Energy profiling of end-users in service and industry sectors with use of Complex Network Analysis

Rosario Portera[1], *Fabrizio* Bonacina[2], *Alessandro* Corsini[2], *Eric Stefan* Miele[1], *Lorenzo* Ricciardi Celsi[3]

[1]Dipartimento di Ingegneria Astronautica, Elettrica ed Energetica, Sapienza Università di Roma
[2]Dipartimento di Ingegneria Meccanica e Aerospaziale, Sapienza Università di Roma
[3]ELIS Innovation Hub, Roma

Abstract. Decarbonization scenarios advocate the transformation of energy systems to a decentralized grid of prosumers. However, in heterogeneous energy systems, profiling of end-users is still to be investigated. As a matter of fact, the knowledge of electrical load dynamics is instrumental to the system efficiency and the optimization of energy dispatch strategies. Recently, a number of clustering algorithms have been proposed to group load diagrams with similar shapes, generating typical profiles. To this end, conventional clustering algorithms are unable to capture the temporal dynamics and sequential relationships among data. This circumstance is of paramount importance in the service and industrial sectors where energy consumption trends over time are possibly non-stationary. In this paper, we aim to reconstruct the annual user energy profile identified through a non-conventional method which combines a time series clustering algorithm, namely K-Means with Dynamic Time Warping, with Complex Network Analysis. For the purpose of the present research, we have used an open database containing the data of 100 commercial and industrial consumers, collected every 5 minutes over a year. From the results, it is possible to identify different patterns of consumer behaviour and similar corporate profiles without any prior knowledge of the raw data.

1 Introduction

More than 80% of World's energy consumption is still powered by fossil fuels (oil, coal, gas) [1]. This circumstance, responsible for climate change acceleration, is driving a revolution in the energy market on both sides, production and demand.
On demand side, applicable strategies entail the use of energy efficiency solutions (management and technologies) irrespective from the specific sector, e.g. residential and

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

industrial sectors are respectively responsible for 17% and 13% of total CO2 emissions in Europe in 2020 [2]. On production side, decarbonisation strategies foster the increase of renewable energy capacity (hydro, wind and solar) along with the improvements of electric grids to properly balance electricity supply and demand.

Instrumental to such an energy scenario evolution, is the profiling of end-users based on data available from so called Internet of Energy [3] to allow grid operators to cope with renewable energy generation capacity and distribution. An adequate characterization of the electrical consumers assumes an important supporting role for stakeholders, in understanding and anticipating the behaviour of electric energy (and in general energy) users or for setting up dedicated services. It enables a more efficient management of resources, creating more accurate forecast models, which allow to adjust the production of energy to the actual demand, manage the load peaks, and identify outliers and losses.

Typically, energy companies assign a load curve to the consumer that is used for the management of energy distribution and billing. This curve is estimated on the base of available energy consumption model for the reference economic sector (agricultural, manufacturing, transport, etc.) [4, 5]. Such energy profiling approaches are not able to take into account any changes in the consumer's habit and electricity usage. In addition, the load curve assigned to a consumer may be wrong at the beginning due to different electricity usage behaviour than the proposed typical consumer group [5]. In fact, load profiles belonging to the same business type often reveal different electricity consumption habits. Therefore, using business sector to categorize consumers can in general result into an inefficient representation [4].

To solve this problem, different energy profiling techniques have been proposed over time, taking into account the energy consumption for a consumer over a period of time, allowing the production, planning and delivery of personalized energy services based on knowledge of consumer profiles [6]. The advent of Industry 4.0 (aka Internet of Energy in the ambit of interest [3]) has introduced a number of solutions which offer a completely new vision for the development of electricity distribution systems. In these contexts, a series of sensor devices (smart meters) monitor and collect electricity consumption data, with a much higher resolution than was possible in the past. Smart meters can be read remotely and are capable of measuring energy consumption multiple times per day, a huge amount of detailed data on building energy consumption is generated at different granularities, allowing more accurate identification of consumer habits [7]. However, in a heterogeneous network, for example in an industrial park where different types of industries with very different electricity consumption habits normally coexist, identifying the energy profile of consumers for the provision of high-value energy services is not easy [8]. The availability of smart data, collected by sensors, allows the use of non-static approaches based on a data driven methodology.

In this scenario, artificial intelligence (AI) algorithms used for data mining play a fundamental role to extract useful information from smart monitoring data by generating a deeper understanding of consumer habits and dynamics.

To this end, unsupervised learning techniques and among others clustering methods, offer one of the most attractive approach to data mining and machine learning. Clustering involves partitioning objects with similar patterns under observation into different groups [9]. In recent years, several clustering algorithms have been proposed to group load diagrams with similar shapes and generate typical load profiles. To mention but a few, models proposed in energy profiling are mainly based on partitional approaches based on techniques as K-means [4, 6, 10, 11], hierarchical approches [5, 12] , and shape-based models as K-shape [7]. Often those models are used in combination with deep learning models, e.g. Autoencoder and SOM (Self-Organizing Map) [8, 9]. However, the literature reveals several limitations peculiar of conventional clustering algorithms which are unable

to capture temporal dynamics and sequential relationships among data [13, 14]. In particular, the weakness of conventional clustering techniques resides in the use of distance functions to find clusters of a predefined shape. In addition, they only identify local relationships between neighbouring data samples, being indifferent to long-distance global relationships [15].
In this article we propose a clustering approach, to identify typical behaviours and find similarities in the energy load profiles, which combines a specific algorithm based on Dynamic Time Warping (DTW) clustering with Complex Network Analysis (CNA). The rationale is to combine DTW proven valid in capturing similarities among time series [16], to CNA which is excellent in unveiling correlations or similarities among individuals part of a network [17–19].
The rest of the paper is organized as follows. In Section 2, an accurate description of the proposed method is given. In Section 3, we describe the case study. In Section 4, we summarize the present work and draw some conclusions.

2 Methodology

The workflow, illustrated in Figure 1, consists of several activities with a macroscopic division in the following two consecutive steps.
Step 1 includes all the analyses focused on single users in order to generate groups of days similar in shape and size. It consists of a first preprocessing for data cleaning and preparation (which includes segmentation and normalization of the daily data), and a final cluster analysis of the time series. To make the results comparable to each other in the multi-user analyses planned in Step 2, the cluster analysis must include all the daily profiles of the different users, and then the results must be broken down by individual IDs.
Step 2, then, implements CNA by combining information from multiple sources (resulting from DTW approach) and create groups of users similar in energy consumption habits. In this phase, the profiles of the time series obtained by the clustering algorithm are mapped in the complex network domain, through the definition of a matrix able to return a synthetic information of the daily consumption of a given user during a year. The matrices of individual users are then represented in the form of a graph and stacked on a multi-level structure where each layer represents a single user. Then a graph similarity algorithm is applied to the multi-user network with the aim of detecting communities of users with similar energy consumption behaviors.
Each part of the proposed approach has been developed in Python 3.7. In particular, the Pandas and Numpy libraries are used for the pre-processing and data preparation phases while the Tslearn library, containing specific algorithms that exploit the DTW distance, is used for the model realization and its evaluation. All the modules related to the complex network analysis have been managed with the NetworkX library [20].

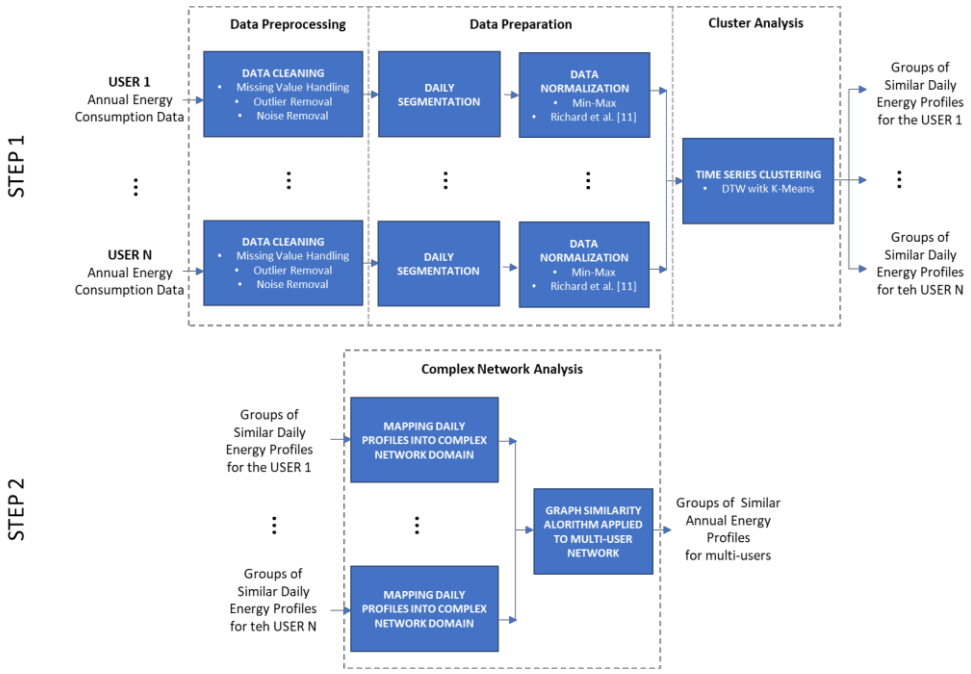

Fig. 1. Building blocks of the proposed method.

2.1 Data Pre-processing

Data pre-processing has been applied in order to deal with the inconsistencies in measured dataset, and to detect and correct eventual corrupt or inaccurate records from the database [21]. Missing values are treated using a linear interpolation [22], while duplicate data, and outliers using are processed using 3-sigma rule [23, 24] and removing time series noise through a low-pass filter.

2.2 Data Preparation

Raw data have been manipulated and transformed by changing the shape or structure of the data make it more suitable for learning algorithms, and using a normalization process [21].

2.2.1 Daily Segmentation

The approach requires using a resample function in the Python pandas library [25]. The raw values with a frequency of 5 minutes were added together in order to obtain the corresponding hourly value. Subsequently the yearlong load profile is separated in a continuous sequence of n daily load profiles. Specifically, for each consumer we generated a dataset containing in each line the hourly consumption of the day.

2.2.2 Data Normalization

For data normalization we tested two different methodologies, eventually choosing the most effective one during the clustering process.
The first method is the *min-max* normalization [26]. It is performed to scale the values so that they fall within a predetermined range. The main advantage of min-max normalization lies in its ability to reserve the relationships among the initial data since it carries out a linear normalization [26].
Concerning the second method we used an approach proposed by Richard et al. [11]. Daily load profiles (D_n) are built from hourly load data ($H_{n,x}$)

$$Y = [D_1 D_2 \dots D_n] \quad (1)$$

Feature vectors are generated for each consumer with 25 dimensions for each day of the year. The first 24 dimensions are the normalized components of the daily load profiles. The last dimension represents the ratio of the norm of the actual daily load profile vector to the greatest norm of all the daily load vectors of the year. This formulation of the feature vectors allows the grouping of similar daily profiles in terms of both shape and average load [11].

$$F_n = [(D_n / ||D_n||) (||D_n|| - max||D_n||)] \quad (2)$$

2.3 Cluster Analysis

Clustering is one of the most common tasks in data mining. The goal is to divide data items into groups according to pre-defined similarity or distance measure. More specifically, clusters should maximize the intra-cluster similarity and minimize the inter-cluster similarity. In the context of time series data mining, the same idea applies. Considering a set of time series, the goal is to find groups of time series that are similar inside the cluster, but are relatively different from times series of other clusters [15].
The K-means algorithm is one of the simplest and most widely used unsupervised learning algorithms to classify a given data set through a certain number of clusters fixed a priori [8]. The K-means algorithm proceeds by combining adjacent data in a specific area and dividing them into several groups. It searches for minimum degree of dispersion by clustering the data entry (every user) in group using iteratively a distance-driven criterion, i.e. minimizing the distance to the K-centroid of the belonging group compared to the one to the other groups [12]. In K-means algorithm, the number of clusters depends on the initial setting of a K-value, here derived following the customary elbow method with the sum of squared errors (SSE) is used as a performance indicator [7, 27].
The choice of distance measure is the other criterion that has a direct impact on the clustering performance [16]. To this end, the most common distance metric is the Euclidean one [16]. However, this metric is unable to capture temporal dynamics and sequential relationships among data, for this reason, we use a different distance, the so-called DTW (Dynamic Time Warping based one [16]. DTW is recognized as one of the most accurate similarity measures for time series data [16], and it aligns two time series using the shortest warping path in a distance matrix. A warping path defines a mapping consisting of a sequence of adjacent matrix. There is a high number of path combinations and the optimal path is the one that minimizes the global warping cost [15].
The evaluation of clustering performance makes use of a Silhouette index, to compare quantitatively the compactness and separation of clusters [28].

2.5 Complex Network Analysis

Starting from the output of the K-means clustering algorithm applied on daily consumption profiles, we mapped every user in the complex network domain, thus defining a multi-user network. In this way, by means of graph similarity algorithms, we were able to compare different users, measure their degree of similarity and, finally, group similar users together using community detection algorithms

2.4.1 Mapping Daily Energy Consumption Clusters into Complex Network Domain

In order to map the output of the daily energy consumption clustering into the complex network domain, we defined a graph G for each user Y with 365 nodes. In particular, each node represents a day of the year, and the weight of non-directed edges is set to unity in case node i and node j refer to energy profiles D_i and D_j of user Y clustered together by the K-means algorithm. Otherwise, the edge weight is set to zero. In this way, by computing a graph for each user, we define a multi-user network composed by N layers (being N the number of end-users).

2.4.2 Graph Similarity applied to multi-user network

The final step compares different users Y by quantifying their similarity in terms of graph topology. For this purpose, we employ a graph similarity algorithm based on the extraction of eigenvalues from the Laplacian of each graph. Specifically, for each user Y we compute the eigenvalues of the Laplacian extracted from the associated graph G and find the smallest λ eigenvalues such that their sum is grater or equal to 90% of the sum of all the eigenvalues.
Then, in order to evaluate the similarity between two users, we compute the sum of the squared differences of their respective λ smallest eigenvalues. The closer this measure is to zero, the more the users are similar.
By computing the similarity between all pairs of users, we defined an NxN similarity matrix S where each entry (i, j) measures the similarity between user Y_i and user Y_j. As a consequence, the NxN connectivity matrix P reads as

$$P = 1 - [(S - S_{min} / S_{max} - S_{min})(S'_{max} - S'_{min}) + S'_{min}] \quad (5)$$

Accordingly, we define a connectivity graph C by taking P as the adjacency matrix. This graph contains a node for each user and edges quantify the degree of similarity between pairs of users.
The community detection is, finally, performed by means of the Louvain algorithm [29]. In order to evaluate the quality of the output, we consider graph modularity as a metric and perform pruning of low similarity edges in C to achieve the best possible score.

3 Case Study

In order to test the energy profiling approach, we have used an open database containing energy consumption data collected from 100 different consumers [30]. Data was sampled every 5 minutes during the year 2012. Table 1 summarizes the list of consumers, final, used for the analysis after removal of any duplicated data entry. As shown in Table 1, energy consumption data refers to four different class of consumers (commercial property,

education, food sales & storage and light industrial) which operate in 10 different sub-industries class.

Table 1. Summary of consumer types.

SECTOR	SUB_SECTOR	Number of consumers	Total number of consumers	SITE_ID
Commercial Property	*Bank/Financial Services*	2	14	22, 30
	Business Services	3		12, 14, 36
	Commercial Real Estate	2		31, 65
	Corporate Office	1		9
	Shopping Center/Shopping Mall	6		6, 8, 10, 29, 32, 44
Education	*Primary/Secondary School*	20	20	88, 92, 99, 100, 101, 103, 116, 137, 144, 153, 186, 197, 213, 214, 218, 224, 228, 236, 259, 275
Food Sales & Storage	*Grocer/Market*	21	21	281, 285, 304, 339, 341, 363, 366, 384, 386, 391, 401, 404, 427, 454, 455, 474, 475, 478, 484, 496, 512
Light Industrial	*Food Processing*	9	11	648, 654, 674, 718, 737, 742, 745, 766, 786
	Manufacturing	1		761
	Other Light Industrial	1		765

After interpolating missing values and removing outliers, the data, were resampled with an hourly frequency. Figure 2 shows, as an example, the annual time series of daily energy demand for two different consumers. It is clear from the plots that the two consumers, the first Education (Primary/Secondary School) and the second Light Industrial (Food Processing), have significantly different types of energy consumption.

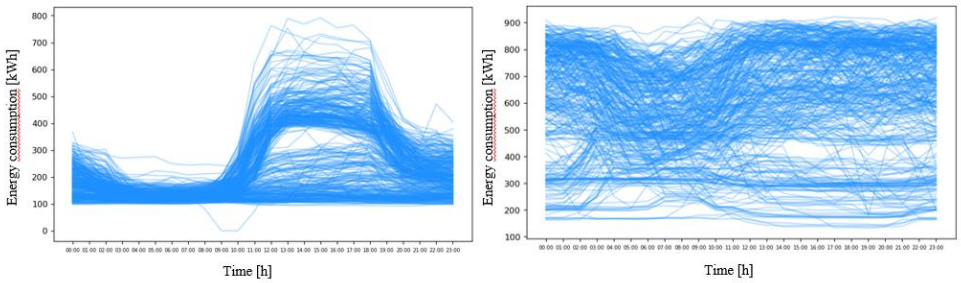

Fig. 2. Daily load curves of the sites ID 103 (left) and ID 737 (right).

Each client's daily series was normalized using two different methods, as explained in Section 2.

In order to select the best method of data pre-processing, after identifying the optimal number k of clusters by elbow method and performing clustering we evaluated the different silhouette indices obtained from the two different procedures. In Table 2, we report the average silhouette index value for each end-user subsector obtained by applying the two different normalization approaches.
As can be seen from Table 2, the Richard et al. [11] method of pre-processing the data provided the best silhouette index values, so we chose these results as input for the next stages of analysis. Specifically, we obtained 14 representative groups from the clustering performed on the daily consumption of all consumers, which were allocated to individual users based on specific IDs. Table 2 also reports the number of daily patterns identifying the typical consumption of each consumer during a year (e.g. the daily consumptions of site ID 31 and ID 65 were grouped, respectively, by 5 and 6 representative clusters).

Table 2. Summary of the number of clusters obtained for each site, comparison of the Silhouette index results for the different sites.

SITE_ID	**Silhouette Indexes for different Normalization Methods**		**Number of clusters**
	Min-Max normalization	**Richard et al.** [11]	
22, 30	0.48	0.52	5, 5
12, 14, 36	0.28	0.36	6, 6, 4
31, 65	0.34	0.42	5, 6
9	0.28	0.41	4
6, 8, 10, 29, 32, 44	0.37	0.50	6, 5, 6, 5, 5, 7
88, 92, 99, 100, 101, 103, 116, 137, 144, 153, 186, 197, 213, 214, 218, 224, 228, 236, 259, 275	0.30	0.43	7, 4, 3, 4, 4, 4, 7, 7, 4, 5, 6, 5, 4, 5, 5, 3, 4, 4, 5, 6
281, 285, 304, 339, 341, 363, 366, 384, 386, 391, 401, 404, 427, 454, 455, 474, 475, 478, 484, 496, 512	0.14	0.33	4, 3, 4, 3, 3, 3, 6, 2, 4, 4, 4, 5, 3, 3, 3, 3, 2, 5, 5, 4, 3
648, 654, 674, 718, 737, 742, 745, 766, 786	0.37	0.47	4, 7, 5, 7, 4, 4, 4, 6, 6
761	0.29	0.49	5
765	0.35	0.37	5

Fig. 3 shows an example of the daily clustering results in terms of the time series of typical profiles obtained for a primary and secondary school (i.e., site ID 103). The results were

sorted by the absolute value of energy demand. Within each cluster, the daily profiles clustered together, and the representative average profile (dotted line) are represented.

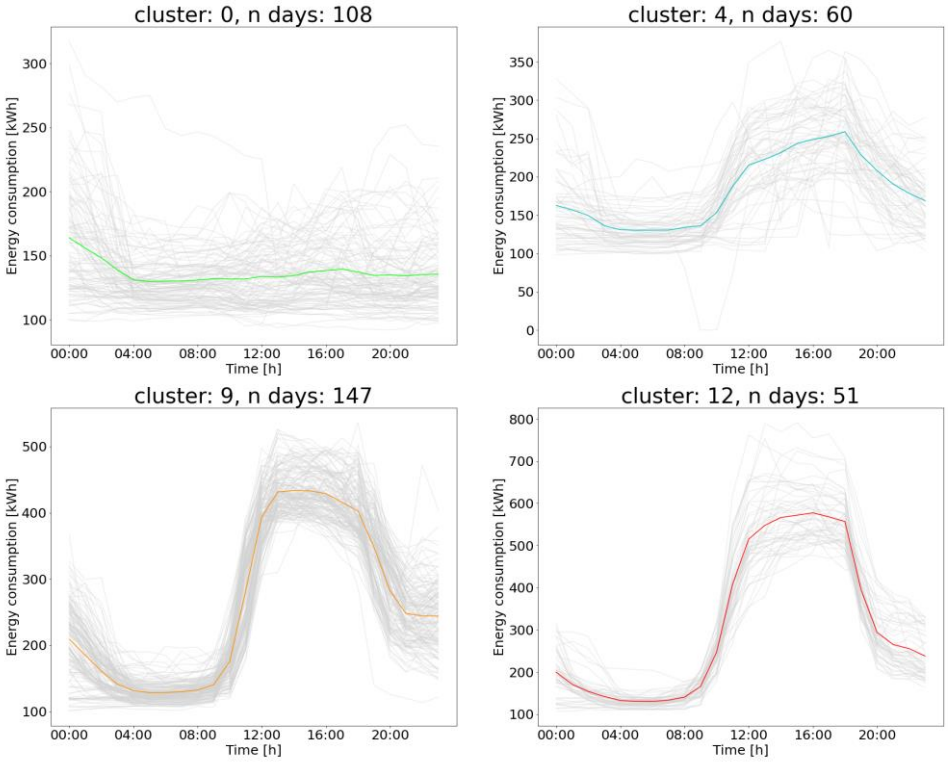

Fig. 3. Daily profiles clusters of site ID 103.

Analysing the results obtained for site ID 103, we can observe 4 main clusters representing the annual consumption patterns of the user. In particular, Cluster ID 0 counts 108 daily curves characterized by an almost constant energy consumption throughout the day. The other Clusters, respectively counting 60, 147 and 51 days each, present a similar pattern, even if shifted or scaled differently, with a low consumption of energy during the morning, followed by a peak during the afternoon. This daily variation is increasingly evident ranging from Cluster ID 4 to Cluster ID 12.

Fig. 4, on the other hand, reports the main statistics of the individual clusters through boxplots, representative of the two different consumers already analysed in overall terms.

The two different graphs show that the consumer ID 737 has a higher average consumption, but also a smaller Inter-Quartile Range for each cluster with respect to Consumer ID 103.

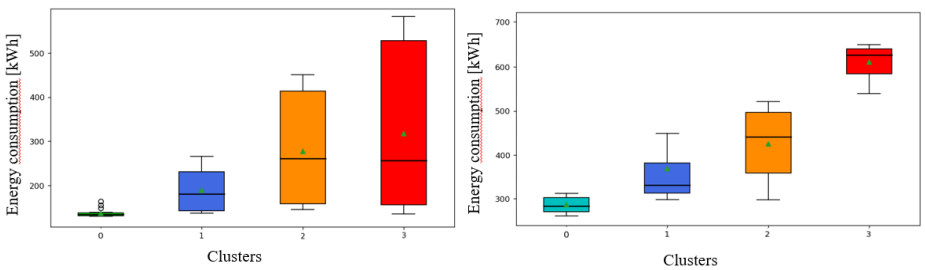

Fig. 4. Energy consumption level of each cluster of sites ID 103 (left) and ID 737 (right).

After obtaining the typical daily profiles, they were used to define a functional graph of each individual user (365 nodes), including key information about annual energy consumption patterns, as described in Section 2.4.1. The functional graphs were then stacked within an N-level structure (multi-user network), where N=66 represents the total number of users. Finally, graph similarity algorithms were applied to the multi-user network, as described in Section 2.4.2, in order to group consumers into communities with similar energy consumption habits.

Fig. 5. presents the Frushterman–Reingold layout applied to the connectivity graph *C* after edge pruning, where nodes are colored based on the output of the Louvain's algorithm. In particular, this representation highlights how similar users are grouped together spatially, but also in terms of cluster membership. Table 3 contains the major communities identified, detailing the consumers included in each of them. As can be seen from the results, graph similarity analysis applied to the multi-user network identified 7 major communities. In addition, 13 users (such as site IDs 14, 65, 88, 259, 366, 391, 427, 454, 478, 718, 742, 761, and 786) were recognized as singular consumers, so they were not associated with any community.

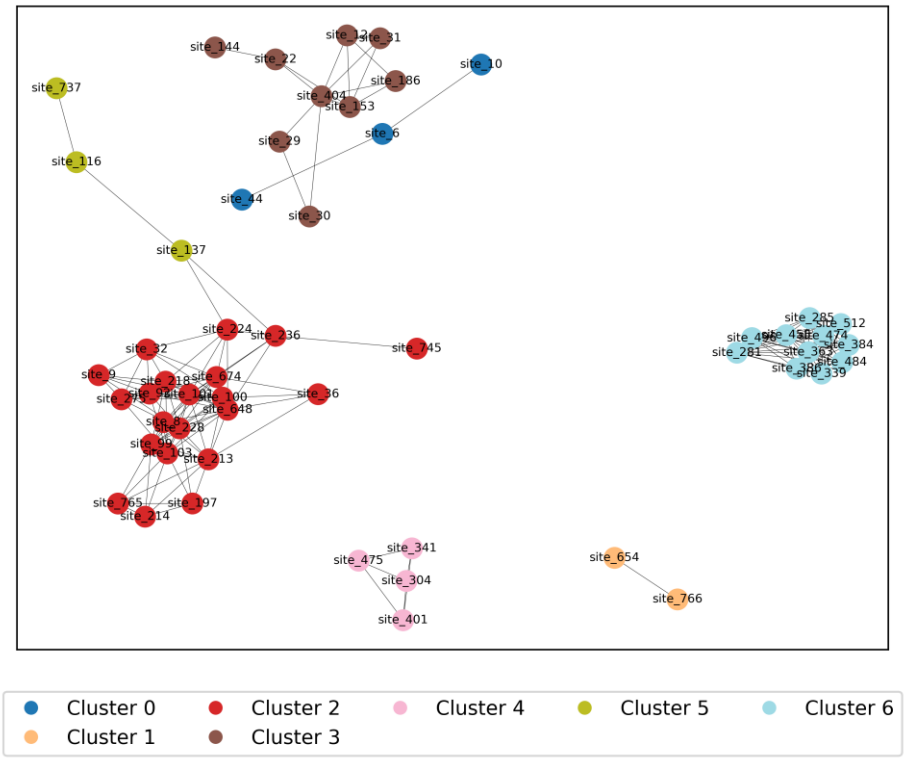

Fig. 5. Community Detection results.

Table 3. Summary of community members.

Comunity ID	SITE_ID	Total number of consumers
0	6, 10, 44	3
1	654, 766	2
2	12, 22, 29, 30, 31, 144, 153, 186, 404	9
3	304, 341, 401, 475	4
4	8, 9, 32, 36, 92, 99, 100, 101, 103, 197, 213, 214, 218, 224, 228, 236, 275, 648, 674, 745, 765	21
5	116, 137, 737	3
6	281, 285, 339, 363, 384, 386, 455, 474, 484, 496, 512	11

It is interesting to note that, while some communities include users belonging to the same subsector (e.g., three shopping centers are grouped together into the Community ID 0, two food industries into the Community ID 1, four food markets into the Community ID 3 and twenty-one primary and secondary schools into the Community ID 4), other communities (e.g., ID 2, ID 4 and ID 6) include mixed users belonging to different subsectors.

The following are examples of some detailed analyses of the results for some of the communities. For each of them, we initially show a representative energy consumption calendar based on a heatmap where cells represent days of the year and colours vary according to the reference daily cluster. Specifically, we first sorted by average energy consumption the 14 clusters obtained from the Step 1 of the proposed methodology (see Fig. 1), and then we classified them with a colour scale ranging from green (low consumption: Cluster ID 0, 1), light blue (medium-low consumption: Cluster ID 2, 3, 4), blue (average consumption: Cluster ID 5, 6 ,7), orange (medium-high consumption: Cluster ID 8, 9, 10) and red (hight consumption: Cluster ID 11, 12, 13).

For each community we also show a comparison of the daily profile clusters of two consumers belonging to different sub-sectors.

Figures 6 and 7 show the results of Community ID 2 in terms of the representative heatmap and daily cluster plot of two different users, respectively.

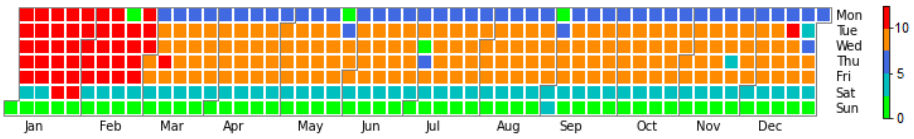

Fig. 6. Daily electricity energy consumption patterns identified in community ID 2.

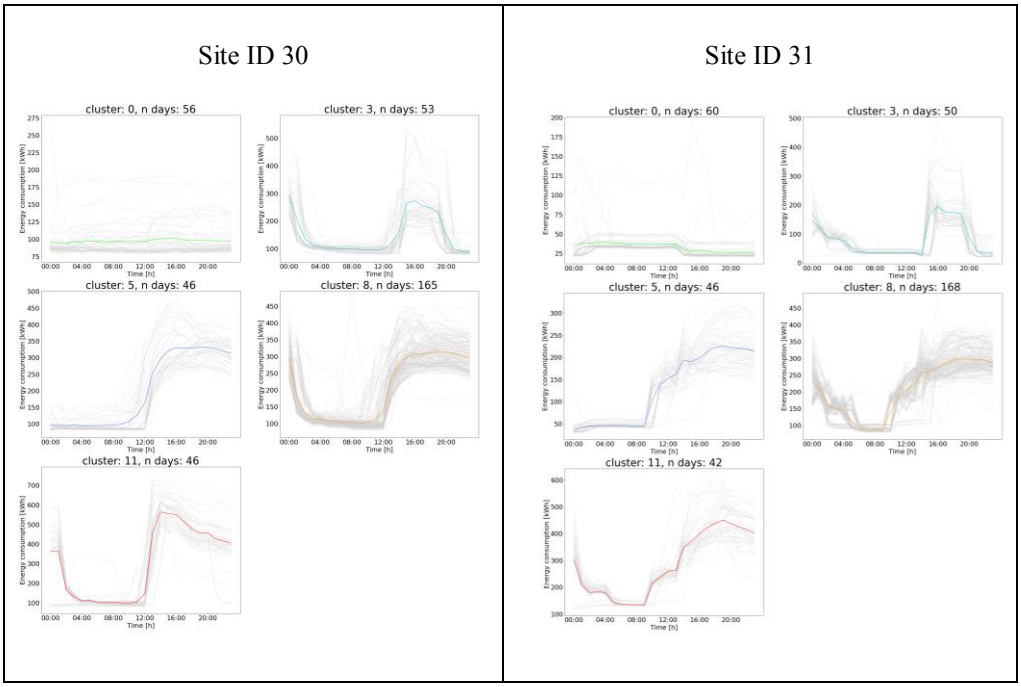

Fig. 7. Daily profiles clusters of site ID 30 (left) and ID 31 (right) belonging to the Community ID 2.

The energy consumption calendar shown in Fig. 6, highlights a higher energy consumption for all users grouped in the Community ID 2 during January and February, while it remains constantly medium-hight during the rest of the year. In addition, an average energy consumption is observed during Monday and a marked reduction during the weekend, with a minimum reached on Sunday.
Fig. 7 compares the hourly energy consumption for a Bank/Financial Services (User ID 30) and a Commercial Property (User ID 31), grouped together in the Community ID 2. Both consumers are characterized by the same five daily representative patterns (i.e., Cluster ID 0, 3, 5, 8 and 11 respectively) each of which includes a similar number of total days.
Moreover, it is possible to see how the low average consumption of the weekend (Cluster 0) typically remains constant throughout the day, while in the winter months with high and medium-high consumption (Cluster 8 and Cluster 11), the daily trend is characterized by a minimum demand around 06:00 and a maximum peak at 18:00.

Figures 8 and 9 show the results of Community ID 4 in terms of the representative heatmap and daily cluster plot of two different users, respectively.

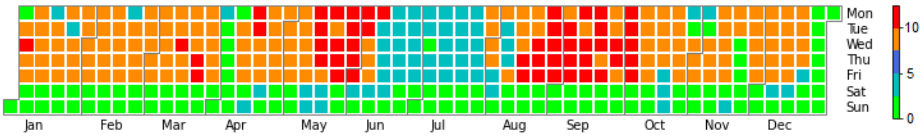

Fig. 8. Daily electricity energy consumption patterns identified in Community ID 4.

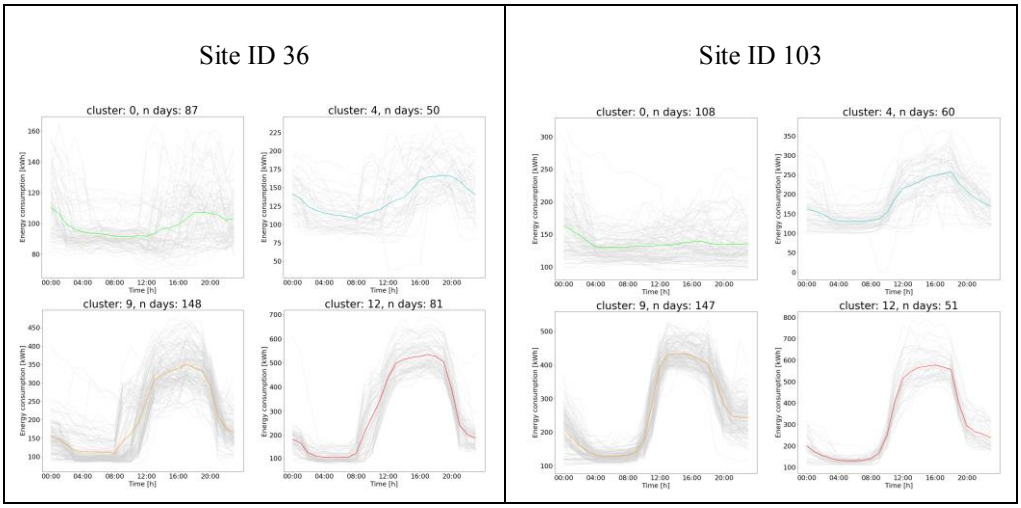

Fig. 9. Daily profiles clusters of site ID 36 (left) and ID 103 (right) belonging to the Community ID 4.

From the energy consumption calendar shown in Fig. 8, we can see a medium-low energy consumption for all users grouped in the Community ID 4 during June and July, a high energy consumption during May, August and September and a mean-high consumption in the other months. In addition, a low energy consumption is observed during the weekend.
Fig. 9 compares the hourly energy consumption for a Business Service (User ID 36) and a Primary/Secondary School (User ID 103), grouped together in the Community ID 4. Both consumers present the same four representative patterns (i.e., Cluster ID 0, 4, 9, and 12 respectively).
These users present a low and constant consumption during the weekends (Cluster 0), but also months with high and medium-high consumption (Cluster 9 and Cluster 12), where daily trends are characterized by a progressive increase in consumption after 8:00, with peak load in the range between 12:00 and 19:00.

Figures 10 and 11 show the results of Community ID 6 in terms of the representative heatmap and daily cluster plot of two different users, respectively.

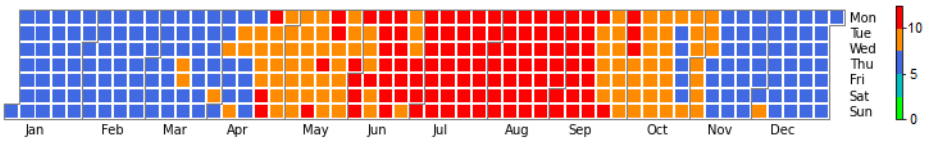

Fig. 10. Daily electricity energy consumption patterns identified in community ID 6.

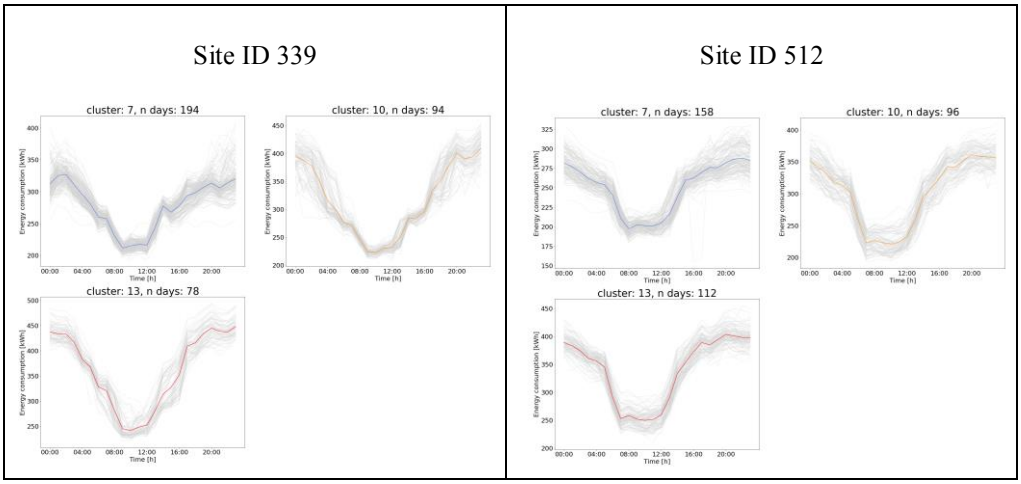

Fig. 11. Daily profiles clusters of site ID 339 (left) and ID 512 (right) belonging to the community ID 6.

The energy consumption calendar in Fig. 10 for Community ID 6, reveals an average consumption strictly connected to the yearly seasonality. In fact, it presents a low consumption during the winter a medium-high consumption during spring and autumn and higher consumptions during the summer, with no specific variations during the weekend as for the previously analysed communities.
Fig. 11 compares the hourly energy consumption for two Food Sales & Storage industrial users (ID 339 and 512) grouped together in this community, characterized by three typical daily consumption patterns (i.e., Cluster ID 7, 10 and 13 respectively).
The daily trend highlights an increase in consumption from 12:00 to 23:00 and a decrease between 23:00 and 11:00.

4 Conclusion

The need to safeguard the environment has accelerated the energy transition towards decarbonisation, the exact profiling of consumers and the identification of their habits is certainly one of the processes that most can lead to a significant reduction in consumption, allowing the optimisation of resources, adapting energy production to actual demand.
In this perspective, machine learning techniques have become a useful tool due to their ability to extract information from data, automatically and without any initial knowledge.
The research presented in this paper developed a new data analysis methodology that combines a time series clustering algorithm, namely K-Means with Dynamic Time Warping (DTW), with Complex Network Analysis (CNA).
The use of an algorithm that exploits a specific distance metric (DTW) for time series allows a better accuracy in the clustering of energy profiles, overcoming the main limitations of the models presented until now in the literature, which are hardly able to capture the temporal dynamics and sequential relationships between the data typical of time series.
The methodology was used on a real dataset containing historical consumption data related to 100 commercial and industrial consumers, during an entire year.

Thanks to the proposed technique we have been able to identify the main characteristics of consumers, tracing their behavior through the identification of the main patterns of daily energy consumption and the periods of the year with higher consumption.
In addition, thanks to CNA we were able to group together consumers with similar behaviors regardless of their macro type of origin.
By combining the proposed profile identification methodology with forecasting techniques, suppliers can predict the energy consumption of each consumer and produce that amount of energy needed for each consumer.

References

1. IEA (International Energy Agency), World Energy Outlook 2018 (2018).
2. IEA (International Energy Agency), Energy Policy Review, European Union 2020 (2020).
3. W. Strielkowski, D. Streimikiene, A. Fomina, E. Semenova, Internet of Energy (IoE) and High-Renewables Electricity System Market Design. Energies. 12, 4790 (2019).
4. S. Ramos, J.M. Duarte, F.J. Duarte, Z. Vale, A data-mining-based methodology to support MV electricity customers' characterization. Energy and Buildings. 91, 16–25 (2015).
5. T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, M. Kolehmainen, Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. Applied Energy. 87, 3538–3545 (2010).
6. O.Y. Al-Jarrah, Y. Al-Hammadi, P.D. Yoo, S.Muhaidat, Multi-Layered Clustering for Power Consumption Profiling in Smart Grids. IEEE Access. 5, 18459–18468 (2017).
7. J. Yang, C. Ning, C. Deb, F. Zhang, D. Cheong, S.E. Lee, C. Sekhar, K.W. Tham, k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. Energy and Buildings. 146, 27–37 (2017).
8. L. Hernández, C. Baladrón, J. Aguiar, B. Carro, A. Sánchez-Esguevillas, Classification and Clustering of Electricity Demand Patterns in Industrial Parks. Energies. 5, 5215–5228 (2012).
9. A. Ullah, K. Haydarov, I. Ul Haq, K. Muhammad, S. Rho, M. Lee, S.W. Baik, Deep Learning Assisted Buildings Energy Consumption Profiling Using Smart Meter Data. Sensors. 20, 873 (2020).
10. J.D. Rhodes, W.J. Cole, C.R. Upshaw, T.F. Edgar, M.E. Webber, Clustering analysis of residential electricity demand profiles. Applied Energy. 135, 461–471 (2014).
11. M. Richard, H. Fortin, A. Poulin, M. Leduc, Daily load profiles clustering: a powerful tool for demand side management in medium-sized industries. 12 (2017).
12. Y. Kim, J.-M. Ko, S.-H., Choi, Methods for generating TLPs (typical load profiles) for smart grid-based energy programs. In: 2011 IEEE Symposium on Computational Intelligence Applications In Smart Grid (CIASG). pp. 1–6. IEEE, Paris, French Guiana (2011).
13. S.K. Popat, Review and Comparative Study of Clustering Techniques. 5, 8 (2014).

14. F. Bonacina, E.S. Miele, A. Corsini, Time Series Clustering: A Complex Network-Based Approach for Feature Selection in Multi-Sensor Data. Modelling. 1, 1–21 (2020).
15. L.N. Ferreira, L. Zhao, Time series clustering via community detection in networks. Information Sciences. 326, 227–242 (2016).
16. A. Javed, B.S. Lee, D.M Rizzo, A benchmark study on time series clustering. Machine Learning with Applications. 1, 100001 (2020).
17. F. Bonacina, A. Corsini, L. Cardillo, F. Lucchetta, Complex Network Analysis of Photovoltaic Plant Operations and Failure Modes. Energies. 12, 1995 (2019).
18. A. Corsini, F. Bonacina, S. Feudo, A. Marchegiani, P. Venturini, Internal Combustion Engine sensor network analysis using graph modeling. Energy Procedia. 126, 907–914 (2017).
19. A.S. da Mata, Complex Networks: a Mini-review. Braz J Phys. 50, 658–672 (2020).
20. https://networkx.org/
21. Z.S Abdallah, L. Du, G.I Webb, Data Preparation. In: Sammut, C. and Webb, G.I. (eds.) Encyclopedia of Machine Learning and Data Mining. pp. 318–327. Springer US, Boston, MA (2017).
22. M. Lepot, J.-B. Aubin, F. Clemens, Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment. Water. 9, 796 (2017).
23. F. Pukelsheim, The Three Sigma Rule, The American Statistician 48 88–91 (1994).
24. https://www.kdnuggets.com/2017/02/removing-outliers-standard-deviation-python.html
25. https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.resample.html
26. Z. Yu, B.C.M. Fung, F. Haghighat, H. Yoshino, E. Morofsky, A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy and Buildings. 43, 1409–1417 (2011).
27. C. Yuan, H. Yang, Research on K-Value Selection Method of K-Means Clustering Algorithm. J. 2, 226–235 (2019).
28. P.J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 20, 53–65 (1987).
29. V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. J. Stat. Mech. 2008, P10008 (2008).
30. https://open-enernoc-data.s3.amazonaws.com/anon/index.html