Energy for Sustainable Development 60 (2021) 67–81

Contents lists available at ScienceDirect

Energy for Sustainable Development

# A scalable framework to measure the impact of spatial heterogeneity on electrification

Simone Fobi [a,*], Ayse Selin Kocaman [b], Jay Taneja [c], Vijay Modi [a]

[a] Columbia University, QSEL, Mechanical Engineering, New York, NY 10027, USA
[b] Bilkent University, Industrial Engineering, Ankara, Turkey
[c] University of Massachusetts Amherst, STIMA Lab, Electrical & Computer Engineering, Amherst, MA 01003, USA

ARTICLE INFO

*Article history:*
Received 10 May 2020
Revised 10 December 2020
Accepted 10 December 2020
Available online 28 December 2020

*Keywords:*
Electrification
Large-scale network design
Connectivity metrics
Power distribution system

ABSTRACT

We propose a scalable computational framework to examine the effects of settlement patterns on the electrification of an entire country. We first propose a data processing strategy to convert structure locations, identified from satellite imagery, to estimated household locations using census data. Then, we present a computational framework that involves a two-level network design algorithm to find an abstract representation of the power distribution system at a national scale involving low voltage (LV) wires, medium voltage (MV) wires, and the transformers between the two levels of the system. Given the system components, we introduce three metrics for per-household connectivity requirements of LV and MV wires, and transformers to interpret our results at the administrative and the sub-administrative unit levels. With our administrative level analysis provided for 9.2 million structures in Kenya, we show that traditional rural/urban classification based on population density may not be enough and is often deceiving in estimating the cost of electrification and a new categorization based on our metrics provides more relevant estimates on the total cost. Moreover, our metrics can help determine the least-cost electrification option (e.g., grid, mini-grid, or stand-alone systems) for expanding access in the sub-administrative unit level and create a platform to perform sensitivity analysis based on different cost components. Our work demonstrates the potential for improvements in universal electrification combining new and more detailed data sources with a scalable planning framework and helps governments achieve Sustainable Development Goal 7 (SDG7) more quickly and at lower cost.


## Introduction

Sustainable Development Goal 7 (SDG7) was adopted in 2015 by the United Nations member states to provide access to “affordable, reliable, sustainable, and modern energy” to all by 2030 (UNDP, 2019). Although there has been significant progress towards reaching SDG7 in recent years, 840 million people still live without electricity as of 2019 (WorldBank, 2019). The lack of access to electricity in developing regions necessitates rapid and informed decision making on electrification options. Among the options available today, isolated or individual customer-scale solar-battery systems, frequently referred to as solar home systems (SHS), do not require any network at all. Networked options, such as a grid connection, rely on one or more large power plants located at multiple points on a network, where transmission lines carry the power over long distances (generally hundreds to thousands of kilometers) on a high-voltage backbone. This backbone in turn feeds a medium-voltage (MV) network, which distributes electricity directly to large consumers and transformers. The transformers drop down the voltage and allow a low-voltage (LV) wire to connect smaller customers in roughly a kilometer radius. The MV and LV network combined with transformers is called the distribution system. In the context of investments for access to grid electricity, this system generally represents the largest fraction of the total system cost and therefore, understanding the requirements of the distribution systems is quite important for proper rural electrification planning.

Determining the best electrification option for a region is particularly challenging especially when a mixture of solutions is possible. In fact, Carvallo et al. show that in places with low electrification rates, hybrid solutions that pair networked systems with standalone decentralized options typically offer an attractive approach to electrification (Carvallo, Taneja, Callaway, & Kammen, 2019). To aid utilities in identifying electrification options, a number of electrification planning tools that are capable of choosing between decentralized and networked options have been developed (Ciller & Lumbreras, 2020). These tools apply least-cost methods to determine the demand points, which may be better served by grid extensions and those whom would gain more benefits from off-grid systems. Depending on the techno-economic model used and the availability of the data, granularity level of these tools varies. Literature suggests that using all consumer locations for large-scale planning imposes strong computational constraints on many models. Thus, the studies aiming for large-scale electrification such as at the country level tend

* Corresponding author.
*E-mail address:* sf2786@columbia.edu (S. Fobi).

https://doi.org/10.1016/j.esd.2020.12.005
0973-0826/

to make simplifications by grouping individual structures into villages or large cells of 1 km (Ciller & Lumbreras, 2020).

When consumer points are aggregated over large areas for planning purposes, it is not possible to understand the impact of the settlement patterns on the components of the distribution systems and this may lead to misleading results when determining the electrification option at the local level. In order to address this problem, in this paper, we first propose a data processing strategy for Kenya to convert structure locations, identified from satellite imagery, to estimated household locations using census data. Then, we present a computational framework that involves a two-level network design algorithm to find an abstract representation of the power distribution system involving low-voltage wires, medium voltage wires, and the transformers between the two levels of the system. Given the system components, we introduce three simple metrics for per-household connectivity requirements of LV wire, MV wire, and transformers to interpret our results at the administrative unit level and the sub-administrative unit level. With our administrative level analysis provided for 9.2 million structures in Kenya, we show that traditional rural/urban classification based on population density is often deceiving in estimating the cost of electrification and a new categorization based on our metrics (combination of MV and LV wire requirements and the number of structures per transformer) provides more relevant estimates on the total cost. Moreover, in the sub-administrative analysis, our metrics can help determine the least-cost electrification option (e.g.,grid, mini-grid, or stand-alone systems) for expanding access and create a platform to perform sensitivity analysis based on different cost components. To the best of our knowledge, there is no focused study that evaluates the value of different connectivity metrics, highlighting their roles and strengths in facilitating the electrification planning process in a scalable manner. In addition, our work shows how these connectivity metrics complement and clarify the composite cost metric, which is usually the only metric reported in many planning studies.

This article adds to the existing knowledge in three ways. First, the paper demonstrates a data processing strategy to estimate the residential connection locations at the country level. Second, the paper proposes a framework for applying large-scale computationally-intensive network optimizations on millions of consumers. Third, the paper introduces three complementary connectivity metrics for evaluating electrification choices agnostic to the network planning approach. The methodology that we put forward can assist the decision-making process in electrification planning and serve as a decision support tool for identifying suitable electrification options. While we present results for Kenya, we believe that this tool can be applied to places with little to no access to electricity. Meeting the targets set in SDG7 requires consideration of multiple consumers across large landscapes with varying settlement patterns; our paper outlines a feasible approach to perform planning at scale to support electrification objectives.

The remainder of the paper is organized as follows: In the Related work section, we present relevant contributions from literature, in A data processing framework section, we discuss the data used for this work and present a method to estimate residential connection locations from building structures identified by satellite images. In A computational framework for distribution systems planning section, we describe the two-level network optimization algorithm used in our framework and our computational improvements. In An analysis on the administrative boundary level and An analysis on the sub-administrative boundary level sections, we show the metrics computed using the two-level network algorithm and their applications at varying resolutions. In the Sensitivity analysis section, we also show the sensitivity of our metrics to cost. Finally in the Conclusion section, we propose feasible extensions to our work and conclude.

## Related work

In a comprehensive review paper by Ciller and Lumbreras (2020), planning tools used for rural electrification are classified into three groups: pre-feasibility studies, intermediate analysis tools and detailed generation and network design tools. Although not all efforts towards rural electrification are presented or used as a software tool in the literature, we review the studies related to our work using the same classification.

Pre-feasibility studies as in Debnath and Mourshed (2018), Mahapatra and Dasappa (2012), Moner-Girona, Bódis, Huld, Kougias, and Szabó (2016), and Zeyringer et al. (2015) estimate the least cost approach for different technology choices using simplifying assumptions, allowing for a first pass at the planning problem. These studies do not typically include network design and are likely to group consumers into villages or cells (e.g., 1 km $\times$ 1 km). Grouping of consumers reduces the computational granularity, and therefore, pre-feasibility studies have lower model complexity, high computational speed, and are valuable for quickly evaluating technology choices over large-scale areas at low resolution given varying generation options. Cost remains the key reported metric of evaluation used with pre-feasibility methodologies.

The studies that are used for intermediate analysis have various complexity levels depending on the network design and the technical details considered. Similar to the pre-feasibility studies, the resolution of the data used in the intermediate analysis studies is low. An intermediate planning approach presented in Parshall et al. (2009) proposes a spatial cost minimization electricity planning model for Kenya to decide between grid-based electrification and off-grid solutions. The model provides the basis for Network Planner (NP), an online decision-support tool that has been developed to explore grid, mini-grid, and off-grid options for rural communities (ModiLabs, 2019) and has been used in national electrification studies of countries such as Senegal (Sanoh & Parshall, 2012), Ghana (Kemausuor, Adkins, Adu-Poku, Brew-Hammond, & Modi, 2014) and Nigeria (Akpan, 2015). Abdul-Salam and Phimister (2016) propose an approach based on hierarchical lexicographic programming that considers both cost efficiency and political economy to give large populations a priority for grid connectivity. Bolukbasi and Kocaman propose a prize collecting Steiner tree approach to choose between grid and off-grid options and to determine the network design for the grid-compatible nodes in a least cost manner (Bolukbasi & Kocaman, 2018). Although these studies offer great value by folding in more modeling complexity, they reduce the computational difficulties by aggregating individual consumers and therefore neglect the effect of settlement distribution. Similar to many electrification planning models, intermediate studies report cost as the key metric of evaluation.

In Ciller and Lumbreras (2020), Reference Electrification Model (REM) (MIT, 2020) is described as the only planning tool that falls under the detailed generation and network design class. REM aims to design a power system configuration evaluating the demand profiles for the individual customers. To overcome the computational burden of a detailed plan using local level data, REM uses a sequential approach to plan the sub-systems in a hierarchical manner. Although it provides a very detailed network configuration, it is acknowledged in Ciller and Lumbreras (2020) that, the network design approach used in REM is not designed for rural electrification planning and may perform poorly when designing small networks.

There are also some studies in the rural electrification literature that use customer or household level data as in REM (MIT, 2020), however, aim for obtaining quick estimates for the network structure and associated costs, rather than being used for detailed implementation. The main objective of these studies is to show that rural settlement patterns – especially in Sub-Saharan Africa – can be diverse and the effect of settlement patterns on the electrification options might be overlooked in the pre-feasibility and intermediate analysis studies due to the aggregated data considered. Using several datasets of structure locations developed from satellite imagery, Zvoleff et al. propose a metric, called the homogeneity index, that serves as a proxy for the degree of dispersion of the structures. They provide solid evidence about the impact of geographic patterns on the cost of energy infrastructure. However, they assume that all identified structures within the images are households and these households can be connected via single level LV network (Zvoleff, Kocaman, Huh, & Modi, 2009). Kocaman, Huh, and Modi (2012) use the same structure locations as Zvoleff et al. (2009) to propose a computationally-intensive two-level (MV and LV) network optimization approach and evaluate the cost of grid extension for the distribution systems in limited-size rural regions. Adkins et al. (2017) use inter-community and inter-household distances as proxies to estimate MV and LV wire lengths.

In this paper, we build upon the approach presented by Kocaman et al. (2012) and present a computational framework to incorporate a large number of connection points into electrification planning, thereby improving modeling capacity at reasonable computational speed. In this direction, our study is the first to propose a detailed data processing strategy to estimate the residential connection locations from hand-labelled structure points. Moreover, we propose a set of per-household connectivity metrics - low-voltage wire, medium-voltage wire and transformers - that can be used to rapidly evaluate electrification choices agnostic to the network planning approach. We show how network outputs from detailed models such as REM (MIT, 2020) can be used to compute our metrics and how these metrics facilitate rapid analyses of the electrification landscape within a country. We discuss all our results for Kenya, for which, to the best of our knowledge, no similar findings are available in the literature.

## A data processing framework

In this section, we first discuss the source of our structure locations data and propose a data processing framework to estimate the household locations.

### *Structure locations*

Our study is principally built upon 11.9 million human-labelled building structures in Kenya from satellite imagery data obtained in 2017. This data was obtained from the Kenya National Electrification Plan - Structures Survey and includes latitude and longitude pairs for each identified structure within the images. No additional information is provided on the structure type or its pertaining attributes such as rooftop type and area.

### *Estimating household locations*

It is quite common for rural households to own multiple structures (shed or outhouse in addition to living quarters), while in more urban locations, multiple households may dwell within the same structure (KNBS, 2014). We propose a method to obtain an estimation of households from human-labelled satellite imagery data. Census data provides the number of households at varying administrative levels. For the case of Kenya, the census provides household counts of each ward. Wards in Kenya (about 1400 in number) represent the smallest administrative unit in Kenya. The household counts from census data, provide only aggregates with no information on household locations. Because the Kenyan census is decennial and there is readily available 2009 Kenya census data, we apply a correction strategy to estimate household counts in 2016. Facebook's 2016 High Resolution Settlement Layer (HRSL), provides population data at a 30 m resolution (Facebook and CIESIN-ColumbiaUniversity, 2016). Given the 2009 population data at the ward level, and using HRSL population data, we estimate a population growth factor $k$ for each ward, which represents the growth a ward has experienced between 2009 and 2016. We assume household counts scale linearly with population, thus we use a 1:1 relation between population growth and the growth in the number of households.[1] Applying this growth factor $k$ to the 2009 ward-level household data, we can estimate the expected number of households in 2016 for each ward. Upon obtaining the 2016 household estimates, a direct comparison can be applied to the 11.9 million structures obtained from satellite images.

Next, we compute a per-ward Structure To Household ratio (STH) that is the ratio of 2017 identified structures to estimated households (obtained from the census data adjusted to 2016). This ratio is frequently greater than 1, as observed by Kenya 2014 DHS results (KNBS, 2014). In this paper, we assume every household in a ward to have the same number of structures; we allow this ratio to vary from ward to ward. Where the STH ratios are higher than 2, we apply a merging algorithm described below. We present our full data processing framework, including estimating household locations and our merging algorithm in Fig. 1.

### *A merging algorithm*

A set-covering algorithm was applied at different radii and the resulting structure counts were compared to each ward's household count. The set-covering problem is an NP-complete problem and aims to find the minimum number of sites and their corresponding location to cover all demand nodes (Garey & Johnson, 1979). Here, we adopt a well-known heuristic approach proposed by Chvatal (1979) to find the reduced set of structures that cover all building structure locations within a radius $r$ of interest. Fig. 1 highlights the merging process when STH are greater than 2. The steps of this approach are as follows:

1) Draw a circle around each building structure location with a specific radius $r$.
2) Count the number of points in each circle.
3) Take the circle with the maximum amount of points (Ties are broken arbitrarily).
4) Eliminate the building structure points 'covered' with the circle in Step 3.
5) Repeat 1–4 with the remaining points until each building structure point is 'covered'.

A merging radius of 20 m was found to be most suitable to match household counts with the adjusted census data, a distance which reduces the 11.9 million human-labelled structures to a merged structure count of approximately 9.2 million. The average STH ratio for all wards is 1.3 with a maximum of 2.6. See the Merging approach section for a more detailed discussion on merging radius. The merged structures and their corresponding locations are subsequently used in the rest of the paper. The paper treats each merged structure as requiring a separate electric connection.

## A computational framework for distribution systems planning

We propose a computational framework to estimate the i) per-structure LV wire requirement; ii) per-structure MV wire requirement needed for each structure to be connected to the network; iii) the number of structures per transformer, and; iv) a per-structure connection cost. In this section, we detail how we compute these four metrics. Motivated by the need to evaluate cost estimates and additional metrics which highlight spatial diversity, this paper adopts a two-level network design (TLND) approach proposed by Kocaman et al. (2012) and proposes a decomposition approach to obtain results over a large spatial extent.

### *A two-level network design approach*

The TLND combines the transformer location problem and the LV and MV network design problems into a single optimization framework by modeling a two-level radial power distribution system. The two-level network connects demand points (in this case post-merged structure locations) via intermediate transformers, which reside on a

[1] From recently released 2019 Kenya census data, we observe roughly 10% difference between population growth and the household count growth from 2009 to 2019.

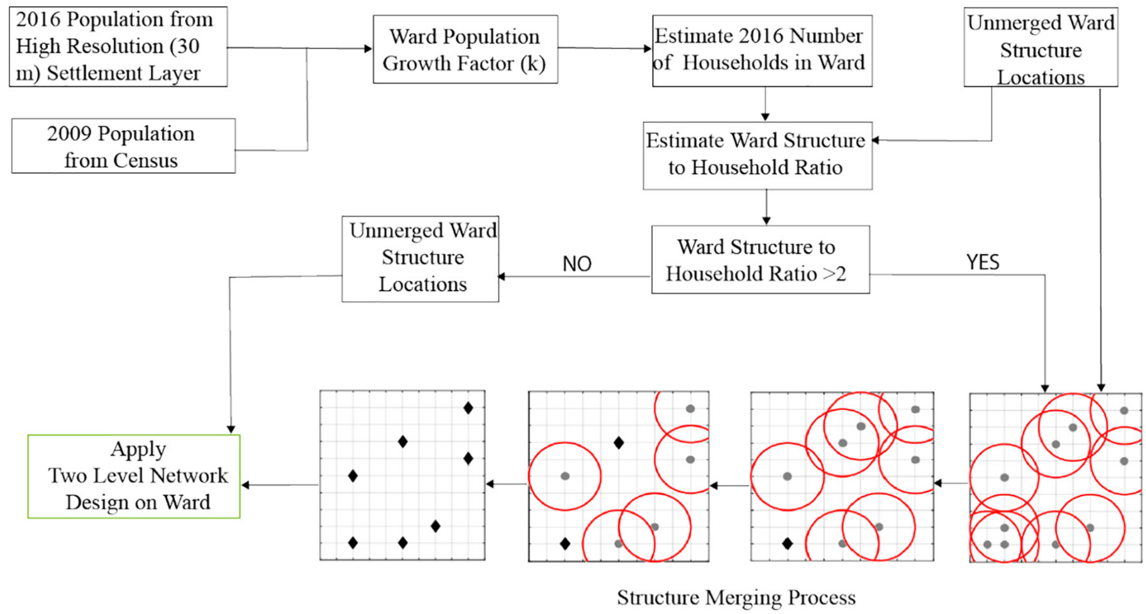



**Fig. 1.** Data processing framework: 2016 population from the High Resolution Settlement Layer and 2009 population census are used to estimate a population growth factor (k), which is used to estimate 2016 household counts. Wards with structure to household ratios > 2 are further processed, where structures are merged using a set-covering merging algorithm. The two level network design is ran on resultant structures.

primary MV network. The merged structure points are connected to the transformers with a secondary multi-point LV network. As in Kocaman et al. (2012), transformers are assumed to be uncapacitated, i.e. they can handle unlimited demand. However, there is a limitation on the distance between a merged structure point and its serving transformer. The TLND does not consider the presence of the legacy grid, high voltage (HV) network,[2] load balancing requirements, or power flow.

Determining the layouts of both LV and MV networks while locating distance-limited transformers that connect them in a continuous space is an NP-hard problem, since the continuous space location-allocation problem is NP-hard (Megiddo & Supowit, 1984). The algorithm proposed by Kocaman et al. (2012) to solve this NP-hard problem leverages an agglomerative hierarchical clustering approach. This bottom-up approach starts with locating a transformer on each demand point (where each demand point represents a singleton cluster) and iteratively decreases the number of transformers as a pair of clusters is agglomerated (merged) in a greedy manner based on a dissimilarity measure. In this paper, the centroid method is used as the dissimilarity measure: the closest pair of transformers which can be replaced by a single transformer located at the centroid of the demand points without violating the distance constraint is merged at each step. The minimum spanning tree problem aims to find a tree (a network containing no cycles) that spans all the points minimizing the total cost of the connection. At any iteration of the clustering algorithm, once the transformer locations are updated, the MV network between them and the source point is found using a minimum spanning tree algorithm with the guarantee of an optimal solution (Prim, 1957). Once the clusters are formed at each iteration of the agglomerative hierarchical clustering approach, the multi-point LV network between the transformers and the demand points is obtained by solving the capacitated minimum spanning tree problem. This problem aims to find a spanning tree rooted at the transformer considering a distance or a number of nodes on each sub-tree emanating from the root point. In the TLND, a distance limit is used on the length of a sub-tree and the problem is solved using Essau and Williams's heuristic approach (Esau & Williams, 1966). The maximum distance between demand points and the transformer is assumed to be 500 m, which is a widely accepted limit for open-wire LV lines. For each step of the agglomerative clustering, the algorithm calculates the minimum spanning tree as the MV network and the capacitated minimum spanning trees within each cluster as the multi-point LV network. The overall cost is computed at each step and the least cost design is outputted.

In order to run the TLND, we also assume that a transformer cost USD 2000, a meter of MV wire cost USD 25, while a meter of LV wire cost USD 10. While we use costs obtained from Kocaman et al. (2012), our TLND can be run with costs that are reflective of any region of interest. Given the cost parameters and the constraints, the objective of the algorithm is to find the number and locations of the transformers and the least-cost layouts of MV and LV networks. In A decomposition approach for large-scale planning section, we demonstrate how we integrate the TLND into the computational framework for estimating the metrics at the country level.

### A decomposition approach for large-scale planning

Planning at a national scale with individual structures result in millions of demand points: in the case of Kenya, 9.2 million merged structure locations need to be considered for planning. Even at the resolution of the smallest Kenyan administrative unit, the median and maximum per-ward merged structure count is 6872 and 32,321, respectively. In response to the significant computational requirements of large-scale optimizations, Navarro and Rudnick (2009) propose micro-optimizations for small zones as an approach to applying network algorithms for large-scale distribution planning. Inspired by this micro-optimization strategy, we devise a framework to run the two level network design algorithm on millions of demand points, without sacrificing spatial heterogeneity.

We develop our computational framework to minimize run-time without sacrificing performance. Our approach considers the smallest administrative unit as the entry point to apply the framework. For Kenya we apply the framework in parallel on each ward. Given a ward, the framework consists of three steps: 1) recursively decompose the ward into cells, 2) parallelize the TLND for all cells, and 3) reconstruct the ward. Fig. 2 shows our computational framework for a

[2] High voltage transmission networks are strongly dependent upon the specific location of central power generation systems.

synthetic ward and its corresponding structures. In Fig. 2(a) we take a ward as shown in i) and check the ward against three predefined parameters $M$, $N$ and $R$. We compute the number of structures in a ward ($m$) and compare it to a predefined threshold ($M$) which represents the maximum number of structures that can be present. Next our approach computes the number of structures for the largest cluster in that ward ($n$). Clustering is performed by the two level network design algorithm to assign structures to a given transformer: by limiting the maximum number of structures in a cluster to a predefined threshold $N$ we are able to reduce the time it takes to design a low voltage network for the structures in the cluster. Similarly, the ward radius ($r$) is computed and compared to a predefined minimum radius $R$, which ensures that the connecting radius of a utility is preserved and the number of structures connected to a transformer is maximized. The radius parameter counterbalances the splitting and prevents the wards from being excessively split. If $r$ is less than $R$, the ward is accepted as a valid cell for the network planning algorithm; if $r$ is greater than $R$, then $m$ and $n$ are compared to $M$ and $N$, respectively.

Taking the example presented in Fig. 2(a)(i), in which the per-cell maximum number of structures $M$ is assumed to be 3, Fig. 2(a)(ii) shows the results of the initial splitting. The split cell that does not meet the constraints is further split until the constraints are met, as shown in Fig. 2(a)(iii). Formally, our recursive split algorithm splits the ward into cells $C_i$ such that they obey the following constraints: 1) the number of merged structures in $C_i$ must be less than a predefined threshold $M$; 2) the number of structures for the largest cluster in $C_i$ must be less than a predefined threshold $N$; and 3) the radius of $C_i$ must be greater than a predefined radius $R$ in meters to allow any further splitting. The predefined parameters of $M$, $N$, $R$, are all user-defined parameters which can be determined a priori by running tests on a small number of wards in order to understand the effect of number of structures, settlement patterns and the search radius on the runtime of the network planning algorithm. We discuss the effect of runtime and our choice of parameters in the Sensitivity to scaling strategy section.

Fig. 3 presents pseudo-code for our splitting algorithm. Once valid cells are obtained, the TLND is applied in parallel. Transformer locations, the low voltage network and a localized medium voltage network are obtained for each $C_i$ cell as shown in Fig. 2(b). The localized medium voltage network does not consider transformers in other cells belonging to the same ward; we address this in a final step by putting cells back together and rerunning the medium voltage computation (minimum spanning tree algorithm) with transformer locations across all cells in the ward. We show in the Sensitivity to scaling strategy section that splitting the ward does not have adverse effects on the obtained results.

**Detailed computing specifications are as follows**: Running 9.2 million structure locations was done on a computer cluster with two Intel Xeon E5-2680 v4 processors with 14 cores each, 128GB RAM and 200 GB local SSD. 17,330 cells were generated for Kenya and the TLND was ran on each cell. With the longest allowable runtime being 21 days, this resulted in 98.8% of cells completing the TLND. Given our framework, 90% of the cells ran in under 12 h with more than 50% of the cells taking less than 1 h to run the TLND. 98% of the cells ran the TLND in under 4 days.

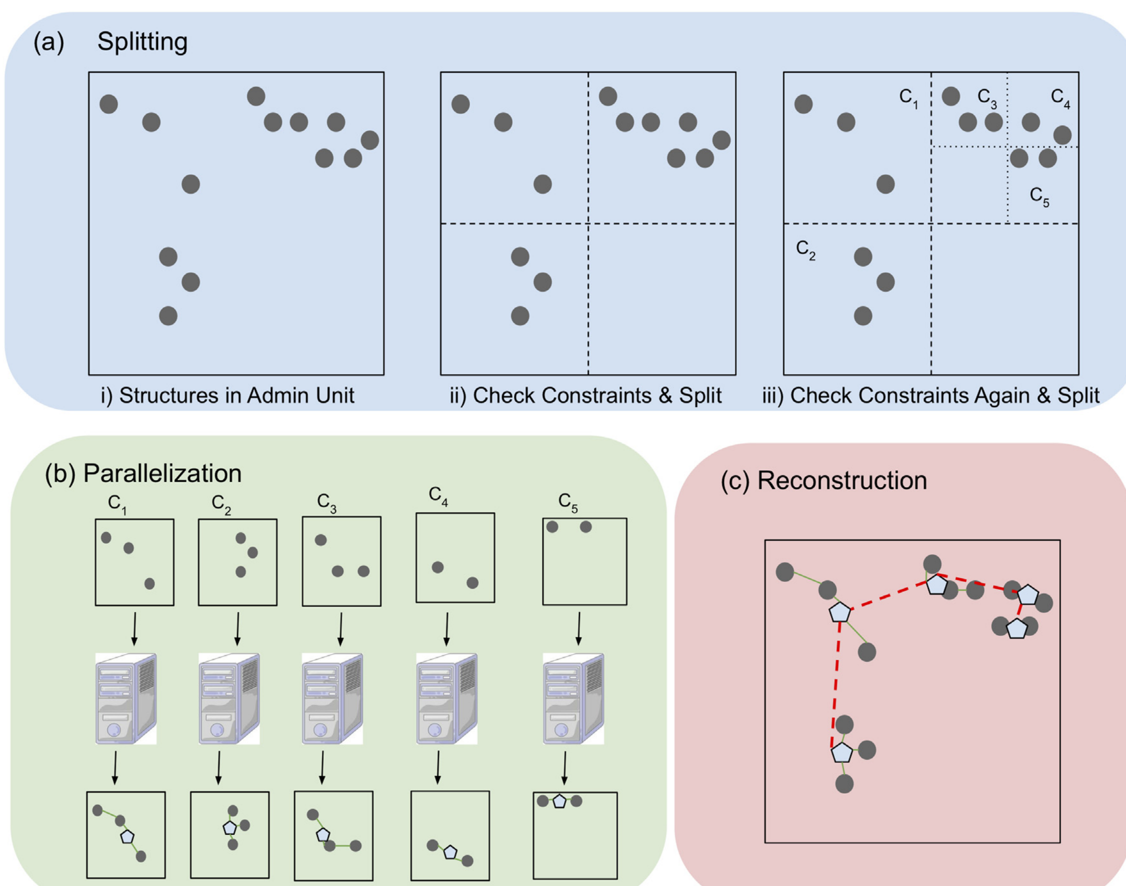



**Fig. 2.** Computational framework for planning using multiple demand points. (a) *Splitting*: A recursive split is used to obtain valid cells for the network planning algorithm. Splitting continues until all three constraints are met (Number of structures in cell < M; Number of structures in largest cell cluster < N; cell radius > R) (b) *Parallelization*: The network planning algorithm is run in parallel on all valid cells to obtain transformer locations, the low voltage network and a local medium voltage network (c)*Reconstruction*: Transformer locations from all cells in a ward are used to compute the medium voltage network for the ward.

**Algorithm** Recursive Split of Structures in Administrative Unit

```
Inputs:
M, maximum number of merged structures in a cell,
N, maximum number of structures in the largest cluster,
R, minimum cell radius.
validcells = empty list
for each adminUnit in allAdminUnits do
    allcells = [adminUnit]
    while the number of cells in allcells is greater than zero do
        for cell in allcells do
            Remove cell from allcells
            m = compute number of merged structures in cell
            n = compute number of structures in largest cluster in cell
            r = compute cell radius
            if r <= R do
                Append cell to validcells
            else do
                if (m < M) and (n <N) do
                    Append cell to validcells
                else do
                    newcells = split cell
                    Append newcells to allcells
```

**Fig. 3.** Recursive split algorithm.

## An analysis on the administrative boundary level

In this section, we first discuss the value of our proposed metrics to measure the impact of spatial heterogeneity on the electrification cost using the smallest administrative unit resolution (i.e. ward). Next, we show the performance of our metrics compared to population density at this resolution. Finally, we discuss the effect of real settlement patterns on our computed metrics.

### *Proposed metrics calculated for Kenya*

Results for each ward are averages across all merged structures within the ward. Here, we do not include the existing grid in Kenya but rather focus on evaluating the impacts of networking given the structures internal to the ward. Fig. 4 shows the average ward level metrics by decile: per-structure low-voltage wire (meters), per-structure medium voltage wire (meters), per-transformer number of structures,

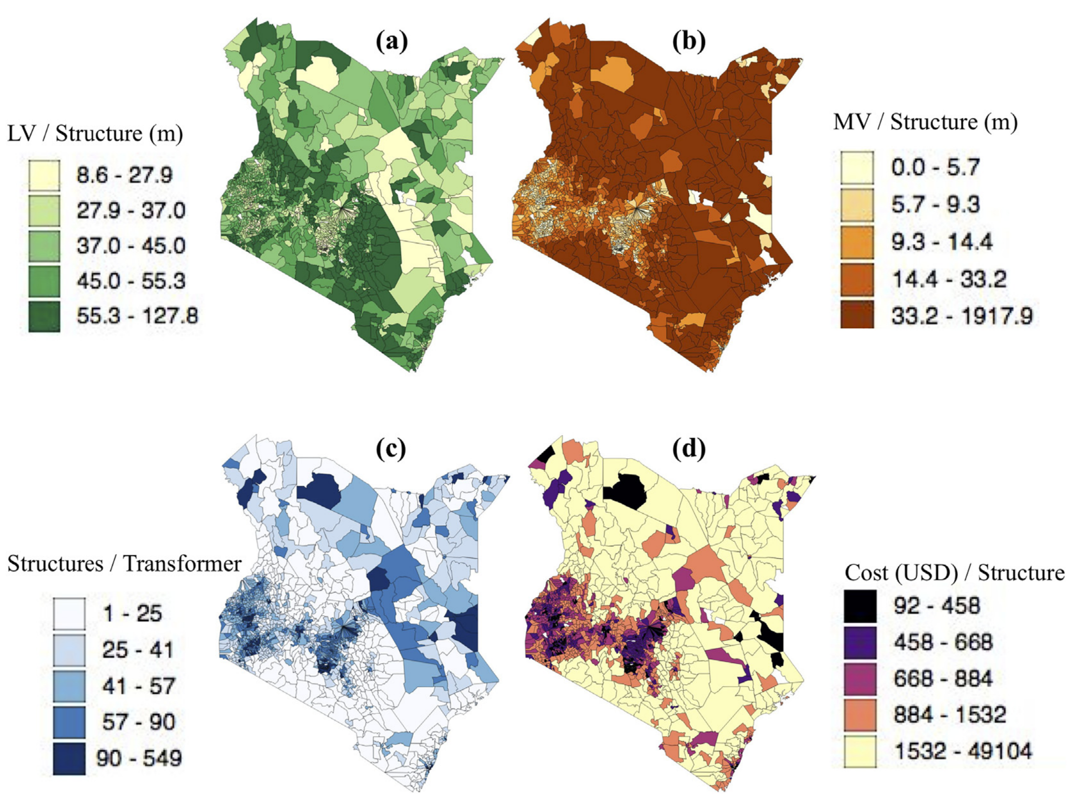



**Fig. 4.** Average ward connectivity metrics for Kenya by decile.

and per-structure cost (USD).[3] Given a desired proximity of structures to each other and to the transformer, our method allows for the quick and easy identification of suitable wards for different types of electrification. For example, an energy provider may be interested in determining which wards have an average distance between merged structures of less than 30 m and correspondingly can be networked through LV connections. As shown in Fig. 4(a), the 30 m threshold corresponds to approximately 25% of the wards – primarily those in Eastern Kenya. Similarly, an energy provider might be interested in wards where transformers are in close proximity to each other and consequently are suited for MV networks. In Fig. 4(b) we show that almost 50% of wards require less than 10 m of MV wire per structure. The ability to specify both LV and MV requirements outside of costs allows planners to quantify the effects of regional geography on network design.[4] Fig. 4(c) shows the average number of structures per transformer. Wards with the highest number of structures per transformer are found in more urban regions in Central Kenya. Generally, number of structures per transformer decreases in more rural regions even though there are a few otherwise rural wards in Eastern Kenya with higher transformer capacity.

Fig. 4(d) shows the average ward per structure connection cost of electricity access: this cost reflects the average combined wire and transformer costs needed to connect a structure in the ward. The connection cost metric shows which wards are suitable candidates for networked grids and which wards are more suited for alternative electrification modes like mini-grids or solar home systems (SHS). Differentiating between wards suited for mini-grids versus those for SHS requires leveraging the 3 other metrics in Fig. 4; the exact cost cutoffs for each technology choice would depend on the price of these alternatives and the utility's cost-sensitivity. The four metrics presented in Fig. 4 capture the complexities of geography-dependent network design, the benefits of which are explored in the next section.

### *Why do we need new metrics?: a comparison with population density*

Population density is a metric that is often used for estimating the location and type (rural or urban) of demand centers. For energy access problems, we observe that rural/urban classification based on population density may not be enough and is often deceiving in estimating the cost of electrification. A new categorization based on a combination of MV and LV wire requirements and the number of structures per transformer provides more relevant metrics to anticipate the total cost and create a platform to perform sensitivity analysis based on different cost components. For this purpose, we compare our metrics against population density to quantify the additional gains which our metrics may offer.

Fig. 5 shows a scatter plot of the per-structure MV requirement as a function of LV requirement. In this figure, each bubble represents a ward, and the bubble sizes show average number of structures per transformer of the ward. The average number of structures per transformer are grouped by quartiles and the quartile ranges are shown in the figure. The coloring in Fig. 5 shows the people per square kilometer (sqkm). As expected, wards with higher population density (i.e. those in blue), tend to be grouped at the lower left hand corner of the figure, with low MV and low LV wire requirements and with higher number of structures per transformer. These wards tend to be more urban, likely with established grids. The upper right hand corner of Fig. 5 contains sparse rural wards with high LV and high MV requirements and low number of structures per transformer. However, it is important to note that not all wards that can be considered rural (based on population density) reside in this quadrant. Given our proposed metrics, these rural wards should be further categorized as nucleated and non-nucleated (or dispersed) rural settlements, given their LV and MV combination. The details of this classification are summarized in Table 1.

A strong observation from Fig. 5 is that there are a number of wards with varying connectivity metrics at similar population densities. To explore this observation, we analyzed two such wards with similar population densities of 120 people per sqkm. Fig. 6 shows both wards in a 30 km$^2$ box for scale but does not show the administrative boundary of the ward. The figure also shows the ward labels and their county name. The figure shows the per-structure LV length, per-structure MV length and the number of structures in brackets, respectively. Upon comparing both wards, we see that ward E in Siaya has very different LV and MV requirements to ward F in Makueni, although they have similar population densities and a similar number of merged structures. LV and MV requirements in ward E are significantly lower because of high structure nucleation, while the LV and MV requirements in ward F are much higher because structures are further away from each other on average. The varied infrastructure requirements of both wards results in an average difference in connection cost of $1341. By using our proposed metrics, we capture more insights on the diversity of wire requirements and by consequence connection costs needed to provide electricity access. We further quantify the dissimilarity in wire requirements for wards with similar population densities in Kenya. For every ward, we identify wards of similar population density (within 10%). We compute the average LV and MV difference between wards with similar population density and the ward of interest. On average, 47% of the wards with similar population density have LV or MV differences greater than 20%. This indicates that using population density as a metric for connectivity would be misleading approximately half of the time. This distribution of system requirements is lost when population or structure density alone is used as the metric of evaluation, or when residential consumption nodes are aggregated to form population centers.

### *Effect of settlement patterns*

Zvoleff et al. (2009) show that geography and by consequence settlement behavior affect network lengths. Similarly, Kocaman et al. (2012) discuss that settlement patterns play a role in the results obtained from the two-level network design. In this section, we aim to understand the effect of real settlement patterns on our computed metrics.

Fig. 7 shows four wards with varying settlement patterns, where each point represents a merged structure (points in close proximity might appear as a single point in the figure). The grey dashed boxes surrounding the structures represent a 25 km$^2$ box. In brackets we report the per-structure LV requirement (m), the per-structure MV requirement (m), and the structure count, respectively for the ward. Fig. 7 (a) and (b) show wards with similar per-structure LV requirements and varying per-structure MV requirements, while Fig. 7(c) and (d) show wards with similar per-structure MV requirements and varying per-structure LV requirements. At similar per-structure LV requirements as seen in Fig. 7(a) and (b), the per-structure MV needed in ward A is 70 times lower than that needed in ward B due to the proximity of clusters. In Fig. 7(b), significant MV is required to connect clusters of structures. These clusters may be villages or communities. However in Fig. 7(a), all structures and their clusters are in tight proximity. The per-structure MV requirement in Fig. 7(b) is even higher due to the smaller number of structures present in ward B when compared to ward A. At similar MV, Fig. 7(c) has one-third the LV requirement of Fig. 7(d). There is an even spread of structures throughout the 25 km$^2$ grid in Fig. 7(d), which influences the per-structure LV requirement. With a higher structure count in Fig. 7(d), it is expected that the per-structure LV requirement would be low as the total LV wire length and cost is spread out among a higher number of structures, however this is not the case. Because structures are more evenly spread out in ward D, the LV wire requirement is high. We observe that nucleation

[3] It is important to note that the two-level network design enforces a limitation of 500 m for connecting structures on the same LV wire (due to voltage drop considerations).

[4] It is important to note that computed wiring requirements are distances as a crow flies, and practical routing considerations might lead to distances which are larger than those presented here. This concern could be addressed by incorporating topology into the methodology.

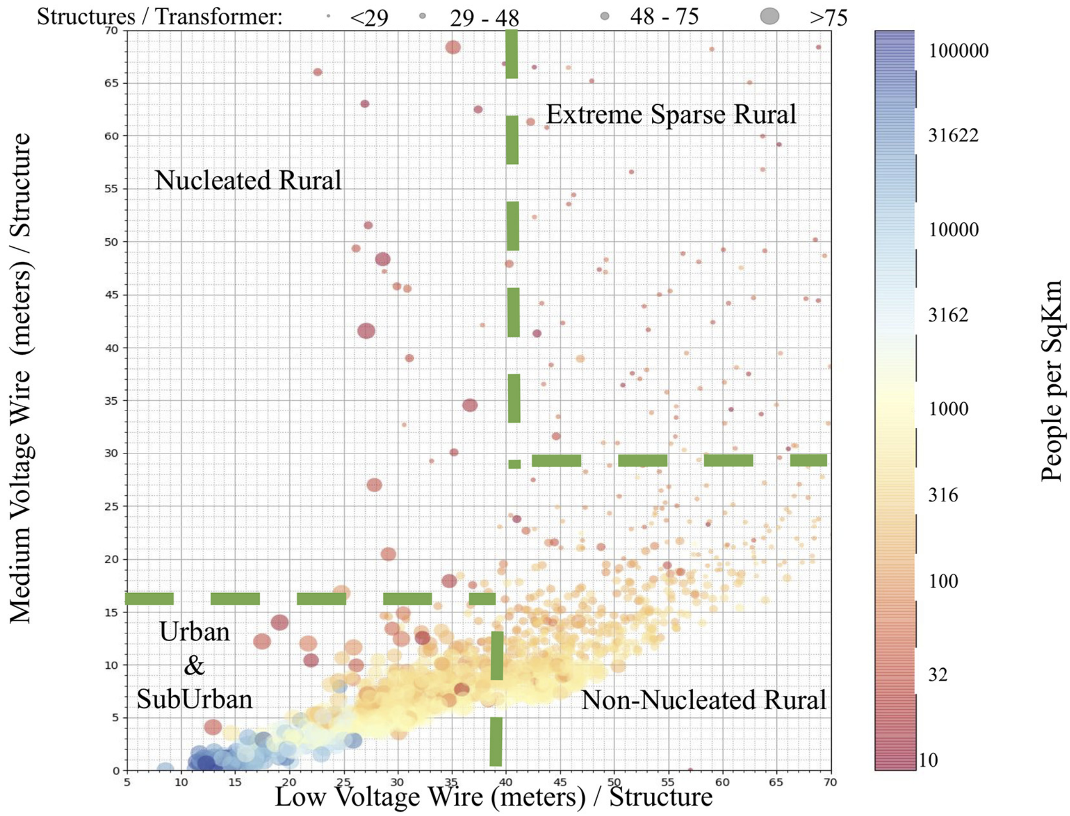



**Fig. 5.** A scatter-plot showing per structure LV wire requirement against per structure MV wire requirements. Each bubble in the figure represents a ward in Kenya and the bubble size indicates the average number of structures per transformer by quartiles. People per sqkm are captured by the coloring of the bubbles. There are multiple wards with similar population densities that have varying MV and LV requirements. Thus our connectivity metrics capture more spatial diversity than population density alone.

of structures drops the per-structure LV requirement while nucleation of clusters (villages, communities) reduces the per-structure MV requirement. We are able to show that our proposed connectivity metrics capture the effects of settlement patterns.

## An analysis on the sub-administrative boundary level

We recognize that decision making about electrification technologies occurs at a granular level and that a single technology choice cannot be assigned to an administrative unit. As a result, we leverage the data and methodology for analysis at sub-administrative boundaries. To explore this in depth, we present the complete network for a sample ward of 7047 structures. Fig. 8(a) shows transformer locations and the MV network for all the structures within the ward. The blue pentagons represent transformer locations, red solid line shows the MV network, and the grey points represent the structures. In Fig. 8(b), we include the LV network (as green dashed lines) for a subset of the ward, showing connections between individual structures and transformers. Given our proposed methodology, the MV and LV network with individual connections can be visualized as demonstrated by the figure. Energy planners can inspect connections across transformers and structures and subsequently aggregate the metrics to a level that is most useful to support their decision making.

With our methodology we can identify which transformer locations and connecting structures can be networked with minimal LV wire. For the same ward, Fig. 10(a) shows the number of structures per transformer at each transformer by quintile. Blue transformers are connected to many structures while red transformers are connected to few structures. In the figure, we observe that transformers with few surrounding grey dots have a lower number of connecting structures, while transformers with many surrounding grey dots have a higher number of connected structures. Fig. 10(b) shows the distribution of structures per transformer for all transformers in the ward. With a ward average of 77.5 structures per transformer, 10% of wards have more than 160 structures per transformer (twice the ward average). The distribution within the ward can be missed when only considering averages of our metrics along administrative boundaries or at lower resolutions. The flexibility to evaluate the proposed metrics at multiple scales allows for deeper evaluation of varying electricity technologies.

Using the same ward, we show that our methodology and metrics can be used to identify opportunities for varying electrification

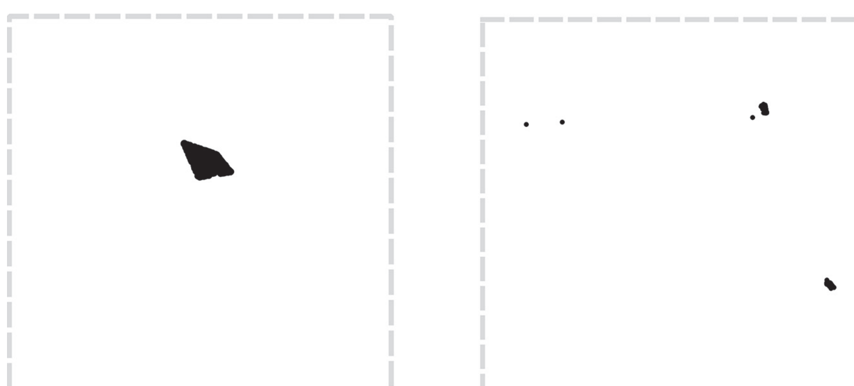
**Table 1**
A new categorization based on a combination of our metrics to anticipate the cost of electrification.

| Category | Proposed metrics | | | Population density |
|---|---|---|---|---|
| | MV/structure | LV/structure | Structures/transformer | |
| Urban & suburban | Low | Low | High | High |
| Nucleated rural | High | Low | High | Low |
| Non-nucleated rural | Low | High | Low | Low |
| Extreme sparse rural | High | High | Low | Low |

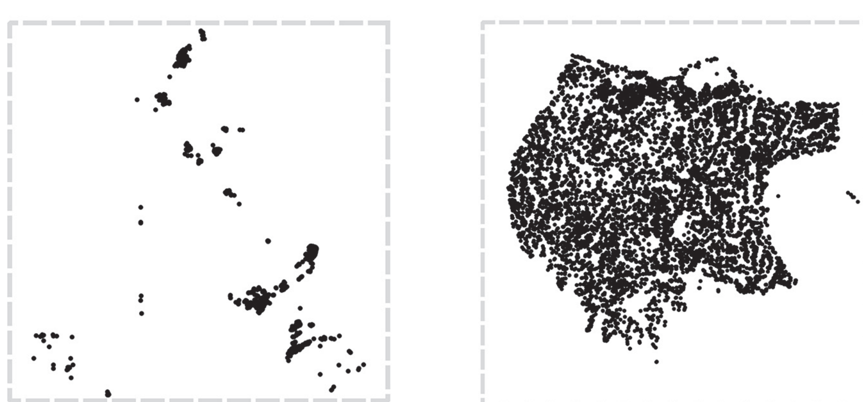
(a) E, Siaya (32.5, 6.9, 7047) (b) F, Makueni (88.4, 34.5, 8610)

**Fig. 6.** Two wards with around 120 people per sqkm are shown. The per-structure LV requirement, per-structure MV requirement, and the structure count of the ward are shown respectively in brackets. The grey boxes surrounding each ward represent 30 km$^2$ area for scale and do no show the administrative boundaries. Figure (a) and (b) show that wards can have similar population densities but varying settlement patterns which can influence the computed metrics.

technologies. Table 2 presents four scenarios that align with the numbers presented in Figs. 9(a) and 10(a). Each scenario shows the combination of two of our metrics which may lead to a different electrification strategy. We refer the reader to both Figs. 9(a) and 10 (a) for spatial visualization. In Table 2, the transformer colors are given in brackets for each scenario. Scenario **1** occurs when there are many structures connected to a given transformer and there is a small LV wire requirement for structures connected to the transformer. With a large number of structures connected to the transformer, the cost of the transformer is spread across multiple



(a) A, Nairobi (25, 2.8, 2194) (b) B, Garissa (27.4, 194.2, 145)



(c) C, TanaRiver (31.1, 39, 1991) (d) D, Embu (96.5, 40.8, 8071)

**Fig. 7.** Four wards with varying settlement patterns are shown. In brackets are the per-structure LV requirement, per-structure MV requirement and the structure count of the ward, respectively. The grey boxes surrounding each ward represent a 25 km$^2$ area. Figure (a) and (b) show similar LV requirements with significantly different MV requirements. Figure (c) and (d) show varying LV requirements at similar MV requirements.

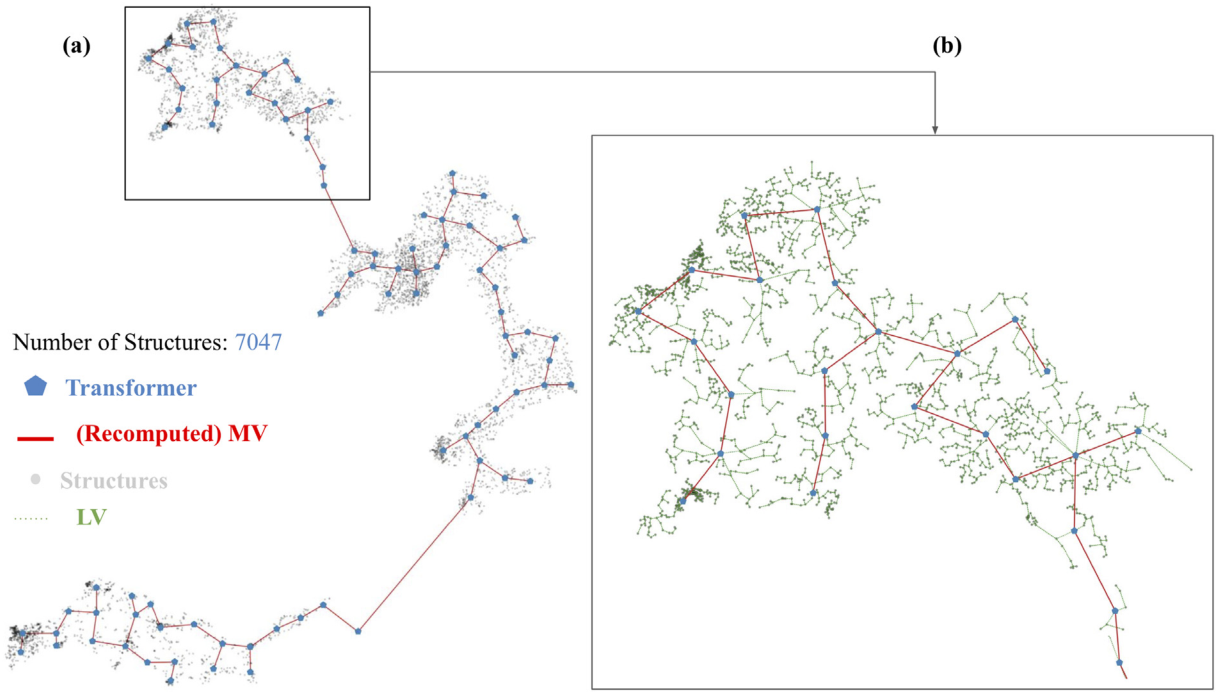



**Fig. 8.** Complete network for a sample ward with 7047 structures. Figure (a) shows transformer placement and the MV network connecting the transformers. Figure (b) includes the LV network for a small section of the ward, showing connections between structures and transformers.

**Table 2**
Scenarios highlighting different electrification strategies which can be identified with our method.

| Scenario | Structures/transformer | LV/structure (m) | Possible system(s) |
|---|---|---|---|
| 1(Purple) | High (blue) | Low (blue) | Grid Extension or Minigrid |
| 2(Black) | High (blue) | High (red) | Solar Home System (SHS) |
| 3(Orange) | Low (red) | High (red) | Solar Home System (SHS) |
| 4(Dark Red) | Low (red) | Low (blue) | Local Generation or Minigrid |

structures, thereby reducing the cost to any individual structure. Coupled with a low LV wire requirement, the choice of electrification is heavily dependent on the per-structure MV wire requirement. A low MV wire requirement suggests a centralized system like grid extension is a viable option for structures connected to these transformers. Scenario **2** shows there are many structures connected to a transformer but the structures are not clustered around the transformer.[5] Although the per-structure transformer cost is low due to high number of connecting structures, the high LV wire requirement becomes a major bottleneck to networking this transformer and the structures associated to it. Solar home systems might prove to be suitable alternatives in this scenario. Scenario **3** presents a worst case scenario from a networking standpoint. Here there are few structures connected to the transformer and the structures are not in close proximity to each other. Similar to scenario 2, solar home systems might be worth considering as the cost to connect structures is high. Scenario **4** represents a case where there are few structures connected to the transformer, but the structures are in close proximity to each other and the associated transformer. In this scenario local generation and distribution through the low cost LV network would seem the most suitable approach. Because our approach uses individual structures, energy providers can explore the implications of networking at multiple resolutions, right down to the individual transformers. We do not show the MV wire metric at sub-administrative boundaries, as the existing grid network is needed in order to assign an MV wire requirement to a given transformer.

[5] Note that we show 7047 structures which may appear as though they are in close proximity but represent multiple kilometers of coverage.

## Sensitivity analysis

We evaluate the robustness of our proposed metrics by performing a cost sensitivity analysis. Table 3 presents our proposed metrics under 3 cost scenarios: i) baseline cost previously discussed, ii) double MV and LV wire cost iii) double transformer cost. The sensitivity analysis is performed on four previously presented wards A through D, first introduced in the Effect of settlement patterns section. From this sensitivity analysis we show that our proposed per structure MV, LV and transformer metrics are stable (less than 3% change) under the three cost scenarios. We also observe that the wire cost is the primary driver of cost. This observation is apparent when doubling transformer cost results in less than 6.5% change in the cost per structure across all four wards, while doubling wire costs, doubles the cost per structure across all wards.

Through this cost sensitivity analysis, we show that our proposed metrics can support infrastructure planning, where the actual unit wire and transformer installation costs (best known by the planner) can be directly multiplied by our metrics to obtain realistic cost estimates to support electricity infrastructure decision making.

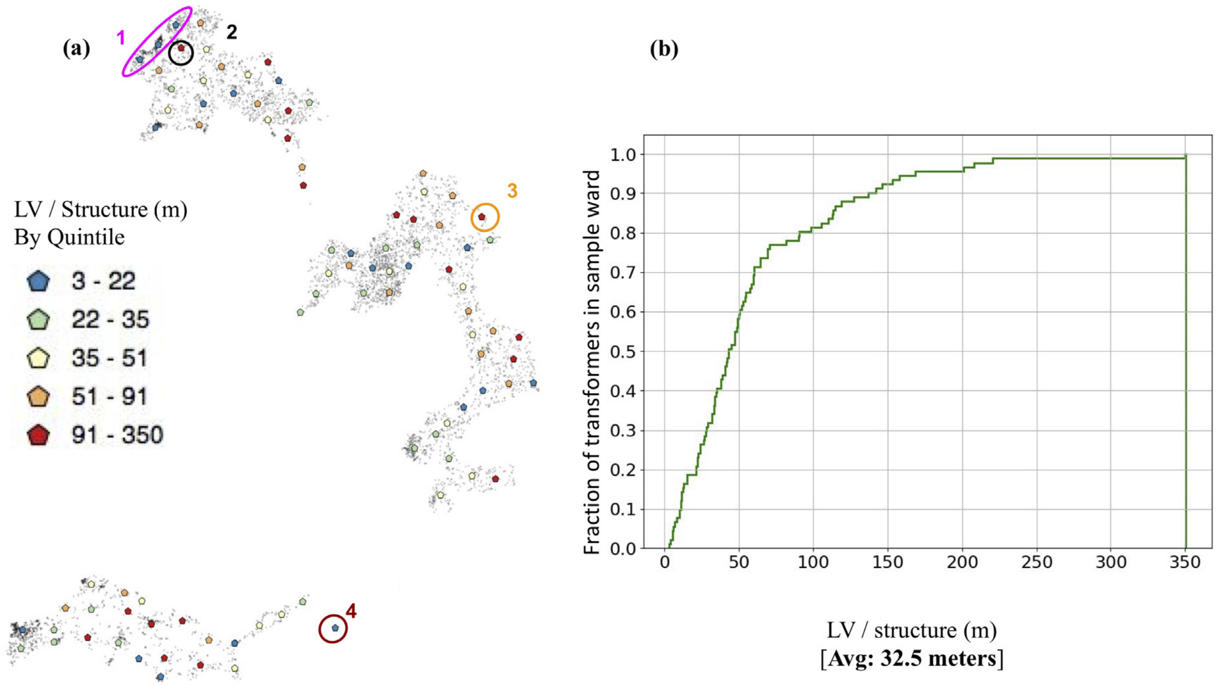



**Fig. 9.** Low Voltage (LV) per structure, for each transformer in sample ward. a) Spatial distribution of LV per structure, binning transformers by quintile. b) CDF of LV per structure for all transformers in ward. The ward average is 32.5 m. Four scenarios are presented, each with different implications for networking. See Table 2 for details.

## Conclusion

In this paper we assess the effects of regional geography and settlements patterns on electrification strategies. By estimating the locations of residential structures through our proposed merging process, we are able to capture settlement behaviors of structures over a whole country. Through our novel computational framework that involves a network design algorithm, we develop a two-level distribution network between the structures. We present a set of connectivity metrics on the wire requirements, number of structures on a transformer, and connection cost on a country level without sacrificing spatial resolution. We show that easily accessible metrics such as population density ignore the interplay between structure locations, and accordingly the true connection cost of a structure.

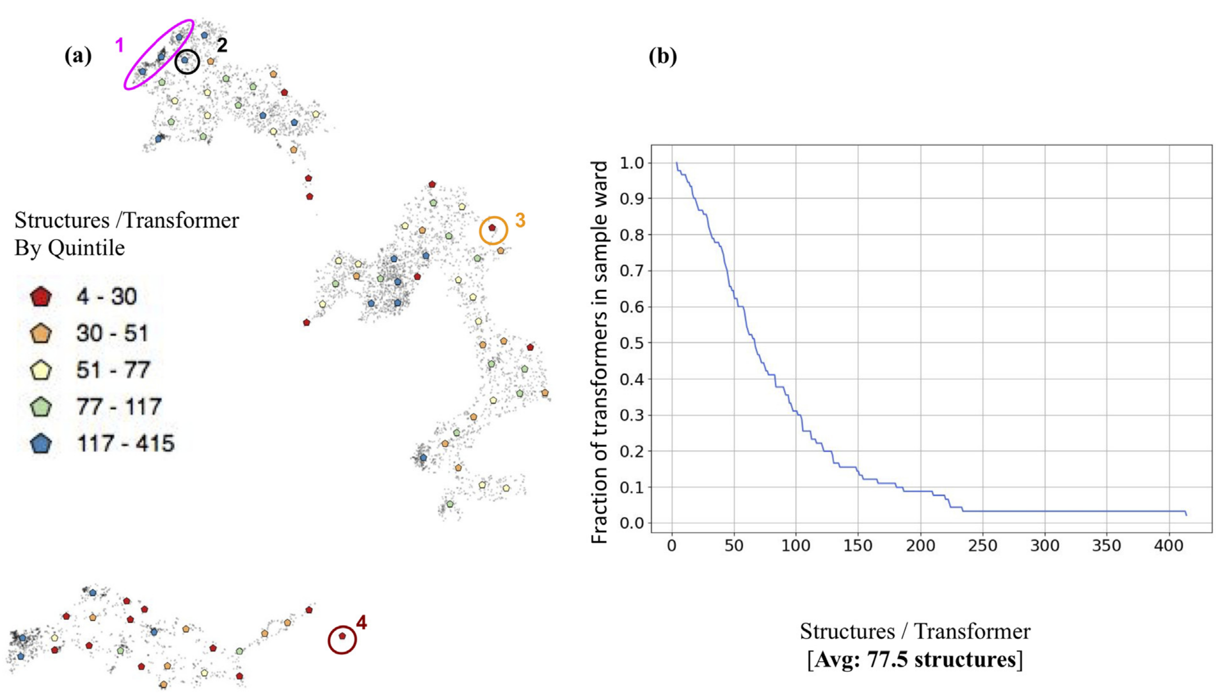



**Fig. 10.** Number of structures per transformer, for each transformer in the sample ward. a) Spatial distribution of structures per transformer, binning transformers by quintile. b) CDF of structures per transformer for all transformers in ward. The ward average is 77.5 structures per transformer. Four scenarios are presented, each with different implications for networking. See Table 2 for details.

**Table 3**
Cost sensitivity analysis under three scenarios i) baseline cost (MV = $25/m, LV = $10/m, Transformer = $2000) ii) 2× MV and 2× LV wire cost, iii) 2× transformer cost. Sensitivity analysis is presented for 4 wards (A,B,C,D) previously in Effect of settlement patterns section

| | | Baseline cost | 2× wire cost | 2× transformer cost |
|---|---|---|---|---|
| LV per structure | Ward A | 25 | 25 | 25 |
| | Ward B | 27.4 | 27.4 | 27.4 |
| | Ward C | 31.1 | 31.1 | 31.1 |
| | Ward D | 96.5 | 96.3 | 96.5 |
| MV per structure | Ward A | 2.87 | 2.89 | 2.89 |
| | Ward B | 194.2 | 194.2 | 194.2 |
| | Ward C | 38.99 | 38.99 | 38.99 |
| | Ward D | 40.85 | 40.92 | 40.85 |
| Structures per transformer | Ward A | 137.12 | 137.13 | 137.13 |
| | Ward B | 29 | 29 | 29 |
| | Ward C | 32.1 | 32.1 | 32.1 |
| | Ward D | 14.7 | 14.6 | 14.7 |
| Cost per structure | Ward A | 336 | 660 | 352 |
| | Ward B | 5198 | 10,326 | 5266 |
| | Ward C | 1348 | 2634 | 1411 |
| | Ward D | 2123 | 4109 | 2259 |

We demonstrate that metrics which capture settlement behavior are crucial when planning efficient electrification on a large scale. Meeting the targets set in SDG7 requires considerations of multiple consumers across large landscapes with varying settlement patterns and our proposed metrics can easily be folded into existing planning approaches to support these objectives. In addition, thanks to its scalability, our framework can support decision making at a granular level by recommending electrification strategies such as solar home systems, mini-grids and grid.

Our future efforts will involve relaxing some of the assumptions made in this work. Relaxing the assumption on uniform consumption would potentially lead to different network outcomes and would allow for variable transformer sizing. We also intend to capture existing grid infrastructures in our planning approach, for all settings have some initial network backbone that influences optimal electrification strategies. Finally, in our current implementation, the two-level network does not account for environmental and topological constraints such as protected areas, rights-of-ways, and elevation. As we believe these constraints would influence the medium voltage computation, we aim to incorporate them in future work.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A

### *Merging approach*

We considered various merging radii to merge the 11.9 million identified building structures.

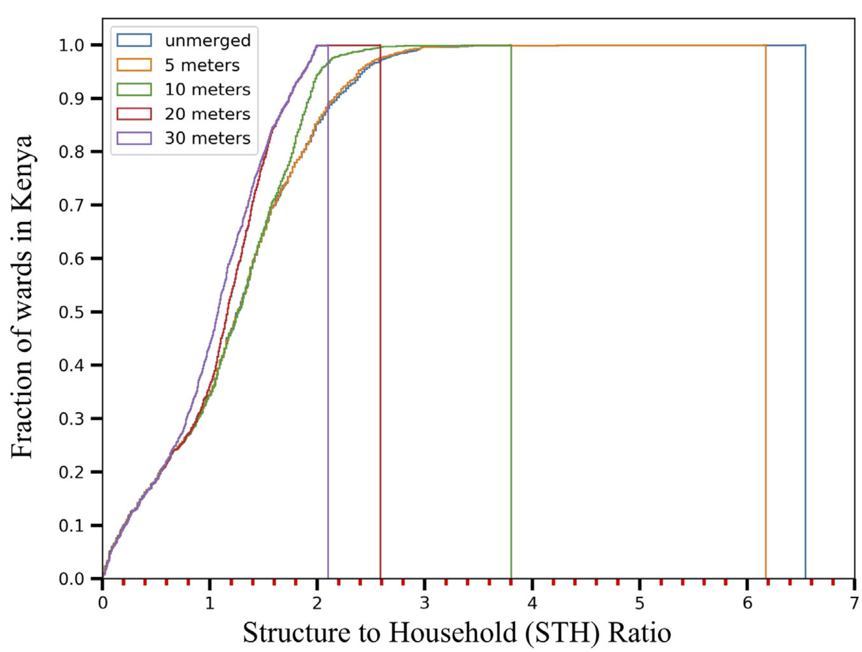



**Fig. 11.** CDF of STH ratios for all wards in Kenya under varying merging radii.

Fig. 11 shows the effect of merging on STH ratio under varying merging radii. We see that the maximum STH ratio is 6.5 for unmerged structures, with multiple wards well above 2 structures per household. This implies that at the worst case, for a specific ward, every household has about 6 structures. We believe this estimate to be wrong as it does not account for other building types (commercial, industrial, etc). For merging radii from 5 to 30 m, we observe a drop in the STH ratio, where at 20 m and 30 m, the maximum STH ratios are 2.6 and 2.1 respectively. We decided on the 20 m merging radius because it reduced the STH ratio for wards with exceedingly high STH ratios, without compromising those wards with STH less than 1. In the case of a merging radius of 30 m (as seen by the purple line), the STH ratios of less than 1 were further depressed.

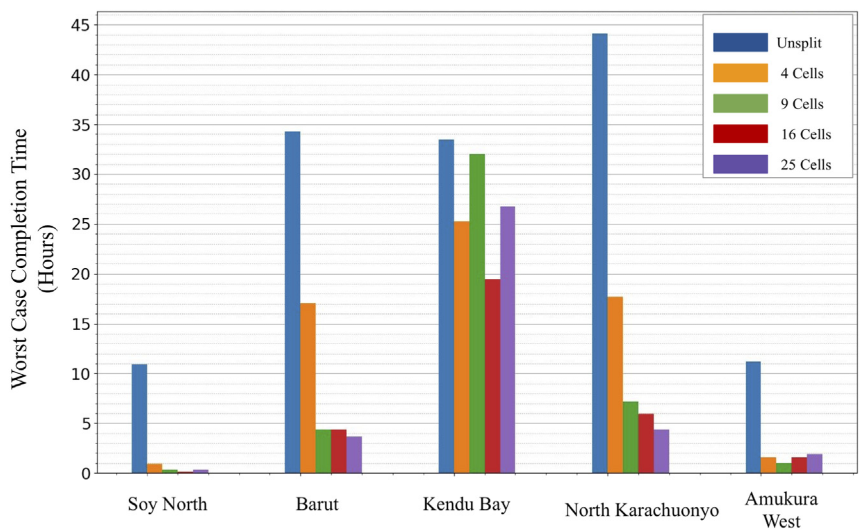



**Fig. 12.** Completion time of the TLND in hours for the cell that took the longest time. Four out of five times, splitting a ward into 4 dropped the completion time by half.

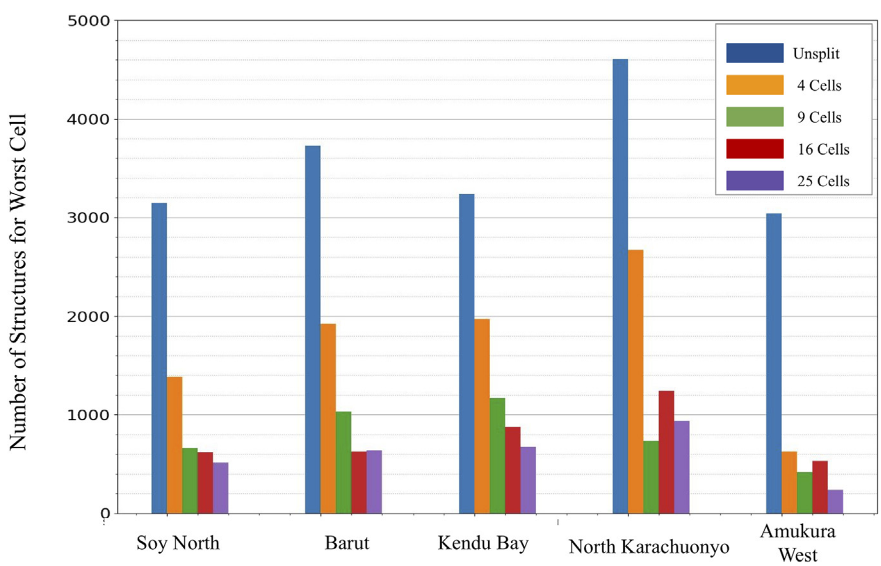


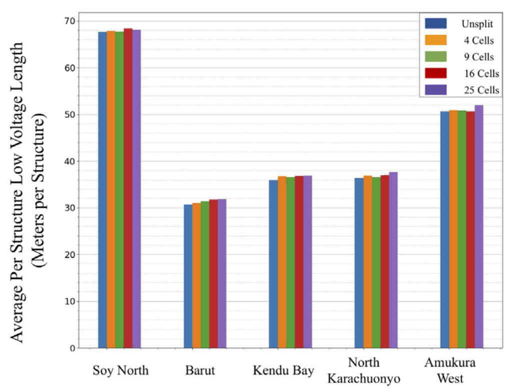

**Fig. 13.** Number of structures for the cell will the longest run time. Splitting decreased the number of structures. However, number of structures is not the only driver of completion time. As in the case of Kendu Bay, spatial layout of structures also influences the computational time.

*Sensitivity to scaling strategy*

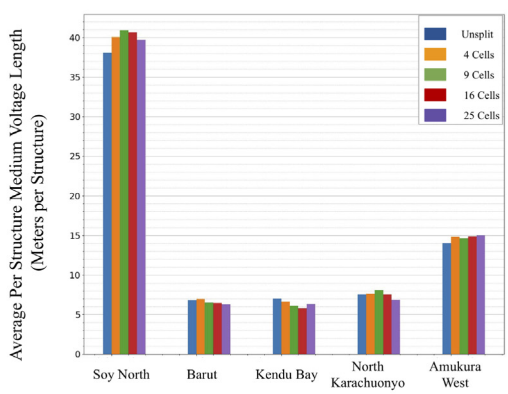
We evaluated our framework by looking at some wards under varying split configurations. The selected wards were split into 4, 9, 16 and 25 cells and the TLND was applied to each cell. The runtime, per-structure low voltage wire requirement, per-structure medium voltage wire requirement and transformer capacity for the split configurations were evaluated against the unsplit ward. In this experiment, we only control the number of cells generated and do not apply limits on the number of structures in the cell or the cell radius. Fig. 12 shows the worst case completion time in hours for five wards split into the aforementioned number of cells. The worse case completion time represents the completion time for the cell that took the longest to run. The computational time is cut by more than half for 4 of the 5 wards when the ward is split into 4 cells. Subsequent splitting further improves the completion time for the 4 wards.

The computational time for Kendu Bay in Fig. 12 oscillates as the number of cells increases, although the worst case always takes less time when the ward is split than when it is left unsplit. To better understand this oscillation, we looked at the number of structures for the cell with the longest runtime in each of the wards. Fig. 13 shows the number of structures under varying splits for the cell with the longest completion time. Capping the number of structures in a cell ($M$) at 3000 structures, significantly decreases the completion time. In our computational framework our choice for the hyper-parameter $M$ was 3000 and thus ensured that large wards were split to cells with manageable number of structures. Revisiting Kendu Bay ward, where completion time oscillated, we observed from Fig. 11 that dropping the number of structures in the cell is not the only contributing factor to completion time. Fig. 13 suggests that the settlement pattern or spatial layout of structures within the cell influences the completion time. It also suggests that without enforcing minimum limits on the cell radius $R$, over-splitting a ward can have negative effects thereby increasing the computational time. Thus we used a minimum cell radius of 500 m to stop over-splitting and capped the maximum number of structures in the largest cluster ($N$) at 300. This ensured computational gains while minimizing degradation in performance of our metrics.

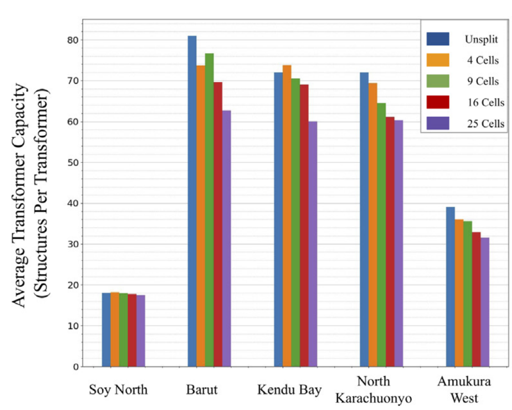
Fig. 14 shows our average connectivity metrics for 5 wards under varying split approaches. The figure also shows the results when the algorithm is run on the whole ward using the **Unsplit** label. These average connectivity metrics are obtained by first summing the metrics across all cells in a ward, then normalizing the sums by the number of structures in the wards. Fig. 14(a) and (b) show that our LV and MV connectivity metrics are not heavily influenced by splitting the ward into cells and applying our reconstruction strategy. However, we notice that the number of structures per transformer varies under different split strategies and tends to drop as we increase the number of cells a ward is split into. From these wards, we observe that transformers tend to be more under-loaded as the number of cells increase. We apply a minimum radius $R$ in our splitting algorithm to prevent excessive splitting, thereby ensuring that the number of structures per transformer is maximized.




(a) LV after Reconstruction




(b) MV after post-processed Recomputation




(c) Number of structures per transformer after Reconstruction

**Fig. 14.** Effect of splitting and MV reconstruction on our proposed connectivity metrics. The two-level network design is applied to each cell. Averages for the ward are reported here.

## References

Abdul-Salam, Yakubu, & Phimister, Euan (2016). The politico-economics of electricity planning in developing countries: A case study of Ghana. *Energy Policy*, *88*, 299–309.

A geospatial framework for electrification planning in developing countries, author = Jason E Adkins and Vijay Modi and Shaky Sherpa and Roy Han and Ayse Selin Kocaman and Naichen Zhao and Chris Natali and Jonathan Carbajal, booktitle = 2017 IEEE Global Humanitarian Technology Conference (GHTC), pages = 1–10, year = 2017, organization = IEEE.

Akpan, Uduak (2015). Technology options for increasing electricity access in areas with low electricity access rate in Nigeria. *Socio-economic Planning Sciences*, *51*, 1–12.

Bolukbasi, Gizem, & Kocaman, Ayse Selin (2018). A prize collecting Steiner tree approach to least cost evaluation of grid and off-grid electrification systems. *Energy*, *160*, 536–543.

Carvallo, Juan-Pablo, Taneja, Jay, Callaway, Duncan, & Kammen, Daniel M. (2019). Distributed resources shift paradigms on power system design, planning, and operation: An application of the GAP model. *Proceedings of the IEEE*, *107*(9), 1906–1922.

Chvatal, Vasek (1979). A greedy heuristic for the set-covering problem. *Mathematics of Operations Research*, *4*(3), 233–235.

Ciller, Pedro, & Lumbreras, Sara (2020). Electricity for all: The contribution of large-scale planning tools to the energy-access problem. *Renewable and Sustainable Energy Reviews*, *120*, 109624.

Debnath, Kumar Biswajit, & Mourshed, Monjur (2018). Challenges and gaps for energy planning models in the developing-world context. *Nature Energy*, *3*(3), 172–184.

Esau, Larry R., & Williams, Kenneth C. (1966). On teleprocessing system design, Part II: A method for approximating the optimal network. *IBM Systems Journal*, *5*(3), 142–147.

Facebook and CIESIN-ColumbiaUniversity. High resolution settlement layer (HRSL) 2016. https://www.ciesin.columbia.edu/data/hrsl/.

Garey, Michael R., & Johnson, David S. (1979). *Computers and intractability – A guide to NP-completeness.*

Kemausuor, Francis, Adkins, Edwin, Adu-Poku, Isaac, Brew-Hammond, Abeeku, & Modi, Vijay (2014). Electrification planning using Network Planner tool: The case of Ghana. *Energy for Sustainable Development*, *19*, 92–101.

KNBS. Kenya demographic and health survey 2014. https://dhsprogram.com/pubs/pdf/fr308/fr308.pdf.

Ayse Selin Kocaman, Woonghee Tim Huh, and Vijay Modi. Initial layout of power distribution systems for rural electrification: A heuristic algorithm for multilevel network design. Applied Energy, 96:302–315, 2012.

Sadhan Mahapatra and S Dasappa. Rural electrification: Optimising the choice between decentralized renewable energy sources and grid extension. Energy for Sustainable Development, 16(2):146–154, 2012.

Megiddo, Nimrod, & Supowit, Kenneth J. (1984). On the complexity of some common geometric location problems. *SIAM Journal on Computing*, *13*(1), 182–196.

MIT. Reference electrification model: A tool for rural electrification planning. https://tatacenter.mit.edu/portfolio/reference-electrification-model-a-tool-for-rural-electrification-planning/, Accessed 2020-01-20.

ModiLabs. Networkplanner. http://optimus.modilabs.org/. Accessed: 2019-03-04.

Magda Moner-Girona, Katalin Bódis, Thomas Huld, Ioannis Kougias, and Sandor Szabó. Universal access to electricity in Burkina Faso: Scaling-up renewable energy technologies. Environmental Research Letters, 11(8):084010, 2016.

Navarro, Alejandro, & Rudnick, Hugh (2009). Large-scale distribution planning—Part I: Simultaneous network and transformer optimization. *IEEE Transaction On Power Systems*, *24*(2), 744–751.

Parshall, Lily, Pillai, Dana, Mohan, Shashank, Sanoh, Aly, & Modi, Vijay (2009). National electricity planning in settings with low pre-existing grid coverage: Development of a spatial model and case study of Kenya. *Energy Policy*, *37*(6), 2395–2410.

Prim, Robert Clay (1957). Shortest connection networks and some generalizations. *The Bell System Technical Journal*, *36*(6), 1389–1401.

Sanoh, Aly, & Parshall, Lily (2012). Ousmane Fall Sarr, Susan Kum, and Vijay Modi. Local and national electricity planning in Senegal: Scenarios and policies. *Energy for Sustainable Development*, *16*(1), 13–25.

UNDP. Goal 7: Affordable and clean energy, 2019. https://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-7-affordable-and-clean-energy.html, Accessed 2019-12-13.

WorldBank. More people have access to electricity than ever before, but world is falling short of sustainable energy goals, 2019. https://www.worldbank.org/en/news/press-release/2019/05/22/tracking-sdg7-the-energy-progress-report-2019, Accessed 2019-12-22.

Zeyringer, Marianne, Pachauri, Shonali, Schmid, Erwin, Schmidt, Johannes, Worrell, Ernst, & Morawetz, Ulrich B. (2015). Analyzing grid extension and stand-alone photovoltaic systems for the cost-effective electrification of Kenya. *Energy for Sustainable Development*, *25*, 75–86.

Zvoleff, Alex, Kocaman, Ayse Selin, Huh, Woonghee Tim, & Modi, Vijay (2009). The impact of geography on energy infrastructure costs. *Energy Policy*, *37*(10), 4066–4078.