investors are uncertain of investment.

- Rural distribution for unserved consumers is expensive: in most of the countries which lack 100% electricity access, the population fraction of people without access to electricity remains very high, with the absolute number typically in the millions [4]. While there is a universal access obligation in many national electricity laws as per the sustainable development 'Goal 7' agreed by many countries in 2015, on-grid connection of these households remains expensive since the disperse nature of rural, un-electrified households means that rural electrification is almost universally a more expensive undertaking than urban

electrification [2]. At the same time, the willingness to pay for reliable electricity by many residential consumers in these countries is generally considered low with some citizens deeming it as a service to be received freely [68] [6] [59]. This raises the all-important issue of the viability gap and how it can be addressed, if the electrification of the all un-served population will be undertaken by the debt-ridden utilities (or the private off-grid developers). The utilities may need to be subsidized or financially driven by external sources if affordable electricity will be provided without further burdening the already debt-ridden utilities in these countries.

- Subsidized, Non-Cost-Reflective Tariffs: In line with social objectives and political realities, electricity tariffs across all tiers in developing countries are heavily subsidized according to the government policy. For instance, in Uttar Pradesh, the largest state in India and one with 55% rural household electricity access, agricultural consumers - despite being consumers of relatively large quanta of power - are highly subsidized in line with government policy since the state is an agricultural economy-based state [3]. Providing this subsidy through un-metered power supply has had significant cost implications on the state utility [39]. In addition, residential consumers are also typically cross-subsidized by industrial and commercial consumers in many developing countries. Overall, it is common that the collected remuneration of the distribution company based on the tariff does not cover the costs without additional compensation.

- Compensation from the government to meet subsidies may be inadequate and their disbursement delayed by bureaucratic approval processes. This follows from the previous point. Delayed subsidy compensation combined with losses and low collection rate incurred may mean that distribution companies struggle to maintain financial viability [70].

- The tariff remuneration may be designed such that the distribution company has no incentive to improve the delivered service. It is not uncommon for the tariff to not be properly designed to reward performance and improved service

delivery [70]. This coupled with delayed or inadequate compensation from the government may not provide any incentive for the distribution company to make extra investments to improve service or extend the grid.

Development Organizations and Financiers

Development organizations have the objective of creating large-scale impact especially in the all-important area of energy access. These organizations are able to provide some of the much needed finance for electrification projects and often have to make decisions such as how much to fund, where to electrify, etc. with these decisions being those that can be supported with the use of REM. They may, for example, want to understand how and where to prioritize funding disbursement to ensure maximum impact.

Another important characteristic of development organizations or banks is the fact that they are able to have diversified impact that cuts across geographical or national borders. These organizations typically engage in global development projects and have experts with understanding of variations in social and political realities across countries. Development organizations also do extensive market research, collecting, publishing and updating data and reports on electrification. They can leverage on these and their alliances with other stakeholders such as governments and the private sector to achieve impact.

Offgrid Microgrid Developers

Off-grid microgrid developers are typically private sector players with the technical and economic capabilities to provide electricity services to underserved or un-served consumers who are not connected to the grid. As such, they are exposed to the various socio-political and regulatory barriers (or lack thereof) which can affect the viability of their business models. The decisions made by these developers also include those on who to electrify and how to electrify from a technoeconomic perspective - questions that can be addressed with the REM methods presented in previous chapters of this thesis. In addition to these, they also face risk-related decisions such as how to

operate in the absence of microgrid regulations and viability when grid is ultimately extended to the consumers electrified by their off-grid project.

Even when microgrid regulations exist, they may still prove unfavorable to the developers. Uttar Pradesh in India, where the state electricity regulatory commission - the UPERC - has made significant effort for universal access by providing regulation on mini-grids in the state, is an interesting example. The microgrid regulation promulgated there categorizes the types of mini-grids that can be regulated into two, with different regulatory environments governing them. According to this regulation, the first category of microgrids are typically for villages designated to be permanently served by off-grid electrification and any developer electrifying these villages would receive state subsidies with these mini-grids set at state regulated tariffs. In the other category, mini-grid operators may be deemed free to negotiate for tariffs they desire, but without access to state subsidies [22]. Grid-arrival remains an issue as under this regulation; when grid-extension arrives it is the burden of the mini-grid developers to either negotiate the sale of their assets to the distribution company or continue operating them at what would probably be unfavorable economic conditions.

In addition to regulatory risks, the off-grid developer may also face other challenges. By investing in off-grid projects to electrify consumers that are otherwise not economically viable, the microgrid stakeholders may have to receive subsidies from the government or other financiers. Such public-private partnership may also come with risks associated with the likelihood of bureaucratic delays in subsidy disbursement, government defaulting or unexpected unfavorable policy change due to political instability.

The challenges with developing sound rural (off-grid) microgrid regulations notwithstanding, it is still important from a government perspective to regulate these alternative electrification solutions. If private microgrid developers are left to negotiate and charge tariffs as they like, economics dictates they will target the most profitable consumers or "low-hanging fruits. This would set back the governments social objectives of achieving 100% access to electricity since the poorest who have the lowest willingness to pay would be left behind. Off-the-grid solutions are however necessary

to complement the grid-extension efforts by the governments and utilities. Thus the government has dual objectives; the creation of an enabling business environment so that the requisite investment can come in for off-grid microgrid developments and, ensuring a regulatory atmosphere that also encourages energy access services to the least economical consumers.

5.2 Recommendations for Viable Business Models

This thesis has so far discussed computation techniques which can greatly enhance electrification planning. The major stakeholders of the electrification landscape in developing countries have also been mapped out, with key challenges faced by each player identified. Achieving universal energy access will require recognition of the different dimensions of the problem presented. More importantly, it will require the establishment of enabling environments for business models which can address the issues to thrive. To conclude this thesis, some recommendations to this effect are presented below.

- It appears that cross-subsidization is a necessary condition for the viability of business models and should be encouraged if universal access is to be achieved. This may be the only way to ensure that the consumers with low willingness to pay are not left behind. Cross-subsidization schemes that cross-subsidize consumers with higher willingness to pay or load sizes with those with lower should be explored.

- Micro-grid regulations are necessary. Minimum quality of service must be established. Carefully designed regulations are required if universal access is to be established. For instance, if, as in the Uttar Pradesh case study raised earlier, micro-grid developers are left to strike mutually agreeable tariffs with consumers without regulatory intervention, then it is very likely that there will be some trade-off in either quality of service or in tariff levels. Since the electricity tariff level is limited by the affordability of the consumers, the economic proclivity

of any business in the absence of regulatory interference, is to cut costs and possibly provide as low quality a service as possible.

- Ex ante regulations must be made without ambiguity and stable. Conditions for grid compatibility of micro-grids must be provided ahead in addition to specifications of economic terms such as allowable tariffs. It has been discussed earlier that uncertainty in regulation can deter necessary investment. A comprehensive analysis of sound practices and pitfalls when regulating the electric power sector can be found in [60].

- Transparency on grid extension plans: It is recommended that publications of grid extension plans are provided in advance so that micro-grid developers understand when grid arrives and the uncertainty associated with investment is reduced.

- Exploration of computation-aided models for least-cost grid extension planning and determination of micro-grid sites: Computer-aided models can help both the regulator and government determine high priority sites for off-grid or on-grid electrification as well as provide symmetric information that aligns private developers with the governments grid extension. The Reference Electrification Model can help address this need.

- Replicating Successful Best Practices: Despite beginning operations under challenging situations such as the high electricity theft losses scenarios described earlier, some utilities have been able to transition to commercially viable business entities through a combination of technological and social practices. For instance, the joint venture distribution utility in Delhi, Tata Power DDL, has been able to do just this [64]. In the off-grid sector, some off-grid developers have had relatively higher success dealing with issues such as payment collection than others. By replicating best practices from other utilities and developers and complementing such efforts with the use of a tool like REM for decision analysis, a utility may be able to provide both grid and off-grid electricity ser-

vices to all consumers in its region with high collection rates and ultimately remain viable.

- It might be a good idea to prioritize off-grid microgrid electrification projects that can drive development and demand growth when disbursing subsidies. It is important that business models must be sustainable enough for long term and this would be achieved if the socio-economic status of the consumers are raised over time such that they are eventually able to pay the true cost of electricity in future. The government or regulatory body would have to define what features in a microgrid project can drive growth and demand before promulgating any regulation on this.

Bibliography

[1] Electrical Planning Software SIMARIS - Power Distribution - Siemens, 2017. `http://w3.siemens.com/powerdistribution/global/en/consultant-support/electrical-planning-software/pages/default.aspx`, Last accessed 2017-02-27.

[2] Sustainable development goal 7. *United Nations*, 2017. `https://sustainabledevelopment.un.org/sdg7`, Last accessed on 2018-16-05.

[3] 44% rural households in UP still powerless, May 2018. `https://www.hindustantimes.com/lucknow/44-rural-households-in-up-still-powerless/story-tNX5EyIjEHH5ao55G3K07L.html`, Last accessed on 2018-16-05.

[4] Access to Electricity: % of Population. *World Bank*, 2018. `https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS`.

[5] International Energy Agency. WEO-2017 Special Report: Energy Access Outlook. Technical report, IEA, 2017.

[6] Majbaul Alam and Subhes Bhattacharyya. Are the off-grid customers ready to pay for electricity from the decentralized renewable hybrid mini-grids? A study of willingness to pay in rural Bangladesh. *Energy*, 139:433–446, November 2017.

[7] M. Alinci, E. Spaho, A. Lala, and V. Kolici. Clustering algorithms in MANETs: A review. In *2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems*, pages 330–335.

[8] E. M. de Almeida and E. N. Asada. NSGA-II applied to the multi-objective distribution system expansion planning problem. In *2015 18th International Conference on Intelligent System Application to Power Systems (ISAP)*, pages 1–6.

[9] J. Amador and J. Domnguez. Application of geographical information systems to rural electrification with renewable energy sources. *Renewable Energy*, 30(12):1897–1912.

[10] J. Amador and J. Domnguez. Spatial analysis methodology applied to rural electrification. *Renewable Energy*, 31(10):1505–1520.

[11] T. Asakura, T. Yura, N. Hayashi, and Y. Fukuyama. Long-term distribution network expansion planning considering multiple construction plans. In *PowerCon 2000. 2000 International Conference on Power System Technology. Proceedings (Cat. No.00EX409)*, volume 2, pages 1101–1106 vol.2.

[12] African Development Bank. Electricity Regulatory Index for Africa 2018. *AfDB*, 2018.

[13] C. Barbulescu, S. Kilyeni, A. Simo, and A. Vernica. Distribution system expansion planning with renewable sources. case study: IEEE 33 test system. In *2015 IEEE Eindhoven PowerTech*, pages 1–6.

[14] Paul Bertheau, Catherina Cader, and Philipp Blechinger. Electrification modelling for nigeria. *Energy Procedia*, 93:108–112.

[15] Joe Bible, Susmita Datta, and Somnath Datta. Chapter 4 - cluster analysis: Finding groups in data. In Krishna Rajan, editor, *Informatics for Materials Science and Engineering*, pages 53–70. Butterworth-Heinemann.

[16] Jeff Bilmes. EE595a Submodular functions, their optimization and applications Spring 2011, Lecture 1, University of Washington, Seattle. 2011.

[17] Yael Borofsky. Towards a transdisciplinary approach to rural electrification planning for universal access in India. Thesis, Massachusetts Institute of Technology, 2015.

[18] Catherina Cader, Philipp Blechinger, and Paul Bertheau. Electrification planning with focus on hybrid mini-grids a comprehensive modelling approach for the global south. *Energy Procedia*, 99:269–276, 2016.

[19] C. Chekuri and A. Ene. Approximation algorithms for submodular multiway partition. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 807–816.

[20] Chandra Chekuri and Alina Ene. Submodular Cost Allocation Problem and Applications. *arXiv:1105.2040 [cs]*.

[21] Sin-Shuen Cheung. The submodular facility location problem and the submodular joint replenishment problem. In Evripidis Bampis and Ola Svensson, editors, *Approximation and Online Algorithms*, pages 71–82, Cham, 2015. Springer International Publishing.

[22] Uttar Pradesh Electricity Regulatory Commission. Uttar Pradesh Electricity Regulatory Commission (Mini-Grid Renewable Energy Generation and Supply) Regulations, 2016, 2016. `http://upneda.org.in/sites/default/files/all/section/MiniGrid%20regulations%202016.pdf`.

[23] E. Cotilla-Sanchez, P. D. H. Hines, C. Barrows, S. Blumsack, and M. Patel. Multi-attribute partitioning of power networks based on electrical distance. *IEEE Transactions on Power Systems*, 28(4):4979–4987.

[24] Turner Cotterman. Enhanced techniques to plan rural electrical networks using the Reference Electrification Model. Master's thesis, Massachusetts Institute of Technology, 2017.

[25] Derya Dinler, Mustafa Kemal Tural, and Cem Iyigun. Heuristics for a continuous multi-facility location problem with demand regions. *Computers & Operations Research*, 62:237–256.

[26] C. Mateo Domingo, T. Gomez San Roman, Sanchez-Miralles, J. P. Peco Gonzalez, and A. Candela Martinez. A reference network model for large-scale distribution planning with automatic street map generation. *IEEE Transactions on Power Systems*, 26(1):190–197.

[27] J. Domnguez and I. Pinedo-Pascua. GIS tool for rural electrification with renewable energies in latin america. In *2009 International Conference on Advanced Geographic Information Systems Web Services*, pages 171–176.

[28] Cailinn Drouin. Geospatial cost drivers in computer-aided electrification planning : the case of Rwanda. Thesis, Massachusetts Institute of Technology, 2018.

[29] R. C. Dugan, J. A. Taylor, and D. Montenegro. Energy storage modeling for distribution planning. *IEEE Transactions on Industry Applications*, PP(99):1–1.

[30] Douglas Ellman. The Reference Electrification Model : a computer model for planning rural electricity access. Master's thesis, Massachusetts Institute of Technology, 2015.

[31] HOMER Energy. Break-even Grid Extension Distance, 2018. `https://www.homerenergy.com/products/pro/docs/3.11/breakeven_grid_extension_distance.html`, Last accessed on 2018-16-04.

[32] M. Ester, H. P. Kriegel, J. Sander, and Xu Xiaowei. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. December 1996.

[33] ETAP. Distribution Management System, 2017. `https://etap.com/packages/distribution/advanced-distribution-management-system/distribution-management-system`, Last accessed on 2017-02-28.

[34] Matteo Fischetti, Ivana Ljubi, and Markus Sinnl. Benders decomposition without separability: A computational study for capacitated facility location problems. *European Journal of Operational Research*, 253(3):557–569.

[35] R. Fronius and M. Gratton. Rural electrication planning software (LAPER). In *16th International Conference and Exhibition on Electricity Distribution, 2001. Part 1: Contributions. CIRED. (IEE Conf. Publ No. 482)*, volume 5, pages 8 pp. vol.5–.

[36] C. Fry and S. Manna. Can we group similar amazon reviews: A case study with different clustering algorithms. In *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*, pages 374–377.

[37] Pavlos S. Georgilakis and Nikos D. Hatziargyriou. A review of power distribution planning in the modern power systems era: Models, methods and future research. *Electric Power Systems Research*, 121:89–100.

[38] L. Gonzlez-Sotres, C. Mateo Domingo, Snchez-Miralles, and M. Alvar Mir. Large-scale MV/LV transformer substation planning considering network costs and flexible area decomposition. *IEEE Transactions on Power Delivery*, 28(4):2245–2253.

[39] M. Gulati and S Pahuja. Direct Delivery of Power Subsidy to Agriculture in India. *The World Bank: SEForAll*, 2015. `https://www.seforall.org/sites/default/files/l/2015/08/SE4All-Direct_Delivery_of_Power_Subsidy_to_Agriculture_in_India.pdf`, Last accessed on 2018-16-05.

[40] Innovation Energie Development (IED). GEOSIM - home, 2017. `http://www.geosim.fr/index.php?page=home`, Last accessed 2017-03-01.

[41] A. Jalali, S. K. Mohammadi, H. Sangrody, and A. R. Karlsruhe. DG-embedded radial distribution system planning using binary-selective PSO. In *2016 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*, pages 996–1001.

[42] N. C. Kalra and P. C. P. Bhatt. Parallel algorithms for tree traversals. *Parallel Computing*, 2(2):163–171.

[43] G. Karypis, Eui-Hong Han, and V. Kumar. Chameleon: hierarchical clustering using dynamic modeling. *Computer*, 32(8):68–75.

[44] Francis Kemausuor, Edwin Adkins, Isaac Adu-Poku, Abeeku Brew-Hammond, and Vijay Modi. Electrification planning using Network Planner tool: The case of Ghana. *Energy for Sustainable Development*, 19:92–101.

[45] Andreas Krause. SFO: A toolbox for submodular function optimization. *J. Mach. Learn. Res.*, 11:1141–1144.

[46] Muralee Krishnan C and Santanu Gupta. Political pricing of electricity: Can it go with universal service provision? *Energy Policy*, 116:373–381, May 2018.

[47] Vivian Li. The Local Reference Electrification Model : comprehensive decision-making tool for the design of rural microgrids. Thesis, Massachusetts Institute of Technology, 2016.

[48] Ke Liao and Diansheng Guo. A clustering-based approach to the capacitated facility location problem. *Transactions in GIS*, 12(3):323–339.

[49] F. Luo, C. Wang, J. Xiao, S. Ge, B. Yu, J. Wang, Y. Li, and S. Wang. A practical GIS-based decision-making support system for urban distribution network expansion planning. In *2009 International Conference on Sustainable Power Generation and Supply*, pages 1–6.

[50] N. N. Mansor. Distribution planning considering network contingencies and switchgear relocation. In *2016 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6.

[51] A. H. Mantway and M. M. Al-Muhaini. Multi-objective BPSO algorithm for distribution system expansion planning including distributed generation. In *2008 IEEE/PES Transmission and Distribution Conference and Exposition*, pages 1–8.

[52] S. M. Mazhari, H. Monsef, and R. Romero. A hybrid heuristic and evolutionary algorithm for distribution substation planning. *IEEE Systems Journal*, 9(4):1396–1408.

[53] Seyed Mahdi Mazhari, Hassan Monsef, and Hamid Falaghi. A hybrid heuristic and learning automata-based algorithm for distribution substations siting, sizing and defining the associated service areas. *International Transactions on Electrical Energy Systems*, 24(3):433–456.

[54] Luis A. A. Meira, Flvio K. Miyazawa, and Lehilton L. C. Pedrosa. Clustering through continuous facility location problems. *Theoretical Computer Science*, 657, Part B:137–145.

[55] Dimitrios Mentis, Magnus Andersson, Mark Howells, Holger Rogner, Shahid Siyal, Oliver Broad, Alexandros Korkovelos, and Morgan Bazilian. The benefits of geospatial planning in energy access A case study on Ethiopia. *Applied Geography*, 72:1–13.

[56] Shahram Mojtahedzadeh, Sajad Najafi-Ravadanegh, and Mahmoud-Reza Haghifam. A framework for optimal clustering of a greenfield distribution network area into multiple autonomous microgrids. *Journal of Power Technologies*, 96(4):219–228.

[57] J. C. Moreira, E. Miguez, C. Vilacha, and A. F. Otero. Large-scale network layout optimization for radial distribution networks by parallel computing: Implementation and numerical results. *IEEE Transactions on Power Delivery*, 27(3):1468–1476.

[58] G. Munoz-Delgado, J. Contreras, and J. M. Arroyo. Joint expansion planning of distributed generation and distribution networks. *IEEE Transactions on Power Systems*, 30(5):2579–2590.

[59] Debajit Palit and Kaushik Ranjan Bandyopadhyay. Rural electricity access in India in retrospect: A critical rumination. *Energy Policy*, 109:109–120, October 2017.

[60] Ignacio J. Prez-Arriaga, editor. *Regulation of the Power Sector*. Power Systems. Springer-Verlag, London, 2013.

[61] Denis RAMBAUD-MEASSON. Introduction on GEOSIM software for rural electrification planning, 2017. `http://www.ied-sa.com/projects/capredeo/upload/FI305.ppt`, Last accessed 2017-03-01.

[62] Lior Rokarch. Chapter 15: Clustering Methods. In *Data Mining and Knowledge Discovery Handbook*. Springer, 2009.

[63] X. Shen, M. Shahidehpour, Y. Han, S. Zhu, and J. Zheng. Expansion planning of active distribution networks with centralized and distributed energy storage systems. *IEEE Transactions on Sustainable Energy*, 8(1):126–134.

[64] M. Singhania and R Kinker. Tata Power Delhi Distribution: Automation vs Manpower. *Vikalpa: The Journal for Decision Makers*, 40(1), 2015.

[65] M-KOPA Solar. Pay-as-you-go solar lighting systems slash household budgets, September 2015. `http://www.m-kopa.com/pay-as-you-go-solar-lighting-systems-slash-household-budgets/`.

[66] Zoya Svitkina and Lisa Fleischer. Submodular approximation: sampling-based algorithms and lower bounds. *arXiv:0805.1071 [cs]*.

[67] Zoya Svitkina and ÉVA Tardos. Facility location with hierarchical facility costs. *ACM Trans. Algorithms*, 6(2):37:1–37:22, April 2010.

[68] Francis Taale and Christian Kyeremeh. Households willingness to pay for reliable electricity services in Ghana. *Renewable and Sustainable Energy Reviews*, 62:280–288, September 2016.

[69] M. Tabarzadi and V. Vahidinasab. A comprehensive expansion planning model for smart electric distribution networks. In *2015 3rd International Istanbul Smart Grid Congress and Fair (ICSG)*, pages 1–5.

[70] P. Tallapragada, M. Shkaratan, A. Izaguirre, J. Helleranta, S. Rahman, and S Bergman. Monitoring Performance of Electric Utilities: Indicators and Benchmarking in sub-saharan Africa. *World Bank, ESMAP*, 2009.

[71] H. H. Wang and L. B. Shi. Optimal distribution network expansion planning incorporating distributed generation. In *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pages 1714–1718.

[72] Kai Wei, Rishabh Iyer, Shengjie Wang, Wenruo Bai, and Jeff Bilmes. Mixed robust/average submodular partitioning: Fast algorithms, guarantees, and applications to parallel machine learning and multi-label image segmentation. *arXiv:1510.08865 [cs]*.

[73] R. Xu and D. C. Wunsch. Clustering algorithms in biomedical research: A review. *IEEE Reviews in Biomedical Engineering*, 3:120–154.

[74] Yong-Jun Zhang, Xi-Tian Zhang, Qin-Hao Li, Li Ran, and Ze-Xiang Cai. Gray theory based energy saving potential evaluation and planning for distribution networks. *International Journal of Electrical Power & Energy Systems*, 57:298–303.

[75] X. Zhu, W. Tang, Y. Liu, H. Yang, W. Cong, and W. Gai. Distribution substation planning method based on weighted voronoi diagram. In *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pages 1898–1902.