The eDREAM project is co-founded by the EU's Horizon 2020 innovation programme under grant agreement No 774478

eDREAM

enabling new Demand REsponse Advanced, Market oriented and secure technologies, solutions and business models

DELIVERABLE: D4.4 Interactive Visualization Framework for improving DR strategies V1

Authors: Antigoni Noula (CERTH), Napoleon Bezas (CERTH), Christina Tsita (CERTH), Dimosthenis Ioannidis (CERTH)

Imprint

D4.4 Interactive Visualization framework for improving DR strategies (Month 16)

Contractual Date of Delivery to the EC:	30.04.2019
Actual Date of Delivery to the EC:	30.04.2019
Author(s):	Antigoni Noula (CERTH), Napoleon Bezas (CERTH), Christina Tsita (CERTH), Dimosthenis Ioannidis (CERTH)
Participant(s):	Contributors: Benjamin Hunter (TU), Dara Kolajo (KIWI), Mircea Bucur (KIWI), Francesca Santori (ASM), Alessio Cavadenti (ASM), Ugo Stecchi (ATOS), Javier Gomez (ATOS), Lourdes Gallego Miguel (ATOS)
Project:	enabling new Demand Response Advanced, Market oriented and secure technologies, solutions and business models (eDREAM)
Work package:	WP4 – Next generation DR Services for Aggregators and Customers
Task:	4.3 – Multi-level Visualization for enhanced user interaction
Confidentiality:	Public
Version:	1.0

Legal Disclaimer

The project enabling new Demand Response Advanced, Market oriented and secure technologies, solutions and business models (eDREAM) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 774478. The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the Innovation and Networks Executive Agency (INEA) or the European Commission (EC). INEA or the EC are not responsible for any use that may be made of the information contained therein.

Copyright

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

KPI	Key Performance Indicator
WP	Work Package
DR	Demand Response
DSO	Distribution System Operator
DB	DataBase
REST	REpresentational State Transfer
DoA	Description of Action
VPP	Virtual Power Plant
ID	IDentification
GUI	Graphical User Interface
CHD	Consumption Histogram Descriptor
PHD	Production Histogram Descriptor
MSD	Multi-Dimensional Scaling
DER	Distributed Energy Resources
MST	Minimum Spanning Tree
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
RES	Renewable Energy Sources
CHP	Combined Heat and Power
AR	Augmented Reality
GPS	Global Positioning System
RFID	Radio Frequency Identification
XML	eXtended Mark-up Language
SDK	Software Development Kit
UWP	Universal Windows Platform
PC	Personal Computer
IDE	Integrated Development Environment
SoC	SoC: System on Chip
HD	High Definition
IMU	Inertial Measurement Unit
eMMC	Embedded Multimedia Card
HPU	Holographic Processing Unit
BIM	Building Information Modelling
DOF	Degrees of Freedom
FOV	Field Of View
RGB	Red Green Blue

Executive Summary

This report provides an overview of the visualization framework description for aggregators, as well as prosumers along with the related visual analytics techniques. The results are based on the work done during the activities of Task 4.3 – Multi-level Visualization for enhanced user interaction. The main part of the visualization framework is a graph analytics platform that supports the analysis of large volumes of energy related data representing consumption/production profile data of the prosumers from the aggregators' portfolios. This deliverable can be considered as a high-level description of the functionalities and related visualization techniques of this platform.

Taking into account the results from the stakeholders' requirements definition process and the architectural requirements, the graph analytics platform is able to provide multi-factor/criteria analysis and detection of spatiotemporal patterns towards the improvement of the aggregator's portfolio management. The adopted techniques are related to the requirements and challenges of Demand Response sector. More specifically, the visualization platform to be developed covers the following high-level functionalities:

- Heat Map visualization for presenting the geographical positioning of prosumers along the indication for the respective level of energy consumption.
- Visual clustering techniques (e.g. k-partite techniques) for prosumers' grouping within the aggregator's portfolio. This functionality provides multi-dimensional analysis and segmentation of prosumers' profiles. The graphical distances between the nodes are considered, in order to dynamically formulate the prosumers' clusters for different purposes (e.g. flexibility requests, participation in energy markets etc .). Correlation with DR-related KPIs is also provided.
- Multi-objective analysis that is an area of multiple criteria decision making and it is concerned with mathematical optimization problems involving more than one objective function to be optimized simultaneously. Multi-objective optimization has been applied in many fields of science, including engineering, economics and logistics, where optimal decisions need to be taken in the presence of trade-offs between two or more conflicting objectives.

1 Introduction

1.1 Purpose

This report provides an overview of the work carried out in the direction of adopting and testing visual analytics techniques for analysing and identifying patterns in large volume of multi-dimensional data. The deliverable presents at a high level the interactive visualization framework and the results of the aforementioned techniques using the dataset from the Terni pilot site.

1.2 Relation to other activities

WP4 uses the outputs of WP2 in terms of requirements and use-cases and designs the eDREAM envisioned next generation of DR (Demand Response) services both in a classical, centralized approach (WP4) taking into account the main models and techniques of WP3 (Figure 1).

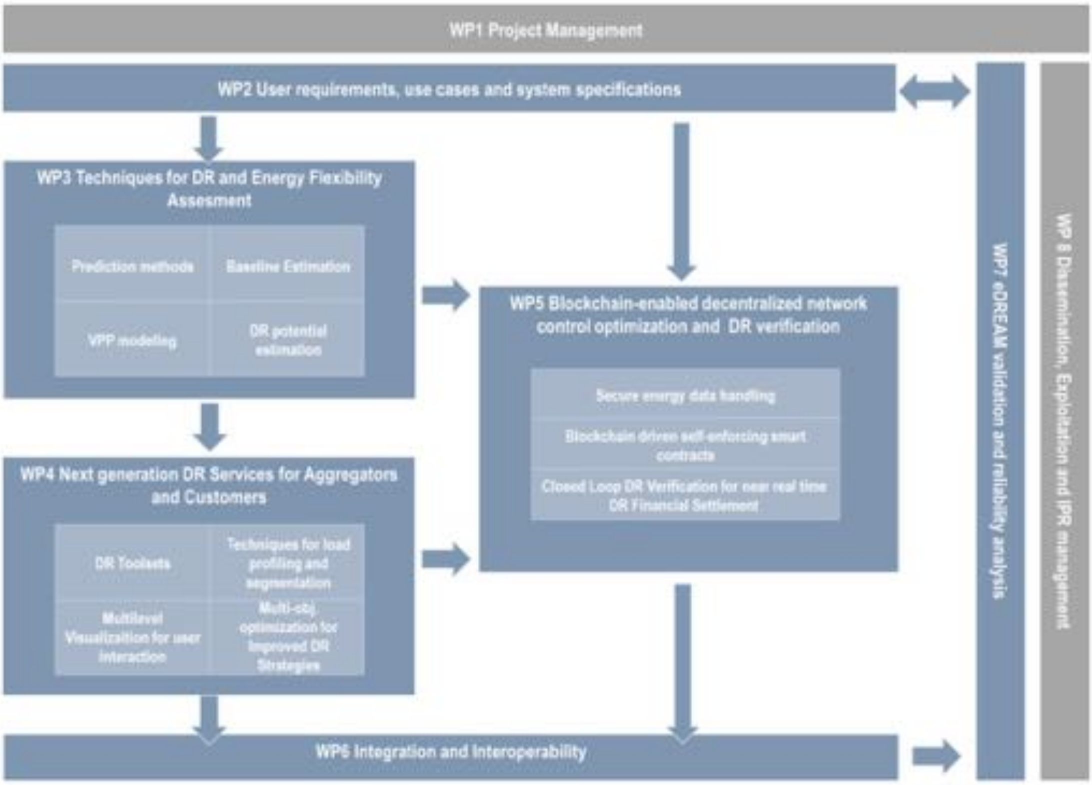


Figure 1 eDREAM pert diagram showing WP4 relations with other deliverables

1.3 Structure of the document

The remainder of the report is organized as follows:

- Section 2 presents the relevant work that has been report in the field of data analytics techniques and the visualization of data analytics;
- Section 3 describes the interactive visualization framework within the eDREAM project, the provided functionalities for the application of Demand Response programs and the relevance with the other components of the platform towards supporting decision making purposes;
- Section 4 presents the algorithmic and mathematical descriptions of the proposed visualization approaches;

- Section 5 mentions the relevance of the visualization framework with the eDREAM use cases and the evaluation of the results;
- Section 6 provides a high-level description of the augmented reality tools and their future integration with the visualization framework;
- Section 7 concludes the deliverable and provides insights for the future work that will be presented in a consolidated form in the next version of the deliverable.

2 Relevant Work on Visual Analytics Techniques

The scope of this section is to provide a high level view of the Visual Analytics framework. The main objective is to provide insights on the algorithmic techniques, but also on the visualization tools that are considered for the provision of a fully enhanced visual analytics tool. The whole analysis is delivered taking into account the energy domain examined in the eDREAM project and more specifically, the management of the prosumer's portfolio profile (energy & flexibility) analysis based on the additional high level business services addressed by the aggregator.

Hence, the overall analysis is delivered as a targeted analysis of the algorithmic approaches and the visualization methods that have been adopted through the design of the eDREAM interactive visualization framework.

At this point, it should be mentioned that an extensive analysis of Data Analysis and Visualization of Data Analytics techniques has already been presented in the respective deliverables of Inertia and Adapt4EE FP7 eu project. However, for the sake of completeness and coherence of the text, some common points are also mentioned here.

2.1 Data Analytics Techniques

In order to provide a concrete and fully enhanced view on the algorithmic techniques addressed in the eDREAM project an overview analysis is considered. Data mining process is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems [1],[2],[3]. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns, such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining) [4],[5]. This usually involves using database techniques, such as spatial indices. These patterns can then be seen as a summary of the input data and may be used in further analysis, as for example in machine learning and predictive analytics.

The overall approach is considered as follows:

1. Selection of data;
2. Pre-processing of dataset;
3. Transformation;
4. Data Mining;
5. Interpretation/Evaluation/Visualization.

From the variety of data mining techniques, the most critical ones are addressed within eDREAM project, in order to fully cover the enhanced graph analytics platform specifications. A short overview is considered:

Anomaly detection/Outlier detection

Anomaly detection, also known as outlier detection, is the search for items or events, which do not conform to an expected pattern. The detected patterns are called anomalies and often are translated to critical information in several application domains. Anomalies are also referred to as outliers, change, deviation, peculiarity, intrusion etc.

Three broad categories of anomaly detection techniques exist. **Unsupervised anomaly detection** techniques detect anomalies in an unlabelled test data set under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the data set. **Supervised anomaly detection** techniques require a data set that has been labelled as "normal" and "abnormal" and involves training a classifier (the key difference to many other statistical classification problems is the inherent unbalanced nature of outlier detection). **Semi-supervised anomaly detection** techniques construct a model representing normal behaviour from a given normal training data set, and then testing the likelihood of a test instance to be generated by the learnt model.

Association rule learning

Association rule learning is a popular and well researched method for discovering interesting **relations between variables** in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness. The term of "trend analysis" is also used to express the identification of common patterns within a large database. Association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time. The disadvantage of this technique is that the extraction of patterns requests a lot of computational work but also sets the basis for the identification of commonalities within the dataset which could be further be exploitable on the analysis of DR programs performance.

Clustering analysis

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters) [6]. It is a main task of exploratory data mining and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval and bioinformatics.

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can, therefore, be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values, such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify data pre-processing and model parameters until the result achieves the desired properties.

The analysis on clustering techniques, as it will be also presented in the deliverable D4.2, set the basis for the concrete and in depth analysis of the eDREAM Visual Analytics techniques. This is an important functionality provided by the tool and a list of scenarios has been selected towards the extraction of clusters and correlations of high interest for the aggregators.

Classification Analysis

In machine learning and statistics, classification is the problem of identifying to which of a set of categories (sub-clusters or segments) a list of observations belongs, on the basis of a training set of data containing observations (or instances) whose category is known.

An algorithm that implements classification, especially in a concrete implementation, is known as a classifier. The term "classifier" sometimes also refers to the mathematical function, implemented by a classification algorithm that maps input data to a category.

In the field of machine learning, classification is considered an instance of supervised learning [10], as for example learning where training set of correctly identified observations is available. The corresponding unsupervised procedure is known as clustering (or cluster analysis), which involves grouping data into categories based on some measure of inherent similarity as presented in the previous analysis.

2.2 Visualization of Data Analytics

2.2.1 Categorization of Visualization Techniques

In eDREAM, we have distinguished the following categories of visualization techniques:

- Standard 1D to 3D graphics;
- Iconographic techniques;
- Geometric techniques;
- Pixel oriented techniques;
- Graphs or hierarchical techniques;
- Combinations of the above focused on DR programs.

A brief description of the above categorization is provided below:

I. Standard graphics are commonly used to view an estimate of certainty about a hypothesis or the frequency distribution of an attribute or to view a data model (e.g. Histograms and Scatter Plot).

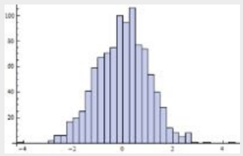
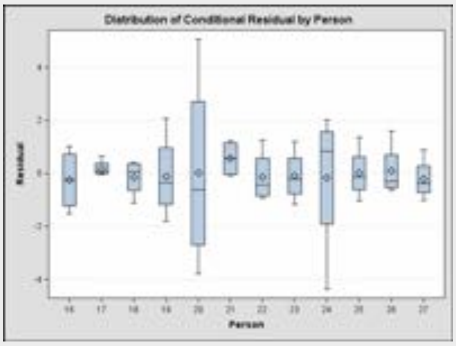
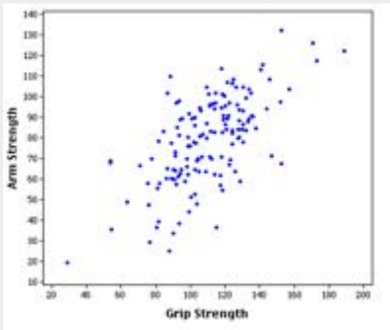
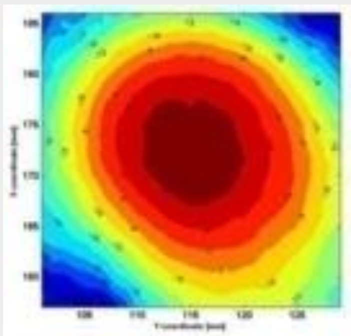
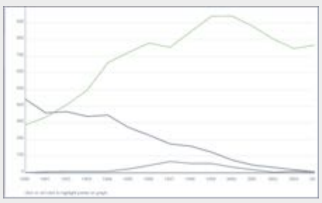
II. In pixel-oriented techniques, each value of an attribute is mapped to a pixel colour and it is placed on the display screen, divided into windows, each corresponding to an attribute. As a final step, they are arranged according to different purposes.

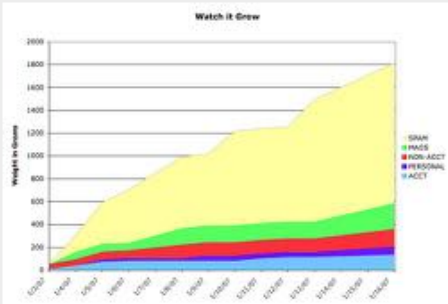
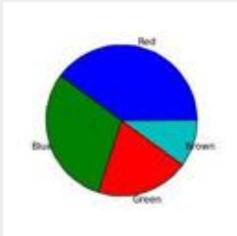
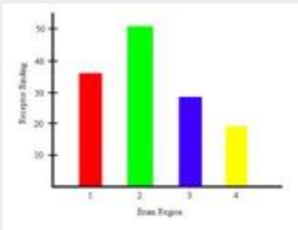
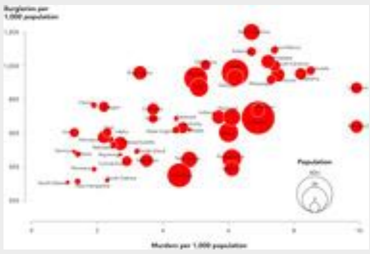
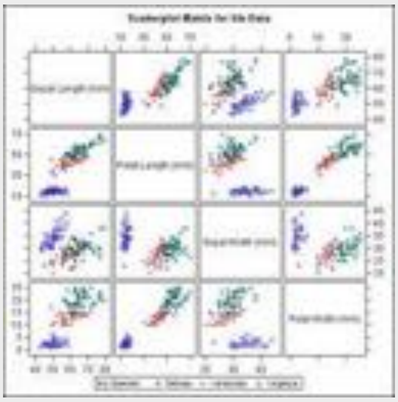
III. Data with a naturally structure of relationships among its elements, as hierarchical or as simple network, may be represented by hierarchical or graph-based techniques, such as Cone Trees, Tree maps, Mosaic Plot, Dimensional Stacking etc.

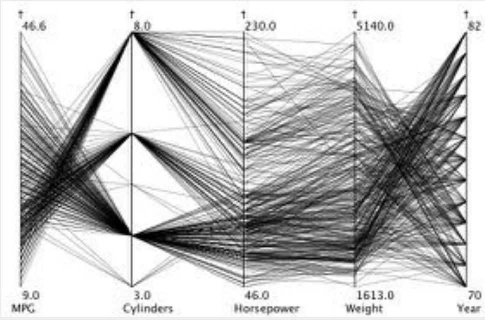
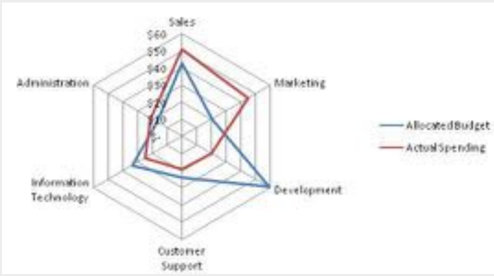
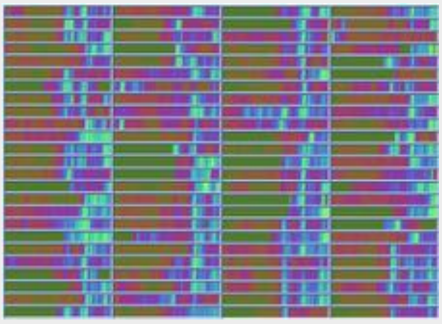
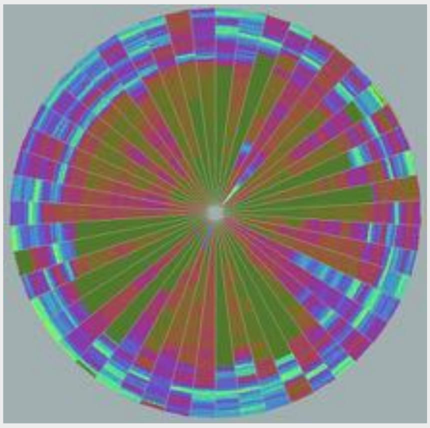
IV. In geometric techniques, multi-dimensional data are mapped into a two-dimensional plane providing an overview of all attributes. As examples Matrix of Scatter Plot, 3D Scatter Plot and Kiviat diagrams can be reported.

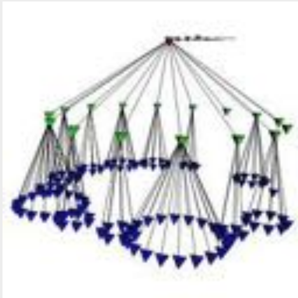
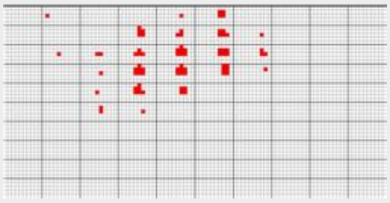
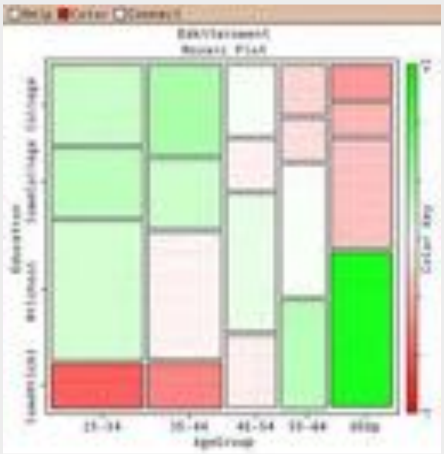
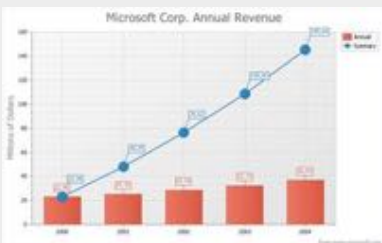
In the table below, there is a presentation of samples of each of these techniques along with a brief description.

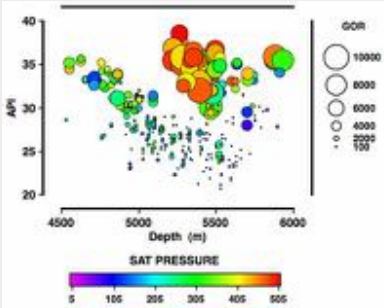
Table 1 Overview of Visualization Techniques

Category	Technique	Sample	Description
1D to 3D	Histogram		A histogram is a graphical representation showing a visual impression of the distribution of data
	Box Plot		A box plot or boxplot (also known as a box-and-whisker diagram or plot) is a convenient way of graphically depicting groups of numerical data through their five-number summaries: the smallest observation (sample minimum), lower quartile (Q1), median (Q2), upper quartile (Q3), and largest observation (sample maximum). A boxplot may also indicate which observations, if any, might be considered outliers.
	Scatter Plot		A scatter plot is a type of mathematical diagram using Cartesian coordinates to display values for two variables for a set of data. The data is displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis
	Contour Plot		A contour plot is a graphical technique for representing a 3-dimensional surface by plotting constant z slices, called contours, on a 2-dimensional format. That is, given a value for z, lines are drawn for connecting the (x,y) coordinates where that z value occurs. The contour plot is an alternative to a 3-D surface plot.
	Lines		A line graph is most useful in displaying data or information that changes continuously over time

	Stack graph		A stack graph is a classic method for visualizing change in a set of items, where the sum of the values is as important as the individual items. For example, a stack graph is excellent for looking at revenue over time across several products. Because stack graphs use areas to convey numbers, they do not work for negative values. And in some situations it might not make sense to add up different data series (say, prices of different stocks over time).
	Pies		A pie chart (or a circle graph) is a circular chart divided into sectors, illustrating proportion. In a pie chart, the arc length of each sector (and consequently its central angle and area), is proportional to the quantity it represents.
	Bars		A bar chart or bar graph is a chart with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column bar chart.
	Bubble Chart		A bubble chart is a type of chart that displays three dimensions of data. Each entity with its triplet (v1, v2, v3) of associated data is plotted as a disk that expresses two of the vi values through the disk's xy location and the third through its size.
Geometrically transformed displays	Scatter Plot Matrix		A scatter plot matrix shows relationships among several variables taken two at a time. Scatter plot matrices can reveal a wealth of information, including dependencies, clusters, and outliers

	Parallel coordinates		To show a set of points in an n-dimensional space, a backdrop is drawn consisting of n parallel lines, typically vertical and equally spaced. A point in n-dimensional space is represented as a polyline with vertices on the parallel axes; the position of the vertex on the ith axis corresponds to the ith coordinate of the point.
	Kiviat diagrams		A Kiviat Diagram is composed of axes extending from a central point. Each axis represents a data category. Each axis is scaled according to certain parameter values. It is used to graphically represent, on a single diagram, how multiple items compare when they are evaluated against more than two variables
Pixel Oriented	Query-independent techniques		The basic idea of pixel-oriented techniques is to map each data value to a coloured pixel and present the data values belonging to one attribute in separate windows. If a user wants to visualize a large data set, the user may use a query-independent visualization technique which sorts the data according to some attribute(s) and uses a screen-filling pattern to arrange the data values on the display.
	Query-dependent techniques		It visualizes the relevance of the data items with respect to a query. Instead of directly mapping the data values to colour, the query-dependent visualization techniques calculate the distances between data and query values, combine the distances for each data item into an overall distance, and visualize the distances for the attributes and the overall distance sorted according to the overall distance.
Graph Based or Hierarchical	Graph		A visualization to derive two-dimensional depictions of graphs arising from various applications

	Cone Tree		An information visualization technique used for visualizing hierarchical information structures
	Tree map		Tree maps display hierarchical (tree-structured) data as a set of nested rectangles. Each branch of the tree is given a rectangle, which is then tiled with smaller rectangles representing sub-branches. A leaf node's rectangle has an area proportional to a specified dimension on the data. Often the leaf nodes are coloured to show a separate dimension of the data.
	Dimensional Stacking		It involves recursively embedding images defined by a pair of dimensions within pixels of a higher-level image
	Mosaic Plot		A mosaic plot is a graphical display that allows you to examine the relationship among two or more categorical variables. The mosaic plot starts as a square with length one. The square is divided first into horizontal bars whose widths are proportional to the probabilities associated with the first categorical variable. Then each bar is split vertically into bars that are proportional to the conditional probabilities of the second categorical variable. Additional splits can be made if wanted using a third, fourth variable, etc.
Combinations suitable for aggregator's portfolio performance visualization	Histogram with Lines		Combination of histogram and lines
	Bars with Lines		Combination of bars and lines

	Bubble chart with colour coding		Combination of bubble chart and lines

In the next sections, each parameter is analysed in association with categories of visualization techniques.

2.2.2 Analysis of the parameters for selection of Visualization Technique

The parameters to be considered when choosing the appropriate visualization techniques are:

- Data types;
- User-task type;
- Scalability;
- Dimensionality;
- Position of the attributes in the graph.

Data type is a determining factor for choosing a visualization technique. This parameter is also used as a criterion for the classification of visualization techniques. In the context of our approach, we identify the following type of data:

- Quantitative: continuous or discrete;
- Qualitative: nominal or ordinal;
- Multidimensional: single dimensional or multi-dimensional;
- Temporal: data represented by time series;
- Spatial: data that are related with spaces within a geographical area.

Task type is another criterion to classify visualization techniques. It refers to the activities that the end user or analyst can peform, according to the setted goals when exploiting the visualization capabilities. The most common tasks are:

- Data Overview: view the whole data collection;
- Correlation among attributes: the degree of relationship among variables can reveal patterns of behaviour and trends;
- Comparison of attributes: compare different set of values;
- Identification of patterns, standards and important characteristics;
- Clusters identification: attributes with similar behaviour;
- Outliers detection: data set with a non-typical behaviour in comparison with the rest of data.

Scalability and dimensionality are characteristics to be observed in the data before applying a visualization technique. To facilitate the analysis of these parameters, a convention is established to classify the scalability and dimensionality of data.

Considering **scalability**, data can be classified as

- Small (10^1 to 10^2 = 10 to 100);
- Medium (10^3 to 10^5 = 1000 to 100000);
- Large volume (10^6 to 10^7 = 1000000 to 10000000), according to the magnitude orders.

Regarding **dimensionality** (number of attributes) the following has been assumed:

- Data with up to four attributes are defined as low-dimensional;
- With five to nine attributes, medium dimensionality;
- With more than ten attributes, high dimensionality.

2.2.3 Data Types in Optimal Visualization Techniques

Techniques of standard 1D-3D category generally represent from one up to few attributes. In most cases, they are used for analysis of quantitative data. All graphs considered in this class are able to display quantitative data. To represent qualitative data, alternative techniques are more suitable. In this case, the histogram is an example whereby it is possible to represent both these data types.

In the relevant literature, there are examples of usage of pixel-oriented techniques on quantitative data. Query-independent techniques, for example, were applied to represent temporal data and query-dependent techniques are commonly used to represent continuous quantitative data.

Hierarchical or graph-based techniques are ideal for displaying data when they have a structure of relationship among themselves or with a structure of hierarchy or simple network.

Iconographic techniques are more appropriate for quantitative data, because icon features vary with the values of represented attributes. Representation of qualitative data presents more technical difficulties, which can be overcome using the display properties, such as the colour of the icons.

Geometric techniques are more flexible, being able to represent quantitative and qualitative data. As for example, the Scatter Plot Matrix, which is formed by a set of scatter plots, is more suitable for continuous quantitative data.

The techniques related to DR programs are more appropriate for quantitative data, both continuous and discrete.

2.2.4 Scalability and Dimensionality in Optimal Visualization Techniques

Implementations of visualization techniques should take into account the limits of dimensionality and scalability of data, in order to provide be capability of a clear overview of data to the end user. As already mentioned, scalability refers to the computational complexity regarding the number of records in an array, as well as the number of features (attributes). The amount of records that can be simultaneously presented is one of the limitations of the visualization techniques. With high number of records, the results show a considerable degree of disorder.

Standard graphics have low dimensionality, because they are intended to represent data with few attributes. In addition, they support the view of a small volume of data, because they are mainly used in the field of statistical studies.

Iconographic techniques are able to handle a larger number of attributes in comparison to the standard graphics. Nevertheless, the visualization generated is best for a small amount of data due to the space occupied by the icons in the screen.

On other hand, geometric techniques may work with an increased number of dimensions and volume, compared to standard 1D-3D graphics and iconographic techniques. But they are outweighed by the pixel-oriented techniques for their capability to represent the largest volume.

Hierarchical techniques or graph-based techniques are usually used to represent the relationship among data, regardless of dimensionality, which can be high or low, but they have the same space constraints like that presented by iconographic techniques.

2.2.5 Role of Task types in Optimal Visualization Techniques

Generally, some techniques are better for certain tasks than others. The task type depends on the objectives of the end-user, the tool to be used and the desired format of results.

Standard 1D-3D techniques are commonly used to view an estimate of certainty about a hypothesis or the frequency distribution about an attribute, such as the usage of histogram. This class also provides graphs to make comparisons and data classifications (as for example box plot, lines, bars) and also to determine the correlation between attributes (as for example using scatter plots).

Iconographic techniques represent each data entry individually, allowing verification of rules and behaviour patterns of the data. Icons with similar properties can be identified and form groups. A representation with a distinct format may characterize an outlier. Icon-based visualizations are multidimensional points that make use of useful dimensional space to detect clusters and outliers.

Geometric techniques provide a good representation of the data, assigning no priorities to represent its attributes. This category of techniques also allows the identification of patterns, rules and behaviours. Therefore, outliers may also be detected, characterized by behaviours outside the common standard. The end user may choose to analyse a group of data that can be identified using the tool. However, it should be mentioned that the groups are not usually immediately identified by employing this type of techniques.

Pixel-oriented techniques can be used in the analysis of relationships among data attributes, so rules and patterns may be identified through observing the correlations among them. Furthermore, the pixels can be arranged to finding clusters.

Hierarchical techniques are useful for exploitation of data arranged in a hierarchical or simple relationship. Through techniques of this category, it is possible to obtain an overview of the data structure and analyse the relationship among the elements. Techniques of this category also allow grouping of data, such as in the form of tree maps.

2.2.6 Overview of Visualization Techniques Analysis

Through the analysis of relationship among the parameters and the visualization techniques, it was observed that each of the techniques has a certain configuration of parameters that reflect the characteristics of data and the objectives of the use of visualization, as described below:

- Data type should be the first parameter to be considered. It is the type of data that determines what kind of visualization technique should be used in a priority order. Qualitative data, for example, will be hardly understood if they are represented by a technique developed to represent quantitative data and vice versa.
- The task type to be performed corresponds to the goals of the end user/analyst during the data exploration. For tasks related to statistical analysis, the graphics 1D-3D may be sufficient, but for tasks of correlation verification, visualization techniques of geometric category may be used and so on.
- Both scalability and dimensionality of data are limiting factors for visualization techniques. Although most of them support multi-dimensional data, these techniques usually differ in the ability to display a certain amount of dimensionality and volume of data. However, other ways of interaction can be

used during the visual exploration to minimize these limitations, as for example the functions of zooming, selection and filter.

2.3 Visualization Techniques for Demand Response Programs

Within the modernized operational framework of electrical power systems, the concept of Demand Response (DR) has been introduced. Towards making a successful transition to the smart grids, the system operators and the utility providers have to deal with the large volume of data resulting from the smart meters and the demand response programs. The system stakeholders should be able to analyze properly the emerging data and identify useful information and patterns, so as to manage their portfolio and determine their strategies in an efficient and time-saving way. Towards this direction the following questions should be answered:

- How to monitor the customer portfolio and identify common behavioral patterns concerning the energy consumption/production;
- How to determine the suitable DR program for the different customers;
- Whether or not the application of a DR program at a specific time interval has a profitable result;
- How to compare different Demand Response programs.

The actions performed during Demand Response events are usually depicted as load data in two dimensions as a load curve. But this representation can not give enough information about the loads that were scheduled along with the respective time interval, how much profit was made, the spatial coordination of DR etc. Below, there are some depictions in the form of heatmaps and three-dimensional (3D) graphs that present the displacement of load during the application of DR events.

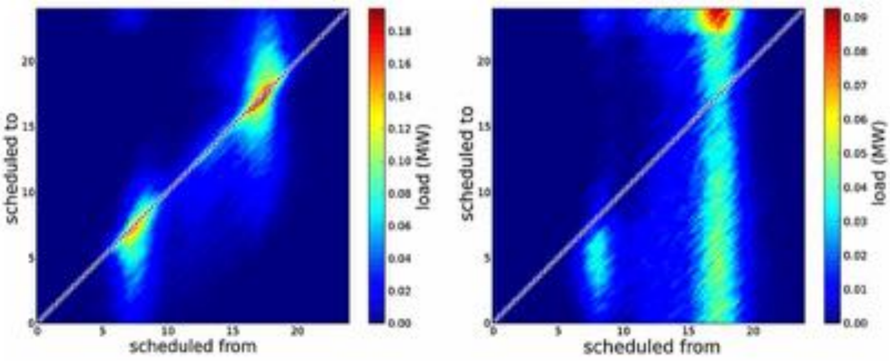


Figure 2 Heatmaps showing the load movement during DR events in two cases [7]

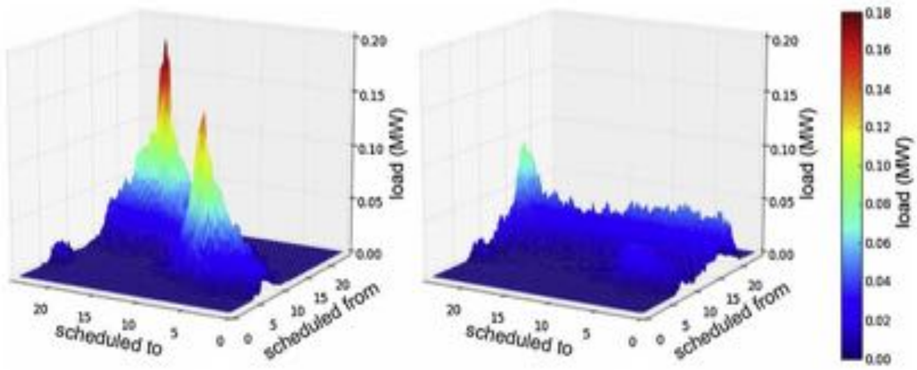


Figure 3 3D Load Graph showing the temporal displacement of the schedulable loads in two cases [7]

3 Interactive Visualization Framework

3.1 Scope of Visual Analytics

The visualization techniques have been developed over the years in order to enable scientists in different fields to cope with the understanding of complex problems through data analysis. Nowadays, data is produced at increasing rates and thus, the requirement to collect and store the data is faster than the ability to analyze it. Towards enabling humans to be involved in the data analysis process, an interactive and easy to handle visualization environment should be created. Within this context, the science of **visual analytics** has been developed that deals with the field of analytical reasoning supported by interactive visual interfaces (Figure 1). The visual analytics tools play an important role in decision-making processes, because they facilitate the communication between the human and the computer. The scientific field of Visual analytics is related to the areas of Visualization and Computer Graphics and benefits from the techniques developed for information retrieval, data mining, data management and knowledge representation.

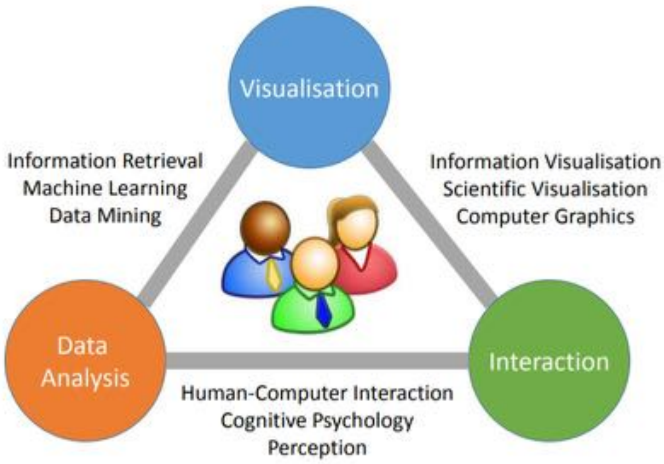


Figure 4 Visual Analytics = Visualization + (Data) Analytics

Generally, the interactive data analysis works (Figure 5) in a loop logic starting with a specified goal. The system stakeholders - actors wants to interact with the system for a specified goal and then the goal is translated to relevant questions. The user organizes and analyzes the data to answer these questions, generate new questions and so on.

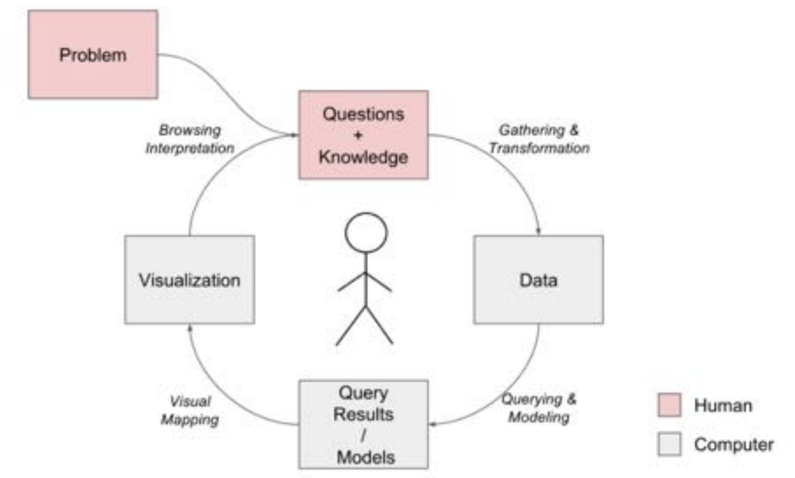


Figure 5 Interactive Data Analysis [8]

The Visual Analytics process includes automatic and visual analysis methods combining them with human interaction, so as to achieve efficient knowledge management. In many application scenarios and use cases, heterogeneous data from different sources need to be integrated before the application of visual and automatic analysis techniques. The integration of this data includes pre-processing tasks, such as data cleaning, normalization, grouping etc. The below Figure 6 depicts all the possible interactions in the Visual Analytics process.

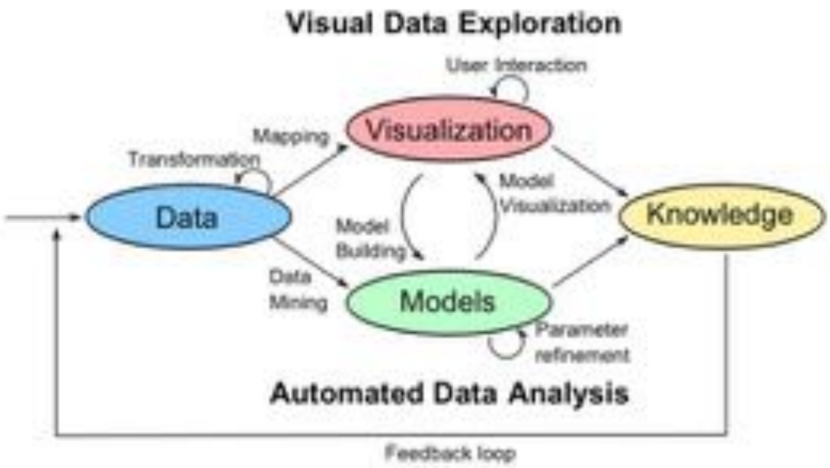


Figure 6 Visual Analytics process [9]

During the automated analysis, data mining methods are used to generate models of the original data. The analyst is able through the visualization to interact with the system by evaluating and refining models, modifying parameters or selecting other analysis algorithms. The user can alternate between visual and automatic methods, so as to achieve continuous refinement and verification of preliminary results. The interactive visualization enables the analyst to identify different aspects of the data, as for example by magninfying different data areas or by exploring different visual data views.

3.2 Visualization Framework of eDREAM platform

The main objective of the visualization framework in the eDREAM context is to enable the aggregators to manage the large volume of energy profile data from their customers towards improving their portfolio management and the application of suitable DR strategies. Actually, this framework will be a web-based suite, as the dashboard of aggregators, providing visual analytics tools, as well as other options for monitoring the assets performance and supporting decision making processes and DR strategies application. The eDREAM visualization component is the interface with the system stakeholders (DSO/aggregators). The following Figure 7 provides a high-level representation of the components functionalities and the connection to the rest of the platform. The Graph-based Analytics component provides visualization and interaction mechanisms to the aggregators for visual clustering, correlation, multidimensional analysis, identification of temporal domain patterns and other options for portfolio management. From the figure, it can be observed that the visualization part receives data from the data storage components, such as historical energy profile data of prosumers, KPIs etc.

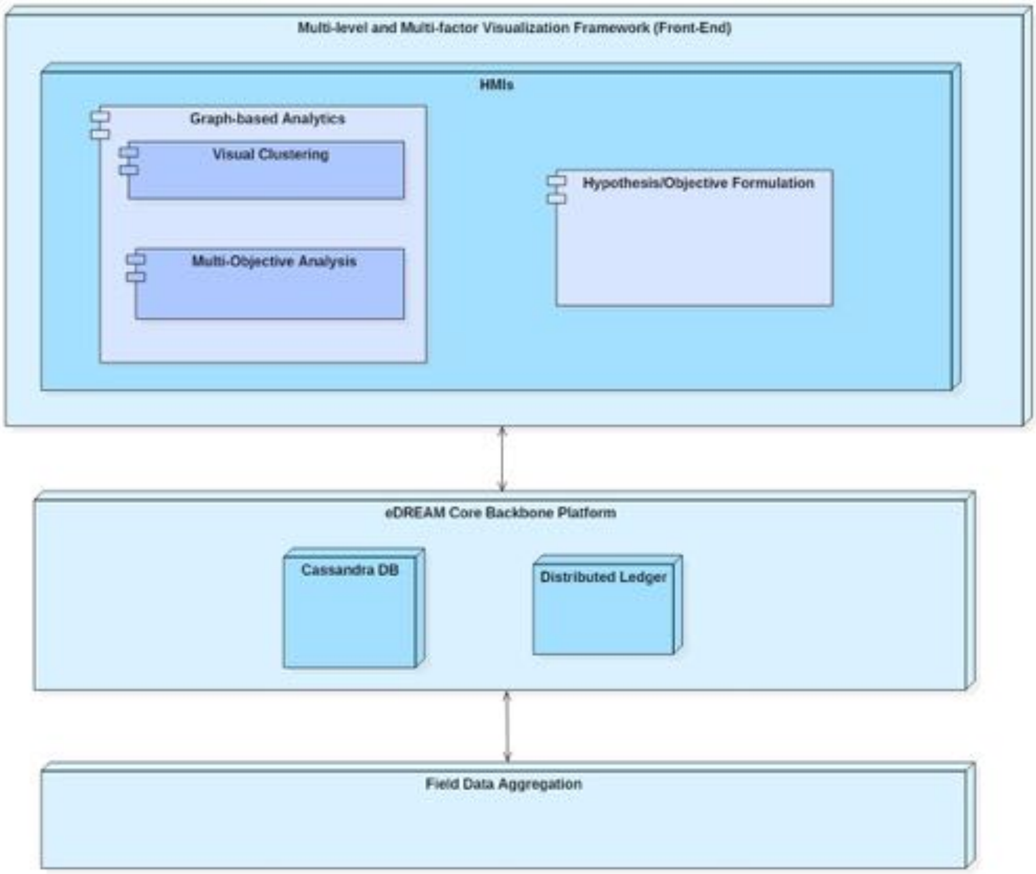


Figure 7 Front-end part of eDREAM platform

The functional requirements of the Graph-based Analytics component, as presented in the deliverable D2.4, are these in the below table:

Table 2 Graph-based Analytics Functional Requirements

Component: Graph-based Analytics	
Functional Requirement ID	***Description***
MF01_BR05_FR01	Receive data for the spatial layout of the grid
MF01_BR05_FR02	Receive flexibility data (actual & forecasted) per each prosumer
MF01_BR05_FR03	Obtain DR related KPIs (e.g. profits/losses, congestion improvement achieved etc.)
MF01_BR05_FR04	Get input settings from the aggregator UI
MF01_BR05_FR05	Receive parameters of DR signals and mapping to the portfolio
MF01_BR05_FR06	Perform data analysis and correlation among the input parameters
MF01_BR05_FR07	Store analyzed data and identified patterns

Thus, considering the identified functional requirements, the **Graph-based Analytics functionalities** cover the following operational patterns:

Normal Operation

- Spatiotemporal analysis:
    I. Geographical representation of prosumers – assets and filtering according to the orientation (north, east etc.) and the type of assets (residential, commercial);
    II. Time range selection.
- Assets and portfolio analysis in terms of KPIs and metrics performance for any selected period of time (consumption, CO2 emissions, cost etc.);
- Outliers detection.

DR Operation

- Spatiotemporal analysis;
- Correlation with critical KPIs, such as profits/losses, congestion reduction achieved etc., in order to determine the effectiveness of the applied DR strategy;
- Outliers detector.

Hypothesis/Objective Formulation Framework (T4.4)

- Dynamical formulation and validation of DR strategies in sub-sets of available historical data;
- Recommendations and services to customers for further smartening/revenue generation of their energy assets and their participation in DR programs.

The interfaces of the graph-based analytics component are mainly depended to data storage components that provide historical data and information. The input/output data and the communication mode of the component are presented below:

Input Data

- Historical data retrieved from either Cassandra DB or Distributed ledger;

Output Data

- Graphs (e.g. k-partite graphs, contour plots, times series of energy data etc.);
- Multi-objective analysis.

Communication:

- REST Interfaces;
- JSON/XML data exchange format;
- Client/server application.

The high-level architectural model of the graph analytics platform following the concept of Figure 6 is depicted below:

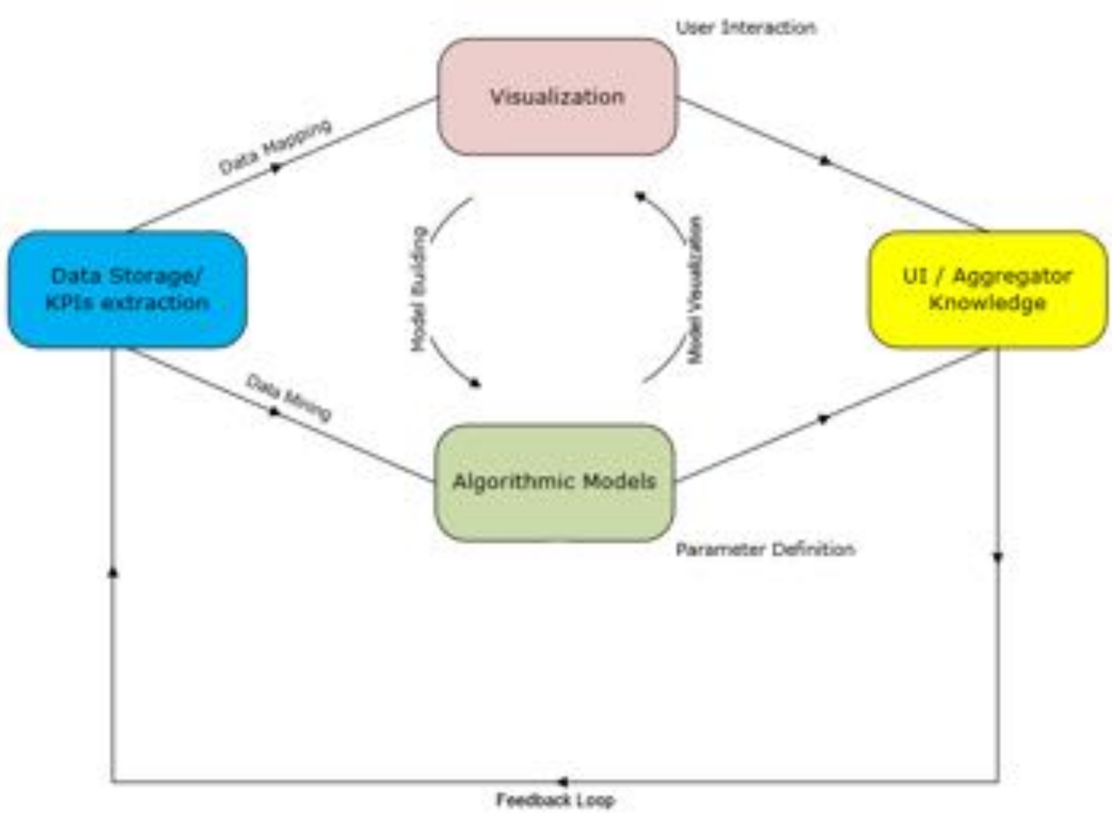


Figure 8 Architectural View of Graph Analytics Platform

As already mentioned in the DoA, the Graph Analytics Platform will be combined with the component for **Optimal DR Scheduling**, that is named **"VPP & DR Services Optimization Engine"** and is the outcome of T4.1, so as to provide the Decision making and DR strategies optimization framework for validation and improvement of the applied DR strategies. The following figure presents a high-level interaction between these two components.

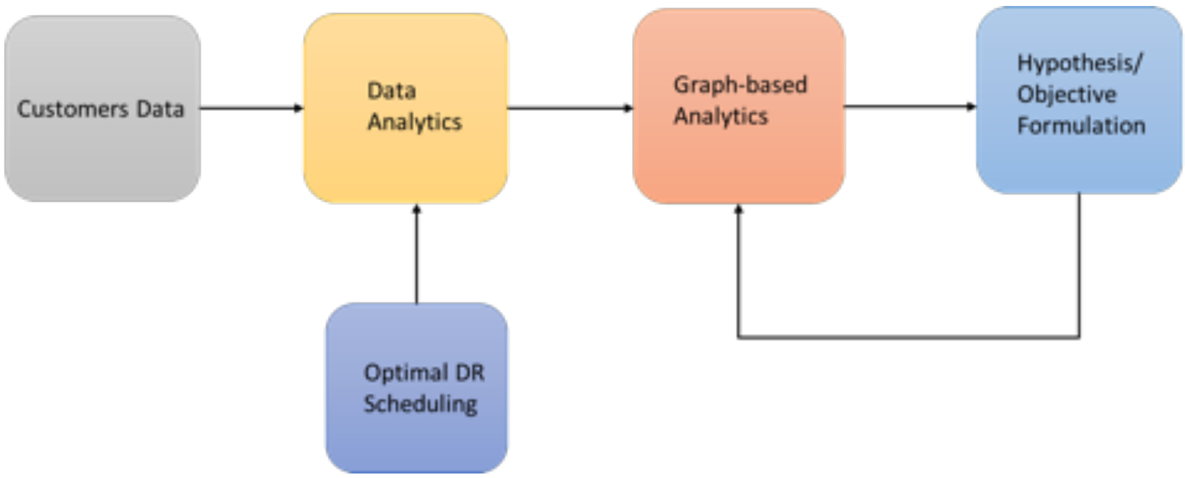


Figure 9 Validation Framework for DR Strategies Scheduling

4 Visualization Approaches

4.1 Heat Map visualization approach

A **Heat Map** is a way of representing the density or intensity value of point data by assigning a color gradient to a raster, where the cell color is based on clustering of points or an intensity value. The color gradient usually ranges from cool/cold colors such as hues of blue, to warmer/hot colors such as yellow, orange and hot red. When mapping the density of points that represent the occurrence of an event, a raster cell is colored red when many points are in close proximity and blue when points are dispersed (if using the color range above). Therefore, the higher the concentration of points in an area, the greater the 'heat' applied to the raster. When mapping intensity, points are assigned to an intensity value, the higher the value, the hotter the color assigned to the raster, with lower intensity values assigned cooler colors.

In the case of energy demand of the prosumers, **the heat map** describes the spatial distribution of electricity consumption in the area of Terni in Italy. The mapping is based on the intensity of the energy values of each prosumer with an interval of fifteen minutes. Thus, as the energy consumption of prosumers is receiving high values, the colors that are assigned to the cluster are getting hotter.

The data for the input in the heat map are the timestamp, the energy consumption of each timestamp respectively the latitude and the longitude of each prosumer. The data that are used as input are in **JSON** format.

To use a heat layer or to create a heat map the addition of the pluginheatmap.js is necessary. A heatmap layer is constructed by an array of points and an object with the following options:

- minOpacity - the minimum opacity the heat will start at;
- maxZoom – zoom level where the points reach maximum intensity (as intensity scales with zoom), equals maxZoom of the map by default;
- max – maximum point intensity, 1.0 by default;
- radius – radius of each "point" of the heatmap, 25 by default;
- blur – amount of blur, 15 by default;
- gradient – color gradient config, e.g. {0.4: 'blue', 0.65: 'lime', 1: 'red'}.

4.2 Graph-based Visualization Approach

This section presents one of the visualization techniques supported by the graph-based analytics component that offers interactive visual characteristics. This module can be considered as the basis for the required visualization and exploration of the datasets from the eDREAM project's pilot sites. These datasets include different kinds of energy related measurements that can give valuable insights for prosumers' consumption, production and flexibility. The tool that is based on this module enables multiple coordinated views, thus providing the aggregators with different visual aspects of the data related to their customers. In addition, the coordination between the views provides efficient possibilities for visual exploration of multiple aspects at the same time.

During the development of the module, the requirement of scalability was taken into account. Nowadays, the companies of energy providers have to handle a large amount of data that is related their customers who are either consumers or producers. As new customers are added to their portfolio, the process of management becomes quite difficult and can be considered as a big data analysis case. Therefore, the identification of patterns becomes more difficult. In addition, the interaction procedure with the user ends in a slower process. This interactive visualization module was developed considering the need for scalability arising from the increasing size of datasets related to energy market stakeholders' portfolios.

Architectural Overview

The architectural overview of the graph-based analytics module is depicted in the following Figure 10, where the internal algorithmic components are depicted. The module receives as input a dataset corresponding to energy measurements of prosumers from the ASM pilot site. The propose of the data analysis is to detect visual patterns related to similar user behaviour in terms of energy consumption and production and the identification of outliers. The information derived from these patterns will provide valuable insights to aggregators on the design of efficient strategies for portfolio management in the long-term. This process will support significantly the improvement of applied DR strategies related to grid and market signals. The graph analytics module currently includes two main kinds of visualizations, the k-partite-graphs and the multi-objective visualization. Other graph formulation techniques are also going to be tested, in order to compare their efficiency. A high-level description of these visual exploration techniques is given with more details in the following paragraphs.

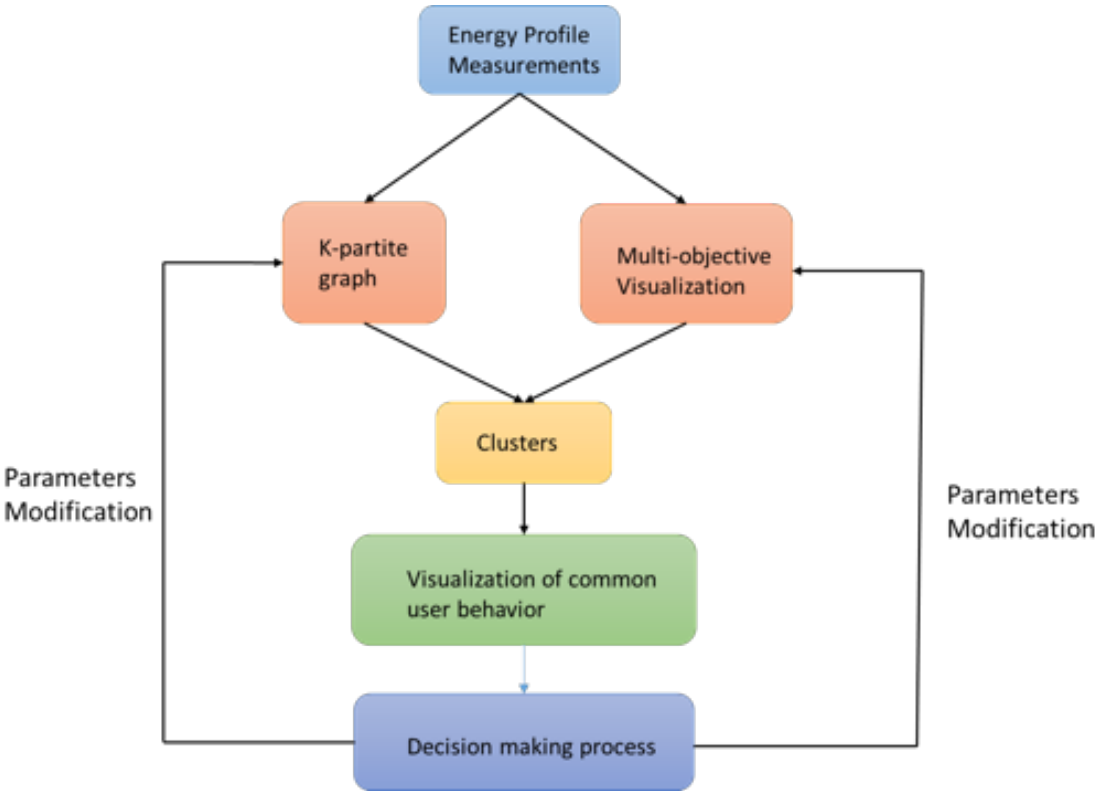


Figure 10 Architectural Overview of Graph-based Analytics

The k-partite graph approach is a visual clustering technique that enables clustering of common behavior in consumption and production of the customer and detection of anomalies. The user can interact with the view to change the subset of the data that is currently visualized. The prosumer's ID along with the related attributes energy consumption, production etc. are depicted as nodes in the visualization framework, thus allowing for detection of patterns. The module provides multiple linked views of k-partite graphs for the different attributes of the prosumer's related measurements. This allows multi-faced perceptron of the data where different coloring is applied to indicate the different attributes of the measurements records.

The multi-objective visualization approach is used to depict the commonalities between the measurements records of different users. The similarities point out similar energy behaviors. The aggregator can use this module to identify normal and unusual energy profiles. Through the multi-objective visualization approach, the aggregator can take advantage of multiple user features extracted from the data views. After this, the module uses multi-objective optimization techniques to compute optimal combinations of them with respect to the visual result. For each data view, an objective function is defined, the minimization of which leads to an optimal presentation of the data that groups users with similar behavior together. The multi-objective visualization technique attempts to simultaneously minimize the objective functions of all views, thus leading to a set of optimal solutions that represents combinations of various trade-offs of the views. An interactive interface enables the aggregator to select different solutions and have an overview of the various aspects of the dataset.

4.2.1 K-partite graph formulation approach for identifying common features

This chapter presents a graph formulation approach with k-partite technique using energy measurements related information and performing visual clustering. More specifically, the prosumers' measurements records from ASM pilot site are used as input data to visualize primarily information related to energy consumption. The common prosumer consumption patterns found in the measurements are clustered by employing this approach, so as to enable the identification of normal and abnormal consumption patterns. It is worth mentioning that the k-partite visualization approach can be utilized to visualize any multi-dimensional dataset comprised of nominal data. A set of data is said to be nominal if the values/observations belonging to it can be assigned to different classes/labels. This corresponds to the distinct nature of the graph nodes, in which each node represents a distinct entity (or specific label).

Architectural Overview

The following Figure 11 depicts the architectural aspect of the k-partite graph formulation approach. The algorithm receives as input measurement records and generates k-partite graphs. The abstract graph representations reduce the size of the graph by removing redundant nodes. Both the k-partite and abstract graph utilize a force directed layout to position the nodes on the display. This approach depicts the common prosumer consumption/production pattern and enables the aggregator to identify anomalies in the measurement records. The end-user can interact with the graph views in order to focus on subsets of the data for further analysis and assumption elicitation.

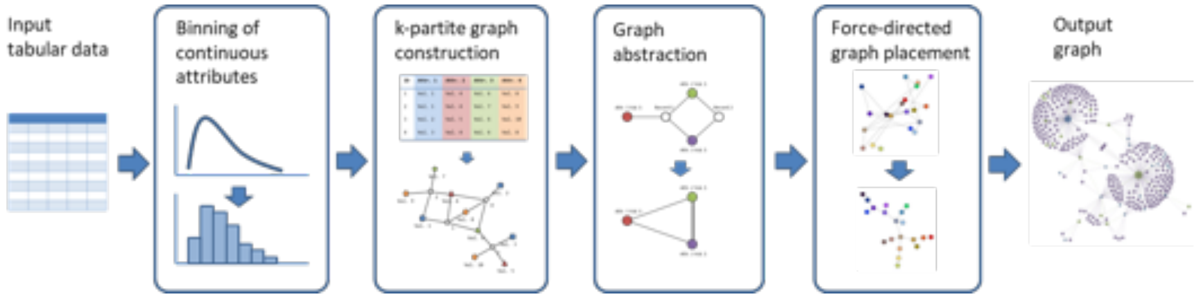


Figure 11 Architectural Overview of the k-partite graph formulation approach

The methodical approach for k-partite graph formulation is structured as follows:

- Initial k-partite graph generation from the prosumers measurement records. The scheme of the nodes of the graph as a visualized pattern is presented in the respective following section;
- After this, an abstract k-partite graph is formulated that is an abstract representation of the initial k-partite graph and reduces its size, while also preserving the significant information. This is an important process, since it reduces visual clutter and enables easier identification of important graph structures and patterns;
- As a final step, a force directed technique is employed that places the nodes to relative positions according to their similarity. This is an important procedure since it helps in the identification of groups of common patterns.

Mathematical Approach – Structuring of k-partite graph

According to the relevant theory, a k-partite graph constitutes a graph, where the set of nodes V can be divided in k discrete groups $V = (V_0, \dots, V_{k-1})$, so that no edge connects the vertices within the same group. More formally, a k-partite graph G is defined as:

$G = \langle V_0 \cup \dots \cup V_{k-1}, E \rangle$, where $V_l = \{n_i | 1 \leq i \leq N_i\}$, $\forall l \in [0, k-1]$, and $E \subset \bigcup_{l=1}^{k-1} \{V_0 \times V_l\}$ (1)

As already mentioned, the k-partite graph approach can be utilized for the visualization on any nominal multi-dimensional dataset. Within eDREAM project, the tested dataset corresponds to energy related measurements of prosumers from the Terni pilot site. This information is depicted in the graph as follows:

- Nodes in V_0 correspond to the ID of prosumers;
- Nodes in $V_{l \neq 0}$ correspond to measurement of consumptions of each prosumer respectively.

The concept of creating a k-partite graph is presented in the following figure based on the features that are related to prosumers measurements. The dataset of raw measurements is presented in Table 3. The dataset involves two features: a) the timestamp and b) the active power absorbed by each prosumer from grid.

Table 3 Measurement records of prosumers from ASM Terni pilot site

Prosumer Id	Timestamp	Consumption
514621011414	05-04-2017 07:00	199 kWh
514625389123	05-04-2017 07:00	97 kWh
514621028753	05-04-2017 07:00	52 kWh
514621077812	05-04-2017 07:00	147 kWh

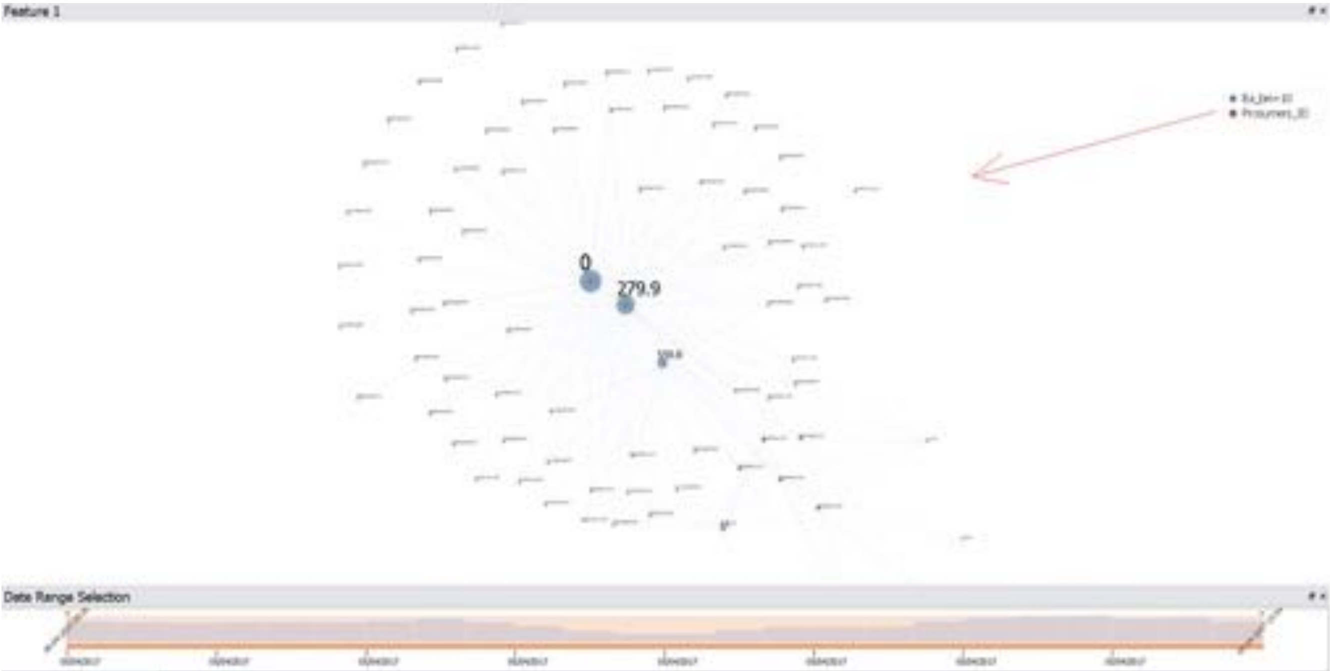


Figure 12 K-partite graph formulation based on the prosumers' measurements related to consumption from ASM TERNI pilot site

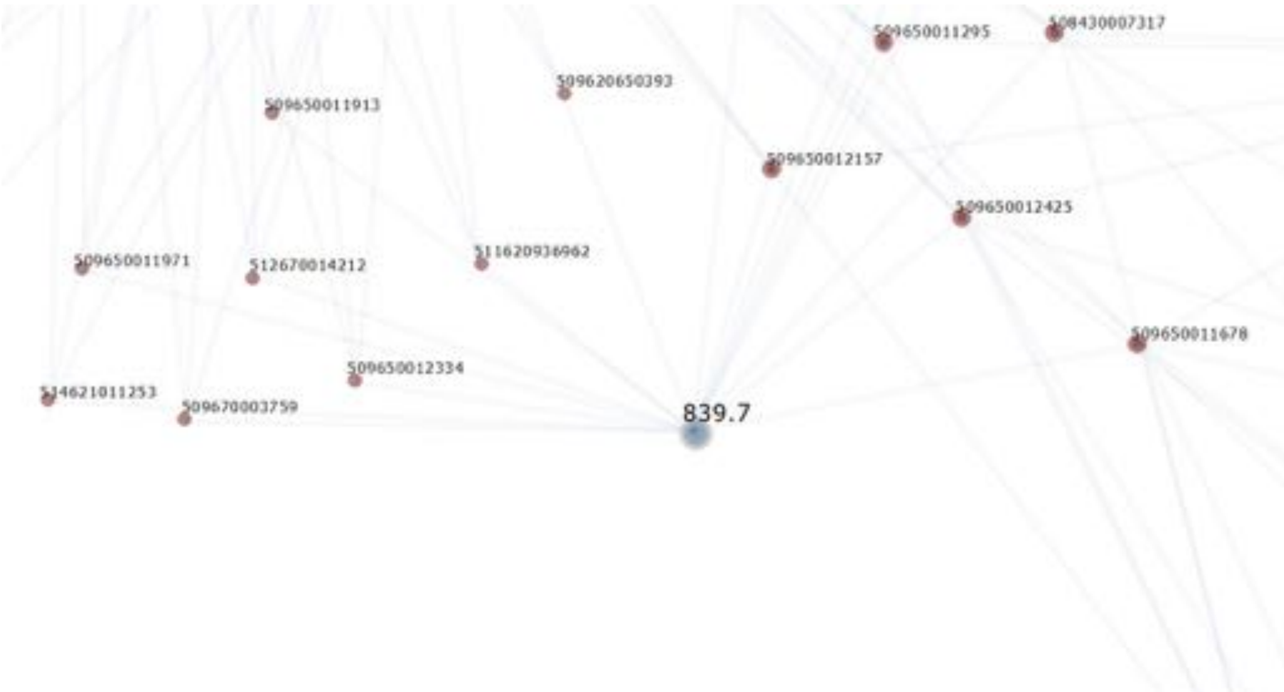


Figure 13 K-partite graph formulation based on the prosumers' measurements related to consumption from ASM TERNI pilot site (zoom in view)

Using the above dataset as an input, a 2-partite graph is created and presented in Figure 12 and Figure 13. In this graph, the measurements of consumption have been divided into 10 bins, where each bin contains the value of consumption that appears the most in the dataset. Each red node represents the ID of the prosumer and each node is divided in k different nodes, where k is the number of bins. In addition, the edges connect the measurements of same consumptions of different prosumers as it is shown in Figure 13. The size of the nodes indicates the degree of the node, which represent the number of times that each feature value appears in the dataset.

This approach considers the positioning of the nodes in the graph based on an algorithm that identifies their similarity. In this way, the structuring of different visualized groups results from the relative placement of the nodes in the screen. Thus, each group in the screen involves a group of multiple nodes that are placed in close proximity to each other. This visualization is very useful for the end-user, because as a human has the ability to identify groups of objects based on their spatial proximity.

4.2.2 Multi-objective graph formulation approach for clustering common behaviours

In an effort to visualize the behaviours of prosumers, there are a number of different features on which the visualization can rely on. For instance, the dataset of ASM contains multiple features describing the energy behaviour of the prosumers: the active energy absorbed from grid, the active energy exported from the prosumer to the grid, the absorbed inductive energy from the grid etc. The selection of one of these features or a combination of them depends on which aspect of the data the end-user (in our case the aggregator) is most interested in. The optimal combination of these multiple features, called modalities, often depends on the subjective preferences of the user.

In this section, the multi-objective visualization approach is presented. Contrary to existing multimodal fusion techniques, which usually aim at finding one optimal combination of the multiple modalities, favouring some of them against others, in multi-objective visualization, no assumption is made that some modalities are more important than others. Instead, all modalities are initially treated separately, and visualization is formed as a different optimization problem for each modality. Then, a multi-objective optimization technique is utilized, in order to simultaneously solve the multiple optimization problems.

<u>Architectural Overview</u>

The architecture of the interactive multi-factor grouping visualization method is depicted in Figure 14. Measurements records are provided as the input to the visualization method. At an initial stage, the measurements records are filtered, so that only the ones within a desired time period are kept. The desired range is set by the operator through the GUI.

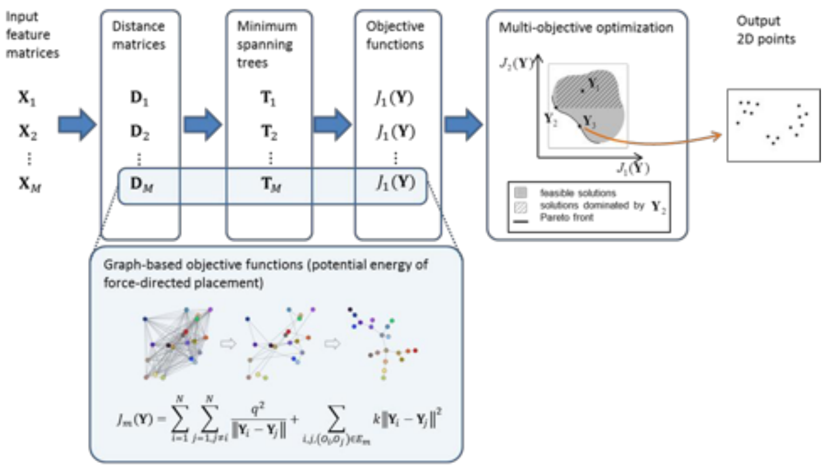


Figure 14 Architectural Overview of the Multi-objective graph formulation approach

Energy data are provided as the input to the visualization method. At the initial stage, the energy data are filtered, so that only the ones within a desired time period are kept. The desired range is set by the operator through the GUI. The filtered data are used to extract multiple features describing different aspects of user behaviour. Hereby, two user descriptors are extracted, namely the Production Histogram Descriptor (PHD) and the Consumption Histogram Descriptor (CHD) for each prosumer.

The individual user descriptors are used to define distances between users, in order to construct individual user graphs. In order to compare two descriptors, either of the PHD or the CHD type or even both together and compute a distance between them, the L1 distance metric is used. Given two vectors $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \dots, y_n)$, both of size N, hereby, two histograms both of N bins, the L1 distance is calculated as follows:

$$d_{L1} = \sum_{i=1}^{N} |x_i - y_i| \quad (2)$$

For the comparison of two histograms descriptors, which may be of different sizes, the one with the fewer bins is padded with trailing zeros, so that the two descriptors become of equal sizes.

The extraction of both descriptors depends on the selection of time period within energy trade-off is considered. This time period is interactively selected by the operator. Thus, the operator, by selecting different time periods, can control the descriptors used and have an overview of the energy behaviour of the prosumers.

For each of the two descriptors a distance matrix is created, which is a square matrix (two dimensional array) containing the distances, taken pairwise, between the elements of a set. When distance is defined as a metric, as for L1 distance, the distance matrix satisfies properties related to the defining properties of a metric. That is, if $M = (x_{ij})$ with $1 \leq i, j \leq N$ is a distance matrix for a metric instance, then:

- The entries on the main diagonal are all zero;
- All of the off-diagonal entries are positive ($x_{ij} > 0 \; if \; i \neq j$);
- The matrix is a symmetric matrix ($x_{ij} = x_{ji}$), and;
- For any i and j, $x_{ij} \leq x_{ik} + x_{kj}$ for all k (the triangle inequality).

Multi-Dimensional Scaling (MDS) is used to display the information contained in a distance matrix. MDS algorithm is used to go from a proximity matrix (similarity or dissimilarity) between a series of N objects to the coordinates of the same objects in a p dimensional space. If p is one or two, then 2D scatter plots of the resulting points are possible.

The data to be analysed is a collection of ***I*** objects, on which a distance is defined,

$\delta_{i,j} := distance$ between i-th and j-th objects. The distances are the entries of the proximity matrix

$$\Delta := \begin{pmatrix} \delta_{1,1} & \delta_{1,2} & \cdots & \delta_{1,I} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{I,1} & \delta_{I,2} & \cdots & \delta_{I,I} \end{pmatrix} \quad (3)$$

The goal of MDS is, given Δ, to find ***I*** vectors $x_1, x_2, \dots, x_n \in R^N$ such that $\left\| x_i - x_j \right\| \approx \delta_{i,j}$ for all $i, j \in 1, \dots, I$, where ||.|| is a vector norm. In other words, MDS attempts to find an embedding from the ***I*** objects into R^N such that distances are preserved. There are various approaches to determining the vectors x_i, usually MDS is formulated as an optimization problem, where $(x_1, x_2, \dots, x_I)$ is found as a minimizer of some cost function, for example $\min_{x_1, x_2, \ldots, x_I} \sum_{i<j} (||x_i - x_j|| - \delta_{i,j})^2$. A solution may then be found by numerical optimization techniques.

Figure 15 shows three different MDS graphs, two individual graphs with respect to each descriptor for consumption and production separately and a MDS graph for prosumers with similar production and consumption behavior as well. As it is evidenced by the graph, prosumers with similar production and consumption behavior are in the same density area in the three MDS graphs.

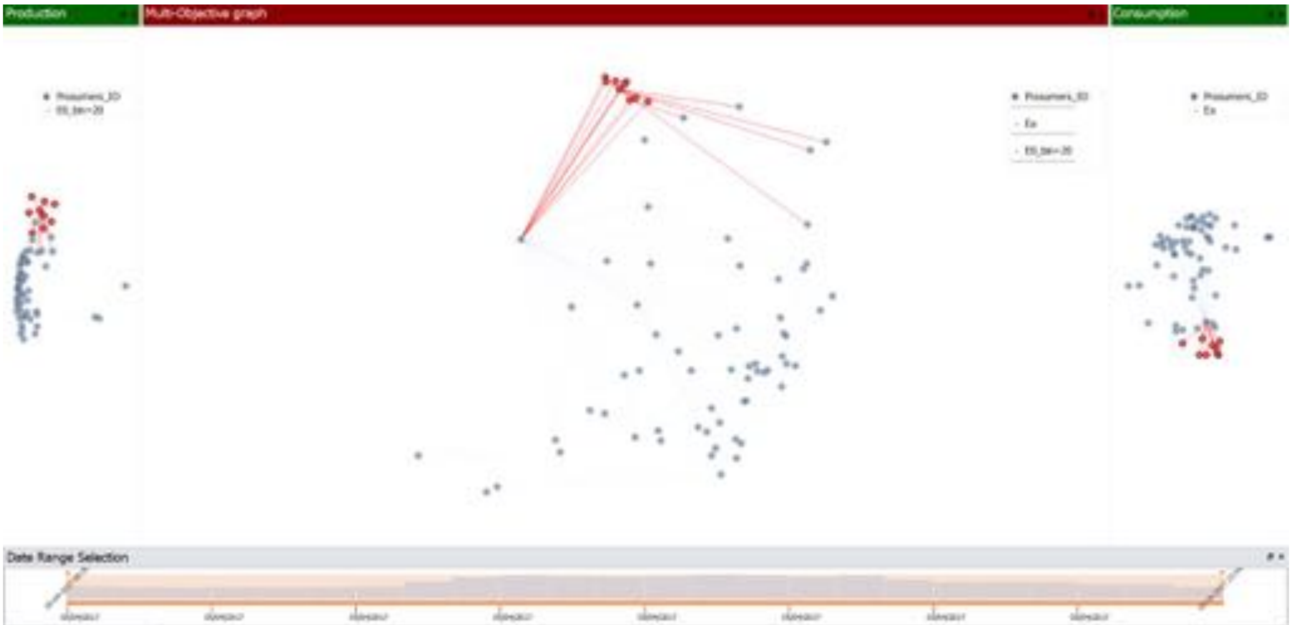


Figure 15 Multi-objective graph formulation approach for indentifying common groups based on the data from ASM TERNI pilot site (examined features are considered the consumption and the production)

Before proceeding with the analysis of results in the context of eDREAM use cases, it is worth mentioning that the aforementioned techniques have been introduced in the NEMESYS FP7 eu project and they have been adapted to the needs of the eDREAM project. As the development activities progressing and advancing, other visual analytics techniques are going to be also examined.

5 Analysis of results and relevance with eDREAM Use Cases

5.1 Relevant Use Cases

Within the eDREAM context, the visualization framework targeting the Demand Response sector will cover the needs of two pilot sites, the Terni ecosystem, which comprises a cluster of DERs operating as a microgrid and offering services to the local DSO and the KiWi Power, which is an utility provider company having a portfolio of prosumers and VPPs. The relevant business scenarios and the respective use cases have already been presented in the deliverable "D2.2: Use Case Analysis and application scenarios description V1", but for the sake of completeness and relevance of the results, high-level descriptions are also mentioned below. In the following Table 4, the use cases that include visual exploration techniques and multi-objective analysis, in order to improve decision making and DR strategies application, are presented:

Table 4 eDREAM Use Cases based on Visual Analytics techniques

eDREAM Use Cases Inventory
HL-UC01: Prosumer DR flexibility aggregation via smart contract
• HL-UC01_LL-UC01: Prosumers enrolment in demand response programs
• HL-UC01_LL-UC02: Contract setting
• HL-UC01_LL-UC05: Flexibility request
• HL-UC01_LL-UC07: Flexibility acceptance
• HL-UC01_LL-UC08: Flexibility provisioning
HL-UC02: Prosumers registration with the energy trading platform
• HL-UC02_LL-UC01: Prosumers registration with the energy trading platform
• HL-UC02_LL-UC04: Transactions validation and financial settlement
HL-UC03: VPP in Energy Community
• HL-UC03_LL-UC01: Prosumers profiling and clusterisation
• HL-UC03_LL-UC02: VPP capability evaluation
• HL-UC03_LL-UC03: VPP for Reserve services
• HL-UC03_LL-UC04: VPP for Frequency services
• HL-UC03_LL-UC05: VPP export evaluation
• HL-UC03_LL-UC06: VPP for Wholesale market – Intraday trading
• HL-UC03_LL-UC07: VPP for Imbalance market

5.2 Dataset Description

The dataset, which is used for testing the visualization techniques, consists of measurements derived from ASM Terni pilot site and concerns the samples from 137 different prosumers collected along the whole 2017. For the sake of accuracy of results, the dataset has been pre-processed and the consumers with incomplete measurements were not included. More specifically, the measurements from month April were used and the data were sampled every 15 minutes. The features that were taken into account during the calculations are the following:

- Prosumer ID;
- Timestamp of the corresponded measurement, where the granularity of the measurements is on 15 minutes' time interval;
- Coordinates of the prosumers (latitude, longitude);
- E_a: The active energy absorbed by the prosumer (absorbed by the grid);
- $E_0 = E_{imm}$: The active energy, which is delivered from the prosumer to the grid (injected to the grid);
- Q_1: The absorbed inductive reactive energy;
- Q_3: The injected inducted reactive energy.

An overview of the dataset is illustrated in the following Figure 16 Dataset Overview:

	Prosumers_ID	Datetime	start_coords_x	end_coords_x	start_coords_y	end_coords_y	Ea	E0	Q1	Q3
8353	514621028979	2017-04-05 00:00:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	64.0	0.0	0.0	0.0
8354	514621028979	2017-04-05 00:15:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	40.0	0.0	0.0	0.0
8355	514621028979	2017-04-05 00:30:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	71.0	0.0	0.0	0.0
8356	514621028979	2017-04-05 00:45:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	73.0	0.0	0.0	0.0
8357	514621028979	2017-04-05 01:00:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	56.0	0.0	0.0	0.0
8358	514621028979	2017-04-05 01:15:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	76.0	0.0	0.0	0.0
8359	514621028979	2017-04-05 01:30:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	54.0	0.0	0.0	0.0
8360	514621028979	2017-04-05 01:45:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	70.0	0.0	0.0	0.0
8361	514621028979	2017-04-05 02:00:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	48.0	0.0	0.0	0.0
8362	514621028979	2017-04-05 02:15:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	42.0	0.0	0.0	0.0
8363	514621028979	2017-04-05 02:30:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	49.0	0.0	0.0	0.0
8364	514621028979	2017-04-05 02:45:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	54.0	0.0	0.0	0.0
8365	514621028979	2017-04-05 03:00:00	42.57013341546...	42.57023341546...	12.65085352094...	12.65105352094...	63.0	0.0	0.0	0.0

Figure 16 Dataset Overview

5.3 Heat Map visualization approach

As already mentioned in the previous section, the Heat Map visualization approach can provide an overall representation of the geographical distribution of customers along with a corresponding indication of the level of consumption. The following Figure 17, Figure 18 and Figure 19 illustrate how the distribution changes depending on the time interval and the observed congestion.

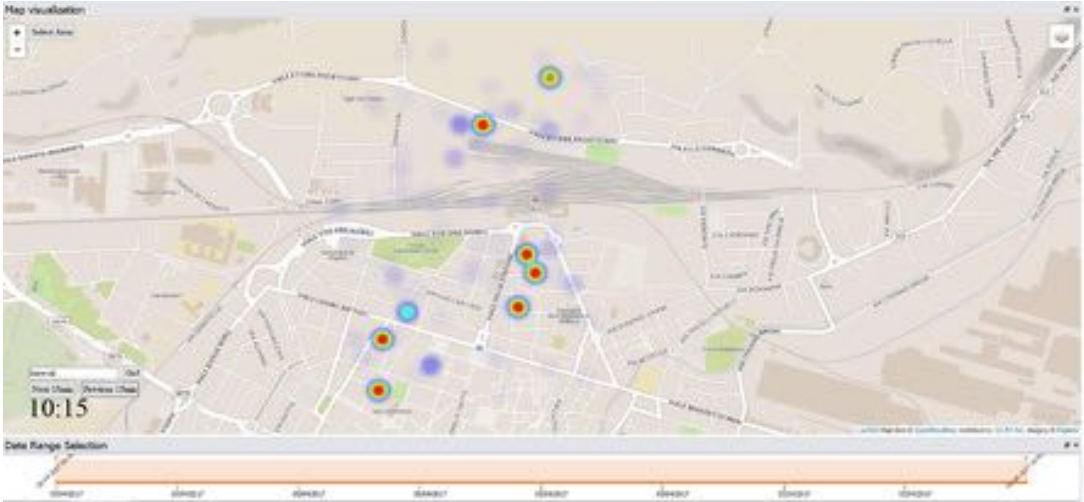


Figure 17 Heat Map representation of prosumers consumption – without significant congestion (TERNI area)

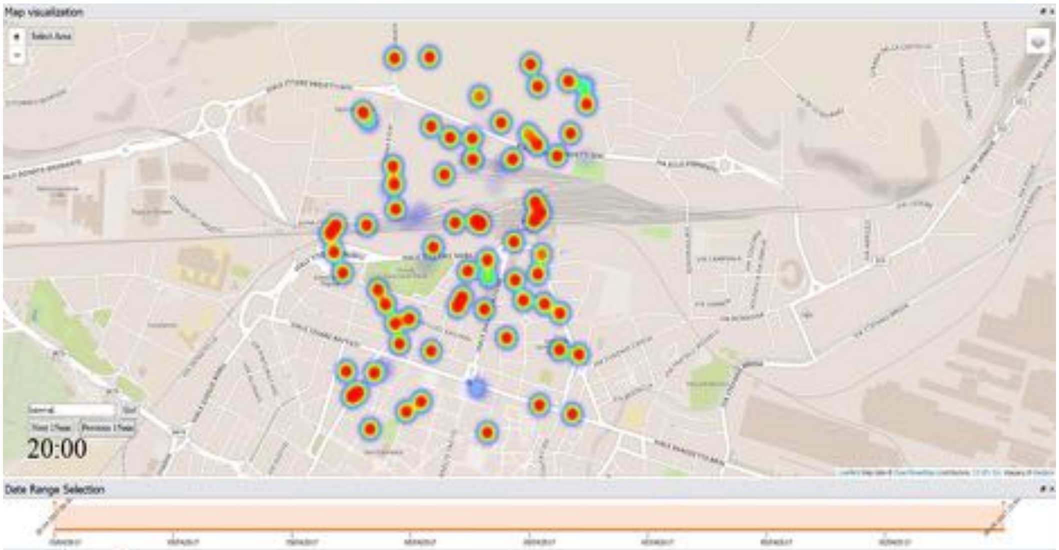


Figure 18 Heat Map representation of prosumers consumption – with significant congestion (TERNI area)

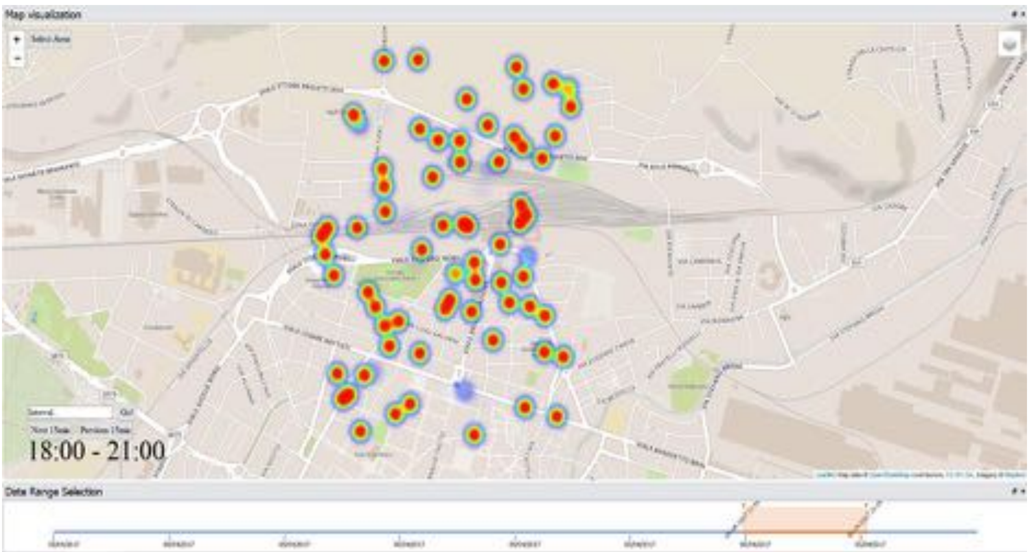


Figure 19 Heat Map representation of prosumers consumption – with significant congestion between the time interval of 18:00-21:00 pm (TERNI area)

5.4 K-partite graph formulation approach

The results from this visualization approach can be used as a preliminary evaluation of the common patterns within the portfolio concerning different features such as consumption, flexibility etc. This technique identifies the values of the selected feature that are more common among the prosumers. Thus, it can provide insights to the aggregator for the prevailing values of specific features and the matching with the prosumers of the portfolio. This representation can be used in the case of a flexibility request, where the aggregator can identify which prosumers can provide relative levels of flexibility. The results in the following Figure 20, Figure 21 and Figure 22 concern the feature of consumption on different days in April 2017.

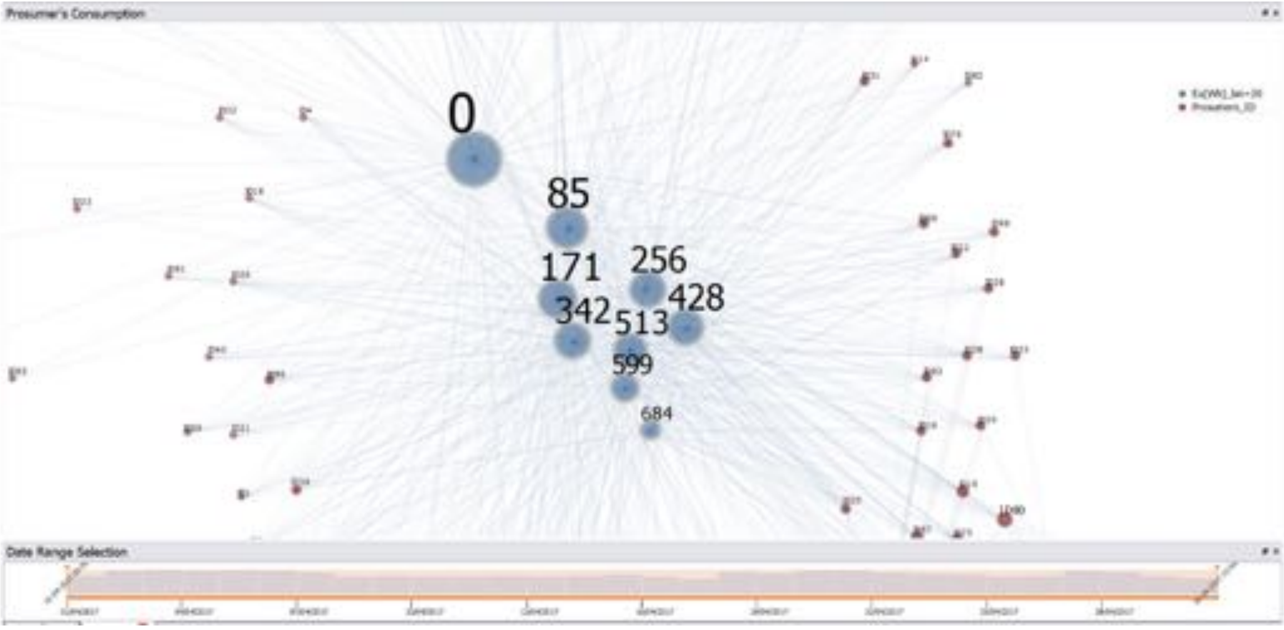


Figure 20 K-partite graph formulation based on prosumers monthly consumption

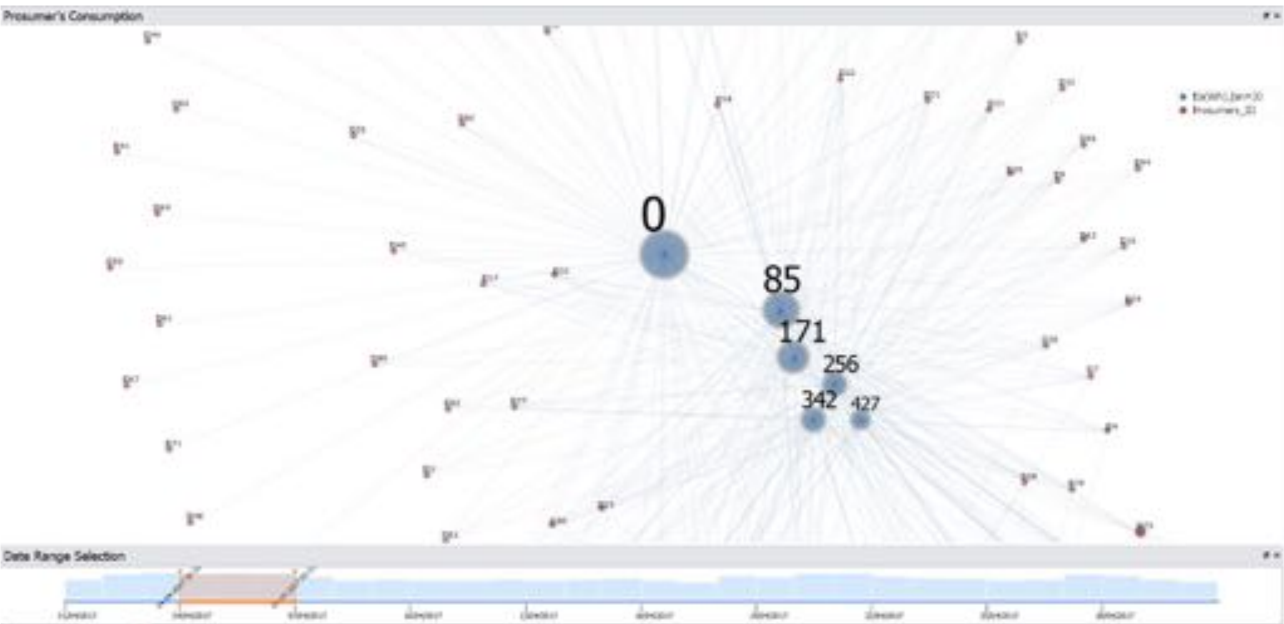


Figure 21 K-partite graph formulation based on prosumers weekly consumption

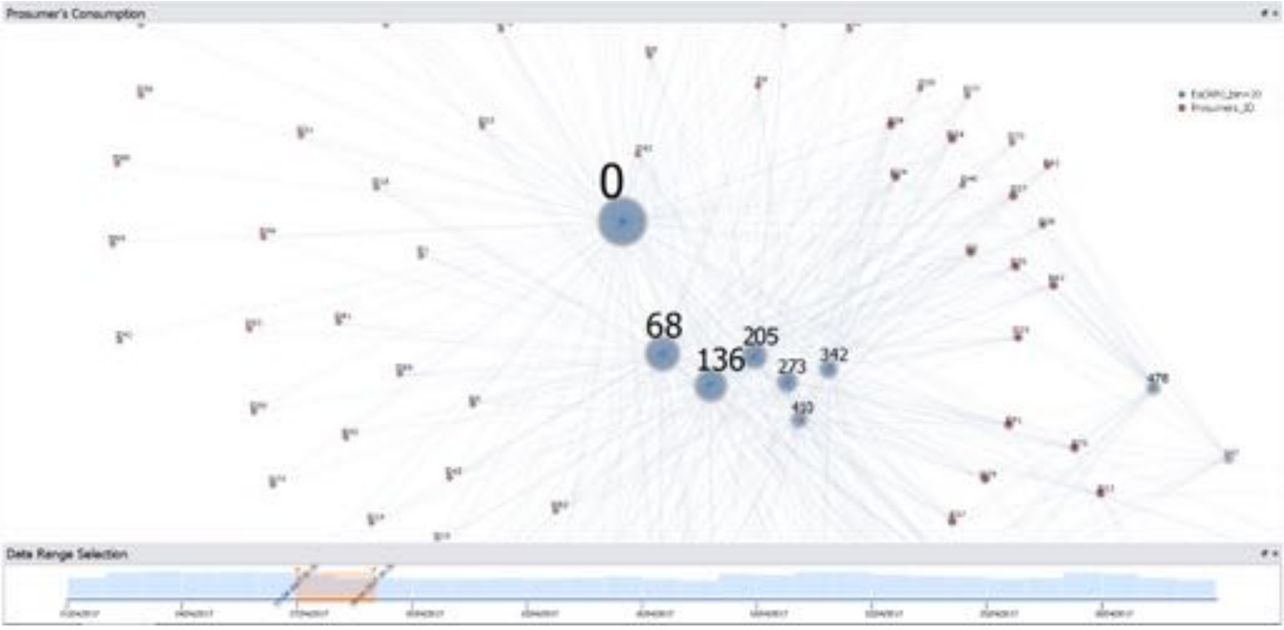


Figure 22 K-partite graph formulation based on prosumers consumption during weekend

As it can be observed, the identified common patterns concerning the selected feature of consumption vary depending on the different time intervals.

5.5 Multi-objective graph formulation approach

This visualization approach is very useful for the correlation between different features, so as to create groups/clusters of prosumers with regard to these features. Therefore, it can be used by the aggregator to handle the VPPs and the other prosumers. As for example, this technique can provide multi-factor insightful correlations concerning the participation of VPPs in ancillary services taking into account features that play an important role each time. The figures below show the patterns resulting from the correlation between the features of consumption and production. The nodes that are close to each other indicate that these prosumers have similar behavior in terms of production and consumption.

Figure 23 Multi-objective graph formulation based on prosumers consumption and production during April

The process of multi-objective optimization, as stated in Section 4.2.2, results in a set of solutions, i.e. the Pareto set, which correspond to various trade-offs among the available modalities. One of these solutions needs to be selected, in order for the corresponding visualization to be presented to the end-user. This solution should represent the best trade-off among the modalities. However, the other solutions are still available for the user, so that he/she can interactively select among them, in order to view different aspects of the visualized dataset.

Figure 24 Multi-objective graph formulation based on prosumers consumption and production during April (identification of common points concerning both consumption and production)

In the Figure 23 and Figure 24, the technique of minimum spanning tree (MST) is employed, in order to create less cluttered visualizations and to reveal the similarities and dissimilarities among the objects, as objects that are similar to each other are connected through small paths on the tree. Since there are M modalities, M minimum spanning trees are constructed, one for each modality. All M trees have the same set of vertices (the multimodal objects), while the set of edges is different for each modality. It can be observed that the prosumers, who have similar patterns both in terms of production and consumption, are represented as nodes close to each other in the middle graph.

Figure 25 Multi-objective graph formulation based on prosumers consumption and production during April (identification of common points concerning both consumption and production)

The graph-analytics platform provides also the capability of creating clusters by employing the k-means or the DBSCAN technique. Thus, by using the k-means process, two clusters have been identified concerning production, consumption and both features together as it is presented in Figure 25.

Figure 26 Histograms of nearby nodes in the Multi-objective graph

In Figure 26, the histograms from nearby nodes are depicted, in order to show that these nodes (prosumers) have similar production values.

Figure 27 Histograms of distant nodes in the Multi-objective graph

On the other hand, the histograms of distant nodes consist of quite different values as it is depicted in Figure 27.

The following Figure 28 presents similar representation as the above concerning the feature of inductive reactive power that is exchanged with the grid. This feature can determine the participation of the prosumers in ancillary services.

Figure 28 Multi-objective graph formulation based on prosumers injected/absorbed inductive reactive energy during April

Other correlations that can provide useful information are those among the KPIs. As for example, the time series of different prosumers KPIs can be considered as the different features of prosumers in the multi-objective optimization approach. These KPIs have been initially presented in the deliverable D2.3. There are still no real values for these indicators, because no corresponding evaluations have been made yet. The following table shows some of these indicators that can be used within this approach.

Table 5 Prosumers KPIs as features in the Multi-objective Visualization approach

Prosumers – related KPIs		
KPI	***Description***	***Metric Unit***
Electricity Savings	Reduction of electricity consumption due to DR (including both shedding and shift periods)	[kWh/DR event]
Peak power reduction	Reduction of the maximum electricity power demand	[kW/DR event]
Economic gain	Overall benefit in national currency (£, €, RON) due to the DR implementation	[euro/DR event]
Share of electrical energy produced by RES/local generation/ CHP	Evaluate the shared energy produced by RES	%
Reduction of green-house gases emissions	reduction of equivalent CO_2 emissions in $kgCO_2eq$ due to the DR implementation	[$kgCO_2eq$/DR event]

6 Augmented Reality Tools

As it is mentioned in the DoA in the description of Task 4.3, the visualization framework will be combined with augmented reality (AR) tools (e.g. Microsoft Hololens), in order to increase the user's capabilities in the energy sector and in the implementation of Demand Response programs. The following sections present some relevant information about these tools.

6.1 Current Techniques

Augmented Reality (AR) is a technology that enhances our physical world by adding layers of digital assets [10]. AR does not replace the real environment with an artificial and virtual one, but instead insert audio, video and graphics to it. AR currently exists in four different types [11]:

- **Markerless AR** – Consist of accelerometer, velocity meter, digital compass and GPS, embedded in the tool or device. This is a widely used for mobile devices location-centric apps such as maps.
- **Marker-based AR** – Uses camera and distinct patterns to produce result. Also known as Image recognition, this system is used by devices to read QR/2D codes.
- **Projection-based AR** – Projects artificial light into real physical environment and allows human interaction. This system combined with laser plasma technology launches interactive hologram.
- **Superimposition-based AR** – Creates a newly augmented view of objects that can be used to replace the initial view.

AR can be used on various devices, such as head-mounted displays, mobile phones and handheld devices. AR consist of a combination of different innovative technology, which includes:

- **General hardware components** – the processor, the display, the sensors and input devices, such as accelerometers, GPS, camera and microphone.
- **Displays** – Optical projection systems, head-mounted displays, eyeglasses, contact lenses, heads up display, virtual retinal displays, EyeTap, Spatial Augmented Reality and handheld displays.
- **Sensors and input devices include** – GPS, gyroscopes, accelerometers, compasses, RFID, wireless sensors, touch recognition, speech recognition, eye tracking and peripherals.
- **Software** – Most of the AR technology developers have their own tailored software. But an Augmented Reality Markup Language (ARML) exists, which is being used to standardize XML grammar for virtual reality. There are also several universal and proprietary software development kits (SDK) which also offer simple environments for AR development. 3D assets may also be constructed using 3D animation and game design tools such as Unity engine, and then copied over to AR SDKs (Software Development Kit).

6.2 Requirements and Interfaces

As the specific features and functionalities of the AR tools are still under planning and development, it is difficult to state exactly, which data these tools will need to access, in order to present useful information to the user.

6.3 Use of AR tools in Demand Response programs

In the context of eDREAM project, AR tools will be utilised to visualize information captured from aerial surveys performed as part of T3.4, as well as data from field components, providing users with the ability to examine a site in real time with visualized information regarding building thermal performance and key DR asset potential. This enables stakeholders and aggregators to quickly and efficiently assess sites for DR service enrolment, or present a site owner or manager with a user-friendly visualization of how DR service enrolment could augment their consumption of energy and reduce costs.

6.4 Hardware Tools

Two main pieces of Augmented reality hardware are currently available, the Microsoft HoloLens and the Daqri smart glasses.

Figure 29 Augmented Reality smart glasses

6.4.1 Microsoft Hololens

The Microsoft HoloLens is a standalone AR headset, developed as part of Microsoft's Mixed reality suite of tools. The HoloLens was designed to be a highly diverse, standalone, user friendly platform. It is designed with integrated access to the Microsoft store, offering UWP (universal windows platform) applications right out of the box, many of which can also be run on a windows 10 PC or tablet. Applications for the HoloLens can also be sourced from 3rd party developers or developed in-house using IDEs such as Microsoft visual studio to suit the application and then be side-loaded onto the device. The below table presents some representative specifications of the Microsoft Hololens.

Table 6 Microsoft Hololens Specifications [12]

Microsoft HoloLens (V1) Specifications	
Optics	See-through holographic lenses (waveguides)
	2 HD 16:9 light engines
	Automatic pupillary distance calibration
	Holographic Resolution: 2.3M total light points
	Holographic Density: >2.5k radians (light points per radian)
Sensors	1 IMU (Inertial Measurement Unit)
	4 environment understanding cameras
	1 depth camera
	1 2MP photo / HD video camera
	4 microphones
	1 ambient light sensor
Human Understanding	Spatial sound
	Gaze tracking
	Gesture input
	Voice support
I/O	Built-in speakers
	Audio 3.5mm jack
	Bluetooth 4.1 LE
	Wi-Fi 802.11ac
	USB Micro-B (file transfer, telemetry and charging)
Power	Integrated Battery with 2-3 hours of active use, and up to 2 weeks of standby time
Processors	Intel Cherry Trail Atom SoC
	Microsoft Holographic Processing Unit (HPU 1.0)
RAM	2GB LPDDR3 (1GB for CPU, 1GB for HPU)
Storage	64GB eMMC
Weight	579g

6.4.2 Daqri Smart Glasses

The Daqri smart glasses were developed as a professional tool for various fields including engineering, medical and military applications, with the main applications being equipment tagging with live data, guided experiences, remote assistance and technical visualisation with BIM (Building Information Modelling) being one of the main use-cases. The specifications of this type of glasses are presented in Table 7 and Table 8.

Figure 30 Daqri Smart Glasses [13]

Table 7 Daqri Smart Glasses Specifications [14]

Hardware	
Weight	Smart Glasses: 335g Compute Pack: 496g
Processor	6th Generation Intel Core m7 Processor (Up to 3.10 GHz) Dedicated vision processing unit for 6-DOF tracking
Optics	Dual LCoS Optical Displays 44° Diagonal FOV Resolution: 1360 X 768 Frame Rate: 90 fps
Connectivity	WiFi 802.11 A/B/G/N/AC 2.4/5 GHz Bluetooth
Battery	Built in rechargeable lithium ion battery 5800 mAh
Storage	64 GB Solid State Drive
I/O Ports	2 USB 3.1 Type C Ports 3.5mm Headphone Jack
Audio	2 Microphones with Active Noise Cancellation
Depth Sensor Camera	Range: 0.4m to 4m Resolution: 640 x 480 Frame Rates: 30, 60, 90 fps
Colour Camera	RGB 1080p HD Camera, 30 fps
AR Tracking Camera	166° Diagonal Wide-Angle Fisheye Lens Resolution: 640 x 480 Frame Rate: 30 fps

Table 8 Daqri Smart Glasses Certifications and Standards [14]

Certifications and Standards	
Agency Certifications	Safety (IEC 60950-1)
	United States (FCC)
	Canada (IC)
	Australia (RCM)
	European Economic Area (CE)
	New Zealand (R-NZ)
Eye Protection	ANSI/ISEA Z87.1 (Eye and Face Protection)
	EN166 1S (Highest Optical Class, Increased Robustness Eye Protection)
Operating Range	Operating Temperature: 0 - 30°C Designed for Indoor and Outdoor Use

7 Conclusions

This report has presented the adopted and tested graph formulation techniques towards the design of a graph analytics platform targeting the challenges of Demand Response sector. The role of the interactive visualization framework was described indicating also the important role of the Visual Analytics techniques regarding the big data handling in the energy sector. The algorithmic and visualization techniques presented in this deliverable will drive the development and implementation of the whole visualization framework. The scope of this deliverable is to provide a first high-level description that covers the issues and techniques related to the development of a graph analytics platform. The correlation of the visualization techniques with the eDREAM use cases and the requirements of Demand Response programs has been also introduced and described. The consolidated description of the platform will be presented in the next version of the deliverable that is reported as "D4.8: Interactive Visualization framework for improving DR strategies V2". Finally, the last section is dedicated to the augmented reality (AR) tools that are going to be integrated with the visualization framework.

References

[1] Ryan, Tim, and Jeffrey S. Vipperman. "Incorporation of scheduling and adaptive historical data in the Sensor-Utility-Network method for occupancy estimation." Energy and Buildings (2013)

[2] J. Thomas and K. Cook, Illuminating the Path: Research and Development Agenda for Visual Analytics. IEEE-Press, 2005.

[3] C. Ware, Information Visualization - Perception for Design, 1st ed. Morgan Kaufmann Publishers, 2000.

[4] J. W. Tuckey, "Exploratory Data Analysis. Addison-Wesley, Reading MA, 1977.

[5] J. J. Thomas and K. A. Cook, A Visual Analytics Agenda," IEEE Transactions on Computer Graphics and Applications, vol. 26, no. 1, pp. 12{19, January/February 2006.

[6] D. A. Keim, J. Kohlhammer, and J. Thomas. (2005) Workshop on visual analytics. http://infovis.uni-konstanz.de/events/ws visual analytics 05/

[7] Timothy M. Hansen, Siddharth Suryanarayanan, Anthony A. Maciejewski, Howard Jay Siegel and Arun V. Modali, "A Visualization Aid for Demand Response Studies in the Smart Grid", Elsevier Inc., April 2015, Vol. 28, Issue 3.

[8] https://medium.com/@FILWD/from-data-visualization-to-interactive-data-analysis-e24ae3751bf3

[9] Thomas, J., Cook, K.: Illuminating the Path: Research and Development Agenda for Visual Analytics. IEEE-Press (2005).

[10] Nayyar, A. et al. (2018) 'Virtual Reality (VR) & Augmented Reality (AR) technologies for tourism and hospitality industry', International Journal of Engineering & Technology, 7(2.21), pp. 156. doi: 10.14419/ijet.v7i2.21.11858.

[11] Yuen, S.C., Yaoyuneyong, G. and Johnson, E. (2011) 'Augmented Reality: An Overview and Five Directions for AR in Education', Journal of Educational Technology Development and Exchange, 4(1), pp. 119-140. doi: 10.18785/jetde.0401.10.

[12] https://docs.microsoft.com/en-us/windows/mixed-reality/hololens-hardware-details

[13] https://daqri.com/products/smart-glasses/

[14] https://assets.daqri.com/documents/DAQRI_Smart_Glasses_Specifications.pdf