The Financial Benefit of Energy Consumption Behavior Diversity Within an Energy Community

Joey de Water
Supervisor(s): Dr. Valentin Robu, Sho Cremers, Peter Zhang
EEMCS, Delft University of Technology, The Netherlands
22-6-2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology, In Partial Fulfilment of the Requirements For the Bachelor of Computer Science and Engineering

Abstract

Energy communities are not yet fully self sufficient, mostly due to financial factors. Efforts are made to reduce these factors. Communities invest in community-owned assets, which provide more savings compared to individually-owned assets. Prosumers share their loads for better energy distribution, this can provide a significant impact. A good predictor for identifying the financial benefit for a community is the diversity of the consumption behaviors of the prosumers. However, an open question is how the diversity exactly affects the community costs. In this paper, we introduce Two-level K-means, an improvement on K-means, and use it on real consumption data to find energy profiles. We use the energy profiles to model communities, varying in diversity. Finally, we provide an analysis of the affects of diversity on costs. Results from the analysis show that an increased diversity factor can provide financial benefit. This is a result of residual demand being compensated by excess energy generated. However, the added financial benefit depends on the composition of the community.

1 Introduction

Interest in renewable energy generation has rapidly risen, following the worldwide effort to fight climate change. With this growing interest came a lot of improvements regarding distributed and decentralized energy systems, i.e. systems with little or no access to a central power grid. These developments created a shift from the larger companies to prosumers, individuals that both consume and produce [Norbu et al., 2021].

For a while, governments of developed countries offered incentives, e.g. feed-in-tariffs, for individuals to adopt distributed energy sources, but soon it became clear that these sources were financially unsustainable to support in the long-term [Nolden, 2015]. As a result, the governments dropped these incentives. Prosumers now wanted to improve their own consumption and production, to minimize their dependency on the central grid.

Over the last few years, more and more energy communities, made up of a number of prosumers, came into existence worldwide. Governments in the UK and EU deploy these communities to support renewable energy generation [Nolden, 2015; Commission, 2022]. In certain developing countries, where there is little to no access to a central power grid, energy communities help people gain access to electricity. Communities in sub-Saharan Africa and even some in South Asia are examples.

Currently, still a lot of improvements are made concerning energy communities, since they are not yet fully self-sufficient. This is mostly due to financial factors. Efforts are being made to reduce these factors. So a lot of progress has been made in improving different types of energy generation, e.g. windmills and solar panels, as well as storage, e.g. new types of batteries. Cooperation within energy communities can provide significant impact, due to prosumers combining their loads [Norbu et al., 2021]. Intuitively, this makes sense due to different consumption behaviors allowing for a better distribution of energy. An example would be two consumers, one consumes most energy during the morning and the other consumes most energy during the evening. This consumption behavior distribution is more even, resulting in more balanced consumption overall for the community. Thus, a good predictor for identifying the financial benefit for a community will be the diversity of the consumption behaviors of the prosumers. However, an open question is how the diversity exactly affects the community costs.

In this paper, we use a widespread technique used in energy demand modeling, i.e. clustering [Jeong et al., 2021], to segment consumption profiles into different classes. Earlier works often use historical consumption profiles to cluster, but depending on the purpose different clustering models perform better [Bidoki et al., 2011] [Xu and Tian, 2015] [Fränti and Sieranoja, 2018]. For our purposes we want clusters in which prosumers' consumption profiles are somewhat similar, so we can control the diversity of the community. Then we use the distribution of these clusters to model multiple communities varying in diversity. Finally, we will use these communities to provide an empirical analysis on how the diversity impacts costs.

The remainder of the paper is as follows: a discussion of related work in this field is presented in section 2. The prosumer community model and clustering method are described in section 3. In section 4 we provide an empirical analysis of the impact of diversity on costs for a large real life dataset. We address the ethical implications of our research in section 5. Ending with the conclusion in section 6.

2 Related Work

The different aspects concerning energy communities, clustering and energy consumption profiles this paper touches upon, are discussed in the paragraphs below.

2.1 Clustering

In [Xu and Tian, 2015], different clustering methods are reviewed. They look at 19 different clustering methods, most notably are K-means, Affinity Propagation and OPTICS, and identify their strengths and weaknesses. In [Bidoki et al., 2011], different clustering methods used for classifying energy consumption profiles are compared and analysed. From the analysed clustering methods, we use K-means to classify energy consumption profiles. In [Fränti and Sieranoja, 2018], the performance of K-means on different datasets is analysed. They find one of the biggest weaknesses of K-means to be incorrect clustering with unbalanced datasets. They briefly consider two improvements to K-means, i.e. better initialization and Repeated K-means. We plan on introducing an improvement to K-means, inspired by Repeated K-means, to address this weakness. In [Bubeck and Von Luxburg, 2007], overfitting and how to avoid it is discussed. Overfitting is another possible weakness of K-means, that we want to avoid during the experiment. In [Kodinariya and Makwana, 2013], different methods for finding the optimal number of clusters in K-means are reviewed. We use the reviewed elbow method during the experiment.

2.2 Energy Community and Energy Consumption

In [Bayliss and Hardy, 2012], the diversity factor is explained and the formula for how to calculate it is given. We need this to compute how diverse consumption behavior in an energy community is. In [Jeong et al., 2021], energy consumption patterns of residential customers are clustered using K-means. The resulting cluster profiles are named according to the shape of consumption behavior. We use the introduced naming scheme for the corresponding profiles we find.

3 Methodology

The community model, community cost, diversity factor, sampling, dataset filtering and clustering, are discussed in the following paragraphs.

3.1 Prosumer Community Modelling

We model an energy community according to the model introduced in [Norbu et al., 2021], consisting of 200 prosumers. We use this model for our experiment, because of the introduced financial benefits. The prosumers, who generate, store, consume and trade energy within the community, collectively own a wind turbine and a battery. Sharing a wind turbine and a battery achieves a lower annual cost because both require a lower optimal capacity for the same services.

Community cost

[Norbu et al., 2021] defines the annual cost c_N for the community N as follows:

$$c_N(T) = \sum_{t=1}^{T} c_i(t) - \sum_{t=1}^{T} v_e(t) + c_d(T) \tag{1}$$

where $c_i(t)$ and $v_e(t)$ are the import cost of energy from the grid and export revenue of energy to the grid by the whole community at time t respectively, and $c_d(T)$ is the deprecation cost of community-owned assets, i.e. wind turbine and battery, for the period T.

Diversity factor

The diversity factor DF shows how diverse energy demands are within a community N. As described in [Bayliss and Hardy, 2012], It is calculated using the following formula:

$$DF = \frac{\sum_{i \in N} \max(d_i)}{\max(D)} \tag{2}$$

where d_i is the demand data from prosumer i and D is the aggregated demand data of the community.

Sampling

We sample from consumption profiles to model an energy community, according to the model discussed above. Then, we calculate the diversity factor and the corresponding cost on the community models, varying in diversity, to find how the diversity factor affects the community cost.

Dataset filtering

During some energy community projects, e.g. [Networks, 2017] and [Networks, 2015], consumption readings are taken at certain intervals over a certain amount of time. For our research, all this demand data needs to be filtered since we want to find the data that is most interesting to us. The most important data is the consumption behavior of prosumers on an average work day.

3.2 Finding Unique Consumption Profiles

A key challenge in classifying diversity is identifying consumption behavior. So, a proven method, used in research [Kodinariya and Makwana, 2013], is clustering to identify this behavior. Here, clustering means the dividing of data points, representing consumption, into a number of clusters such that the data points in the same cluster are more similar to each other than they are to data points in another cluster.

K-means

In this paper, we use the K-means algorithm for clustering. K-means is an unsupervised learning algorithm [Kodinariya and Makwana, 2013]. The algorithm divides data into K, a pre-defined number of, clusters. It works as follows:

1. Specify the number of clusters K.
2. Select K data points randomly as the initial cluster centers.
3. Assign each data point to the closest cluster center by calculating the euclidean distance between the data point and all centers.
4. Find the new centers by calculating the average of all data points that belong to each cluster, i.e. the Euclidean distance.
5. Repeat steps 3 - 4 until the cluster centers no longer move.

The consumption profiles are clustered by minimizing the following objective function:

$$f = \sum_{i=1}^{K} \sum_{\mathbf{x} \in g_i} |\mathbf{x} - \mu_i|^2 \tag{3}$$

where g_i is the i-th cluster, $\mathbf{x}$ is the data point in g_i, μ_i is the center of g_i and K is the number of clusters. The distance from data points to the cluster centers is measured by the Euclidean distance:

$$dist(\mathbf{x}, \mu_i) = |\mathbf{x} - \mu_i|^2 \tag{4}$$

K-means improvement: Two-level K-means

One of the biggest weaknesses of K-means is incorrect clustering with unbalanced datasets [Fränti and Sieranoja, 2018]. This becomes a problem when applying K-means clustering to energy consumption profiles, since it is very common for the majority of prosumers to display very similar consumption behavior and only a small minority displaying a significant deviation [Bidoki et al., 2011]. To address this problem, we introduce an improvement to K-means, inspired by Repeated K-means introduced in [Fränti and Sieranoja, 2018]. Namely, two-level K-means, which consists out of two phases. First, we apply K-means to identify representatives by overestimating K. Then, we apply K-means again on the identified representatives to identify the unique clusters.

A weakness that remains is the possibility of overfitting [Bubeck and Von Luxburg, 2007]. Overfitting is the idea that a model is fitted exactly or too closely to a dataset, e.g. taking $K = n$, where n is the number of datapoints.

Finding the optimal number of clusters K

Now we have a method for identifying the different consumption behaviors, we need to find a value for the optimal number of clusters K. We will use the elbow method to find this value.

Elbow method

The elbow method is a visual method, since we have to look at the elbow plot to find the optimal K value. As described in [Kodinariya and Makwana, 2013], the idea is to run the K-means algorithm on different values of K, starting with $K = 2$, increasing K by 1 at each step. At every time step, we calculate the inertia, i.e. the average distance of each data point to their closest cluster center. For a number of values of K the inertia drops significantly. At some value for K the inertia starts decreasing in a linear fashion when K is increased further, i.e. the inflexion point. This is the optimal K value.

4 Experimental Setup and Results

In the following section the experimental setup and results are discussed.

4.1 Dataset

For this experiment, we are using a collection of energy demands of households connected to a smart grid during a trial, i.e. Low Carbon London [Networks, 2015]. The dataset contains the energy demands of 5567 households over a timespan of 2.5 years, recorded every 30 minutes. The data collection has been performed by UK power networks.

Dataset filtering

We filter the dataset to find data that represents the average day. This means we have to filter out the days that are different from the average day. The first days we filter on are weekdays. This means we filter out holidays and weekends, since on these days individuals do not have to go to work. We count Friday as weekend since not every individual has to work on that day. This leaves us with data for Monday to Thursday for the entire timespan of 2.5 years.

Now we filter on are specific calendar days, i.e. seasons. Seasons have a lot of influence on consumption behavior. During summer, when the temperature is high, most individuals consume less energy. During winter, when the temperature is low, most individuals consume more energy, e.g. for electrical heating. Since we are interested in energy consumption, we choose to filter on winter days. This leaves us with demand data on winter-weekdays.

4.2 Unique Consumption Profiles

To find the consumption profiles, we use the K-means algorithm. Before we apply this algorithm on the data, we first

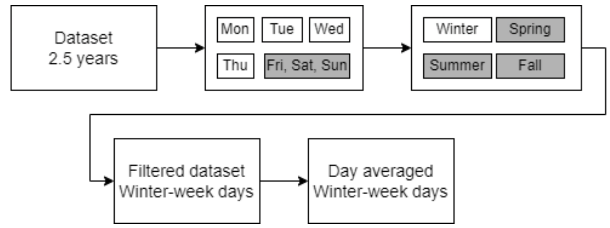


Figure 1: Dataset filtering

normalize the data, because we are interested in the consumption patterns, not the consumption amounts.

Finding the optimal K for K-means

Now we need to find the optimal number of clusters K. We perform the elbow method on the data. In figure 2, we see the elbow value to be between 8 and 12. We do not overestimate

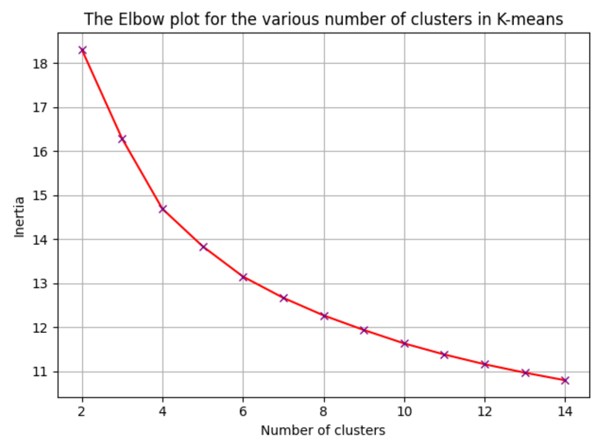


Figure 2: Elbow method on normalized data

$K = 12$, since we want to avoid overfitting K [Bubeck and Von Luxburg, 2007]. Instead, We overestimate $K = 10$.

K-means

Using the overestimated value $K = 10$, we find 10 clusters. We show the cluster means in figure 3. We find clusters 0 and 2 which follow the same pattern as evening peak consumers, 3 and 8 which follow the same pattern as energy saving consumers, 4 and 9 which follow the same pattern as morning peak consumers and 5 and 6 which follow the same pattern as day working consumers. The cluster sizes can be found in table 1 and the individual cluster means can be found in appendix A. We have similar profiles, thus we continue with the second phase of our method, two-level K-means.

Finding the optimal K for Two-level K-means

Now we perform the elbow method for two-level K-means. In figure 4, we see the elbow value to be 6.

Two-level K-means

Using $K = 6$, we now find 6 unique consumption profiles. See figure 5. We can classify the profiles, following the naming introduced in [Jeong et al., 2021]. We show the name and size of each found profile in table 2.

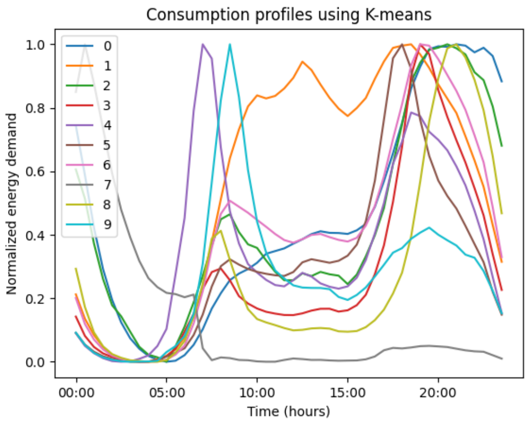


Figure 3: K-means

Profile	Cluster(s)	Combined size
Evening peak	0, 2	1577
Home working	1	811
Energy saving	3, 8	773
Morning peak	4, 9	461
Day working	5, 6	1582
Night peak	7	47

Table 1: K-means clusters

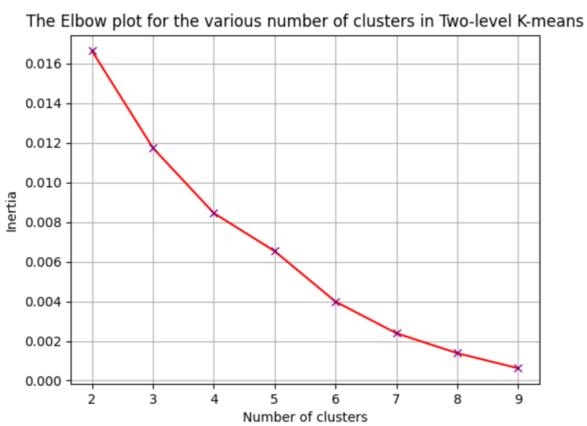


Figure 4: Elbow method Two-level K-means

Cluster	Profile	Size
0	Day Worker	2827
1	Owl	47
2	Morning Peak	151
3	M-pattern	310
4	Evening Peak	1577
5	Energy Saving	339

Table 2: Two-level K-means clusters

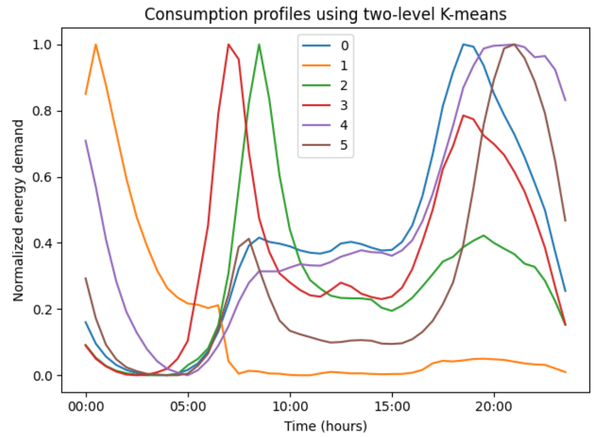


Figure 5: Two-level K-means

The Day Worker profile, as can be found in figure 6a, shows the consumption behavior of the average consumer that works during the day. The consumer uses energy in the morning before going to work and in the evening after coming back home. In the evening energy consumption peaks, most likely due to lower temperature and more use of electronic devices.

The Owl profile, as shown in figure 6b, shows the consumption behavior of a consumer that consumes energy as an owl. Just like an owl the consumer uses energy from night to daybreak.

The Morning Peak profile, as presented in figure 6c, is almost the same as the Day Worker profile. It differs from the Day Worker profile in where the peak is located. Instead of in the evening, energy consumption peaks in the morning.

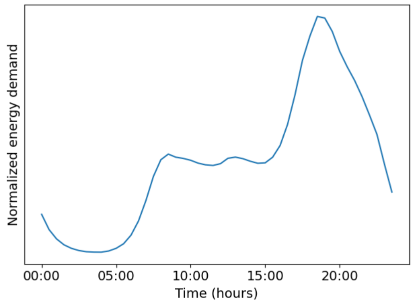
The M-pattern profile, as can be seen in figure 6d, shows the consumption behavior that has is M shaped.

The Evening Peak profile, as shown in figure 6e, shows consumption behavior of a consumer that works during the day, but keeps consuming energy during the night longer than a Day Worker.

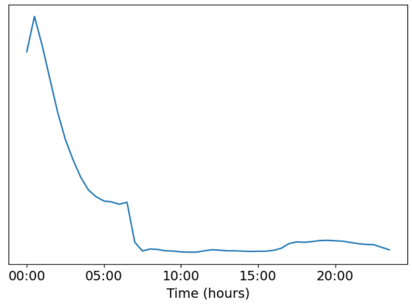
The Energy Saving profile, as can be found in figure 6f, shows consumption behavior of a consumer that consumes less energy during the afternoon.

4.3 Community Modelling

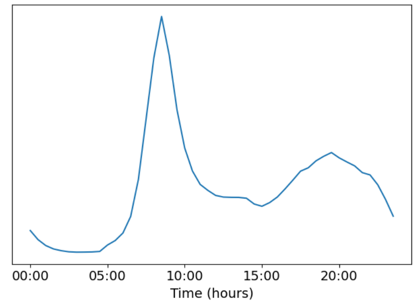
We sample from the consumption profiles we found to find the influence of the diversity factor on the cost. We model different communities for varying diversities. We do this by modelling a community, as described in 3.1, of 200 prosumers who have the exact same consumption behavior. This community has a diversity factor of exactly 1. So, to increase the diversity factor slightly, we introduce a different consumption behavior by replacing one of the 200 prosumers with another consumption profile. This results in a diversity factor higher than 1. We do this iteratively until all the prosumers have been replaced, to keep increasing the diversity factor. For example, we model a community of 200 Day Workers, with a diversity factor of 1.000. Now, we replace one of the 200 Day Workers with 1 Owl. This results in a community of 199 Day Workers and 1 Owl with a diversity



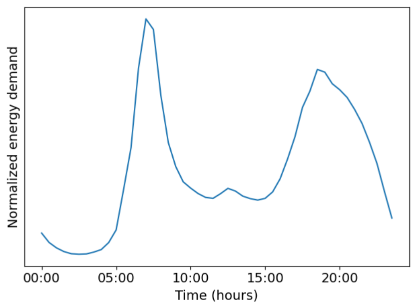
(a) Day Worker



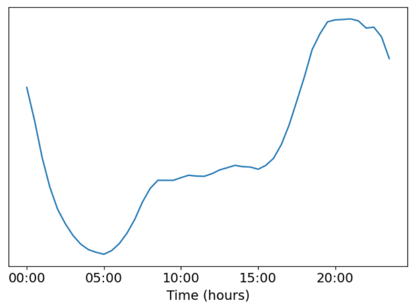
(b) Owl



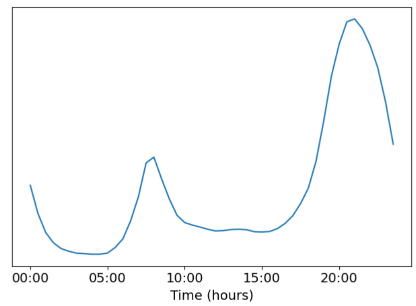
(c) Morning Peak



(d) M-pattern



(e) Evening Peak



(f) Energy Saving

Figure 6: 6 unique consumption profiles

factor of 1.001. However, the change in diversity and the effect of the diversity factor is very volatile, since it depends on the exact composition of the community. So, to discover a general pattern, we repeat this experiment with 6 different starting communities, representing every found profile.

4.4 Influence of the Diversity Factor on the Cost

Having done the modelling, we can now see how the diversity factor affects the cost. In figure 7, we can see that a diversity factor of 1.00 has a mean energy cost of around £ 0.091 per kWh and a diversity factor of 1.30 a mean energy cost of around £ 0.088 per kWh. This reduction in costs is a result of residual demand being compensated by excess energy generated by the community. For communities with a low diversity factor, the effect of a single prosumer that can compensate the residual demand is higher since there are many periods of time where residual demand can be compensated by a prosumer with a different consumption profile. However, for communities with a high diversity factor, most of the residual demand is already being compensated by other prosumers in the community. In such communities, the added value of a prosumer with different consumption behavior is small since it is less likely that the prosumer's excess energy and the communities' residual demands align. Even if such a prosumer's excess generation could compensate the residual demands perfectly, most of the cost benefits due to diversity are already achieved by other prosumers. This can be seen in figure 7 at a diversity factor of 1.15. From this value forward the added financial benefit becomes lower.

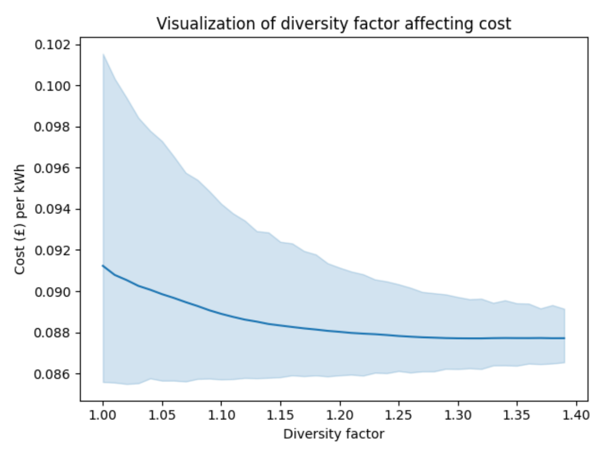


Figure 7: Community costs with growing diversity factor

5 Responsible Research

The experiment is performed as described in section 3. Every performed step can be reproduced following its description. Furthermore, the research is performed in a thoughtful manner. We have presented work, data and ideas of others with the appropriate medium of presentation, i.e. citations. All cited work can be found in the reference list. No data is fabricated or falsified and all relevant observations are reported. Finally, there is no conflict of interest.

6 Conclusions

In this paper, we investigated the influence of energy consumption behavior diversity on community energy costs. We discussed using K-means to identify consumption profiles from the Low Carbon London dataset [Networks, 2015], which contains energy demands of 5567 households over a timespan of 2.5 years. The data was filtered on the winter-weekdays and normalized. We introduced a novel improvement to K-means to address one of the biggest weaknesses of K-means, i.e. incorrect clustering with unbalanced datasets, namely Two-level K-means. We found unique consumption profiles by clustering representatives from the first phase of Two-level K-means. We used the found consumption profiles to model energy communities, according to the model introduced in [Norbu et al., 2021]. Finally, we analysed the influence of the diversity factor on the community cost.

Results from the analysis show that an increased diversity factor can provide financial benefit. This is a result of residual demand being compensated by excess energy generated. The added financial benefit depends on the composition of the community. For a community with a low diversity factor, the added value of a prosumer with deviating consumption behavior is high. For a community with a high diversity factor, the added value is small since it is less likely the prosumer's excess energy and the communities' residual demands align. Even if such a prosumer's excess energy generation could compensate the residual demands perfectly, almost all of the cost benefits due to diversity are already achieved by other prosumers.

In future work, our experiment can be extended with different clustering methods to find all the different unique clusters. Other clustering methods could find more or different unique clusters. Different energy community models and sampling methods are other directions which can be promising. Another interesting question is what the best community composition is based on the most common and unique consumption profiles. Finally, this experiment can be performed on other datasets, e.g. energy community projects in other environments.

References

[Bayliss and Hardy, 2012] Bayliss, C. and Hardy, B. (2012). Chapter 23 - distribution planning. In Bayliss, C. and Hardy, B., editors, *Transmission and Distribution Electrical Engineering (Fourth Edition)*, pages 939–985. Newnes, Oxford, fourth edition edition.

[Bidoki et al., 2011] Bidoki, S. M., Mahmoudi-Kohan, N., and Gerami, S. (2011). Comparison of several clustering methods in the case of electrical load curves classification. In *16th Electrical Power Distribution Conference*, pages 1–7.

[Bubeck and Von Luxburg, 2007] Bubeck, S. and Von Luxburg, U. (2007). Overfitting of clustering and how to avoid it. *Preprint*, pages 1–39.

[Commission, 2022] Commission, E. (2022). Energy communities. accessed at 06/05/2022.

[Fränti and Sieranoja, 2018] Fränti, P. and Sieranoja, S. (2018). K-means properties on six clustering benchmark datasets. *Applied Intelligence*, 48(12):4743–4759. Copyright - Applied Intelligence is a copyright of Springer, (2018). All Rights Reserved; Last updated - 2018-11-08.

[Jeong et al., 2021] Jeong, H. C., Jang, M., Kim, T., and Joo, S.-K. (2021). Clustering of load profiles of residential customers using extreme points and demographic characteristics. *Electronics*, 10(3).

[Kodinariya and Makwana, 2013] Kodinariya, T. M. and Makwana, P. R. (2013). Review on determining number of cluster in k-means clustering. *International Journal*, 1(6):90–95.

[Networks, 2017] Networks, S. . S. E. (2017). Thames valley vision.

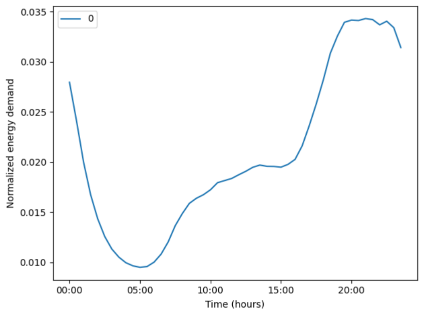
[Networks, 2015] Networks, U. P. (2015). Smartmeter energy consumption data in london households.

[Nolden, 2015] Nolden, C. (2015). *Performance and Impact of the Feed-in Tariff Scheme: Review of Evidence*. UK Department of Energy and Climate Change (DECC).

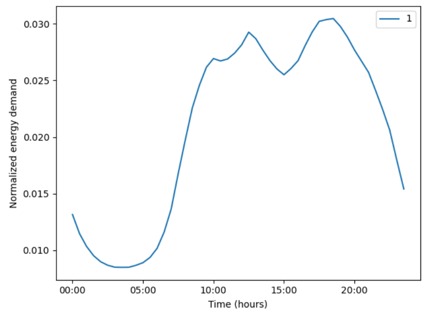
[Norbu et al., 2021] Norbu, S., Couraud, B., Robu, V., Andoni, M., and Flynn, D. (2021). Modelling the redistribution of benefits from joint investments in community energy projects. *Applied Energy*, 287:116575.

[Xu and Tian, 2015] Xu, D. and Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2):165–193.

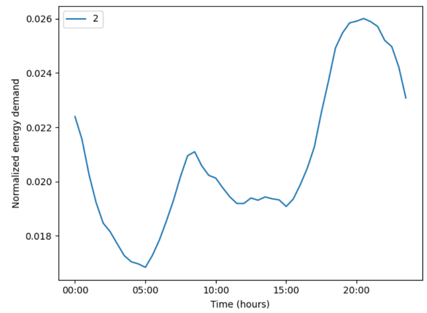
A K-means: Overestimated Cluster Means



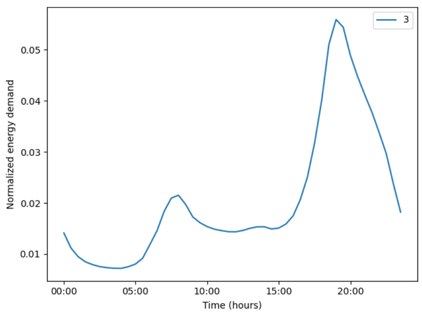
(a) K-means profile 0



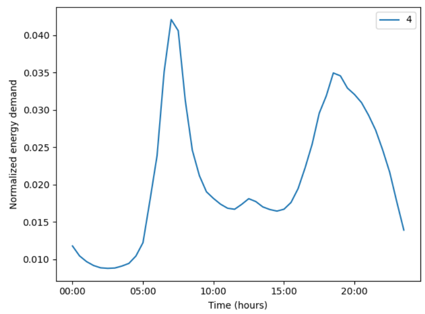
(b) K-means profile 1



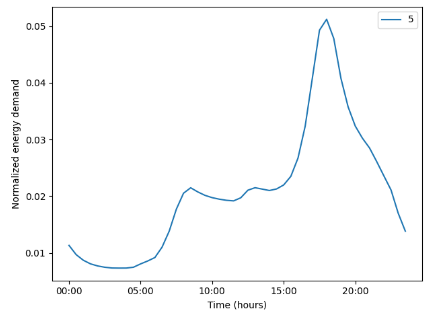
(c) K-means profile 2



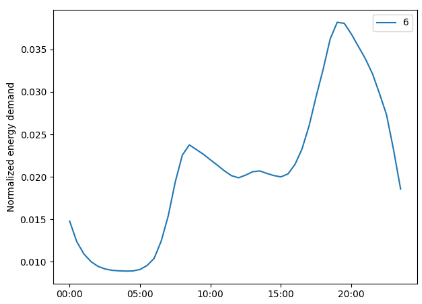
(d) K-means profile 3



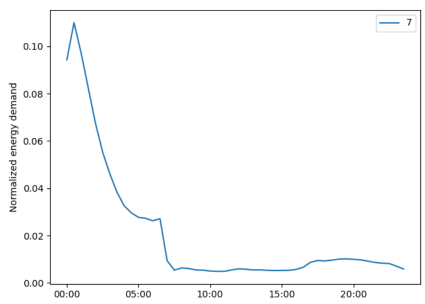
(e) K-means profile 4



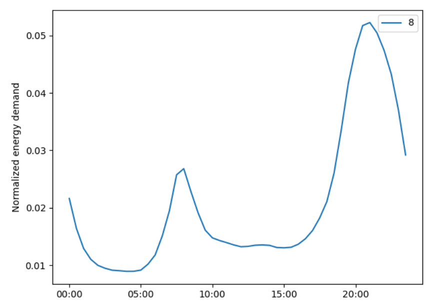
(f) K-means profile 5



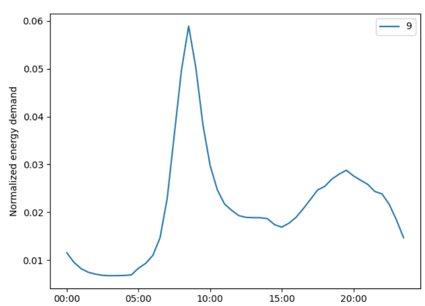
(g) K-means profile 6



(h) K-means profile 7



(i) K-means profile 8



(j) K-means profile 9

Figure 8: 10 representatives