

7-2019

## Discovering Communities for Microgrids with Spatial-Temporal Net Energy

Shangyu Xie

Han Wang

Shengbin Wang

Haibing Lu

Yuan Hong

*See next page for additional authors*

---

**Authors**

Shangyu Xie, Han Wang, Shengbin Wang, Haibing Lu, Yuan Hong, Dong Jin, and Qi Liu

---



# Discovering communities for microgrids with spatial-temporal net energy



Shangyu XIE<sup>1</sup>, Han WANG<sup>1</sup>, Shengbin WANG<sup>2</sup>, Haibing LU<sup>3</sup>,  
Yuan HONG<sup>1</sup> , Dong JIN<sup>1</sup>, Qi LIU<sup>4</sup>

**Abstract** Smart grid has integrated an increasing number of distributed energy resources to improve the efficiency and flexibility of power generation and consumption as well as the resilience of the power grid. The energy consumers on the power grid, e.g., households, equipped with distributed energy resources can be considered as “microgrids” that both generate and consume electricity. In this paper, we study the energy community discovery problems which identify energy communities for the microgrids to facilitate energy management, e.g., load balancing, energy sharing and trading on the grid. Specifically, we present efficient algorithms to discover such communities of microgrids considering both their geo-locations and net energy (NE) over any period. Finally, we experimentally validate the performance of the algorithms using both synthetic and real datasets.

**Keywords** Smart grid, Microgrid, Community discovery, Net energy (NE), Clustering

## 1 Introduction

The smart grid infrastructure enables the integration of renewable energy resources at the individual consumer level [1]. It creates a paradigm where any individual consumer in the grid can also be a power supplier. This facilitates the creation of microgrids. Microgrids are the localized grids that can be separated from the larger power grid to operate autonomously and be self-sufficient in the power. A microgrid typically consists of renewable (wind turbines, solar panels, etc.) and/or non-renewable (micro-turbines, fuel cells, etc.) energy resources, energy storage devices, and energy consuming devices/appliances, all of which are connected through a power and communication network [2]. A microgrid can operate in a grid with the connected or islanded mode. In the islanded mode, it could be connected to other microgrids or operate independently.

CrossCheck date: 4 April 2019

Received: 25 October 2018 / Accepted: 4 April 2019 / Published online: 20 July 2019

© The Author(s) 2019

✉ Yuan HONG  
yuan.hong@iit.edu

Shangyu XIE  
sxie14@hawk.iit.edu

Han WANG  
hwang185@hawk.iit.edu

Shengbin WANG  
wangs@tcnj.edu

Haibing LU  
hlu@scu.edu

Dong JIN  
dong.jin@iit.edu

Qi LIU  
qiliu@uri.edu

<sup>1</sup> Department of Computer Science, Illinois Institute of Technology, 10 W 31st Street, Chicago, IL 60616, USA

<sup>2</sup> School of Business, College of New Jersey, 2000 Pennington Rd., Ewing, NJ 08628, USA

<sup>3</sup> Department of Information Systems and Analytics, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053, USA

<sup>4</sup> College of Business, University of Rhode Island, 7 Lippitt Road, Kingston, RI 02881, USA

Therefore, microgrids can provide energy independence to individual communities or entities who intend to manage their own power generation and distribution [3]. Moreover, microgrids can provide resilience against large-scale failures across the grid. They can continue to operate if large-scale blackouts occur [3].

With autonomous energy, microgrids may fully or partially feed their local demand. Numerous microgrids would have great flexibility to utilize their local energy to collaboratively advance the energy management in the power grid, e.g., load balancing [4, 5], energy sharing [6, 7], and load shifting [8]. Thus, it is desirable to discover microgrid communities that can efficiently implement their cooperation in the grid [9–11]. For instance, the grid can identify communities for a mixed set of microgrids, some of which request external power supply while the others have excessive electricity, such that the microgrids within each community can supply their demand load by themselves regularly or when power outage occurs in the main grid.

More specifically, based on every microgrid’s local energy amount (supply) and its local consumption amount (demand load), we can simply derive its net energy (NE) as the amount of supply minus the demand load, which can be either positive or negative at specific time. If the NE of a microgrid is 0 in  $[T_1, T_2]$ , we can simply skip it or assign it to the nearest community. Thus, in this paper, we only consider the microgrids whose NE is either positive or negative. Clearly, a microgrid with positive at time  $t$  means that it has excessive electricity at time  $t$ ; otherwise, it requests external power supply at time  $t$ . In addition, we denote the time series NE of a microgrid  $m_i$  over a period  $[T_1, T_2]$ , where  $T_1 < T_2$ , as  $\forall t \in [T_1, T_2], e_i(t)$ , which can be either positive or negative. Then, some energy communities with respect to time interval  $[T_1, T_2]$  can be defined as follows.

### 1.1 Energy communities

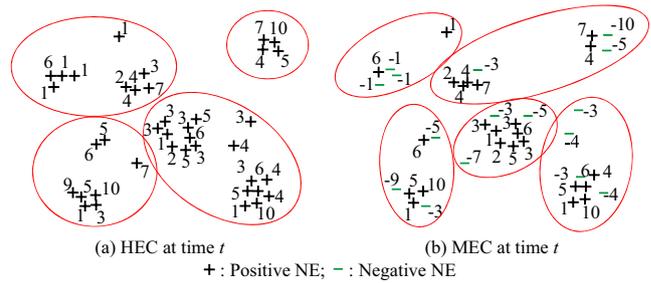
#### 1.1.1 Definition 1: homogeneous energy community (HEC)

A group of microgrids whose NE are exclusively positive, or exclusively negative at any time in  $[T_1, T_2]$ .

In this case, all the microgrids in the community can feed themselves using their local energy, or all the microgrids in the community request external supply. On the contrary, if the microgrids in the community have different NE status (positive and negative) at any time over the period  $[T_1, T_2]$ , we define such community as the follows.

#### 1.1.2 Definition 2: mixed energy community (MEC)

A group of microgrids whose NE are mixed with positive and negative at any time in  $[T_1, T_2]$ .



**Fig. 1** Energy communities of microgrids on power grid (“+” represents positive NE while “-” represents negative NE)

Hence, we can categorize the energy community discovery problems [10] based on their inputs (the NE of all the microgrids is homogeneous or mixed between time  $T_1$  and  $T_2$ ): ① HECs discovery; ② MECs discovery. Figure 1 presents the examples for two different energy communities in the grid at a specific time, respectively. Note that if  $T_1 = T_2$ , HECs and MECs are obtained for a specific time instead of a time interval.

Furthermore, we define a special form of MEC in which all the microgrids’ local energy can fully supply the overall demand of the community.

#### 1.1.3 Definition 3: self-sufficient energy community (SEC)

A mixed energy community whose total NE is non-negative at any time in  $[T_1, T_2]$ .

Since classic clustering algorithms (e.g.,  $K$ -means, DBSCAN) can be tailored to discover HECs by integrating the NE amounts [10], we focus on the MEC discovery and SEC discovery.

### 1.2 Related work

As the important building blocks on the grid, microgrids have attracted significant interests in both industry and academia in the past decade. In such context, many recent research are conducted to design microgrids and/or energy management schemes so as to improve the performance of the power grid such as load management techniques [12], demand response solutions [13], and home automation [14]. More specifically, [15] and [16] propose techniques for establishing microgrids in the power grid based on different criteria such as cost minimization [15] and power flow optimization [16]. In addition, the analysis of data collected from distributed microgrids (e.g., demand load, energy generation and storage) has advanced the energy management of the grid and microgrids [17]. Such applications include short term load forecasting for microgrids [18], load restoration for microgrids [19], load shifting [8], energy trading [20, 21], etc.

Moreover, some cooperative models among distributed microgrids have been investigated in multiple applications, e.g., optimizing the power loss via a unified microgrid

voltage profile [22], eliminating the central energy management unit and price coordinator via localized smart devices [23], distributed energy dispatch and demand response [24], privacy preserving energy management among networked microgrids [25], and load management via sharing local electricity [6, 26]. In this paper, we develop techniques to identify communities of microgrids which can directly implement all these cooperative applications within each energy community to further improve the grid performance.

### 1.3 Contributions

Community discovery problems generally group data objects which share similar characteristics or are close to each other, e.g., detecting communities of individuals who have similar interests on the social network [27], and analyzing the spatial datasets to identify geographical communities [28]. The energy community discovery problems are significantly different from the aforementioned prior community discovery problems studied in other contexts. The key difference is that the criteria of grouping two microgrids into the same energy community should consider not only the spatial distances on the power grid but also their NE amounts of time series. Moreover, additional constraints may apply in the problems, for example, MECs and SECs may require all the microgrids in each community to balance their demand and supply, and to bound the overall NE within a small number or even as 0 [4]; SECs require a nonnegative overall NE for each community. In addition, both energy consumption and generation of microgrids (e.g., wind and solar) are generally stochastic, thus the energy communities (e.g., MECs and SECs) may vary over time. To the best of our knowledge, these have not been investigated and tackled in literature. To address these issues, this paper has the following primary contributions.

- 1) We define the energy community discovery problems for MECs and SECs as well as the proposed new algorithms to effectively and efficiently generate MECs and SECs.
- 2) We discuss how to realize MECs and SECs in the current energy management system, and define some utility metrics to evaluate their performance.
- 3) We conduct comprehensive experiments to validate the performance of our approaches using both synthetic and real-world microgrid datasets.

The rest of this paper is organized as follows. Sections 2 and 3 illustrate how to discover MECs and SECs, respectively. Section 4 discusses how to realize the discovered

communities in the current energy management system on the power grid. Section 5 demonstrates the experimental results. Finally, Section 6 presents the concluding remarks and discusses the future work.

## 2 Discovering MECs

Among thousands of microgrids on the power grid, some of them may have excessive energy while some others may request the energy from external resources, e.g., the main grid. Therefore, adjacent microgrids can share or trade their locally generated electricity for avoiding wasting excessive energy while ensuring better reliability and resilience of power supply [6, 26]. Such microgrids can form an energy community to occasionally feed their local energy demands, e.g., via trading, which are beneficial to both the power grid and themselves. Clearly, the NE of the microgrids in the communities is mixed with negative and positive, thus called as MECs.

The ideal case of the discovered MECs is that all the microgrids in the same MEC are geographically close to each other while balancing their demand and supply of each MEC within a tight margin [4, 9, 29] (then microgrids can fully consume their local energy). We now propose an algorithm to identify such MECs on the grid towards this goal.

Specifically, we denote each NE microgrid  $m_i$  at time  $t$  as  $e_i(t)$ , which can be either positive or negative. While grouping two microgrids, e.g.,  $m_i$  and  $m_j$ , into an MEC, besides the spatial distance between them on the grid  $Dis(m_i, m_j)$ , we also have to consider their NE  $e_i$  and  $e_j$  towards the load balancing of their community. The overall demand and supply at different time should be balanced (ideally, equal to each other). For example, if one microgrid has an NE  $e_i$  while the other has an NE demand  $-e_i$ , such two microgrids can supply their demands using their local energy. Thus, we define a new measure namely “NE distance” of two microgrids  $m_i$  and  $m_j$  in time interval  $[T_1, T_2]$  as:

$$NE(m_i, m_j) = \sum_{t \in [T_1, T_2]} |e_i(t) + e_j(t)| \quad (1)$$

If  $\forall t \in [T_1, T_2]$ ,  $e_i(t) + e_j(t) = 0$  holds, we have  $NE(m_i, m_j) = 0$ . If  $\forall t \in [T_1, T_2]$ ,  $e_i(t) = e_j(t)$  holds, however, we have  $NE(m_i, m_j) = 2 \sum_{t \in [T_1, T_2]} |e_i|$ . The NE distance

differs from other distance measures used in traditional community discovery problems due to its unique feature: two opposite values, e.g.,  $e_i$  and  $-e_i$ , are measured as “close”.

**Algorithm 1** Discovery of two-threshold MECs

**Input:**  $\xi$ : maximum threshold of the NE distances  
 $\xi'$ : maximum threshold of the spatial distances  
**Output:** MECs

- 1: **while** any ungrouped microgrid  $m_i$  in  $m_1, m_2, \dots, m_N$  **do**
- 2:   initialize a new MEC with  $m_i$ :  $c_j = \{m_i\}$
- 3:   **for** each ungrouped microgrid  $m_k$  **do**
- 4:     compute MEC  $c_j$ 's net energy at time  $\forall t \in [T_1, T_2]$ :  $E_j(t)$   
       and its centroid's geo-location  $\mu_j$
- 5:     **if**  $NE(\mu_j, m_k) \leq \xi$  and  $Dis(\mu_j, m_k) \leq \xi'$  **then**
- 6:        $c_j = c_j \cup m_k$  (add  $m_k$  to the MEC  $c_j$ )
- 7:       update  $\forall t \in [T_1, T_2]$ ,  $E_j(t)$  and  $\mu_j$
- 8:     **end if**
- 9:   **end for**
- 10: **end while**
- 11: considering each MEC  $c_j$  as a microgrid with net energy  $E_j(t)$  at time  $t$  and geo-location  $\mu_j$ , repeat lines 1-7 to hierarchically merge the MECs based on  $\xi$  and  $\xi'$  until convergence

Therefore, the difference of the overall supply and demand of every MEC is bounded/balanced at different time by  $\xi$ , and the spatial distance between any microgrid and its MEC centroid is bounded by  $\xi'$ .

For the MEC discovery, we define two maximum distance thresholds for the normalized NE distances and the normalized spatial distances, respectively:  $\xi, \xi' \in [0, 1]$ . Then, we propose a new agglomerative algorithm [30] to identify MECs by utilizing  $\xi$  and  $\xi'$  to specify the criteria for bounding the differences between the overall supply and demand of each community and the spatial distances between the microgrids in each community. Specifically, we let each microgrid find its nearest microgrid (with an NE distance no more than  $\xi$  and a spatial distance  $Dis(\mu_j, m_k)$  no more than  $\xi'$ ) to form an MEC, update the MEC centroid geo-location and NE, and then hierarchically merge "small MECs" to form "large MECs" for better resilience. The merging process terminates if the NE distance between any two MECs' centroids exceeds  $\xi$  or their spatial distance exceeds  $\xi'$  as shown in Algorithm 1.

**3 Discovering SECs**

Many real-world applications require that the microgrids in each MEC can fully supply their demand with their local energy, e.g., large-scale blackouts. Therefore, it is also desirable to discover the SECs with nonnegative NE [11].

Specifically, given  $N$  microgrids  $m_1, m_2, \dots, m_N$ , we denote the number of SECs for the  $N$  microgrids as  $K$ . Then, denoting  $K$  SECs as  $c_1, c_2, \dots, c_K$ , we can define binary variables  $\forall i \in [1, N], \forall j \in [1, K], x_{ij} \in \{0, 1\}$  to indicate if the microgrid  $m_i$  is included in SEC  $c_j$  or not: if  $x_{ij} = 1, m_i \in c_j$ ; otherwise,  $m_i \notin c_j$ .

**3.1 Optimization-based SEC discovery**

If the aggregated NE of the given microgrids is non-negative in  $[T_1, T_2]$ , we can formulate an optimization problem to discover SECs. We first consider the clustering constraints. Note every microgrid can only be assigned to exactly one SEC. This creates a group of clustering constraints  $\sum_{j=1}^K x_{ij} = 1, \forall i \in [1, N]$ .

Secondly, recall that the NE of any SEC should be non-negative at any time  $t \in [T_1, T_2]$ . This criterion creates another group of clustering constraints:

$$\sum_{i=1}^N [e_i(t)x_{ij}] \geq 0, \forall t \in [T_1, T_2], \forall j \in [1, K].$$

Then, we can summarize clustering constraints of SECs as:

$$\begin{cases} \text{s.t.} \\ \sum_{j=1}^K x_{ij} = 1 & \forall i \in [1, N] \\ \sum_{i=1}^N [e_i(t)x_{ij}] \geq 0 & \forall t \in [T_1, T_2], \forall j \in [1, K] \\ \forall i \in [1, N], x_{ij} \in \{0, 1\} \end{cases} \quad (2)$$

**3.1.1 Problem formulation**

If all the binary variables satisfy all the constraints in (2), all the output energy communities would be SECs. Thus, we can solve the constraint satisfaction problem (CSP) without an objective function to find out feasible solutions for SECs. Note that such CSP problem is NP-hard due to the involvement of a large number of binary variables.

More importantly, besides the constraint satisfaction problem, we can formulate the SEC discovery problem by minimizing the overall load on the transmission lines (energy loss in transmission) in all the SECs. Then, we can denote the energy loss rate as  $\theta$ . For example, transmitting an amount of energy 100 W, the load on 1 unit distance is 100 $\theta$  W. Given  $m_i$  with positive NE at time  $t$  as  $e_i(t)$  and any other microgrid  $m_s$  with negative NE at time  $t$  as  $e_s(t)$ , we define the amount of energy from  $m_i$  to  $m_s$  at time  $t$  as  $y_{is}(t)$ . Thus, the overall load on the transmission lines can be represented using the model in [26]:

$$\sum_{t=T_1}^{T_2} \sum_{j=1}^K \sum_{i=1}^N \sum_{s=1, s \neq i}^N [x_{ij}x_{sj}y_{is}(t)\theta \cdot Dis(m_i, m_s)] \quad (3)$$

If  $x_{ij} = 1$  and  $x_{sj} = 1$  ( $m_i, m_s \in c_j$ ), then the load of the power flow from  $m_i$  to  $m_s$  at time  $t$  is derived as  $y_{is}(t)\theta \cdot Dis(m_i, m_s)$ . If  $x_{ij}$  or  $x_{sj} = 0$  (they are not in the same community), there is no power transmission from  $m_i$  to  $m_s$ , and the load is 0. Then, the overall load on the



transmission lines can be aggregated as (3). Meanwhile, there are two additional sets of power flow constraints:

$$\left\{ \begin{array}{l} \text{s.t.} \\ \sum_{s=1, s \neq i}^N [x_{ij}x_{sj}y_{is}(t)] \leq e_i(t) \quad \forall t, \forall i \in [1, N] \\ \sum_{i=1, i \neq s}^N [x_{ij}x_{sj}y_{is}(t)](1 - \theta) \geq |e_s(t)| \quad \forall t, \forall s \in [1, N] \\ y_{is}(t) \geq 0 \quad \forall t, \forall i, \forall s \in [1, N] \end{array} \right. \quad (4)$$

where the above two sets of constraints ensure that the overall outgoing energy of every microgrid with positive NE is no greater than its current excessive energy, and the overall incoming energy of every microgrid with negative energy is no less than its current demand, respectively [26].

In summary, we consider (3) as the objective function, and combine (2) and (4) as constraints.

### 3.1.2 Tabu search based algorithm

Due to the NP-hardness of the optimization problem, we propose a Tabu search [31] based meta-heuristic algorithm to solve the problem. Specifically, the algorithm first specifies a range for the number of SECs  $K \in \{K_{\min}, K_{\min+1}, \dots, K_{\max}\}$ , and arbitrarily partitions all the microgrids into  $K$  groups based on their geo-locations. Then, for every  $K \in \{K_{\min}, K_{\min+1}, \dots, K_{\max}\}$ , the algorithm iteratively searches the neighboring solutions to make the number of SECs reach  $K$  where “moving a microgrid from one group to another nearest group” is defined as one of its neighboring solutions. After obtaining a set of candidate neighboring solutions (different moves), the neighboring solution can mostly improve the objective function (reduce the load with the greatest amount), then replace the current solution with the neighboring solution. To improve the performance of searching performance, the following criteria are integrated in the algorithm.

- 1) An initial community assignment should be specified in Tabu search, e.g., assigning all the microgrids to random communities based on their geo-locations.
- 2) To avoid the solutions getting stuck in local optimum while searching SECs for every  $K$ , a Tabu list is defined with length  $S$  which stores  $S$  most recent solutions that replaced the previous solution. Then, in the searching process, if any neighboring solution is found in the Tabu list, the searching process continues without visiting such neighboring solution.

- 3) Among all the SECs, select the SEC with the highest NE (positive) at most times in  $[T_1, T_2]$ , and then move each microgrid with the positive NE to the corresponding nearest non-SEC, so that a set of candidate neighboring solutions can be found.

The load based objective function cannot be reduced for the current  $K$ . Then, the algorithm moves to the next  $K \in \{K_{\min}, K_{\min+1}, \dots, K_{\max}\}$ . Among all the discovered SECs for all  $K \in \{K_{\min}, K_{\min+1}, \dots, K_{\max}\}$ , the best solution (with the minimum overall load on the transmission lines while satisfying all the constraints) will be selected as the output SECs.

## 3.2 A two-phase algorithm for discovering SECs

Besides the optimization-based approach which formulates the optimization problem and solves the problem with a Tabu search based algorithm, we present a two-phase algorithm to discover a subset of microgrids to form the SECs. Note that, if the overall NE of all the given microgrids are negative in  $[T_1, T_2]$ , the constraints in the optimization-based approach cannot be satisfied simultaneously to form the SECs for all the given microgrids. Instead, the proposed two-phase heuristic algorithm can still effectively discover SECs out of the given microgrids.

Specifically, among all the  $N$  microgrids, we denote the set of microgrids with positive NE at any time in  $[T_1, T_2]$  as  $M^+$ , and the set of microgrids with any negative NE in  $[T_1, T_2]$  as  $M^-$ . Then, the two phases are illustrated as follows.

Phase 1: the algorithm first clusters all the microgrids in  $M^+$  based on their geo-locations, where each cluster can be considered as a “merged microgrid” with aggregated positive NE. In this phase, we extend the  $K$ -means algorithm [32] to cluster such microgrids based on their geo-locations by specifying different  $K \in \{K_{\min}, K_{\min+1}, \dots, K_{\max}\}$ . Then, the algorithm applies different  $K$  values to  $K$ -means and chooses the best clustering result – the minimum sum of squared errors (SSE) of the spatial distances [30] in all the clustering results.

Phase 2: Denote the clustering result of  $M^+$  as  $c_1^*, c_2^*, \dots, c_K^*$ , and the NE of any cluster  $\forall j \in [1, K]$ ,  $c_j^*$  at time  $t$  can be aggregated as  $\sum_{m_i \in c_j^*} e_i(t)$ . Then,  $\forall j \in [1, K]$ ,  $c_j^*$

iteratively adds the nearest ungrouped microgrid of its centroid in  $M^-$  until its NE drops close to 0 at any time in  $[T_1, T_2]$

Finally, the updated  $c_1^*, c_2^*, \dots, c_K^*$  are identified as  $K$  different SECs. The details of the two-phase algorithm are given in Algorithm 2. Note that Algorithm 2 involves all the microgrids in  $M^+$  in the SECs, but may not involve

all the microgrids in  $M^-$  (depending on the NE of the microgrids in  $M^+$  and  $M^-$ ). Furthermore, the NE of most self-sufficient communities can be well balanced to form “zero NE” communities [9].

---

**Algorithm 2** Discovery of two-phase SECs

---

**Input:**  $M^+$ : set of microgrids with positive net energy  
 $M^-$ : set of microgrids with negative net energy  
 $\{K_{\min}, K_{\min+1}, \dots, K_{\max}\}$ : possible values for  $K$

**Output:** SECs

```

1: for  $K = K_{\min}, K_{\min+1}, \dots, K_{\max}$  do
2:   run  $K$ -means for all microgrids in  $M^+$  based on their geo-
   locations to obtain  $c_1, c_2, \dots, c_K$ 
3: end for
4: choose the best clustering result with the minimum SSE for different
    $K: c_1^*, c_2^*, \dots, c_K^*$  (best  $K$ )
5: for  $j \in [1, K]$  do
6:   compute the centroid of  $c_j^*$  as  $\mu_j^*$ 
7:   while  $\forall t \in [T_1, T_2], \sum_{m_i \in c_j^*} e_i(t) \geq 0$  do
8:     find  $\mu_j^*$ 's nearest ungrouped microgrid in  $M^-: m_k$ 
9:      $c_j^* = c_j^* \cup m_k$  (add  $m_k$  to the SEC  $c_j$ )
10:    update the geo-location of  $\mu_j^*$  and  $c_j^*$ 's net energy:
11:    for each  $t \in [T_1, T_2]$  do
12:       $\sum_{m_i \in c_j^*} e_i(t) + = e_k(t)$ 
13:    end for
14:  end while
15: end for
16: return the updated  $c_1^*, c_2^*, \dots, c_K^*$  as SECs

```

---

### 4 Realizing MEC and SEC

After discovering MECs and SECs, microgrids could cooperate with each other by sharing/trading their local energy [6, 20, 26]. Since every microgrid can only be either a power supplier or consumer [6] at any specific time, MECs and SECs are implemented as a bipartite graph on the power grid. In each MEC or SEC, the power might be routed from any microgrid with positive NE to any microgrid with negative NE.

Note that the structure of the bipartite graph may change over time (e.g.,  $M_1$  might be a supplier at time  $T_1$  and it may become a consumer at time  $T_2$ ). Also, the connection between every pair of microgrids can be available via the power transmission network of the main grid [1, 26]. As illustrated in Section 3.1.1, the optimal energy transmission solution (power flow) within each community can be obtained using the model in [26] (which is simplified from the optimization model in Section 3.1.1):

$$\begin{cases} \min \sum_{\forall i, \forall s} y_{is}(t) \\ \text{s.t.} \\ \sum_{\forall s} y_{is}(t) \leq e_i(t) & \forall i \\ \sum_{\forall i} y_{is}(t)(1 - \theta) \geq |e_s(t)| & \forall s \\ y_{is}(t) \geq 0 & \forall i, \forall s \end{cases} \tag{5}$$

Note that SECs are always feasible in the above problem (due to their relatively large amounts of excessive energy). If MECs cannot find an optimal solution (overall demand exceeds overall supply in any MEC), the main grid will fill the gap [26]. Similarly, we can also identify some utility measures for evaluating MECs and SECs.

- 1) Average distance between every pair of power supplier (positive NE) and consumer (negative NE): shorter distance could reduce the energy loss during transmission from the power supplier to the power consumer. Since the structure of the bipartite graph may change over time, we still use the metric of the (spatial) SSE of all the communities to measure such average distance.
- 2) The average NE of each MEC or SEC by taking into account each microgrid’s NE at different time in  $[T_1, T_2]$  denotes  $|\mathit{l}|$  as the number of timestamps utilized for energy community discovery. We identify MECs and SECs based on the energy status of microgrids over a longer period  $[T_1, T_2]$  (a larger  $|\mathit{l}|$ ), which would reflect more accurate results of the communities.
- 3) The load on transmission lines: MECs and SECs have better utility if such load is lower.

## 5 Experiments

### 5.1 Experimental setup

Our experimental simulations are conducted on the synthetic data generated from three real-world datasets: a spatial dataset and two power generation and consumption datasets. Firstly, the spatial dataset of 115475 cities/towns in the US is collected by the US geological survey on 7 July, 2012 and is available in National Imagery and Mapping Agency [33]. Secondly, two power generation and consumption datasets are collected in [34] in East Midlands, UK, and in Massachusetts [35], US. Specifically, [34] collects 22 dwellings’ power consumption over 2 years. Reference [35] collects a low resolution dataset (Umass smart\* home dataset) with 443 households’ power consumption on 2 April, 2011. And it collects a high



resolution dataset (Umass smart\* microgrid dataset) with three microgrids' power generation and consumption over 3 months in 2012. In the Umass smart\* microgrid dataset, both solar panels and wind turbines are installed.

In our experiments, we generate synthetic datasets based on the real-world spatial dataset, and the time-series generation and consumption datasets: ① we aggregate all the generation and consumption datasets with the frequency of one reading per 15 min; ② to test the MECs, we generate two synthetic datasets by sampling 50000 microgrids' power generation and consumption over 1 month based on the microgrid dataset in [35], and then randomly assigning geo-locations in the spatial dataset [33] to the 50000 microgrids; ③ to test the SECs, we use the data in ② MEC discovery to evaluate the two-phase algorithm. To compare the optimization-based approach and the two-phase algorithm, we select 10000 microgrids with a high percent of microgrids with positive NE out of the 50000 microgrids with both generation and consumption, ensuring that the optimization-based algorithm can find a feasible solution.

We use Euclidean distance to measure the spatial distance between any two microgrids on the grid. Both the Euclidean distances and the NE distances are normalized into  $[0, 1]$  in all the experiments.

## 5.2 MEC discovery

Recall that the NE of all the 50000 microgrids (overall power generation minus overall power consumption) is negative. To test the effectiveness of Algorithm 1 in two different cases: ① positive NE, and ② negative NE, we extract two subgroups of microgrids from the 50000 microgrids, each of which includes 20000 microgrids, mixed with positive and negative NE at 2880 different time. For the simplicity of notations, these two subsets of microgrids are named as "positive" and "negative", respectively. Note that the "positive" means all the microgrids are mixed with positive and the negative NE (and the overall NE of all the microgrids is positive); "negative" means all the microgrids are also mixed with positive and negative NE (but the overall NE of all the microgrids is negative).

Firstly, we implemented Algorithm 1 with  $\xi \in [0.03, 0.3]$ , where the normalized spatial distance threshold  $\xi'$  is fixed as a reasonable value 0.05. Then, Fig. 2a shows the average, maximum and minimum NE of all the communities generated from "positive" where  $\xi \in [0.03, 0.3]$ . As  $\xi$  increases from 0.03 to 0.3, the allowed maximum differences between the overall demand and overall supply in every MEC increase significantly. The average, maximum and minimum NE then increase as  $\xi$

increases. Thus, the demand and supply of the MECs become better balanced with an NE closer to 0. On the contrary, Fig. 2c demonstrates the results for "negative", which presents a reverse trend as "positive", but still tend to better balanced load (NE also becomes closer to 0) as  $\xi$  decreases.

Secondly, we also had some other findings in the MEC discovery by utilizing microgrid time series NE over different lengths of periods (varying number of timestamps  $l$ ). As shown in Fig. 2b and 2d, as the NE of microgrids over a longer period (larger  $l$ ) is utilized in the MEC discovery, the average NE of the identified MECs can have both increasing and decreasing trends. This is because larger  $l$  can possibly lead to involving either more or less microgrids in every MEC (i.e., NE distance of two microgrids might be large in the short term but small in the long term, and vice-versa). Then, we cannot determine whether the number of microgrids in each MEC can increase or decrease as  $l$  increases in Fig. 2b and 2d. Furthermore, also in Fig. 2b and 2d, larger  $\xi$  would lead to a higher average NE (positive) and lower average NE (negative). This is because larger  $\xi$  (the threshold of NE distance) allows more microgrids to be clustered in every MEC.

Thirdly, we also measure the geo-locations of the microgrids in the MECs. On one hand, we have examined the (spatial) SSE of the discovered MECs by utilizing microgrid time series NE over different length of periods (different  $l$ ). As shown in Fig. 3a, for any  $l$ , larger  $\xi$  leads to higher SSE of MECs since microgrids in the same MEC would be less cohesive if more microgrids are clustered with a larger  $\xi$ . Meanwhile, larger  $l$  (more timestamps) results in lower SSE of MECs. This means less microgrids are clustered in each MEC as  $l$  increases. Indeed, this fact cannot be observed from Fig. 2b and 2d. Even if larger  $l$  gives more average number of microgrids in each MEC, since such mixed microgrids can have either positive or negative NE, more microgrids in each MEC do not necessarily make the NE of the MECs (positive case) higher nor make the NE of the MECs (negative case) lower. This matches the observations in Fig. 2b and 2d.

On the other hand, we fix  $\xi = 1$  and  $\xi' = 0.05$  in Algorithm 1, which then removes the constraint of NE distances and turns into a traditional agglomerative clustering problem based on geo-locations. Then, we compute the (spatial) SSE in the above case as the benchmark SSE ( $SSE_0$ ) and test how the spatial distances (SSE) within each MEC vary for different levels of balanced load (different  $\xi$ ). More specifically, we fix  $\xi' = 1$  (Algorithm 1 only specifies the maximum NE distance threshold  $\xi$  and removes the constraint of spatial distances), generate the

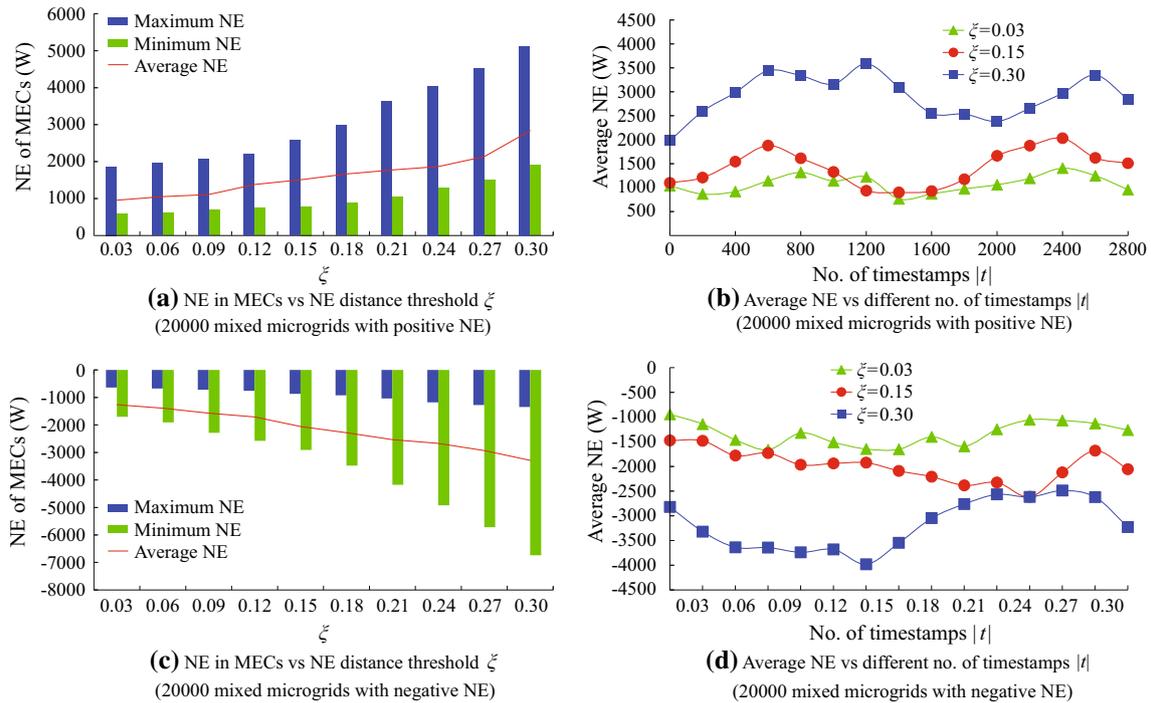


Fig. 2 MEC discovery

MECs with  $\zeta \in [0.03, 0.3]$  for two inputs “positive” and “negative”, respectively, and compute the corresponding (spatial) SSE for each MEC. Then, we define a new measure SSE ratio as  $\frac{SSE}{SSE_0}$  and plot it in Fig. 3b. Clearly, the (spatial) SSE increases as  $\zeta$  declines – an MEC with better balanced load includes the furthest microgrids from each other if the spatial distances within each MEC are not bounded (since  $\zeta' = 1$ ).

Finally, we let  $\theta = 0.0001$  per normalized distance of 0.1, randomly simulate five substations, and derive the average distance between each of the 50000 microgrids and its nearest substation. Then, we compare the overall load on transmission lines at 2880 timestamps for 50000 microgrids in two cases (with or without MECs). Table 1 also shows that such energy loss can be greatly reduced with MECs.

### 5.3 SEC discovery

We implement both the optimization-based approach and the two-phase algorithm to discover the SECs. For the optimization-based approach, we solve the optimization problem using the proposed Tabu Search [31] based algorithm (the length of Tabu list was set as  $S = 10$ ). If the algorithm cannot find a feasible solution within 10000 seconds, the algorithm will be terminated. As mentioned earlier, to compare the two approaches, we have generated a synthetic dataset for 10000 microgrids with mixed NE

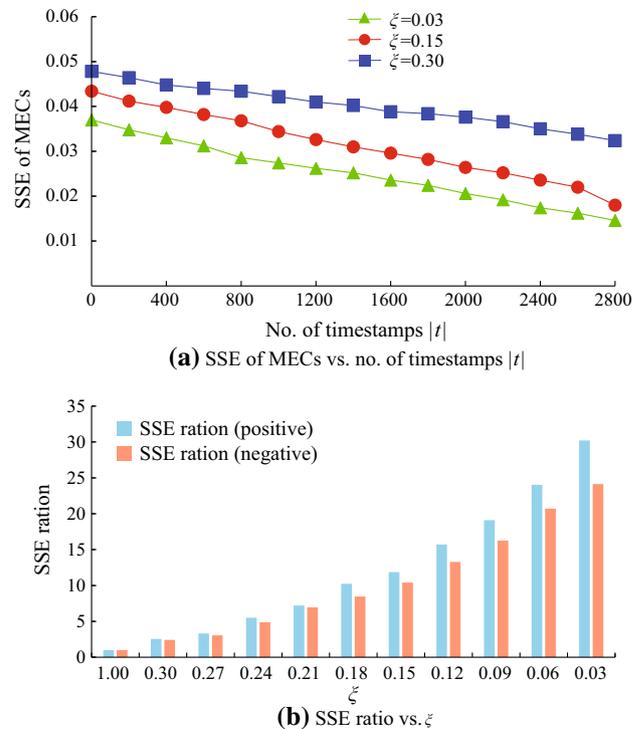


Fig. 3 Spatial SSE in MECs

(more microgrids with positive NE in  $[T_1, T_2]$ ). Tables 2 and 3 present the experimental results of these two approaches. We have the following observations.



**Table 1** Load on transmission lines (MEC discovery)

Load	With MECs	Without MECs
0.03	102795	3265520
0.12	291694	3265520
0.21	487679	3265520
0.30	720731	3265520

Firstly, both approaches are effective to discover SECs. Optimization-based approach can assign all the microgrids to the corresponding SECs as long as the all the constraints are satisfied. However, as a heuristic algorithm, when  $|t| \geq 900$ , the two-phase algorithm cannot involve all the microgrids in the SECs (feasible solution indeed exists as solved by the optimization-based approach). Among all the

microgrids, the two-phase algorithm has missed some microgrids with negative NE in  $[T_1, T_2]$  as  $|t| \geq 900$ . Then, the average NE of all the SECs discovered by the two-phase algorithm is greater than that of the optimization-based approach (as  $|t| \geq 900$ ).

Secondly, the SECs discovered by the optimization-based approach are more cohesive than that discovered by the two-phase algorithm (smaller SSE), since the optimization-based approach minimizes the SSE out of all the  $K$  values. In addition, we use  $K$ -means to simulate five substations of the main grid, and derive the average distance to the main grid (nearest substation) for the 10000 microgrids, which represents the average transmission distance (from the main grid to microgrids). Then, we find out that utilizing SECs for sharing local energy can significantly reduce the energy loss in the transmission, since SSE (the average transmission distance using SECs) is far

**Table 2** SEC discovery (optimization-based approach, 10000 microgrids)

No. of $ t $ timestamps	Average NE (all the SECs in $[T_1, T_2]$ )	No. of SECs: best $K$ (found by Tabu search)	No. of microgrids in all the SECs	Average no. of microgrids in the SECs	Microgrids (with positive NE at all times in $[T_1, T_2]$ )	Microgrids (with negative NE at any time in $[T_1, T_2]$ )	SSE (average transmission distance using SECs)	Average distance to main grid (if no SECs)
1	732	100	10000	100	6588	3412	0.097	0.247
300	704	110	10000	90.9	6588	3412	0.097	0.247
600	656	110	10000	90.9	6588	3412	0.097	0.247
900	621	120	10000	83.3	6588	3412	0.097	0.247
1200	587	120	10000	83.3	6588	3412	0.097	0.247
1500	543	130	10000	76.9	6588	3412	0.097	0.247
1800	488	130	10000	76.9	6588	3412	0.097	0.247
2100	432	130	10000	76.9	6588	3412	0.097	0.247
2400	381	140	10000	71.4	6588	3412	0.097	0.247
2700	324	140	10000	71.4	6588	3412	0.097	0.247

**Table 3** SEC discovery (two-phase algorithm, 10000 microgrids)

No. of $ t $ timestamps	Average NE (all the SECs in $[T_1, T_2]$ )	No. of SECs: best $K$ (found by $K$ -means)	No. of microgrids in all SECs	Average no. of microgrids in SECs	Microgrids (with positive NE at all times in $[T_1, T_2]$ )	Microgrids (with negative NE at any time in $[T_1, T_2]$ )	SSE (average transmission distance using SECs)	Average distance to main grid (if no SECs)
1	732	80	10000	125.0	6588	3412	0.124	0.247
300	704	80	10000	125.0	6588	3412	0.126	0.247
600	656	80	10000	125.0	6588	3412	0.129	0.247
900	681	80	9577	119.7	6588	3019	0.112	0.247
1200	717	80	9103	113.8	6588	2515	0.119	0.247
1500	743	80	8672	108.4	6588	2084	0.118	0.247
1800	748	80	8557	107.0	6588	1969	0.109	0.247
2100	758	80	8390	104.9	6588	1802	0.106	0.247
2400	774	80	8115	101.4	6588	1527	0.113	0.247
2700	789	80	8046	100.6	6588	1476	0.108	0.247

**Table 4** Load on transmission lines (SEC discovery)

Load	Optimization	Two-phase	Without MECs
1	345	498	2412
600	1401	1747	57649
1500	8815	9414	242705
2700	16765	18609	565504

less than the average distance to the main grid (0.097/0.108 vs. 0.247). Also, Table 4 shows that the load on transmission lines can be significantly reduced using the SECs discovered by both approaches.

Thirdly, for both approaches,  $K$  is selected as  $\{50, 60, \dots, 200\}$ , which is a reasonable set of values for 10000 microgrids (6588 microgrids in  $M^+$ ). Then, the average number of microgrids with positive NE in each community varies from 32.94 to 131.76. Tables 2 and 3 show that the optimization-based approach identifies more SECs than the two-phase algorithm. For any  $l$ , the number of SECs identified by the two-phase algorithm is fixed (since the best  $K$  is determined only by the microgrids' geo-locations with positive NE in  $[T_1, T_2]$ , in the first phase). However, the optimization-based approach may identify different numbers of SECs if different  $l$  are considered.

## 6 Conclusion and future work

Energy communities formed by distributed energy resources (microgrids) could facilitate the power grid to advance energy management and enable microgrids to find peer microgrids to cooperate (e.g., sharing/trading energy). In this paper, we have proposed a series of approaches to identify different energy communities for the microgrids such as mixed energy communities and self-sufficient energy communities. We have also validated the effectiveness and efficiency of the approaches using real-world spatial dataset as well as the power generation and consumption datasets.

In the future, we will investigate and solve some other variants of energy community discovery problems for microgrids and we will try to incorporate such preferences into the energy community discovery problems. In addition, besides integrating all the energy generation and consumption over a period into the MECs and SECs discovery, we will explore stochastic optimization models for energy community discovery based on the prediction of the future power generation and consumption, which is expected to improve the efficiency of the energy

community discovery algorithms. Finally, energy community discovery requests data collection from all the microgrids, which may compromise their privacy [36]. It is also interesting and challenging to propose privacy preserving energy community discovery techniques which enable the cooperation of microgrids while protecting their local information [5, 7].

**Acknowledgments** This work is partially supported by the National Science Foundation (NSF) (No. CNS-1745894) and the WISER ISFG grant. It is also partly sponsored by the Air Force Office of Scientific Research (AFOSR) (No. YIP FA9550-17-1-0240) and the Maryland Procurement Office (No. H98230-18-D-0007).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- [1] Goel S, Hong Y, Papakonstantinou V et al (2015) Smart grid security. Springer, New York
- [2] Roughan T (2013) Workshop on microgrid technologies and applications. National Grid
- [3] Lasseter R (2002) Microgrids. In: Proceeding of the IEEE PES winter meeting, New York, USA, 27–31 January 2002, 4 pp
- [4] Kennedy J, Ciuffo P, Agalgaonkar A (2012) Intelligent load management in microgrids. In: Proceeding of the IEEE PES general meeting, San Diego, USA, 22–26 July 2012, 8 pp
- [5] Xie S, Hong Y, Wan P (2019) A privacy preserving multiagent system for load balancing in the smart grid. In: Proceeding of the 2019 international conference on autonomous agents and multiagent systems, Montreal, Canada, 3 pp
- [6] Saad W, Han Z, Poor V et al (2012) Game-theoretic methods for the smart grid: an overview of microgrid systems, demand-side management, and smart grid communications. IEEE Signal Proc Mag 29(5):86–105
- [7] Hong Y, Wang H, Xie S et al (2018) Privacy preserving and collusion resistant energy sharing. In: Proceeding of 2018 IEEE international conference on acoustics, speech and signal processing, Alberta, Canada, 15–20 April 2018, 5 pp
- [8] Liang X, Li X, Lu R et al (2013) UDP: usage-based dynamic pricing with privacy preservation for smart grid. IEEE Trans Smart Grid 4(1):141–150
- [9] Carlisle N, Geet O, Pless S (2009) Definition of a “zero net energy” community. Technical report of National Renewable Energy Laboratory, Golden, USA
- [10] Hong Y, Goel S, Lu H et al (2017) Discovering energy communities for microgrids on the power grid. In: Proceedings of 2017 IEEE international conference on smart grid communications, Dresden, Germany, 23–27 October 2017, 7 pp
- [11] King M (2012) Community energy: planning, development and delivery. International District Energy Association, Massachusetts, USA
- [12] Ahmadi M (2013) Optimizing load control for a residential microgrid in a collaborative environment. Dissertations, University of Texas, Arlington



- [13] Sarker M, Ortega-Vazquez M, Kirschen D (2015) Optimal coordination and scheduling of demand response via monetary incentives. *IEEE Trans Smart Grid* 6(3):1341–1352
- [14] Sobe A, Elmenreich W (2013) Smart microgrids: overview and outlook. *CoRR*. <https://www.arxiv.org/abs/abs/1304.3944>. Accessed 25 Mar 2019
- [15] Erol-Kantarci M, Kantarci B, Mouftah H (2011) Reliable overlay topology design for the smart microgrid network. *IEEE Netw* 25(5):38–43
- [16] Dall'Anese E, Zhu H, Giannakis G (2013) Distributed optimal power flow for smart microgrids. *IEEE Trans Smart Grid* 4(3):1464–1475
- [17] Logenthiran T, Srinivasan D, Shun T (2012) Demand side management in smart grid using heuristic optimization. *IEEE Trans Smart Grid* 3(3):1244–1252
- [18] Amjady N, Keynia F, Zareipour H (2010) Short-term load forecast of microgrids by a new bilevel prediction strategy. *IEEE Trans Smart Grid* 1(3):286–294
- [19] Xu Y, Liu W (2011) Novel multiagent based load restoration algorithm for microgrids. *IEEE Trans Smart Grid* 2(1):152–161
- [20] Wang H, Huang J (2018) Incentivizing energy trading for interconnected microgrids. *IEEE Trans Smart Grid* 9(4):2647–2657
- [21] Zhang C, Wu J, Zhou Y et al (2018) Peer-to-Peer energy trading in a Microgrid. *Appl Energy* 220:1–12
- [22] Maknouninejad A, Qu Z (2014) Realizing unified microgrid voltage profile and loss minimization: a cooperative distributed optimization and control approach. *IEEE Trans Smart Grid* 5(4):1621–1630
- [23] Asr N, Ojha U, Zhang Z et al (2014) Incremental welfare consensus algorithm for cooperative distributed generation/demand response in smart grid. *IEEE Trans Smart Grid* 5(6):2836–2845
- [24] Fioretto F, Yeoh W, Pontelli E et al (2017) A distributed constraint optimization (DCOP) approach to the economic dispatch with demand response. In: *Proceedings of the 16th conference on autonomous agents and multi-agent systems, São Paulo, Brazil, 8–12 May 2017*, 9 pp
- [25] Hussain A, Bui V, Kim H (2018) A resilient and privacy-preserving energy management strategy for networked microgrids. *IEEE Trans Smart Grid* 9(3):2127–2139
- [26] Hong Y, Goel S, Liu W (2016) An efficient and privacy-preserving scheme for P2P energy exchange among smart microgrids. *Int J Energy Res* 40(3):313–331
- [27] Yang J, McAuley J, Leskovec J (2013) Community detection in networks with node attributes. In: *Proceeding of the 2013 IEEE international conference on data mining, Dallas, USA, 7–10 December 2013*, 6 pp
- [28] Coscia M, Giannotti F, Pedreschi D (2011) A classification for community discovery methods in complex networks. *Stat Anal Data Min* 4(5):512–546
- [29] Arboleya P, Gonzalez-Moran C, Coto M et al (2015) Efficient energy management in smart micro-grids: ZERO grid impact buildings. *IEEE Trans Smart Grid* 6(2):1055–1063
- [30] Tan P, Steinbach M, Kumar V (2005) *Introduction to data mining*. Pearson Education, London
- [31] Glover F (1990) Tabu search: a tutorial interfaces. *J Appl Anal* 20(4):74–94
- [32] MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: *Proceedings of Berkeley symposium on mathematical statistics and probability, University of California, Berkeley, California*. <https://www.projecteuclid.org/euclid.bmsp/1200512992>. Accessed 25 Mar 2019
- [33] *Traveling Salesman Problem* (2005) <http://www.math.uwaterloo.ca/tsp/>. Accessed May 2005
- [34] Richardson I, Thomson M, Infield D et al (2010) Domestic electricity use: a high-resolution energy demand model. *Energy Build* 42(10):1878–1887
- [35] Barker S, Mishra A, Irwin D et al (2012) Smart\*: an open data set and tools for enabling research in sustainable homes. In: *Proceedings of 2012 Workshop on data mining applications in sustainability, Beijing, China*, 5 pp
- [36] Hong Y, Liu W, Wang L (2017) Privacy preserving smart meter streaming against information leakage of appliance status. *IEEE Trans Inf Forensics Secur* 12(9):2227–2241

**Shangyu XIE** is a Ph.D. student at Illinois Institute of Technology, USA. He received his Bachelor degree in Electrical Engineering from Shanghai Jiao Tong University, China. His research interests are data privacy and security.

**Han WANG** is a Ph.D. student at Illinois Institute of Technology, USA. She received her M.Sc. degree in Software Engineering from Huazhong University of Science and Technology, China. Her research interests are data privacy and security.

**Shengbin WANG** is an Associate Professor at the College of New Jersey, USA. He received his Ph.D. degree from Rutgers University, USA. His primary research interests include supply chain operations modeling, optimization, and algorithm design.

**Haibing LU** is an Associate Professor at Santa Clara University, USA. He received his Ph.D. degree from Rutgers University, USA. His research interests are the confluence of privacy/security, data analytics, and information systems.

**Yuan HONG** is an Assistant Professor at Illinois Institute of Technology, USA. He received his Ph.D. degree from Rutgers University, USA. His research interests include data privacy, security and optimization. He is a senior member of the IEEE.

**Dong JIN** is an Assistant Professor at Illinois Institute of Technology, USA. He obtained his Ph.D. degree from the University of Illinois at Urbana-Champaign, USA. His research interests include modeling and simulation of large-scale systems, smart grid security, software defined network, and cyber-security.

**Qi LIU** is an Assistant Professor at the University of Rhode Island, USA. She received her Ph.D. degree from Rutgers University, USA. Her research interests include data analytics, fraud detection, and auditing.