A prize collecting Steiner tree approach to least cost evaluation of grid and off-grid electrification systems

Gizem Bolukbasi, Ayse Selin Kocaman*

Department of Industrial Engineering, Bilkent University, 06800, Ankara, Turkey

ABSTRACT

The lack of access to electricity in developing countries necessitates spatial electricity planning for guiding sustainable electrification projects that evaluate the costs of centralized systems vis-a-vis decentralized systems. Heuristic approaches have been widely used in such electrification problems to find feasible, cost effective solutions; however, most of the time global optimality of these solutions is not guaranteed. Our paper through its modeling approach provides a new methodology to find the least cost solution to this electrification problem. We model the spatial network planning problem as Prize Collecting Steiner Tree problem, which would be a base for a decision support tool for rural electrification. This new method is systematically assessed using both randomly generated data and real data from rural regions across Sub-Saharan Africa. Comparative results for the proposed approach and a widely used heuristic method are presented based on computational experiments.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

The International Energy Agency estimated that nearly 1.1 billion people do not have access to electricity in 2017 and a majority of those are in Sub-Saharan African countries [1]. In this region, electricity coverage is estimated to be approximately 30%, where some countries such as Chad, Liberia, and Sudan had less than 10% electricity penetration rate in 2014 [2]. However, developments in health, education, environmental sustainability, and agriculture depend on the electricity coverage [3]. Therefore, lack of access to electricity is a fundamental problem to be solved, which requires spatial electricity planning projects that evaluate the costs of centralized approaches vis-a-vis decentralized approaches in these countries.

In centralized systems, electricity is produced at large scale facilities and distributed to the users through a transmission and distribution network. These systems require large amount of infrastructural investments and generally utilize fossil fuels [4]. On the other hand, the emissions arising from the usage of fossil fuels are the main reason behind the global warming and various health problems. Moreover, dependence on fossil fuels is ultimately unsustainable and leads to issues of energy security. These facts

explain why decentralized energy systems are becoming increasingly important for effective electrification of underdeveloped world. Unlike centralized energy systems, decentralized energy systems are more commonly running with renewable energy sources such as solar, wind and geothermal power. These systems are especially appropriate for isolated rural communities since they can be sized at various scales. Moreover, decentralized systems can be designed either as off-grid (standalone systems) or grid-connected systems [5]. Stand-alone decentralized systems are suitable for remote locations where the grid cannot penetrate; however, they have some drawbacks such as low capacity factor (due to intermittency of solar/wind), high battery costs and finite capacity to store electricity. On the other hand, at grid-connected decentralized systems, connectivity to the grid enables setting up relatively large-scale systems such as regional or national scale and hence, they can operate at high load factors improving the economic viability of the operation. Grid connectivity also helps reduce the intermittency of renewable sources since geographic aggregation can smooth the variability of different renewable sources.

Although energy leapfrogging, a concept that implies skipping fossil fuel based centralized systems and moving directly to cleaner and more efficient decentralized systems, has been also widely discussed in the literature for the long-term solutions of rural electrification process [6–8], an extensive number of studies focusing on rural electricity planning incorporating both centralized and decentralized systems can be found in the literature. A

---

* Corresponding author.
E-mail address: selin.kocaman@bilkent.edu.tr (A.S. Kocaman).

model based analysis of global rural electrification scenarios that assesses future trends in electrification has been provided in Ref. [9] and it has been stated that for Sub-Saharan African countries, the potential in mini-grid and off-grid technologies is expected to be high since the grid electrification will have high infrastructural costs due to low population densities in rural areas. A geospatial analysis has been performed in Ref. [10] to examine the cost trade-offs between grid and off-grid electrification options in areas with different settlement patterns. In Refs. [11,12], it has been shown that the grid extension and off-grid electrification complement each other. One example of this result is obtained for Kenya by comparing grid extension with stand-alone photovoltaic systems [13].

The main goal of the studies focusing on the evaluation of grid and off-grid systems is to consider alternative options based on factors such as demographic conditions, investment costs, geographic features of the regions or the greenhouse gas emissions and to find the least cost scenario. Considering all these parameters and using large data sets make these optimization problems difficult to solve and leads the researches to alternative heuristic approaches. A hierarchical lexicographic programming based algorithm for politico-economics of electricity planning using a case study of unelectrified communities in Ghana has been used in Ref. [14]. Minimum spanning tree (MST) problem has been widely adopted for the grid expansion planning in similar problems [15]. In the minimum spanning tree problem, the aim is to connect all nodes in a set in a way to minimize the total cost of connections. Variations of the well-known MST algorithms have been also developed to incorporate the isolated nodes that would be electrified with a decentralized system into the decision making process [16,17]. While evaluating the cost of grid connected and isolated systems, spatial distribution of the demand nodes are very important. Zvoleff and Kocaman [18] and Kocaman et al. [19] have developed heuristic approaches that evaluate the impact of settlement patterns in local level (demand nodes like home, school etc.) while calculating the cost of single and two-level distribution networks, respectively.

Another heuristic based approach has been proposed in Ref. [20] for guiding the electrification projects of the regions with low electricity coverage. This approach has been updated to be used also for the regions that do not have any existing electricity coverage and used as the basis of a decision support mechanism to explore grid and off-grid electrification options in rural communities. This open-source mechanism is called Network Planner and can be reached online [21]. This mechanism is capable of testing different scenarios, demonstrating performance comparisons and maps. Network Planner has been used in national electrification studies of countries such as Senegal [22] and Ghana [23]. The underlying solution approach in Network Planner is an adaptive and fast heuristic approach that can provide solutions in a short amount of time; however, its optimality performance (i.e. the closeness of the results to the optimal solution) is unknown. Heuristic approaches can find quick solutions to large problems, yet they do not provide any guarantee on the solution quality. Solution approaches that guarantee a solution quality either find the best solution of the problem or provide an optimality gap, a measure on the closeness to optimality. However, solution times of these approaches are usually longer compared to the heuristic ones. Therefore, there is a trade-off between the solution quality and the solution time. Considering the large investment requirements of the electrification projects, it is very important to know the quality as much as the solution time of the approaches to be able to use them as a decision support tool. In this direction, Abdul-Salam et al. [24] presented a non-linear discrete modeling approach for solving the problem proposed in Ref. [20] optimally; however, they could

obtain solutions only for small instances of the problem (maximum 40 nodes). In this study, we approach to this problem as a Prize Collecting Steiner Tree problem for the first time in the literature and propose a mixed-integer linear programming (MILP) model that can be solved with the existing MILP solvers for larger instances. Existing state-of-the-art algorithms specifically developed for the PCST problem can also be used to quickly solve the real life examples [25]. This modeling approach provides a new methodology to find the least cost solution of electricity networks that are designed as a combination of centralized and decentralized options rapidly and optimally.

The outcome of this research is expected to be an important decision support tool for the electrification of the underdeveloped and developing countries, having the potential of contributing to the socio-economic development of these countries. This study is expected to make a significant contribution to the energy literature. It includes both theoretical modeling that can solve an important social problem in the field optimally for large instances and applications with real case studies, for which the data has also been provided [26].

2. Problem statement

In this study, we work on a spatial network planning problem that would be a basis for the decision support tools for rural electrification. In this problem, two options will be considered, as shown in Fig. 1, for the rural communities that may be of varied sizes, such as a country, region, or city. The first option is to design an isolated decentralized system that generates electricity using clean renewable energy sources. The other option is to connect the community to the grid where the electricity is generated in a centralized way and transmitted using transmission lines. Depending on the total electricity demand of the community, estimating the cost of a centralized system is more difficult than of a decentralized option, because it requires considering the spatial distribution of the communities and the optimal placement of the networked infrastructure between them.

Three main cost accounts are considered for the decision-making process of rural electrification. The first two of these costs are related to grid option and are referred to as grid internal and grid external costs for the communities. The external costs are for medium or high voltage lines that connect the communities on the network to each other and to a central source. The internal costs consist of the transformers that convert medium voltage to low voltage and the cost of the low voltage lines that provide distribution of the electricity from the transformers to the demand points such as households, hospital, school etc. While grid-connected systems have both internal and external costs, a system that is not connected to the grid, such as an isolated solar or wind power system, will include costs related to electricity generation and storage. This decentralized system can be called an off-grid system. The problem is to provide each community with electricity access in a way that the total cost of the overall system is minimized. This will be achieved either by connecting communities to the grid or by suggesting an isolated decentralized option such as a solar energy system for them.

If the grid internal cost within a community is higher than the cost of an isolated system, then it would never be cost effective for this community to be electrified with the centralized grid option. If the cost of isolated system is higher for a community, the difference between the cost of the isolated system and the grid internal cost determines the amount that this community can afford to be connected to the grid via transmission lines. Our solution approach aims to determine the communities that will have off-grid option, and the network topology between the communities that will be
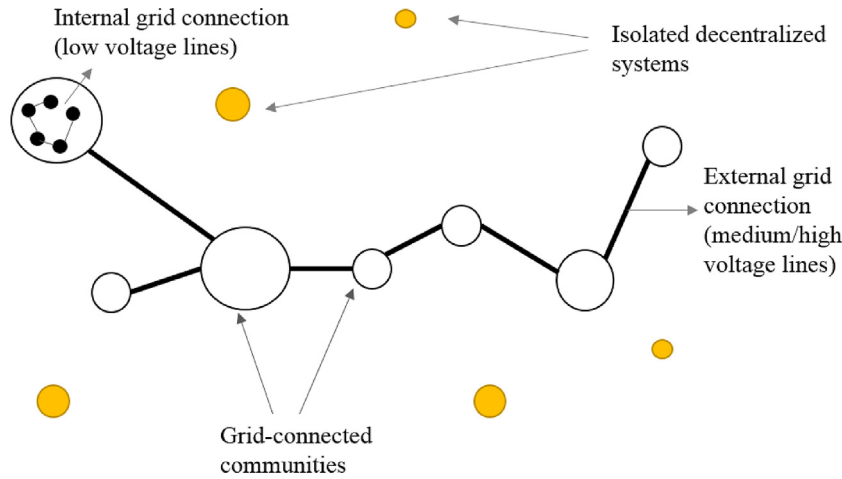
Fig. 1. Systems components considered in the problem.

electrified with the grid option, minimizing the cost of the entire system.

3. Solution approach

The heuristic approach used in Network Planner to solve the problem defined in Section 2 is known as the modified Kruskal's algorithm. In this section, we firstly review the modified Kruskal's algorithm and present our solution approach, which can provide optimal results in a reasonable time.

3.1. An existing approach: Modified Kruskal's algorithm

As an input, Network Planner takes the data that are specific to the respective communities such as the cost, population, finance, average distance between houses etc. As a result, it calculates the total cost of the decentralized systems (off-grid) and the cost of the network for the centralized system (grid). Since the cost of the centralized option includes two parts (internal grid cost and external grid cost), selecting the communities which will be connected to the grid is not an easy task. If the cost of the decentralized option is lower than the internal grid cost (cost of low voltage lines) for a community, then it would never be cost-effective for this community to be connected to the grid as the connection to the other grid-connected communities also requires external grid cost. Therefore, this community can be directly assigned to the off-grid option. However, if the internal grid cost is lower than the cost of decentralized option for a community, some external grid cost may be compensated with the difference. If the difference between the cost of the decentralized option and grid internal cost is sufficient for a community to be connected to other communities on the network via medium-voltage lines, then this community can be considered as a grid-compatible node, i.e. this node can be connected to the network. To define the maximum length of medium voltage line that can be afforded by each community, a metric called *MVmax* is introduced in Ref. [20] as follows:

$$MVmax = (Cost\ of\ decentralized\ system - Internal\ grid\ cost) \\ /Unit\ cost\ of\ medium\ voltage\ (MV)\ line$$

The *MVmax* values calculated for each community are given as inputs to the heuristic approach. In the Network Planner, Kruskal's algorithm [27], which optimally solves the Minimum Spanning Tree (MST) problem, is modified to include the *MVmax* values into account. In Minimum Spanning Tree problem, there is a given connected graph with positive edge weights and the aim is to find a minimum weight set of edges that connects all of the nodes. Kruskal's algorithm is a famous method, in which as a first step, distances between all points are sorted in an ascending order. The algorithm starts with an empty network and repeats itself by adding the shortest available edge to the network as long as it does not create a cycle. The algorithm stops when all of the points are connected with a tree. In the modified version of Kruskal's algorithm, *MVmax* criteria is added as an additional constraint. At each iteration, the *MVmax* of the two points considered to be added to the network is compared with the distance between the two points, and if the *MVmax* of both points is sufficient to meet the distance, the edge between these points is added to the network. Whenever a new connection is added, the *MVmax* of the new set, *k*, containing the connected points *i* and *j* is updated as follows:

$$MVmax_k = MVmax_i + MVmax_j - distance_{ij}$$

Please note that at the end of this approach, one might end up with several trees. In that case, largest tree can be considered as the main grid and the small ones can be designed as decentralized systems or, in some cases, as mini-grids. Modified Kruskal (MK)'s algorithm explained above is a very fast solution method that gives practical solutions. However, it is a heuristic method without a proven optimality bound and therefore, its reliability is questionable. Our new solution approach, that is explained in the next section, assures the optimal solution and therefore permits to evaluate the mathematical reliability of MK method.

3.2. Proposed approach: Prize Collecting Steiner Tree

In previous sections, the rural electrification problem that aims to minimize the cost of the entire system by combining centralized and decentralized isolated systems is presented and the widely used modified Kruskals' algorithm is explained in details. One of the most important contributions of our study to the literature is to define the mentioned rural electrification problem as a Prize Collecting Steiner Tree problem and provide solutions with known optimality performance.

In the general version of the Prize Collecting Steiner Tree problem, there is a prize to be earned specific to each node if the node is included in the tree. The goal in the problem is to collect the highest prize while reducing the cost of the tree that connects the nodes. In this sense, there is a trade-off in the problem. In the Goemans and Williamson (GW) version of the problem, a penalty is

defined instead of a prize for each node [28]. If a node is not included in the tree, the penalty determined for the node should be added to the system cost. Therefore, the problem is to minimize the cost of the tree and the penalty of the ones that are not included in the tree in the GW version. Note that a Steiner tree may contain non-terminal nodes that do not have associated penalty/prize. These nodes are referred to as Steiner points. In our problem, all nodes have non-negative penalties as there is an associated cost of decentralized systems, which will be paid unless the nodes are included in the tree.

In the general formulation of Prize Collecting Steiner Tree Problem, one is given a non-directional graph $G = (V, E)$, a non-negative edge cost $c(e)$ for each edge $e \in E$, a non-negative node prize/penalty $p(v)$ for each node $v \in V$. The Goemans-Williamson version of the PCST problem aims to find a subgraph $T' = (V', E')$ of G, $V' \subseteq V$, $E' \subseteq E$ that minimizes the objective function $c(T')$, which includes the cost of the edges in the tree and the penalty of the nodes not in the tree as follows:

$$c(T') = \sum_{e \in E'} c(e) + \sum_{v \notin T'} p(v) \tag{1}$$

We model the PCST problem using a single commodity flow formulation [29]. The non-negative edge costs correspond to the cost of external grid system and the penalties correspond to the differences between the decentralized system costs and internal grid costs of the communities. Let us denote the index set 1, …,N by $\mathcal{N}$, where N is the number of demand points and define the parameters and the decision variables used in the model as follows:

Parameters:
$CoffGrid_i$: The cost of a decentralized system to be installed at point $i \in \mathcal{N}$
$CgridInternal_i$: The cost of distribution network within point $i \in \mathcal{N}$ (internal grid cost)
$CgridExternal_{ij}$: Medium or high voltage line cost between points $i \in \mathcal{N}$ and $j \in \mathcal{N}$
Variables:
x_i: 1 if point $i \in \mathcal{N}$ is connected to the network, 0 otherwise
s_i: 1 if point $i \in \mathcal{N}$ has a decentralized system, 0 otherwise
f_{ij}: flow amount from point $i \in \mathcal{N}$ to point $j \in \mathcal{N}$
u_{ij}: 1 if point $i \in \mathcal{N}$ is connected to point $j \in \mathcal{N}$, 0 otherwise
v_i: 1 if point $i \in \mathcal{N}$ is selected as the source node, 0 otherwise
ϕ_i: flow amount coming to point $i \in \mathcal{N}$

$$\min \sum_{i=1}^{N} s_i CoffGrid_i + \sum_{i=1}^{N} x_i CgridInternal_i + \sum_{i=1}^{N}$$
$$\times \sum_{j=1}^{N} u_{ij} CgridExternal_{ij} \tag{2}$$

subject to

$$x_i + s_i = 1 \quad i \in \mathcal{N} \tag{3}$$

$$\sum_{i=1}^{N} v_i \leq 1 \tag{4}$$

$$N v_i \geq \phi_i \quad i \in \mathcal{N} \tag{5}$$

$$\sum_{j=1}^{N} f_{ij} + x_i = \sum_{l=1}^{N} f_{li} + \phi_i \quad i \in \mathcal{N} \tag{6}$$

$$N u_{li} \geq f_{li} \quad l, i \in \mathcal{N} \tag{7}$$

$$\sum_{l=1}^{N} u_{li} + v_i = x_i \quad i \in \mathcal{N} \tag{8}$$

$$u_{ij} \leq x_i \quad i, j \in \mathcal{N} \tag{9}$$

$$x_i \in \{0, 1\} \quad i \in \mathcal{N} \tag{10}$$

$$s_i \in \{0, 1\} \quad i \in \mathcal{N} \tag{11}$$

$$v_i \in \{0, 1\} \quad i \in \mathcal{N} \tag{12}$$

$$u_{ij} \in \{0, 1\} \quad i, j \in \mathcal{N} \tag{13}$$

$$f_{ij} \geq 0 \quad i, j \in \mathcal{N} \tag{14}$$

$$\phi_i \geq 0 \quad i \in \mathcal{N} \tag{15}$$

In the objective function of the above model, the total cost of the overall system that has three components is minimized. The first component includes the total off-grid costs of the nodes that are not connected to the grid. The second and the third components of the objective function are related to the grid cost. The grid cost has two parts: grid internal costs and the grid external costs. Grid internal cost includes the cost of the low voltage lines that connect demand points such as households, schools, hospitals to the grid. The external grid cost is related to the medium voltage or high voltage lines between the nodes. The objective is minimized subject to the following constraints: Constraint (3) guarantees that each community has either a decentralized system or connection to the grid. Constraints (4)–(9) model the tree structure between the connected nodes. In (4), it is assured that there will be only one source node. Constraint (5) guarantees the flow amount on the source node. The constraint (6) provides the flow balance between two consecutive nodes that will be included in the tree. Constraint (7) guarantees that if an edge between two nodes is not included to the tree, there is no flow between them. The constraint (8) assures that if a node is selected as a source node, it can not be accessed through a predecessor node. The constraint (9) guarantees that if a node is not included in the tree, its outgoing edges cannot also be included in the tree. Constraints (10)–(15) are the domain constraints.

4. Numerical analysis

The modified Kruskal's algorithm has been widely used in national and local electrification projects for developing and underdeveloped countries to provide feasible results that offer the cost-effective balance between decentralized and centralized systems. In this heuristic approach, global optimality is not guaranteed. Our new approach, on the other hand, provides the optimal solution to the same problem. We use these optimal solutions of the PCST formulation as a benchmark to evaluate the performance of modified Kruskal's algorithm, which we refer as MK in the rest of the paper. In this section, the computational experiments required for the comparative analysis of those two approaches are presented.

Firstly, we tested these two solution approaches by using synthetic data and secondly, we made further analysis by the help of real life instances obtained from Network Planner and provided in Ref. [26]. Experiments are performed on a dual core computer with Intel(R) Core(TM) i7-4510U CPU @ 2.00 GHz, 2001 Mhz. modified Kruskal's algorithm is implemented in MATLAB R2016a and PCST model is solved by CPLEX Studio IDE 12.6.

4.1. Experiments with synthetic data

We created test instances at different sizes to test the PCST approach and compare it with the MK approach. We generated our own sample data rather than using the ones from literature since our problem includes only terminal nodes and the existing data sets that are frequently used in the literature for PCST problems contain both non-terminal and terminal nodes. Both the distance and off-grid costs are generated randomly using uniform distribution. This analysis provides information about mathematical performance of MK. Moreover, experiments demonstrate the computational time required for the PCST approach. In Table 1, the size of the instances used in the computational experiments can be found in the first column. In the second and third columns, percentages of grid-compatible nodes obtained by MK and PCST methods are presented, respectively. In the following six columns, decentralized system cost, centralized system cost and total cost of both methods are reported. In the following columns, the CPU time requirement of the both approaches and cost difference percentages are provided. Note that some of the instances did not give optimal results in our 3 h time limit when we use PCST method. For these results, the optimality gap reported by CPLEX 12.6 can be found next to the result in the Total Cost column.

Given that PCST approach provides the optimal solution to the problem, results in Table 1 show that MK approach is practical and reliable for electrification projects since the results are close to the optimal solution and CPU times are small even for the large instances. It can also be concluded that PCST formulation can be preferred especially for small sized instances since there is no significant difference in terms of the solution time. It should also be underlined that, in this study we directly solve the model via CPLEX. However, alternative methods that solve PCST problem optimally and in a faster way can be found in the literature [25].

4.2. Experiments with case studies from Network Planner

In this section, we test the performances of the two approaches using the real data obtained from the Network Planner tool. Note that Network Planner is an open source and online platform, in which all the data required such as internal costs of centralized

systems, cost of unit MV-line, off-grid, and mini-grid costs, co-ordinates of each node can be found. We have chosen six instances to make a comparative analysis. Diversity in terms of sizes and geographic features are our main criteria while choosing our instances. The results of the two methods can be found in Table 2. Similar to Table 1, the size of the instances used in the computational experiments can be found in the first column. In the second and third columns, percentages of grid-compatible nodes obtained by MK and PCST methods are presented, respectively. In the following six columns, decentralized system cost, centralized system cost and total cost of both methods are reported. In the following columns, the CPU time requirement of the both approaches and cost difference percentages are provided. Note that for some instances, optimal results could not be obtained within our 3 h solution time limit in our computational setting. For these results, the optimality gap reported by CPLEX is indicated under the result in the Total Cost column.

Table 2 shows that the optimal solution of the PCST problem for some instances could not be obtained within our 3 h time limit given our computational environment. However, we observe that the results with positive optimality gaps are still better than the result of MK approach except for one instance. In that instance, the MK approach provides 0.04% better result than PCST result with 4.96% optimality gap. The optimal solution of the instances resulted with an optimality gap in our time limit and computational environment can be easily obtained in better computational settings or without a time limit. We observe that MK approach provides a solution for all instances in units of seconds. However, the solution quality of these results are unknown unless we compare these results with the results of PCST approach. On the other hand, PCST approach provides results in units of hours and the solution time highly differs based on grid percentage in the final solution. Since the tree topology between the grid-compatible nodes makes the problem very complicated, the solution time of the model is also affected by the grid-compatible nodes. For example, the optimal result for 94-node instance is not obtained within 3 h time limit, whereas the optimal result for 141-node instance is obtained. As the number of grid-connected nodes is higher in the 94-node hli-ninstance, the solution process includes more decisions about the connections made between these nodes.

Another interesting results that Tables 1 and 2 show is that MK approach is more conservative in terms of the number of grid connections. To examine the reason behind the potential differences between the solutions in detail, we have chosen an instance and investigated the steps of MK approach iteratively to detect, where in contrast to the PCST solution, the MK chooses the decentralized option. In other words, this analysis aims to explain the behavioral tendencies of MK approach that makes it result with

Table 1
Comparative analysis of two methods for randomly generated instances.

Number of nodes	Grid percentage (%)[a]		Decentralized System Cost		Centralized System Cost		Total Cost		CPU Times (seconds)		Cost Diff.(%)[b]
	MK	PCST	MK	PCST	MK	PCST	MK	PCST	MK	PCST	
5	40	40	45	45	19	19	64	64	4.5	0.1	0
10	50	50	50	50	63	63	113	113	4.0	0.2	0
20	55	60	97	86	112	120	209	206	6.8	1.4	1.9
50	46	56	210	183	217	234	426	417	5.4	7.4	2.3
100	40	58	326	206	275	374	601	580	8.4	76.3	3.6
200[c]	43	62	2836	1648	2950	3886	5786	5534 (0.95%)	13.6	10,800	4.6
300	49	62	2538	1728	3577	4153	6115	5881	22.4	5135.2	4.0
500[c]	44	61	3052	1972	3407	4233	6459	6205 (5.6%)	81.5	10,800	4.1

[a] Grid percentage = Number of grid-compatible nodes/Total number of nodes.
[b] Cost Difference = (Total cost of MK- Total cost of PCST)*100/(Total cost of PCST).
[c] The result is not optimal, percentages within the parenthesis show optimality gap reported by CPLEX. Time limit is 3 h.

Table 2
Comparative analysis of two methods for real life instances.

Number of nodes	Grid percentage (%)[a]		Decentralized System Cost		Centralized System Cost		Total Cost		CPU Times (seconds)		Cost Diff.(%)[b]
	MK	PCST	MK	PCST	MK	PCST	MK	PCST	MK	PCST	
21	90.5	90.5	677,478	677,478	4,205,352	4,205,352	4,882,830	4,882,830	6.8	6.1	0
94[c]	23.4	70.2	2,247,649	760,543	2,118,953	3,473,336	4,366,602	4,233,878 (4.39%)	9.3	10,800	3.14
141	7.1	14.2	7,004,089	5,824,077	32,405,490	33,502,098	39,409,579	39,326,175	10.5	2578.8	0.21
102	43	58	3,549,189	2,322,924	3,339,708	4,104,273	6,888,897	6,427,197	17.0	91.8	7.18
230[c]	91.3	100	2,418,715	0	62,920,938	65,316,844	65,339,653	65,316,844 (30.81%)	21.2	10,800	0.04
274[c]	84.7	89.1	1,398,681	1,087,633	68,124,741	68,463,021	69,523,422	69,550,654 (4.96%)	19.6	10,800	−0.04

[a] Grid percentage = Number of grid-compatible nodes/Total number of nodes.

[b] Cost difference = (Total cost of MK- Total cost of PCST) * 100/(Total cost of PCST).

[c] The result of PCST is not optimal, percentages under the results show optimality gap reported by CPLEX. Time limit is 3 h.

different solutions than PCST. We have worked on 102-node instance to investigate the reasons behind the differences in solutions. The final trees obtained by both methods are different as presented in Fig. 2. In Fig. 3, we zoom into the right-bottom corner of the 102-node hlininstance. Remember that MK approach connects two communities only if *MVmax* values of the both end points of an edge are larger than or equal to the length of the edge. Therefore, it is obvious that *MVmax* value gets larger when two communities merge and form a cluster. This means that their

power for buying more MV lines may grow when they come together with other clusters. However, the Kruskal's algorithm sorts the edges in increasing order of their lengths and tries to connect the shortest edge possible unless that edge creates a cycle. Since in the modified Kruskal's algorithm the edge to be considered for connection is also determined based on the lengths of the edges, the priority is given to the distances before the *MVmax* values. Therefore, some edges are considered while the *MVmax* values of the end points of the edge are small and this causes some of the
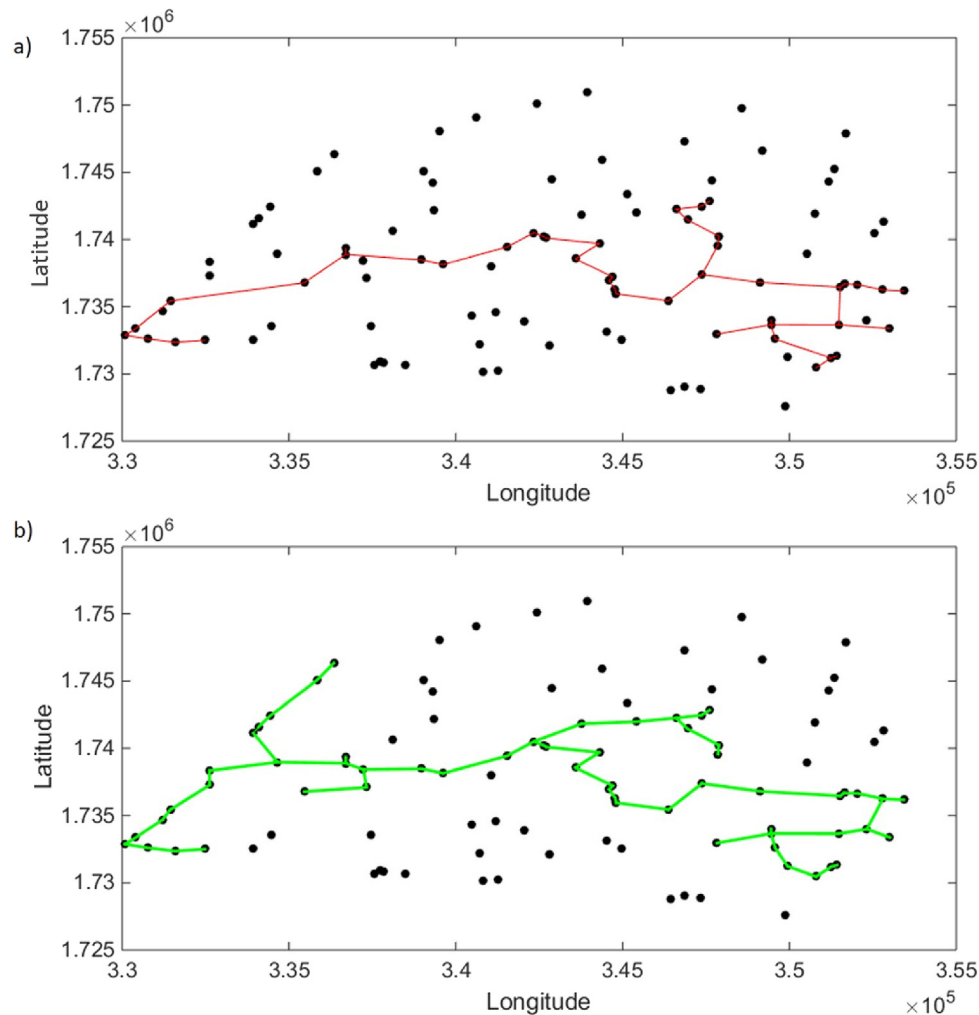
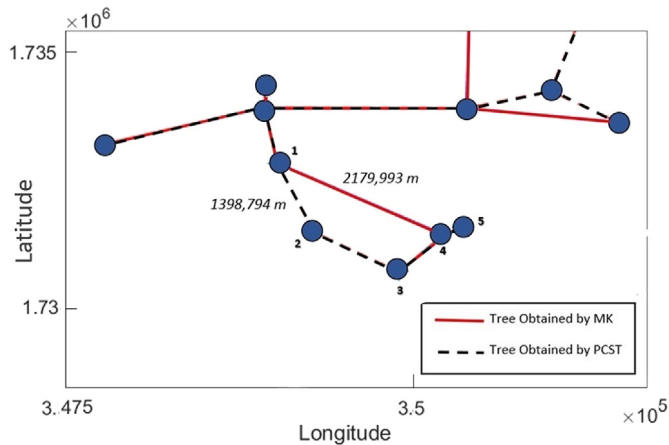**Fig. 2.** Final grid obtained for 102-node instance with a) MK, b) PCST methods.

Fig. 3. Difference between the tails of two trees obtained by MK and PCST approaches on 102-node instance.

edges not to be included in the tree (although in the optimal solution they might a part of the tree). In this example, the distance between node 1–2 is smaller than 1–4 with values of 1398.8 and 2178.8 m, respectively. However, in Fig. 3 it can be observed that MK approach chooses to connect 1–4 instead of 1–2 because when edge 1–2 was considered, the length of the edge 1–2 was greater than *MVmax* of node 2. Recalling the fact that *MVmax* is a metric that is getting larger proportionally to the number of nodes included in the tree, *MVmax* of node 2 could have been larger if it had made another connection before. Likewise, initially the *MVmax* of node 4 was 1670, however it reached the value of 3926 by getting connected with nodes 5 and 3 before the algorithm considers its connection with node 1. The distance between node 1–4 is 2179.8 m and it was impossible for node 4 to be connected to 1 if the edge between them is considered before node 4 made other connections.

4.3. Sensitivity analysis on distance parameters

We have conducted a sensitivity analysis to observe how results alter with variations of distances (or with different costs for

external network). To assess the potential impact of variations on total cost and grid percentage, we created a 100-node test instance and its reduced forms as shown in Fig. 4. The results of both MK and PCST methods are summarized in Table 3.

In Table 3, in the first two columns total costs obtained by MK and PCST approaches are presented. In the following two columns, how much the total costs decrease with respect to distance reduction is presented. The percentage based cost comparison of MK and PCST approaches can be found under the cost difference column. Finally, the grid percentages of the solution trees of both methods are presented in the last two columns. As can be expected, when distances between the communities are reduced (i.e. communities are clustered), the total costs of the final systems decrease and the number of nodes connected to grid increases for both methods as the *MVmax* values of the communities will be more than the lengths of the higher number of edges. Similar to previous results, MK approach tends to connect less communities to the grid than PCST approach. However, as the clustering of the communities increases, the difference between the solutions of the both approaches reduces. In the 100% grid connection percentage case, both methods are supposed to find a minimum spanning tree between the communities and this tree will be the optimal tree that spans all points. Therefore, it is expected that MK approach will provide solutions close to the optimal result for the cases with high grid connection.

Our results show that PCST model may be used for electricity planning problem for large scaled unelectrified regions including up to 500 nodes given our computational setting. However, the number of nodes is not the only factor that determines the level of computational complexity and CPU time. Costs and the level of dispersion of the communities are some other factors contributing to them. Analyzing the spatial distribution of the communities can give us an idea about how the level of clustering affects the computational complexity.

5. Conclusion

A balanced combination of decentralized and centralized electrification systems can be considered as the best way for the electrification of developing countries in terms of both cost effectiveness and environmental responsibilities. However, due to computational complexity of the problem, especially for large
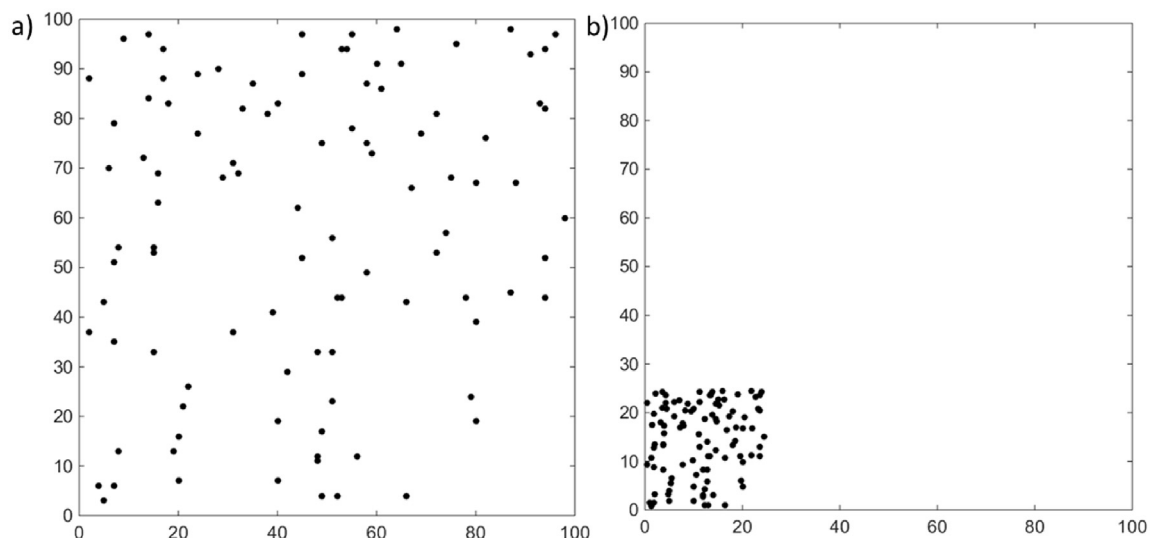

Fig. 4. a) Randomly generated 100-node sample (Base Scenario), b) 100-node sample with distances reduced by 75%.

Table 3
Comparative analysis of two methods with respect to distance variations.

	Cost		Decrease in Cost (%)		Cost Difference (%)	Grid percentage	
	MK	PCST	MK	PCST		MK	PCST
Base Scenario	601.1	580			3.6	40%	58%
Reduce distances by 25%	472.8	465.3	21.3	19.8	1.6	74%	80%
Reduce distances by 50%	328.1	320.9	30.6	31.0	2.3	83%	92%
Reduce distances by 75%	164.6	164.1	49.8	48.9	0.3	95%	96%

scaled instances, solving the problem optimally requires a lot of time. Therefore, heuristic approaches are widely used since they can provide practical solutions in a reasonable amount of time. Yet, most of the time, they do not guarantee global optimality of the solution. In this paper, we focused on the problem introduced by Parshall et al. and the heuristic solution approach called modified Kruskal's algorithm [20]. Alternatively, we presented the Prize Collecting Steiner Tree model of the underlying problem, which yields the optimal solutions to the problem.

Our computational experiments were based on both randomly synthetic and real life instances. In this paper, our aim is to share the patterns and implications observed through our limited number of experiments. The empirical analysis showed that it is possible to find the optimal solution of the least cost planning problem of centralized and decentralized electrification systems within a reasonable time using the PCST approach. Given that infrastructure planning problems are long term and capital intensive problems, most of the time, it is worth the wait to obtain the best solution. On the other hand, we observed that the existing MK approach can quickly provide solutions which are close to the optimal solution.

Finally, recalling the ecological harm caused by traditional electrification methods, cost should not be the only objective to be considered. The new electrification planning policies for developing countries should therefore contribute both to cost-effective sustainable development and to climate change mitigation. Therefore, as a future work developing multi-objective optimization models for this problem would permit taking into account other factors such as carbon emission and health issues.

References

[1] IEA. World Energy Outlook. http://www.iea.org/weo2017/. Accessed: 2018-06-19.
[2] World Bank. Access to electricity (% of population). http://data.worldbank.org/indicator/EG.ELC.ACCS.ZS. Accessed: 2018-06-19.
[3] Davidson Ogunlade R, Sokona Youba. A new sustainable energy path for African development: think bigger act faster. Energy and Development Research Centre, University of Cape Town Cape Town; 2002.
[4] Hiremath RB, Shikha S, Ravindranath NH. Decentralized energy planning; modeling and application -a review. Renew Sustain Energy Rev 2007;11(5): 729—52.
[5] Kaundinya Deepak Paramashivan, Balachandra P, Ravindranath NH. Grid-connected versus stand-alone energy systems for decentralized power a review of literature. Renew Sustain Energy Rev 2009;13(8):2041—50.
[6] Szabó Sandor, Bódis K, Huld T, Moner-Girona M. Sustainable energy planning: leapfrogging the energy poverty gap in Africa. Renew Sustain Energy Rev 2013;28:500—9.
[7] Levin Todd, Thomas Valerie M. Can developing countries leapfrog the centralized electrification paradigm? Energy Sustain Dev 2016;31:97—107.
[8] Amankwah-Amoah Joseph. Solar energy in sub-saharan africa: the challenges and opportunities of technological leapfrogging. Thunderbird Int Bus Rev 2015;57(1):15—31.
[9] Van Ruijven Bas J, Schers Jules, van Vuuren Detlef P. Model-based scenarios for rural electrification in developing countries. Energy 2012;38(1):386—97.
[10] Adkins Jason Edwin, Modi Vijay, Sherpa Shaky, Han Roy, Kocaman Ayse Selin, Zhao Naichen, Natali Chris, Carbajal Jonathan. A geospatial framework for electrification planning in developing countries. In: Global humanitarian technology conference (GHTC), 2017 IEEE. IEEE; 2017. p. 1—10.
[11] Urpelainen Johannes. Grid and off-grid electrification: an integrated model with applications to India. Energy Sustain Dev 2014;19:66—71.
[12] Dagnachew Anteneh G, Lucas Paul L, Hof Andries F, EHJ Gernaat David, de Boer Harmen-Sytze, van Vuuren Detlef P. The role of decentralized systems in providing universal electricity access in sub-saharan africa—a model-based approach. Energy 2017;139:184—95.
[13] Zeyringer Marianne, Pachauri Shonali, Schmid Erwin, Schmidt Johannes, Worrell Ernst, Morawetz Ulrich B. Analyzing grid extension and stand-alone photovoltaic systems for the cost-effective electrification of Kenya. Energy Sustain Dev 2015;25:75—86.
[14] Abdul-Salam Yakubu, Phimister Euan. The politico-economics of electricity planning in developing countries: a case study of Ghana. Energy Pol 2016;88: 299—309.
[15] Lambert Thomas William, Hittle DC. Optimization of autonomous village electrification systems by simulated annealing. Sol Energy 2000;68(1): 121—32.
[16] Levin Todd, Thomas Valerie M. Least-cost network evaluation of centralized and decentralized contributions to global electrification. Energy Pol 2012;41: 286—302.
[17] Deichmann Uwe, Meisner Craig, Murray Siobhan, Wheeler David. The economics of renewable energy expansion in rural Sub-Saharan Africa. Energy Pol 2011;39(1):215—27.
[18] Zvoleff Alex, Selin Kocaman Ayse, Tim Huh Woonghee, Modi Vijay. The impact of geography on energy infrastructure costs. Energy Pol 2009;37(10): 4066—78.
[19] Kocaman Ayse Selin, Huh Woonghee Tim, Modi Vijay. Initial layout of power distribution systems for rural electrification: a heuristic algorithm for multi-level network design. Appl Energy 2012;96:302—15.
[20] Parshall Lily, Pillai Dana, Mohan Shashank, Sanoh Aly, Modi Vijay. National electricity planning in settings with low pre-existing grid coverage: development of a spatial model and case study of Kenya. Energy Pol 2009;37(6): 2395—410.
[21] NetworkPlanner. http://optimus.modilabs.org/. Accessed: 2018-06-19.
[22] Sanoh Aly, Parshall Lily, Fall Sarr Ousmane, Kum Susan, Modi Vijay. Local and national electricity planning in Senegal: scenarios and policies. Energy Sustain Dev 2012;16(1):13—25.
[23] Kemausuor Francis, Adkins Edwin, Adu-Poku Isaac, Brew-Hammond Abeeku, Modi Vijay. Electrification planning using Network Planner tool: the case of Ghana. Energy Sustain Dev 2014;19:92—101.
[24] Abdul-Salam Yakubu, Phimister Euan. How effective are heuristic solutions for electricity planning in developing countries. Soc Econ Plann Sci 2016;55: 14—24.
[25] Ljubić Ivana, Weiskircher René, Pferschy Ulrich, Klau Gunnar W, Mutzel Petra, Fischetti Matteo. An algorithmic framework for the exact solution of the prize-collecting steiner tree problem. Math Program 2006;105(2):427—49.
[26] Ayse Selin Kocaman. Data for Least Cost Evaluation of Grid and Off-grid Electrification Systems, Mendeley Data, v2. https://doi.org/10.17632/mxh7h63grr.2. Accessed: 2018-18-06.
[27] Kruskal Joseph B. On the shortest spanning subtree of a graph and the traveling salesman problem. Proc Am Math Soc 1956;7(1):48—50.
[28] Johnson David S, Minkoff Maria, Phillips Steven. The prize collecting Steiner tree problem: theory and practice. SODA 2000;1:4.
[29] Haouari Mohamed, Layeb Safa Bhar, Sherali Hanif D. Tight compact models and comparative analysis for the prize collecting Steiner tree problem. Discrete Appl Math 2013;161(4):618—32.