Efficient Methods for Approximating the Shapley Value for Asset Sharing in Energy Communities

Sho Cremers
CWI, National Centre for Mathematics and Computer Science
Amsterdam, The Netherlands
sho.cremers@cwi.nl

Valentin Robu
CWI, National Centre for Mathematics and Computer Science
Amsterdam, The Netherlands
valentin.robu@cwi.nl

Daan Hofman
Delft University of Technology
Delft, The Netherlands
d.hofman@student.tudelft.nl

Titus Naber
Delft University of Technology
Delft, The Netherlands
t.t.naber@student.tudelft.nl

Kawin Zheng
Delft University of Technology
Delft, The Netherlands
k.zheng-2@student.tudelft.nl

Sonam Norbu
Heriot-Watt University
Edinburgh, UK
sn51@hw.ac.uk

ABSTRACT

With the emergence of energy communities, where a number of prosumers invest in shared renewable generation capacity and battery storage, the issue of fair allocation of benefits and costs has become increasingly important. The Shapley value has attracted increasing interest for redistribution in energy settings - however, computing it exactly is intractable beyond a few dozen prosumers. In this paper, we examine a number of methods for approximating the Shapley value in realistic community energy settings, and propose a new one. To compare the performances of these methods, we also design a novel method to compute the Shapley value exactly, for communities of up to several hundred agents by clustering consumers into a smaller number of demand profiles. We compare the methods in a large-scale case study of a community of up to 200 household consumers in the UK, and show that our method can achieve very close redistribution to the exact Shapley values but at a much lower (and practically feasible) computation cost.

CCS CONCEPTS

• **Hardware** → *Smart grid*; • **Theory of computation** → **Algorithmic game theory**; *Approximation algorithms analysis.*

KEYWORDS

energy community, shared assets, Shapley value, approximation, stratification

ACM Reference Format:

Sho Cremers, Valentin Robu, Daan Hofman, Titus Naber, Kawin Zheng, and Sonam Norbu. 2022. Efficient Methods for Approximating the Shapley Value for Asset Sharing in Energy Communities. In *The Thirteenth ACM International Conference on Future Energy Systems (e-Energy '22), June 28-July 1, 2022, Virtual Event, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3538637.3538861

e-Energy '22, June 28-July 1, 2022, Virtual Event, USA

ACM ISBN 978-1-4503-9397-3/22/06.
https://doi.org/10.1145/3538637.3538861

1 INTRODUCTION

Recent years have seen a shift towards decentralised energy systems, where prosumers (consumers with their own local renewable generation capacity and storage) satisfy more of their own energy needs from renewable energy generated from local sources. This led to the renewed interest in *energy communities*, where group of prosumers buy a shared generation resource (such as wind turbine) or a community battery. Such communities are often located on the low-voltage (LV) network behind a local transformer, interacting with the utility provider as a single entity. The prosumers share the outputs of jointly-owned energy assets, as well as the energy bill for the aggregate residual demand. Clearly, one key challenge in this setting is the redistribution of costs and benefits in a *fair* way.

Coalitional game theory has long studied such redistribution problems in a wide variety of systems [1]. A key concept is the Shapley value [13], which has recently received substantial attention in energy applications. However, a key challenge is that its computation is intractable beyond a small number of agents.

Prior papers dealt with this in several ways. Most consider experimental models with up to ~20 agents to keep computations tractable [2, 12, 14]. Yet, realistically sized energy communities have more members, e.g., there are usually 50-200 consumers behind a substation/LV transformer in Europe [11], or potentially even more sharing an asset such as a large community wind turbine.

Another approach is to approximate Shapley value by sampling. Sampling-based approaches do have merit, and in this paper, we implemented the most advanced sampling-based method we are aware of, that of O'Brien et al. [10]. However, time complexity is still an issue; the larger the setting, the more samples are needed for a reasonable approximation. Hence, new methods to approximate the Shapley value in larger settings are clearly needed. As one of the contributions of this paper, we introduce a novel redistribution method that approximates the Shapley value well within polynomial time, and compare it to existing methods.

Determining the "ground truth", i.e. computing the exact Shapley value, to compare methods still remains a challenge. Prior approaches, like O'Brien et al. [10], use a setting of only 20 agents to compute the true Shapley value, as it becomes infeasible for larger communities. Yet, an approximation method that does poorly for a small number of agents may actually do well for a realistically sized setting of 100-200 agents. Another important contribution of this paper is that we develop a method to efficiently compute the *exact* Shapley value for larger communities by clustering the agents in a much smaller number of consumption profiles.

Finally, as part of our contributions, we implemented our method in a realistic community case study, both in terms of demand, generation and battery data used, and in terms of size, granularity and duration. We used the community model in Norbu et al. [9], which draws data from the Thames Valley vision trial and provides a highly realistic case study to provide confidence in the robustness of our experimental comparison results.

2 COMMUNITY ENERGY MODEL

Community model. An energy community $\mathcal{N}$ consists of a set of $|\mathcal{N}| = N$ prosumers, a shared battery and renewable energy source (RES). In this study, a lithium-ion battery and Enercon E-33 wind turbines [3] were considered as the community's energy storage system and RES, respectively. Each prosumer i in the community has a half-hourly power demand profile represented as $d_i(t)$ at time step t. The final time step of the operation is denoted as T. In this study, the data contains half-hourly demands and generation during a 1 year period, and hence $T = 365 \times 48 = 17520$. The demand of the community at time t, $d_{\mathcal{N}}(t)$, is the sum of the demands of the agents in the community at t. Furthermore, a community has a generated power by the jointly owned local RES, $g(t)$, and the power of the battery, $p^{\text{bat}}(t)$, which is negative when charging and positive when discharging. Finally, a community buys power from the utility grid if the assets do not meet the power demand, else it sells any excess power generated to the grid.

Community cost calculation. The total cost of the community during time $\{1, ..., T\}$, $c_T(\mathcal{N})$ is the sum of the three components:

$$c_T(\mathcal{N}) = c_T^{\text{grid}}(\mathcal{N}) + c_T^{\text{wind}}(\mathcal{N}) + c_T^{\text{bat}}(\mathcal{N}) \tag{1}$$

where $c_T^{\text{grid}}(\mathcal{N})$ is the cost of importing energy from the grid, determined by the battery control algorithm by Norbu et al. [9] given the inputs $d_{\mathcal{N}}(t)$, $g(t)$, and $p^{\text{bat}}(t-1)$. The variables $c_T^{\text{wind}}(\mathcal{N})$ and $c_T^{\text{bat}}(\mathcal{N})$ represent the costs of the community wind turbine and battery respectively, and are calculated similarly to [9]. Maximum capacities of the wind turbine and the battery both increase linearly with the community size, hence the costs also increase linearly. Furthermore, a battery depreciation model is used to estimate the community battery lifetime more precisely.

The community cost can be computed for any subset of agents, and thus the cost contribution of an agent to a group can be determined by comparing the cost of the group with and without the agent. Furthermore, the community cost calculation can be seen as a cost function for a set of prosumers with demands. The notation of the community cost is simplified to $c(\mathcal{N})$ w.l.o.g., because time horizon $T = 1$ year is used to compute costs in the rest of the paper.

3 REDISTRIBUTION METHODS

One way to express the Shapley value that is particularly useful for our approach is by using the concept of *stratum*, given as follows.

$$\phi_i(c) = \frac{1}{N} \sum_{j=0}^{N-1} \sum_{\substack{\mathcal{S} \subseteq \mathcal{N} \setminus \{i\}, \\ |\mathcal{S}|=j}} \frac{j!(N-1-j)!}{(N-1)!} (c(\mathcal{S} \cup \{i\}) - c(\mathcal{S})) \tag{2}$$

Here, the marginal contribution of agent i to the subcoalition $\mathcal{S}$ is denoted as $c(\mathcal{S} \cup \{i\}) - c(\mathcal{S})$, which is how much i adds to the cost of $\mathcal{S}$ by joining. Then, the marginal contribution is multiplied by the relative frequency of $\mathcal{S}$ in the stratum. A stratum j is a set of all possible subcoalition with size j. From this, the expected marginal contribution of agent i to a stratum 0 (empty subcoalition) up to stratum $N-1$ (the whole community except i) can be computed. Finally, the Shapley value is equivalent to averaged expected marginal contributions over the strata.

Yet, computing Shapley exactly has a very large time complexity (exponential to the number of agents in the community), which makes it intractable very quickly as the community size increases.

3.1 Last Marginal Contribution

While the Shapley value requires exponential number of steps to compute the marginal contributions of every possible subcoalition, it is possible to use the marginal contribution principle to design a much simpler scheme that considers the marginal contribution of each agent w.r.t to the other $N-1$ [6, 9]. Formally, let the cost of an agent i in the community $\mathcal{N}$ be simply the marginal contribution of agent i to the rest of the community, defined as the following.

$$MC_i = c(\mathcal{N}) - c(\mathcal{N} \setminus \{i\}) \tag{3}$$

Costs based on the above equation do not hold the same property as the Shapley values in which the sum of individual cost is equivalent to the total community cost [13]. Hence, the last marginal cost needs to be normalised. The final redistributed cost $\overline{MC_i}$ of agent i according to the marginal contribution method is given as:

$$\overline{MC}_i = c(\mathcal{N}) \frac{MC_i}{\sum_{q \in \mathcal{N}} MC_q} \tag{4}$$

The time complexity of this method is $O(N)$, so, while simple, it is a very computationally efficient method.

3.2 Stratified Expected Values

The marginal contribution method only takes into account the marginal contribution of the last stratum. Here, we propose a novel Shapley redistribution scheme that goes a step further and considers the expected marginal contribution for *every stratum* while still avoiding the huge combinatorial cost of the exact Shapley method. We call this the *stratified expected values* method.

Formally, for each agent i, an agent profile p_{-i} that has average demands from the rest of the agents in the community is created. The main idea is that, since p_{-i} has the average demand of the rest of the community for every time step, computing the marginal contribution from a set of agents with such a demand profile can approximate the *expected* marginal contribution of that stratum.

The cost of agent i based on the stratified expected values method SEV_i is calculated as the following.

$$SEV_i = \frac{1}{N} \sum_{j=0}^{N-1} c(\{1, ..., j\} \cup \{i\}) - c(\{1, ..., j\}) \quad \text{such that } d_1 = ... = d_j = d_{p_{-i}} \tag{5}$$

Like the marginal contribution method, the sum of individual costs does not equal the community's total cost. Hence, a normalisation step is required to compute $\overline{SEV}_i$, similarly to Eq. 4.

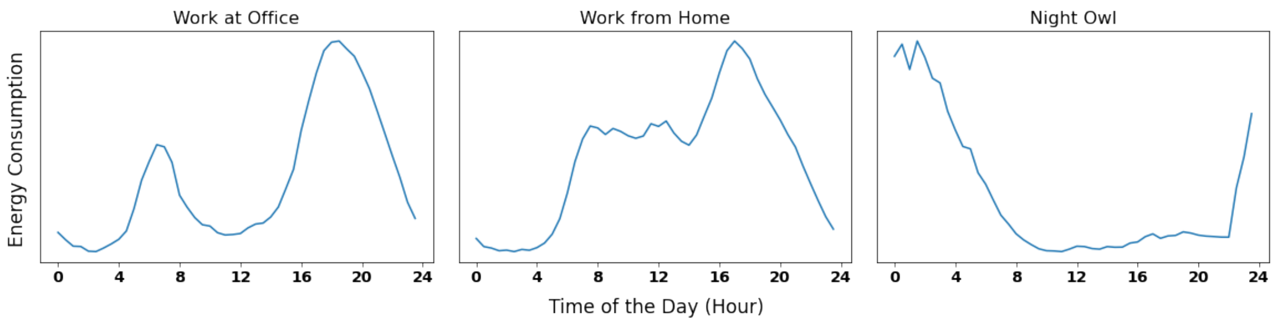


Figure 1: Average daily energy demands of consumer profiles, work at office, work from home, and night owl

The time complexity of computing the individual costs with this method is $O(N^2)$. While this is more than that of the marginal contribution method, it is still tractable for relatively large community sizes and potentially gives us a better approximation.

3.3 Adaptive Sampling Shapley Approximation

The previous redistribution methods were deterministic, providing the same numerical results every time the redistributed costs are calculated, given the demands of the agents remain the same. We also compared the performance of a state-of-the-art, random sampling Shapley approximation method. Specifically, we implemented the adaptive sampling method introduced by O'Brien et al. [10]. This method samples a subcoalition randomly from a stratum and computes the marginal contribution, repeating this step for M samples preset by the user. After every sample, the expected marginal contribution and its estimated standard deviation (SD) of the stratum are updated. The selection of the stratum at the next sample is dependent on the SDs of the strata, where strata with larger spread are more likely chosen. Finally, the mean of all expected marginal contributions of the strata is computed as the cost.

The time complexity of this redistribution scheme is $O(N \cdot M)$. Note that, in principle, the number of samples required to approximate the Shapley value well increases faster as the community size increases, hence it is the case that $M \gg N$. M was set to 1000 in this study, assuring multiple samples from each stratum.

4 EXACT SHAPLEY VALUES FOR K CLASSES

In order to have the "ground truth" consisting of the exact Shapley values for the redistribution methods above, we present a method that significantly reduces the computation time by limiting the number of classes of agents in the community. Such a method allows to compute the exact Shapley values for a *large* community as a benchmark to compare the performances of the redistribution methods, which previous works have not looked at.

Assume that a community $\mathcal{N}$ still consists of N agents, with now K classes of demand profiles in the community such that every agent belongs to one class, and all the agents in the same class have equal half-hourly demands. Given that N_k is the size of class k, the number of unique subcoalition in the community is $(N_1 + 1) \times ... \times (N_K + 1)$, since from each class k you can have 0 to N_k agents being part of the subcoalition. This reduces the number of cost calculations compared to the original settings of N unique agents of 2^N, when K is small.

What allows to compute the Shapley values efficiently for a limited number of classes is the *(multivariate) hypergeometric distribution* [4]. The probability mass function of the hypergeometric distribution allows to compute the probability of a certain set of agents from the K classes to be selected ahead of the chosen agent at the specific stratum. The Shapley value of an agent can then be computed by replacing the relative frequency of a subcoalition in Eq. 2, with the hypergeometric function. Additionally, the Shapley value is only required to be computed once per class due to the symmetry axiom [13], making it much more tractable for the computation of the whole community. More detailed time complexity comparisons of exact Shapley computation and the redistribution methods are presented in Appendix A.

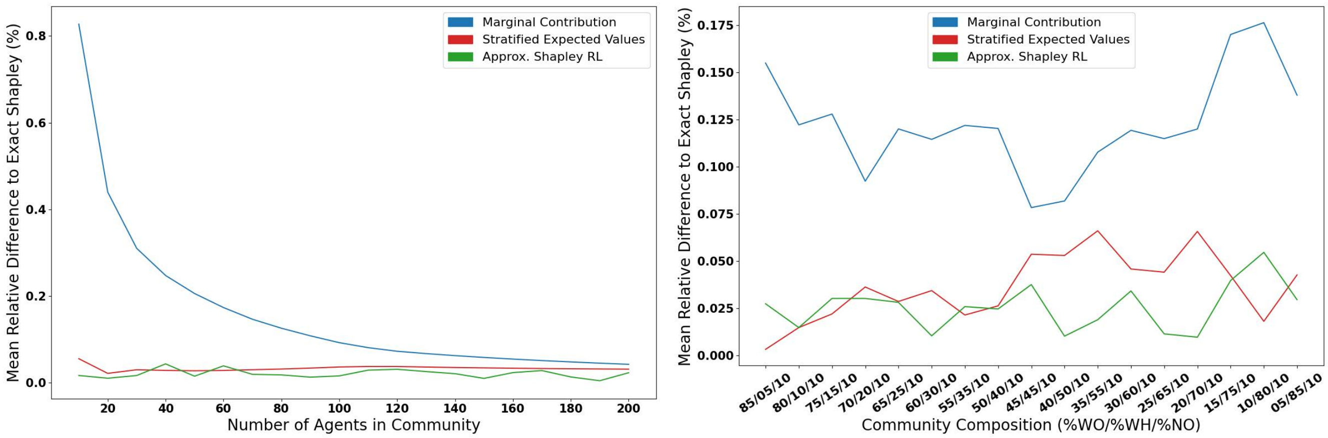
5 EXPERIMENTAL COMPARISON

5.1 Experimental Setup

Dataset and parameters. For the experimental comparison, the energy demands of 200 households in a realistically-sized energy community in the UK sharing a wind turbine and battery from Norbu et al. [9] were used. This data was collected in the Thames Valley Vision project[1]. The demands and generation data correspond to a whole year, with a half-hour time step. An import tariff of 16 UK pence/kWh and an export tariff of 0 pence/kWh were used.

Demand profiles. Clustering energy consumers into a number of classes according to their daily consumption profile is a well-established practice in energy demand modelling [5, 7, 15]. Here, prosumers were grouped using k-means clustering into a limited number of classes from their consumption behaviours over a day during the winter weekdays. From these clusters, three groups showed distinct behaviours, which are presented in Fig. 1. The "work at office" class shows a small morning consumption peak followed by minimal consumption during the day and a final large peak in the evening, representing households that are absent during the day. The "work from home" class shows a flatter consumption during the day followed by a large evening peak, resembling households that are home during the day. The last class, named "night owl", has almost no demand during the day, but has high demand

[1]https://ukerc.rl.ac.uk/DC/cgi-bin/edc_search.pl?WantComp=147



(a) Different Community Size **(b) Different Compositions**

Figure 2: Performances of redistribution methods with varying (a) community size, and (b) composition of the consumer profile.

overnight. This behaviour is also observed in the work by Jeong et al. [5], and only makes up a very small fraction of the community.

Setup and performance measure. Communities with three consumer profiles were used to compare how well the redistribution methods approximate the Shapley values. In the first setting, the composition of the consumer profiles in the community was kept constant with increasing size of the community N. In the second setting, N was kept constant, but the composition changed.

The performances of the methods were compared using the percentage *relative difference* to the Shapley value, defined as follows.

$$RD_\phi(\hat{\phi}_k) = \frac{|\hat{\phi}_k - \phi_k|}{\phi_k} \times 100 \quad (6)$$

where $\hat{\phi}_k$ is the energy cost of agent of class k from a particular redistribution method. The variable ϕ_k is the cost redistributed to class k according to the Shapley value. Given the relative difference of individual agents in the community, the relative difference to the exact Shapley of the *whole* community is simply computed as the *average* relative difference of all the agents in the community.

5.2 Results

Fig. 2 shows the results of the experiments described previously. Fig. 2a shows the performances with increasing community size, from $N = 10$ up to $N = 200$, and the composition of the community set to "work at office" 70%, "work from home" 20%, and "night owl" 10%. Fig. 2b shows the performances with different compositions of the community. N was set to 100, and the ratio of "night owl" class in the community was set to 10%. Initially, "work at office" was the majority of the community with 85% belonging to this class. Then, the ratio was decreased after each step and "work from home" increased, until it was the majority with 85% belonging to this class.

The results indicate that, while marginal contribution method performs poorly for small community size, all three methods approximate Shapley well for a large community size (100-200 agents).

As expected, the stratified expected values method outperforms the marginal contribution method. Furthermore, there is a minimal difference in the performances between the stratified expected values and the adaptive sampling methods, even though the sampling method have a larger time complexity (assuming $M \gg N$).

Although it approximates the exact Shapley values very well, a potential disadvantage of sampling methods is that the redistributed costs can vary every time the algorithm is run. In practice, the random output of the method can have an undesirable effect on the *perceived fairness* of the redistribution, as prosumers with the same demand profile can result in being assigned slightly different costs.

6 CONCLUSIONS & FURTHER WORK

While the use of Shapley value is popular in redistributing costs in energy communities, previous works tend to overlook its applicability in a large, realistically-sized setting. We proposed a new method that efficiently approximates the Shapley value and compared its complexity and performance with two existing methods in a large setting. Additionally, to develop a "ground truth" benchmark for comparing these approximations, we proposed a novel method to compute the Shapley value *exactly* even for large population sizes by clustering agents into a smaller number of consumption profiles.

Our results showed that all three cost redistribution methods approximated the Shapley values well for large communities. Furthermore, our proposed stratified expected values method outperformed the simpler marginal contribution method, and performed comparatively with the adaptive sampling method which requires more computational resources.

Many promising directions exist for future explorations. On the experimental side, we plan to investigate a greater demand dataset of ~6000 households and corresponding renewable generation in the London area[2]. This will allow for further validation of the robustness of the approximation methods. On the modelling side, looking into the case of physical capacity constraints (voltage, power) on the local distribution network, similarly to [8], would be interesting. Such a case can lead to changes in coalitional games and the computation of fairly redistributed costs with Shapley value.

[2] https://www.kaggle.com/datasets/jeanmidev/smart-meters-in-london

REFERENCES

[1] Georgios Chalkiadakis, Edith Elkind, and Michael Wooldridge. 2011. *Computational Aspects of Cooperative Game Theory.* Morgan Claypool.

[2] Adriana Chiş and Visa Koivunen. 2019. Coalitional Game-Based Cost Optimization of Energy Portfolio in Smart Grid Communities. *IEEE Transactions on Smart Grid* 10, 2 (2019), 1960–1970. https://doi.org/10.1109/TSG.2017.2784902

[3] Enercon. 2010. ENERCON Wind energy converters Product overview. https://docs.wind-watch.org/Enercon.pdf

[4] K. G. Janardan and G. P. Patil. 1972. A Unified Approach for a Class of Multivariate Hypergeometric Models. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* 34, 4 (1972), 363–376. http://www.jstor.org/stable/25049821

[5] Hyun Cheol Jeong, Minseok Jang, Taegon Kim, and Sung-Kwan Joo. 2021. Clustering of Load Profiles of Residential Customers Using Extreme Points and Demographic Characteristics. *Electronics* 10, 3 (2021). https://doi.org/10.3390/electronics10030290

[6] Anna Kulmala, Marius Baranauskas, Amir Safdarian, Jussi Valta, Pertti Järventausta, and Tomas Björkqvist. 2021. Comparing Value Sharing Methods for Different Types of Energy Communities. In *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*. 1–6. https://doi.org/10.1109/ISGTEurope52324.2021.9640205

[7] Jungsuk Kwac, June Flora, and Ram Rajagopal. 2014. Household Energy Consumption Segmentation Using Hourly Data. *IEEE Transactions on Smart Grid* 5, 1 (2014), 420–430. https://doi.org/10.1109/TSG.2013.2278477

[8] Sonam Norbu, Benoit Couraud, Valentin Robu, Merlinda Andoni, and David Flynn. 2021. Modeling Economic Sharing of Joint Assets in Community Energy Projects Under LV Network Constraints. *IEEE Access* 9 (2021), 112019–112042. https://doi.org/10.1109/ACCESS.2021.3103480

[9] Sonam Norbu, Benoit Couraud, Valentin Robu, Merlinda Andoni, and David Flynn. 2021. Modelling the redistribution of benefits from joint investments in community energy projects. *Applied Energy* 287 (2021), 116575. https://doi.org/10.1016/j.apenergy.2021.116575

[10] Geaórid O'Brien, Abbas El Gamal, and Ram Rajagopal. 2015. Shapley Value Estimation for Compensation of Participants in Demand Response Programs. *IEEE Transactions on Smart Grid* 6, 6 (2015), 2837–2844. https://doi.org/10.1109/TSG.2015.2402194

[11] Giuseppe Prettico, Flavia Gangale, Anna Mengolini, Alexandre Lucas, and Gianluca Fulli. 2016. DISTRIBUTION SYSTEM OPERATORS OBSERVATORY - European Electricity Distribution Systems to Representative Distribution Networks. https://doi.org/10.2790/701791

[12] Amir Safdarian, Poria Hasanpor Divshali, Marius Baranauskas, Antti Keski-Koukkari, and Anna Kulmala. 2021. Coalitional Game Theory Based Value Sharing in Energy Communities. *IEEE Access* 9 (2021), 78266–78275. https://doi.org/10.1109/ACCESS.2021.3081871

[13] Lloyd S Shapley. 1953. A value for n-person games. *Contributions to the Theory of Games* 2, 28 (1953), 307–317.

[14] Niklas Vespermann, Thomas Hamacher, and Jalal Kazempour. 2021. Access Economy for Storage in Energy Communities. *IEEE Transactions on Power Systems* 36, 3 (2021), 2234–2250. https://doi.org/10.1109/TPWRS.2020.3033999

[15] Kai-le Zhou, Shan-lin Yang, and Chao Shen. 2013. A review of electric load classification in smart grid environment. *Renewable and Sustainable Energy Reviews* 24 (2013), 103–110. https://doi.org/10.1016/j.rser.2013.03.023

APPENDICES

A COMPUTATIONAL COMPLEXITIES OF REDISTRIBUTION METHODS

	Time Complexity	
Algorithm	**Unique**	**K Classes**
Exact Shapley	$O(2^N \cdot (N-1))$	$O(N^K \cdot (K-1))^*$
Marginal Contribution	$O(N)$	$O(K)$
Stratified Expected Values	$O(N^2)$	$O(K \cdot N)$
Approx. Shapley RL	$O(N \cdot M)$	$O(K \cdot M)$

*Upper bound time complexity

Table 1: Time complexity per algorithm

Table 1 shows the time complexities of exact Shapley values and the three approximation methods used in this study for two scenarios; when the community of size N consists of unique demand profiles and when the number of demand profiles is limited to K classes.

In the case of N unique demands, it was explained previously that it requires 2^N steps to compute the Shapley value of an agent. To compute the Shapley values of the whole community, it is required for $N-1$ agents since the value of the last agent can be determined by simply subtracting the sum of the rest of the agents' values from the total cost. This is due to the efficiency property of the Shapley value, in which the sum of the redistributed values equals the total value [13].

When the community is restricted to K classes, the number of unique subcoalition in the community is $(N_1+1) \times ... \times (N_K+1)$. Considering w.l.o.g that the classes are ordered by the size, i.e., $N_1 \geq N_2 \geq ... \geq N_K$, and assuming there are at least two non-empty classes, i.e. $K \geq 2$, then it holds that $N_i + 1 \leq N, \forall i = 1, ..., K$. The time complexity of the Shapley value for the agent is hence upper bounded by N^K. Due to the symmetry property of the Shapley value [13], it is only required to be computed once per class. Furthermore, it is required to compute $K-1$ times with the same reasoning as in the unique demand profiles scenario.

For the approximation methods, the cost needs to compute once for each class when limited to K classes, and hence the time complexities of the methods drop by the factor of replacing N with K.