Final Report

Complex Opinion Visualization

Group A	Group B
N.J. de Rijk	M.S. van Genderen
n.j.derijk@students.uu.nl	*m.s.vangenderen@students.uu.nl*
P.J. Hollander	A.A.L. van Uden
p.j.hollander@students.uu.nl	*a.a.l.vanuden@students.uu.nl*

Supervisor: T. Mtsendlintze (Tamara)

Research based on: Applying argumentation to structure and visualize multi-dimensional opinion spaces
Published on: 6 December, 2018
Journal: Argument Computation, vol. 10, no. 1, pp. 23-40, 2019
Authors: Betz, G., Hamann, M., Mchedlize, T., von Schmettow, S.

Date: February 04, 2022

1 Abstract

The visualization of opinions is difficult as opinions are complex compositions of multiple dimensions. To investigate complex opinions, OpMAP was created. This paper presents research conducted into how OpMAP can be extended to support complex, high-dimensional data and subsequently the interpretation of complex opinions. This research is separated two distinct but connected parts. The first part (Part A) is focused on the implementation of complex opinions to investigate the possibility of OpMAP featuring complex, high-dimensional data in regard to the religion topic of the European Values Study (EVS) data set. The EVS data set was reduced in dimensionality with t-SNE and subsequently clustered with DBSCAN. However, the final integration with OpMAP was unsuccessful. The second part (Part B) investigates the improvement and subsequent usability of OpMAP for cluster interpretation. Development was focused on improving readability and usability of the interface by improving the organization of the available information. This was done by restructuring interaction mechanics of cluster selection, adding detailed statistics of the data, and changing the use of color. With these changes, a user study was conducted to find out whether the participants are able to interpret the visualization. The results of this user study show that the participants were not able to do this well. Potential causes are the limited information about the clustering and limited interaction of the interface.

2 General Introduction

Every human has opinions. They are at the center of what defines us and they influence our behaviour every day. In order to understand humans better, one must look at the opinions of humans. During the last two decades, researchers have focused on investigating opinions and other relevant concepts such as sentiment and attitude in studies categorized as sentiment analysis or opinion mining [25]. Although often used interchangeably, opinion mining differs from sentiment analysis as a sentiment is a feeling one person may have and an opinion is a view of a person which may include sentiment [25].

Research on opinion mining has been predominantly focused on text mining and review mining [12] [15] [36]. These are tasks where topics and sentiments are extracted from a corpus of reviews. The outputs of these tasks are often the topics of the concerned texts or the sentiments a user or person holds towards a topic. However, little attention has been paid to visualizing opinions [15]. The reason for this is that opinion visualization is difficult as opinions are complex compositions of multiple dimensions [11] [29].

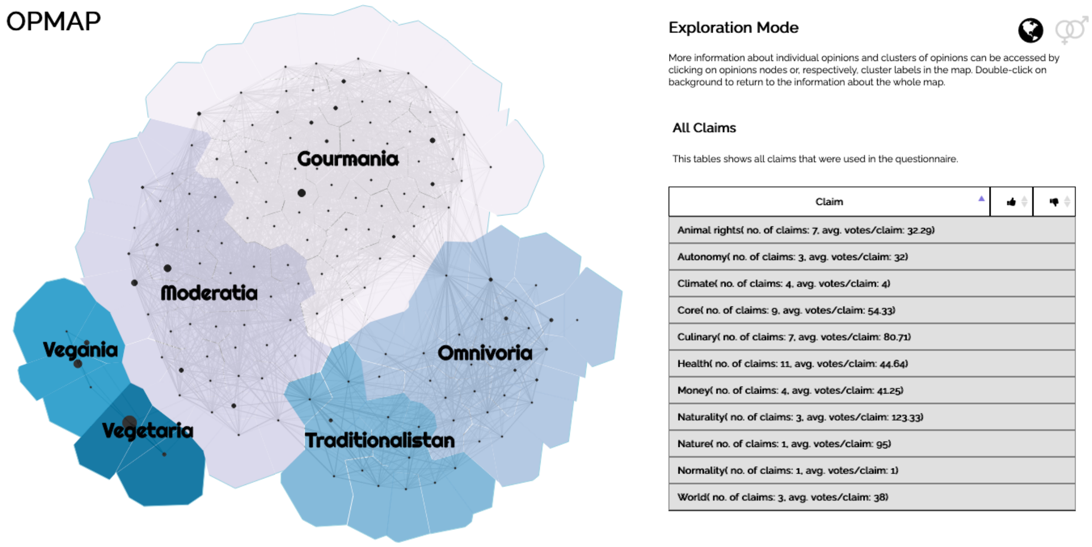


Figure 1: Screenshot of the original OpMAP interface

To investigate complex opinions, OpMAP was created [11]. OpMAP is ”a tool for visualizing large scale, multi-dimensional opinion spaces as geographic maps” [11]. OpMAP visualizes opinions with a weighted graph and a force-directed layout as shown in figure 1. Individual opinions are structured in opinion vectors to allow for quantitative comparison between opinions. Each node represents an opinion vector and the edges between nodes the extent to which they cohere to each other. Consequently, the set of nodes is clustered to generate so-called 'countries' representing groups of like-minded individuals (e.g. nodes that share similar opinions)[11]. The OpMap visualization technique has shown that opinions of a structured argumentation can be visualised using an opinion network [11]. However, the modeling of opinions through opinion vectors has only been tested on binary data (i.e. yes/no answers) and a limited amount of data points.

To explore the applicability of techniques given in OpMAP to other types of data, the present study attempts to expand the OpMAP interface to aid in the interpretation of large, non-binary data. The general research question (RQ) that will be investigated is defined as follows:

RQ: *How can OpMAP be extended to support the interpretation of complex opinions?*

Complex opinions are defined as opinions that include non-binary arguments and are constructed of a large set of arguments. An example of a data set that contains complex opinions is the European Values Study (EVS) data set [19]. This data set features different topics with different questions. The choice was made to investigate one topic specifically as investigating opinions on a multitude of topics is impossible at this stage. Therefore, the chosen topic to investigate further is religion as there are 21 questions pertaining to religion which would indicate one could have a complex opinion about the topic. Also, these questions are of different types (scalar, yes/no, categorical) which further demonstrates the complexity of an opinion one could have one the topic.

The research question is further divided in to several sub-questions to structure the research into two separate, but connected, parts. The first part (Part A) is focused on the implementation of complex opinions to investigate the possibility of OpMAP featuring complex, high-dimensional data by involving the religion topic of the EVS data set. The sub-question (RQa) that guides this part is:

RQa: *Can OpMAP be extended to feature complex, high-dimensional data?*

Additionally, the second part (Part B) investigates the improvement and subsequent usability of OpMAP for cluster interpretation. The sub-question (RQb) that guides this part is:

RQb: *How can OpMAP be improved to facilitate better understanding of clustering presented in OpMAP?*

The remainder of this project report is structured as follows. Section 3 reports on the investigation of Part A of the research (i.e. opinion mining on the EVS data set). Section 4 reports on the investigation of Part B of the research (i.e. the usability and improvement of OpMAP). Both Section 3 and 4 have further defined their own sub-questions to guide the overall research questions. Also, their respective methodology, results, and conclusions are presented in their respective sections. Section 5 discusses the overall results of the study and presents its general conclusions.

3 Preparation and analysis of EVS dataset

3.1 Introduction

Our research concerns complex opinion visualization. Since opinions are complex compositions of multiple dimensions, they are hard to visualize. At this point in time, research is primarily concerned with visualizing reviews in which opinions are embedded [12, 15, 29]. Chen et al. proposed that future work should be aimed at visualising complex opinions (i.e. multi-dimensional) [15].

Previous research has attempted to create tools for visualising complex opinions using geographic maps [11, 15]. Recent improvements have enabled the integration of geographic maps with opinion visualisation developing so-called "opinion maps" [11]. However, these opinion maps have only been created for relatively small data sets and usually one type of variable (e.g. boolean). Therefore, the technique is in its infancy. To build on this work, we aim to use this technique with larger and more complex data sets. Larger and more complex data sets are inherently more difficult to visualise as they include more variables of different types. Such characteristics increase the necessity of data filtering, data reduction, and clustering throughout the visualisation process [11]. We aim to verify the usability of the opinion map technique on the 2017 version of the European Value Study (EVS) data set [19]. This data set contains the opinions on various topics of more than 24.000 citizens distributed over 34 countries in Europe.

The research question that will guide Part A is: *"How can OpMAP be extended to support the interpretation of complex opinions?"* The sub-questions that guide this part of the research are:

- *If we model the answers of survey participants as multi-dimensional vectors (i.e. opinions), does this space have a pronounced clustering?*
- *Can we visualize this clustering in 2D?*
- *How can the clusters be interpreted?*
- *How can we integrate the results with OpMAP?*

3.2 Literature Review

Dimensionality reduction (DR) transforms high-dimensional data (i.e. a multitude of variables) into a meaningful representation of reduced dimensionality [33]. This way the granularity is not completely lost and is reduced to a dimension a human is capable of understanding. And since the goal of the research is to investigate similar opinions on religion and clustering them together, dimensionality reduction is applied to achieve two dimensions. Afterwards it is possible to visualize them in a way that humans can understand.

In essence, DR techniques convert high-dimensional data sets of form of $X = x1, x2, ..., xn$ into two or three dimensions where $Y = y1, y2, ..., yn$ and Y is a map of the low-dimensional representation of the data [32]. The functional goal of DR techniques is to preserve as much of the structure of the data as possible. Each DR technique has its own algorithm for structure preservation. Van der Maaten, Postma, Van den Herik (2009) present a taxonomy indexing the DR techniques [33] and their structure preservation techniques. The taxonomy can be observed in Figure 2.

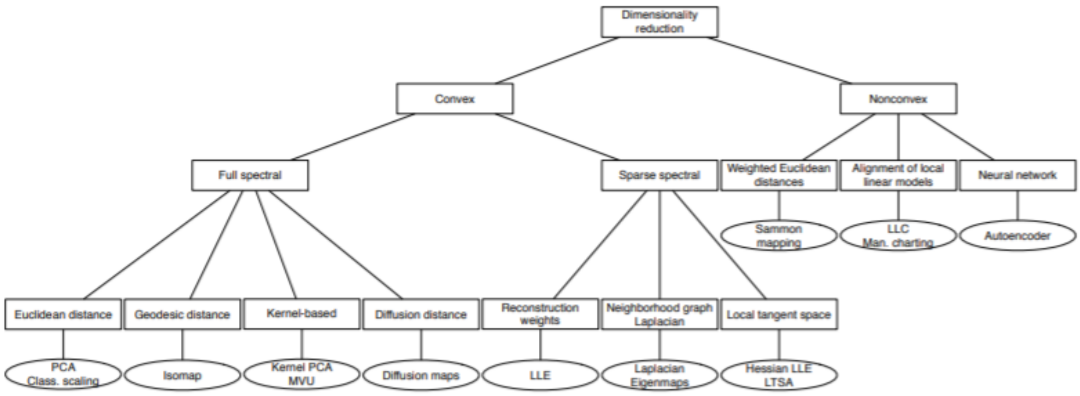


Figure 2: Taxonomy of dimensionality reduction techniques

In essence, each DR technique has its own structure preservation technique that is either focused on preserving the local or global structure. Techniques such as Principal Components Analysis (PCA) [22] attempt to preserve the global structure of the data by distancing points that are dissimilar based on a linear distance between them. There are also techniques such as t-distributed Stochastic Neighbor Embedding (t-SNE) [33] which preserve the local structure of the data by looking at the non-linear distance of pairwise similarities between points.

When investigating the possible DR techniques that preserve the local structure of the data, seven techniques were considered that were also presented in the taxonomy in Figure 2. These DR techniques were: (1) Sammon mapping [27], (2) curvilinear components analysis [17], (3) Stochastic Neighbor Embedding [21], (4) Isomap [30], (5) Maximum Variance Unfolding [34], (6) Locally Linear Embedding [26], and (7) Laplacian Eigenmaps [10].

In recent years, several studies have investigated the performance of non-linear DR techniques on artificial and real-world data sets and they found that these techniques only perform well on artificial data sets [21, 24]. For this reason, Van der Maaten / Hinton (2002) developed t-SNE, an improved iteration of their Stochastic Neighbor Embedding [33]. This is the DR technique that we will be using on the real-world, high-dimensional EVS data set.

In essence, t-SNE functions by performing a binary search for the value of sigma that produces a probability distribution with a fixed perplexity (i.e. a value provided by the user). Perplexity is defined as a target number for the number of neighbors for a central point. This function is performed in high-dimensional space and afterwards these points are mapped to a low-dimensional space by using a t-distribution with a single degree of freedom. This t-distribution is optimized with the Kullback-Leibler divergence. This divergence calculates a gradient for each point in the low-dimensional space where each point is pulled or pushed into a certain direction in space [33].

3.3 Methodology

To answer the main research question, it is imperative to visualize and cluster the complex opinions of the EVS data set. This section highlights which steps were performed to visualize the complex opinions of the EVS data set and what alternatives were available.

3.3.1 Preparatory steps

Data collection The EVS data set was collected from the European Values Study website [19] and the version from 2017 was used.

Data cleaning and preparation The EVS data set has questions on many different topics. Since this study is only focused on opinions on religion, questions were selected by hand that pertained to religion. After manual selection, 21 questions were chosen, namely: *v6, v9, v36, v51, v52, v53, v54, v55, v56, v57, v58, v59, v60, v61, v62, v63, v64, v93, v115, v134, v196*.

The data set consists of various types of questions and answers are of varying types (boolean, categorical and ratio). They therefore had to be converted to the appropriate variable types when loaded into a *Pandas DataFrame*.

10 questions (v6, v36, v53, v54, v55, v63, v64, v115, v134, v196) with 4 different numerical scales for answers, were normalized into a range of 0-1.

Rows containing questions that were answered with *8* (i.e. don't know) or *9* (i.e. no answer / not available) were dropped as it is expected that every participant answers all questions from the survey for completeness' sake [19].

Categorical questions (v52, v56, v62) were encoded as is.

8 Yes / No questions (v9, v51, v57, v58, v59, v60, v61, v93) were encoded as boolean variables.

3.3.2 Visualization

Exploratory Data Analysis To understand the EVS data set that resulted from the data cleaning and pre-processing tasks, an exploratory data analysis was performed. The resulting Pandas DataFrame has 31.574 rows and 21 variables (10 scalar in range [0-1], 3 categorical and 8 boolean [True/False]).

In this Pandas DataFrame, every row represents an entire opinion on religion. Because every row consists of answers to 21 questions, they are complex opinions. In this case, every questions represent a dimension of an opinion. Therefore, the curse of dimensionality arises as the space increases 21-fold which results in a sparse distribution of data [23]. To combat this phenomenon and such that a human is able to understand the opinions still, a dimensionality reduction (DR) technique needs to be applied [33].

Dimensionality reduction Regarding the EVS data set and dimensionality reduction, it is more important to preserve the local structure of the data as herein lies the possibility to look at the (dis)similarity of composite opinions at a local level. For this reason, t-SNE was selected as the dimensionality reduction technique.

Applying t-SNE Having cleaned, pre-processed, and normalized the EVS data set, it is possible to serve as input for the t-SNE algorithm. The t-SNE algorithm has several hyperparameters and the most important ones are defined as follows:

- Perplexity: The number of nearest neighbors used in the learning algorithm. Often in the range [5, 50]. Larger datasets often require a larger perplexity. This hyperparameter can be optimized [6].
- Learning rate: Dictates the step size of the gradient updates (i.e. gradient calculated with Kullback-Leibler divergence). Often in the range [10, 1000] or *auto*. This hyperparameter can be optimized [6].
- Random state: A seed that promotes repeatability [6].
- Number of components: The resulting number of dimensions [6].

Strictly speaking, hyperparameter tuning is not possible for the t-SNE algorithm as the output is a visual. Therefore, hyperparameters were chosen iteratively with the help following criteria:

- The resulting plot must not visualize the nodes as a 'ball'.

- The resulting plot must not visualize the nodes as a 'dense cloud' with just a few outliers.

The reasoning behind these criteria is that the t-SNE algorithm must be able to preserve local structure while separating neighborhoods of points. When nodes are visualized in the same neighborhood (i.e. a ball; no discernible distance between groups of points), the plot is not meaningful. When nodes are primarily visualized on top of each other (i.e. a dense cloud; no discernible distance at all between points), the plot is not meaningful [6].

Having iterated through possible combinations of hyperparameters and having reviewed the visualized plots, the resulting hyperparameters were as follows:

- Perplexity = 125
- Learning rate = 1500
- Random state = 234
- Number of components = 2

3.3.3 Clustering

After reducing the dimensionality of the EVS data set to 2D, an unsupervised clustering algorithm must be applied to achieve clusters of similar opinions. The clustering algorithm must be unsupervised as there is no ground truth available (i.e. there are no labels regarding the actual opinion of people in the data set).

The following unsupervised clustering algorithms were considered:

- K-Means: This algorithm identifies *k* number of centroids and allocates nodes to the nearest cluster based on the average distance to the centroids [4].
- Agglomerative Clustering: A bottom-up hierarchical clustering algorithm that assigns each node to its own cluster and the pairs of nodes are merged/agglomerated until the desired amount of clusters is reached [2].
- Density-based spatial clustering of applications with noise (DBSCAN): A density-based algorithm that finds the highest density areas within the given space and clusters points together around the high density areas while marking outliers that lie in low-density areas [3].
- Ordering Points To Identify Cluster Structure (OPTICS): A soft clustering algorithm similar to DBSCAN. OPTICS is an iteration upon DBSCAN as

it adds the concepts of *core distance* and *reachability distance*. Core distance pertains to the minimum value of a radius required to classify a given point as a core point. The reachability distance defines the maximum distance between points based on the euclidean distance if one of them is a core point. OPTICS does not yield a hard assignment of clusters but only produces a plot based on the reachability distance so that the clusters can be assigned manually [5].

Having reviewed these four unsupervised clustering algorithms, the choice was made to use DBSCAN. The reason for this is that t-SNE has preserved the local structure of the data by looking at pairwise distances of the data. Therefore, only the distance between points matters and not between clusters. If the K-Means algorithm would be applied, it would assign nodes to clusters based on the average distance to centroids rather than looking at the individual nodes. Agglomerative clustering was not performed because we did not know how many clusters were optimal to interpret the clusters. OPTICS was not performed because it was not feasible to assign clusters by hand given the timeframe of the project. And DBSCAN is best suited for the data given that DBSCAN looks at the density of points (i.e. looking at local structure) and marking outliers that lie in low-density areas. This seemed to be the most robust technique for this research.

Applying DBSCAN Strictly speaking, hyperparameter tuning is not possible for the DBSCAN algorithm as the output is a visual. Therefore, hyperparameters were chosen iteratively with the help of the following criteria:

- There must be clear separation between clusters.
- The number of clusters can not be too low as more opinions on religions exist than the number of religions there are.
- The number of clusters can not be too high so that they can be investigated manually. Albeit that the Silhouette Coefficient is used to investigate intrinsic measures of the clusters, manual investigation is used to validate the Silhouette coefficient.

The DBSCAN algorithm has several hyperparameters and the most important ones are defined as follows:

- Epsilon-value: The maximum distance between two samples for one to be considered as in the neighborhood of the other. This is not a maximum bound on the distances of points within a cluster [3].

- Minimum amount of samples: The number of samples in a neighborhood for a point to be considered as a core point. This includes the point itself [3].

Having iterated through possible combinations of hyperparameters and having reviewed the visualized plots, the resulting hyperparameters were as follows:

- Epsilon = 5
- Minimum amount of samples = 150

Cluster quality evaluation To evaluate the quality of clustering resulting from DBSCAN, the silhouette coefficient is applied and the points in the cluster are investigated by looking at the data. The silhouette coefficient is a method that evaluates the inter-cluster quality (i.e. the separation between clusters in a given space). The silhouette coefficient ranges from [-1, 1] where -1 suggests that clusters have been assigned wrongly, 0 suggests that the distance between clusters is not significant and is not distinguishable, and 1 suggests that clusters are separated significantly from each other and distinguishable.

By manually looking at the data, we are able to verify the intra-cluster quality (i.e. coherence).

3.4 Results

Visualization After applying the t-SNE algorithm to the cleaned and pre-processed EVS data set, plots were made using the Python Seaborn library. The resulting plot can be observed in Figure 3. As can be seen, the t-SNE algorithm has assigned data points into a 2D space with adequate spacing between nodes and groups of points.

To get an idea of how the data was partitioned in the 2D space, a visualization was made to show the data points by religion. The resulting plot can be observed in Figure 4. As can be seen, religions have the tendency to clump together. In this sense, the outliers are *Muslim*, *Evangelical*, and *Hindu*.

Based on these plots and the criteria posed in Section 3.3, it can be said that the criteria are being met. More specifically, the resulting plots are not visualized as a ball and the resulting plots are not visualized as a dense cloud. Therefore, the plots are meaningful enough to continue with.

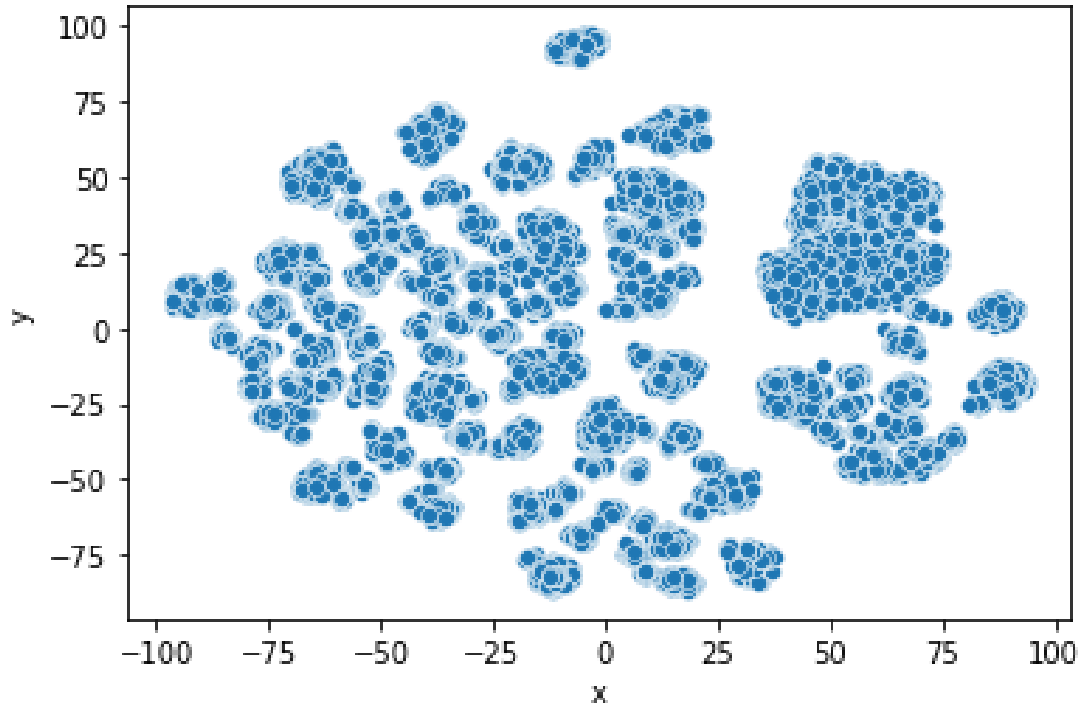


Figure 3: Dimensionality reduction output of EVS data set by t-SNE

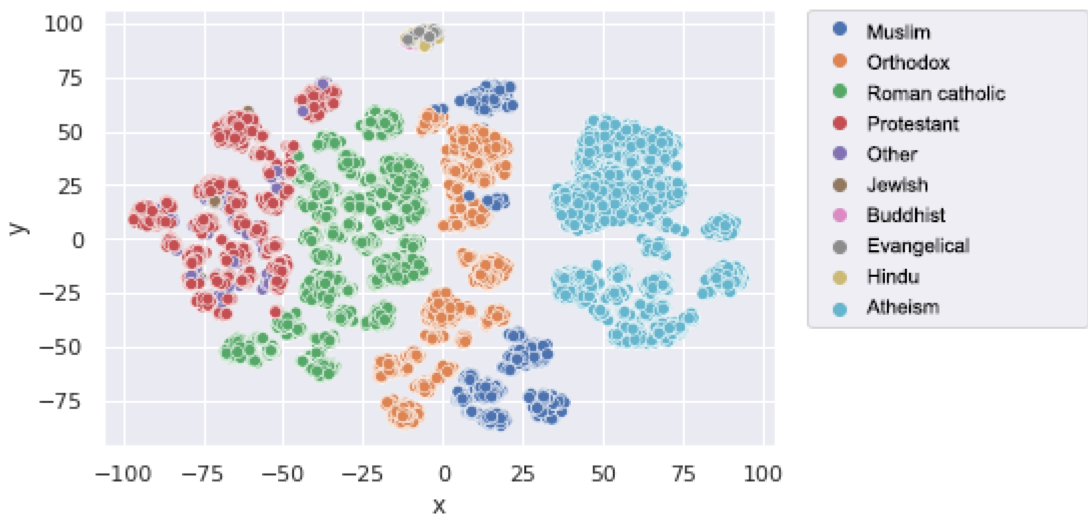


Figure 4: Dimensionality reduction output of EVS data set by t-SNE grouped by religion

Clustering After reducing the dimensionality of the EVS data set to 2D with t-SNE, the DBSCAN algorithm was applied with the hyperparameters as mentioned in Section 3.3. The resulting plot can be observed in Figure 5.

Based on this plot and the criteria posed in Section 3.3, it can be said that the criteria are being met. More specifically, there exists clear separation between clusters, the number of clusters is not low (i.e. 45), and the number of clusters is not too high (i.e. it is feasible to investigate 45 clusters manually).

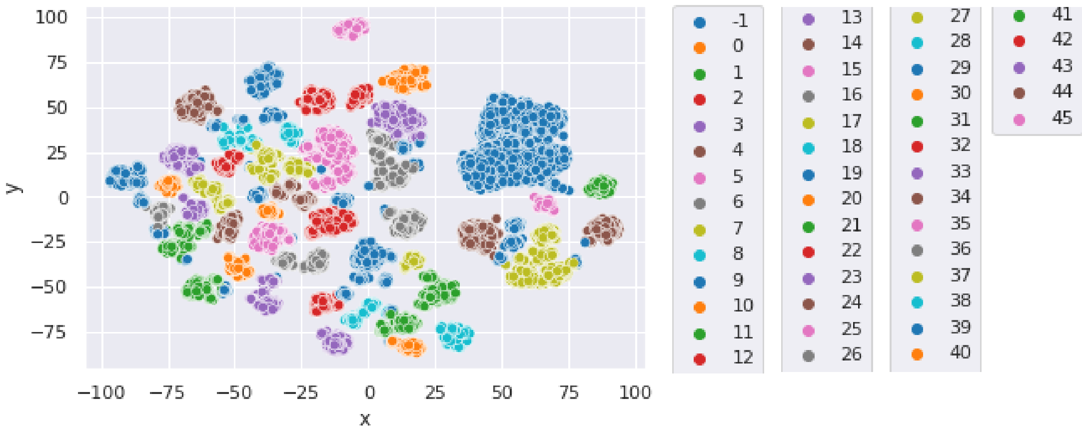


Figure 5: Unsupervised clustering algorithm output of EVS data set by DBSCAN

Cluster Quality To investigate the inter-cluster quality of the clusters generated by DBSCAN, the Silhouette Coefficient [7] was calculated.

The resulting Silhouette Coefficient was: 0.487. This means that the clusters generated by DBSCAN are separated significantly from each other such that they can be easily distinguished.

Clustering Interpretation A starting point for investigating the clustering generated by the DBSCAN algorithm would be to investigate religions as a whole. We started with the assumption that people belonging to specific religions would have the same views on religion. However, when comparing Figure 4 with Figure 5, this assumption can be rejected. It can be seen that groups of points clumped togetherby t-SNE such as Atheists and Orthodox religous people were later split up by DBSCAN. Furthermore, groups such as Protestants, Roman Catholics, and Muslims have all been divided into one or more clusters.

Based on these observations and accounting for the fact that the DBSCAN algorithm yielded distinguishable and clearly separated clusters, we have interpreted that the religions in Europe are quite dissimilar in nature.

Integration with OpMAP Having acquired the aforementioned results, it was the question if these results can be integrated with OpMAP. To achieve this, we established three challenges that we needed to address:

- Converting the clusters achieved with the DBSCAN algorithm to a format interpretable by OpMAP
- Extending OpMAP's functionality to feature a large amount of data
- Extending OpMAP's functionality to feature more types of variables (i.e. scalar and categorical)

Converting the clusters achieved with the DBSCAN algorithm to a format interpretable by OpMAP The input format that is used by OpMAP is a *.json* file containing the following structure:

```
{
"nodes":[
{"id": "1", "opinion": ["!23", "24", ], "vweight": 1.020962, "proponents":
[{"age": "25-34", "gender": "m", "education": "College graduate",
"residence": "DE", "diet": "restricted"},
...
],
"links":[
{ "source": "1", "target": "2", "value": -0.99833},
   ...
]}
```

As can be seen, opinions are categorized as boolean values, meaning that a person either holds a certain value or does not hold a certain value. Furthermore, proponents are values that characterize the person holding values. These can be *age, gender, education, diet*.

The first step to convert the clusters achieved with the DBSCAN algorithm to a format interpretable by OpMAP is to convert points to nodes and to calculate pairwise distances to serve as links between nodes.

The following code was used to calculate and process pairwise distances (i.e. Euclidean distance) between nodes:

```
from scipy.spatial.distance import squareform
from scipy.spatial.distance import pdist
import json
```

```
pairwise = pd.DataFrame(
    squareform(pdist(data[['x', 'y']].to_numpy())),
    columns = data[['x', 'y']].index,
    index = data[['x', 'y']].index
)

pairwise.applymap(lambda x: np.nan if x > 10 or x==0 else x)

pairwise.applymap(lambda x: 5 - x)
```

This code takes the coordinates of each point and calculates the Euclidean distance between every other node in the data set. To prevent links with very high distances or distances of 0, an *.applymap* function was used to convert these distances to null. Afterwards, we needed to convert distances to weights. Since distances are in the range of [0, 10], subtracting x from 5 would yield the inverse, and therefore an appropriate substitute for a link's weight.

To generate nodes that serve as input for OpMAP, a NetworkX [1] graph was built from which data was extracted. The code to build the graph and export the node data is as follows:

```
import networkx as nx
import numpy as np
import string

G = nx.from_pandas_adjacency(pairwise)

from networkx.readwrite import json_graph

graph_data = json_graph.node_link_data(
    G, {"link": "edges", "source": "source", "target": "target",
    "weight": "value"}
)

with open('nodes.json', 'w') as fp:
fp.write(
    '[' +
    ',\n'.join(json.dumps(i) for i in graph_data['nodes']) +
    ']\n')
```

Extending OpMAP's functionality to feature a large amount of data As it stands, OpMAP's functionality is built upon drawing SVG graphics on the screen when the local OpMAP web server is started. For the diet data set this does not come at a cost as the data set is relatively small (i.e. 200 nodes). When visualizing 1.000 or more nodes, SVG's performance is inadequate to the task. We sought to replace SVG with alternative techniques such as Canvas or WebGL. However, due to OpMAP's complex code structure, this was not achieved.

To achieve an intermediate result, we built a prototype version of how we would address this issue. We clustered the output of the t-SNE algorithm with a variation of the DBSCAN algorithm called DBSCAN Reduce [13]. This algorithm is the same as DBSCAN with the slight variation that it only keeps the centermost points. With this algorithm, we reduced the complexity of the original data set with about 30.000 rows/opinions to 300 without losing a lot of the data's structure. Then we executed the aforementioned steps again to gain the nodes and edges in the JSON format. Afterwards, we visualized the reduced data set with the SVG technique. Unfortunately, we were not able to integrate this result with OpMAP. The prototype can be observed in Figure 6. The full code for this prototype is available in Appendix C.

Extending OpMAP's functionality to feature more types of variables (i.e. scalar and categorical) Unfortunately, due to a lack of time, we have not been able to investigate how OpMAP's functionality can be extended to feature more types of variables (i.e. scalar and categorical).

3.5 Limitations

The initial goal of this research was to fully integrate the EVS into OpMAP. This would require multiple steps.

Firstly, data cleaning and exploration would be performed. This part of the research was successfully completed. It did however take much longer than anticipated because the dataset turned out to be quite complex.

Secondly, experiments were performed to determine which dimensionality reduction methods were appropriate to use with this dataset. This step was also completed successfully, although also taking longer than anticipated. The DBSCAN plots could be improved by using the cluster size to vary the size of individual nodes.

Thirdly, the dataset would be integrated into OpMAP. The dataset at this point contained around 32,000 rows. All these data points would need to be rendered by D3.js. However, as the current implementation is based on SVG to draw graphics, this is not feasible. Using SVG, D3.js is only capable of drawing

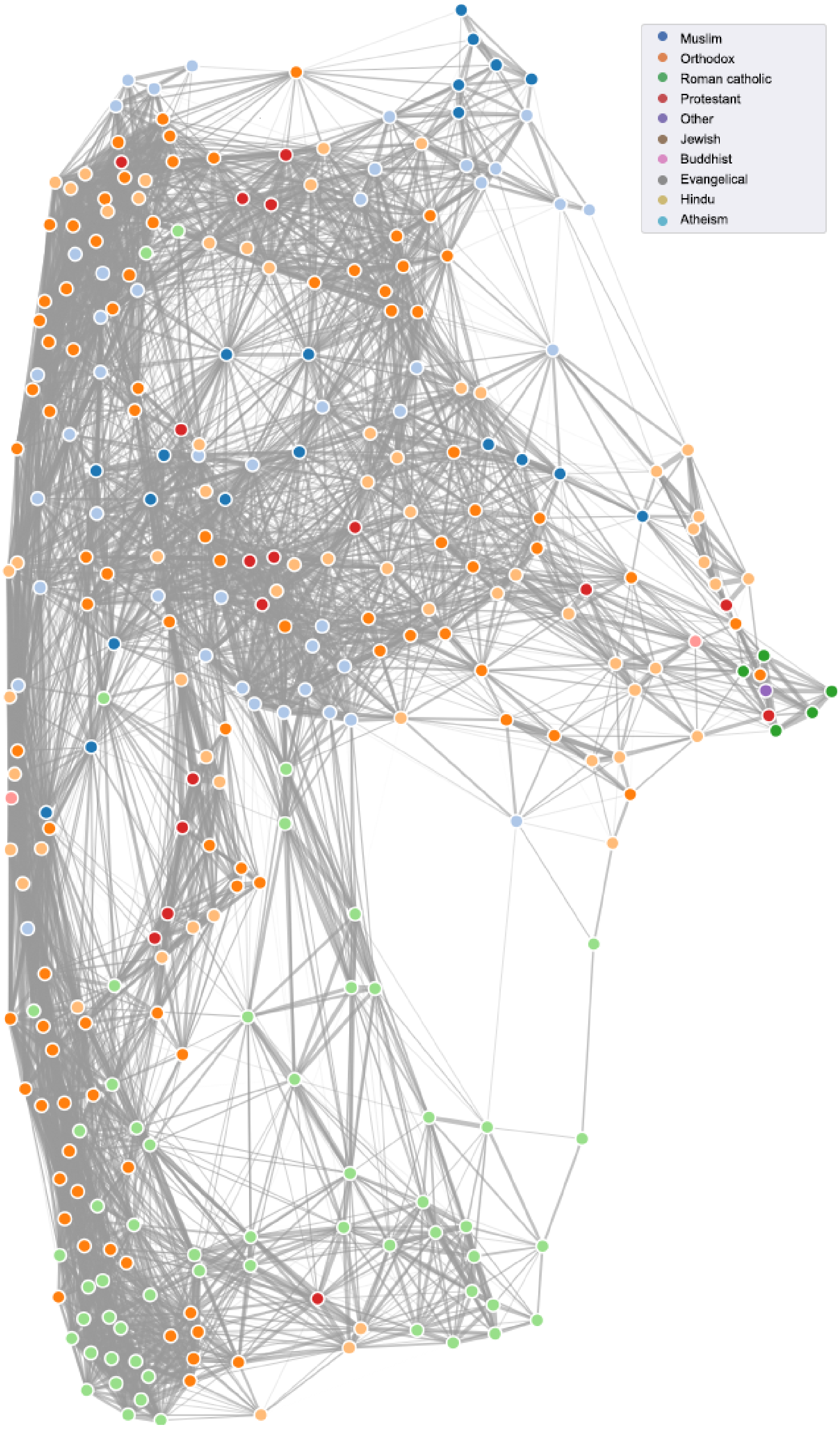


Figure 6: Prototype of OpMAP-based network of reduced EVS data set with DBSCAN Reduce

up to around 1,000 data points. After first trying to work around this by reducing the resolution of our data by grouping nodes together. We realised that this was not the right approach. A bad implementation should not be mitigated by complex sampling and data clustering operations but by improving the implementation. Improving the OpMAP implementation was outside the scope of this project. It could however be achieved by migrating from SVG to Canvas rendering. HTML5 Canvas rendering should be capable of rendering up to around 10,000 data points at 60 frames per second. Millions of data points can be rendered smoothly by implementing WebGL. A hybrid approach mixing these three techniques is also possible [18]. After these initial steps have been completed, the OpMAP could be made interactive by e.g. allowing the user to select different years in which the survey was conducted or to zoom into the map and explore the dataset on a deeper level.

A last limitation which should be mentioned is that it is not feasible to describe every intermediate step we took in this report. However, we have therefore created a Python notebook with elaborate annotations which offers a much better format to understand these steps. These have been included in Appendix TODO X and should be regarded as part of this report.

3.6 Conclusion

This part of the research aimed to investigate how OpMAP can be extended to feature complex, high-dimensional data. To assess this, the EVS data set was used to mine opinions on religions by applying a dimensionality reduction technique, visualizing the points in 2D, and subsequently clustering the points with an unsupervised clustering algorithm.

Returning to the research question *If we model the answers of survey participants as multi-dimensional vectors (i.e. opinions), does this space have a pronounced clustering?*, it can be concluded that after applying the t-SNE algorithm, the 2D space and the points therein already have a pronounced clustering. The clusters are separated fairly well and are distinguishable. Although t-SNE is not a clustering algorithm, the points are partitioned in such a way that the criteria mentioned in Section 3.3 are met. Therefore, the results is meaningful. The pronounced clustering is further illustrated by Figure 4 where the groups of opinions predominantly coincide with the religion the people in the group abide by.

Returning to the research question *Can we visualize this clustering in 2D?*, we applied the DBSCAN algorithm to achieve the result presented in Figure 5. It can be concluded that it is indeed possible to visualize a meaningful clustering in 2D by applying DBSCAN on a data set that is reduced to 2 dimensions with t-SNE.

Returning to the research question *Can we interpret the clusters?*, it can be concluded that it is indeed possible to interpret some of the clusters presented in Figures 4 and 5. The interpretations were as follows:

- Atheists and Orthodox people have a multitude of different opinions of religion and are inherently dissimilar.
- Protestants, Roman Catholics, and Muslims were also split up in one or more groups due to them having dissimilar opinions within their respective groups.
- Views on religion in Europe are quite dissimilar as the clusters are distinguishable and clearly separated visually. This is also highlighted by a Silhouette Coefficient of 0.487.

Returning to the research question *How can we integrate the results with OpMAP?*, it can be concluded that the first and most important steps to integrate the results of this research with OpMAP have been taken. It is possible to convert the results to a JSON file containing the nodes, edges and their respective weights. Furthermore, a prototype was developed to allow OpMAP to visualize a reduced number of nodes by reducing the complexity of the EVS data set. However, it was not possible to extend OpMAP's functionality to support variables of scalar and categorical types. This ultimately prohibitied our full integration with OpMAP.

4 Interactive system OpMAP

4.1 Introduction

As discussed in the general introduction, developers of the OpMAP interactive system have shown that the opinions in a structured argumentation can be visualized using an opinion network [11]. The existing opinion network, representing the so-called 'veggie debate', consists of six separate clusters. These clusters appear to be interpretable into six different types of overarching opinions [8]. Opinions are based upon the acceptance (up-voted) or rejection (down-voted) of claims. In short, answers are summarized in vectors, vectors are represented as nodes, and similar vectors result in proximity of nodes. Currently, the interactive interface allows users to fill in the questionnaire and see their resulting position in the network based on their own opinion [11].

The current OpMAP interface visualizes the opinion network and clusters are marked. However, it is not yet specified how these clusters are to be interpreted. The interpretation of clustering in the original OpMAP visualization was done by topic experts that were sufficiently familiar with the data [11]. The interpretations are shown using cluster (or 'country') labels and by providing short descriptive explanations of the clusters. The provided labeling and explanations give a general indication of the distribution of opinions, but they provide limited aid for self-guided in-depth data analysis of the claims underlying the compound opinions. Users have limited tools to verify the labels and explanations making it difficult to further investigate the opinion network.

Existing research has already explored techniques to visualize relevant features of the clustering to aid interpretation of the clusters [31]. These techniques focus on the clusters themselves by extending the existing network with novel views of the network. For example, by coloring clusters based on their characteristics in a variance view or correlation view [31]. Although such techniques have shown to be highly informative and aid the interpretation of clustering, some domain expertise is required. The OpMAP interface was designed for the general audience, as becomes apparent from its interactivity, and it cannot be expected that this audience is familiar with terms as 'variance' and 'local dimensionality'. Hence, a more straightforward representation of the compositions of clustering would be more informative for users of the OpMAP interface.

An example of a simpler representation of the composition of clusters could be the distribution of votes among claims. Currently, all claims are shown in a sortable data table (see figure 1). When clicking a cluster ('country') or individual node, the corresponding claims and their respective vote counts become visible in the table. The available raw data (e.g. vote count) allows for in-depth analysis of, for example, most important claim of the cluster or most accepted claim in the cluster. However, the current opinion network and data table provide limited support for making such inferences. As a result, users would have to sort the table and step-by-step compare the claims to gain more insights resulting in information overload [20]. Hence, the goal of the current paper is to extend the existing OpMAP interface such that the analysis and interpretation of clusters is better supported. The goal is to add novel visualization techniques while preserving the original raw data. In other words, to prevent data loss or biases in representation, data aggregation or filtration should be limited. On the other hand, the data visualization should be sufficient in size to prevent visual noise or overload due to too large image perception [20]. The current goal can be phrased in the following research question:

RQ1: *How can detailed statistics of the clusters be visualized to aid the inter-*

pretation of the clusters?

Equally important is a proper verification of the usability of the OpMAP interface. The original study reported a small-scale and informal user study [11]. Hence, the present paper aims to verify the novel design of the OpMAP interface using a qualitative survey method. The goal is to gain more insight in the overall impression of the OpMAP interface and the ability to perform data exploration using the opinion network, data table and the additional data visualization method. In the original study, the researchers were able to interpret the clusters into meaningful 'countries' representing a subset of the opinions. However, the researchers had more experience with the data through the development of the interface. It is not yet clear if users that see the data for the first time via OpMAP will be able to infer such information to a similar extent. In other words, the second goal of the current study is to verify the interpretability of the opinion network. This can be phrased into a second research question:

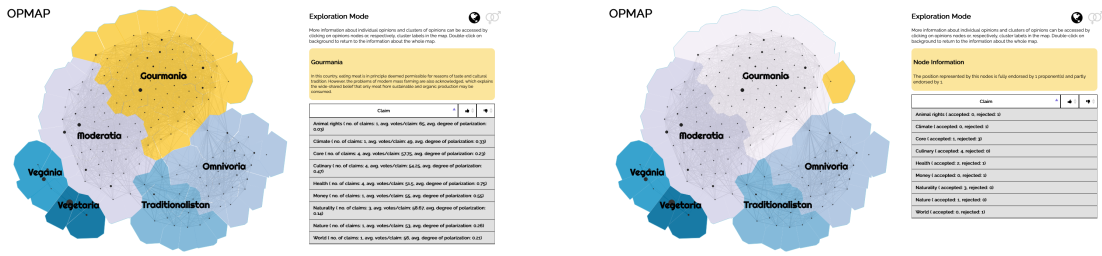
RQ2: *Are users able to interpret the clusters in of the opinion network?*

It is expected that the additional visualization method as developed for answering the first research question will enable users to better explore the data and thus, interpret the clusters of the opinion network.

4.2 Methodology

The original OpMAP interactive system consists of the opinion network and one table showing all claims and their respective number of up votes and down votes as shown in figure 1. Country clusters are labeled according to the aggregated opinion represented in that cluster. When clicking on a country label, a short explanation of the opinion of that country and an overview of the claims of that country are shown on the right (see figure 7a). The interface also provides interaction with individual nodes showing the votes on claims of individuals (see figure 7b).

The design of the novel interactive system was developed in close consultation with one of the original creators of OpMAP. Development was focused on improving readability and usability of the interface by improving the organization of the available information. Moreover, a novel visualization method, a bar chart, was introduced to display some information in more detail. A complete overview of the novel interface design is shown in figure 10.



(a) Country selection (b) Node selection

Figure 7: Screenshots of the original OpMAP interface after selection of a country (top) and a node (bottom)

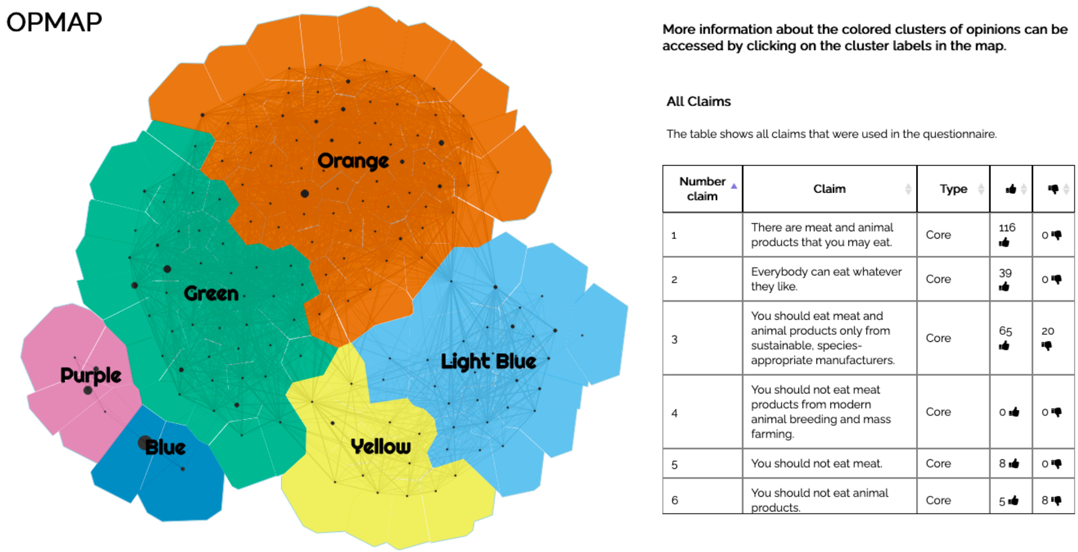
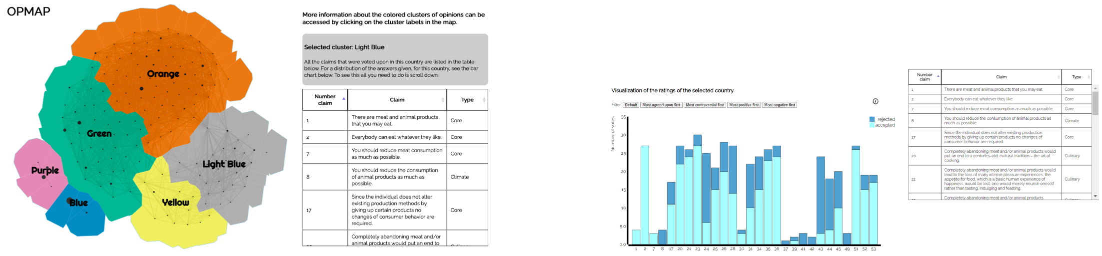


Figure 8: Screenshot of the novel OpMAP interface

Interaction The coloring of the clusters is changed from hues in blue to a color-blind friendly color scheme [16]. The change in coloring is a minor technical development but greatly enhances overall usability by allowing more users to understand and work with the OpMAP interface [35]. Furthermore, all labeling and additional information on the clusters has been removed from the interface design. This allows for a user study to verify if users are able to interpret the clusters themselves with the help of the available data. Consequently, clusters are labeled by their color and the text box contains guidelines on how to explore the information (see Figure 9a). When users select a country cluster, the relevant claims are shown in the table and the bar chart showing the vote distributions appears below the opinion network (see Figure 9b).

Data Table The design of the original data table only allowed limited interactivity and interpretation of the information. Claims are organized by abstract type tags (e.g. 'core') which makes it difficult to find a specific claim if the type is not known (see figure 1). Also, such organization does not allow efficient sorting nor cross-referencing between interactive system elements. Consequently, two columns were added to the table; one for the type of the claim and one for its question code. This design enables users to sort the table on both type of claim and on more complex queries without compromising the readability of the table. Question codes enable cross-referencing of claims between various visualization methods (e.g the opinion network, the table, and additional charts). Users should be able to easily find a question by its respective question code.

Stacked bar chart A novel interactive subsystem was introduced by adding a stacked bar chart below the opinion network to the interface. The goal was to minimize summarizing the data beforehand to ensure that the information is presented neutrally to the user. However, with many data points, readability and understandability might become an issue. Stacked bar charts allow for a high-quality presentation of large amount of data points consisting of various levels/categories [20]. In the current design, the stacked bar chart is linked with the opinion network. It only shows the claims that are appear either positively or negatively in the opinions in the selected cluster. Moreover, the coloring of the bar chart is similar to that of the selected cluster to ease the cross-referencing between the interactive elements. For each claim, the amount of up-votes and down-votes is displayed. The more votes, the larger the respective rectangle within the bar. In addition, the interactivity of the chart was increased by allowing the user to sort the displayed claims. Sorting is based upon; level of agreement, level of controversy, amount of positive votes, and amount of negative votes. Level of agreement first shows claims that were highly agreed upon; e.g. for which the majority of votes contained one answer option. In contrast, level of controversy puts claims that had a mixed answer distribution within the cluster first; e.g. for which half of the votes were negative, and half positive. Sorting on the amount of positive or negative votes sorts the claims based most up-votes and down-votes respectively. Using sorting buttons allows users to interact with the data on their own terms without overcrowding the interface with a large amount of charts. Lastly, the position of the data table is fixed on the screen ensuring that is visible while reading the bar chart.



(a) Country selection (b) Stacked bar chart

Figure 9: Screenshots of the original OpMAP interface after selection of a country (top) and a node (bottom)

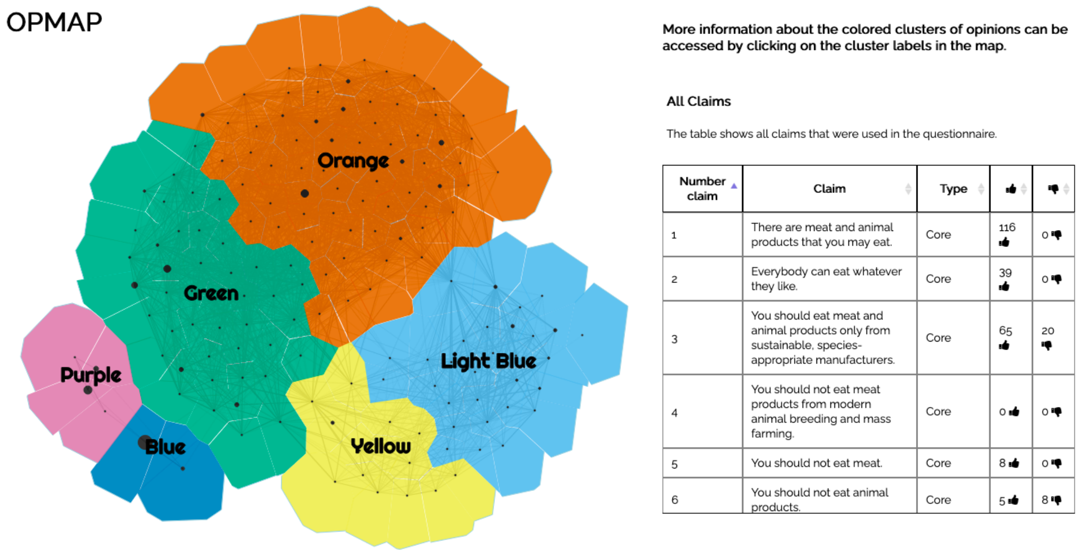


Figure 10: Screenshot of the novel OpMAP interface

4.3 Evaluation Method

Design The goal of the user study is to gain more insight in complex data analysis using the OpMAP interface. Hence, in the present user study, users are asked both quantitative and qualitative questions about the data and the visualization. Users are also asked about their impression of the general design of the interface. Moreover, the usability of the interface is verified using the Systems Usability Scale.

The answers about the data and opinion network are compared to the original descriptions of the clusters provided by a topic expert. This makes it possible to verify if the users are able to come to the same conclusions and interpretations as the experts. Similar answers imply that the visualization has successfully aided the users in retrieving the requested information. On the other hand, less extensive information implies that the OpMAP interface does not yet enable the users to retrieve the requested information. If the answers are completely different, it could also be argued that users gain other insights that were not thought to be part of the OpMAP interface by the previous researchers.

Participants Participants are recruited through a convenience sampling method; each group member approached colleagues, friends and family members for a total of 11 participants. There were no restrictions on demographic variables such as age and level of education as the visualization should be able to be interpreted regardless of background knowledge. Participants were on average 27,8 years old (between 21 - 49). Around 80% was male and 20% female. One participant (participant 4) is excluded from the results, as this person only answered the first question and left the rest blank.

Procedure The user study is implemented as an online survey. The full questionnaire is attached in Appendix B. Time and meeting restrictions did not allow for in-person interviews. The survey starts with an informed consent and instructions for downloading and running the OpMAP software. Participants are given a short introduction to the visualization and are instructed to explore the OpMAP interface for at least 5 minutes before continuing to the questionnaire. The questionnaire consisted of five parts: 1) general questions about the opinion network, 2) questions about the interpretation per cluster, 3) general impression of the interface, 4) the Systems Usability scale, and 5) demographics. On completion of the questionnaire, participants were thanked for their participation. Participants did not receive any reward or payment for their participation.

Measures The Systems Usability Scale is a generally acknowledged method for measuring usability of a tool [9]. The whole process of the SUS consists of three main steps: questionnaire deployment, score calculation, and classifying the scores. The questionnaire contains ten questions from the original SUS, see Figure 11 [14]. Each question is answered by giving a number between 1 and 5. Where 1 stands for strongly disagree and 5 for strongly agree. We have adapted these questions so that they apply to the visualization (see Appendix B).

Predefined questions regarding the exploration and interpretation of the data include the interpretation of clustering (e.g. 'how can this cluster be interpreted?'), the interpretation of the interactive visualization (e.g. 'how are opinions distributed for Dutch people?'), and inferences on the visualization (e.g. 'what is the most important opinion for this country?').

No.	Original Item
1	I think that I would like to use this system.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in the system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

Figure 11: 10 questions of the original SUS

4.4 Results

The results of the questionnaire are discussed in this section. The first step in analyzing the data was to clean the raw data and then divide it into the five parts as described in §4.3 Procedure. For an overview of the questions per part, see Appendix B. The fifth part will not be discussed in this section, because this has already been done in §4.3 Participants. The other four parts will be discussed below.

General questions about the visualization To analyse this part the answers to these questions are coded and then compared with the researchers' interpretation. The results of ten participants are compared with the interpretation of the researchers. Table 1 shows how the three questions of this part were answered by the participants compared to the researchers' answers. An answer is partly correct if it gets part of the answer right, but not all parts. So, for example, if it contains the correct answer that two clusters are equal because they contain the same kind of opinion, but not because the nodes that are next to each other on the edge are connected to each other. This shows that all ten participants had the correct answer to the question: "Which cluster represents the opinion of the majority of the participants?". The other two questions show that they have actually thought about this because they are talking about the amount of people in the clusters.

The question "Which clusters are (somewhat) similar to each other? Why? " was answered correctly by only one participant. This participant answered: "*Green and orange, as they have many connections*", which may be a bit short,

Answer type	Correct	Partly correct	Incorrect
Which cluster represents the opinion of the majority of the participants?			
Researcher	Orange		
Participants	10	0	0
Which clusters are (somewhat) similar to each other? Why?			
Researcher	The clusters that are equal to each other are those that are next to each other in the network. These have many connections with each other. The nodes in the cluster have many connections to the nodes of the neighboring cluster.		
Participants	1	1	8
Which clusters are (somewhat) different to each other? Why?			
Researcher	For example, the cluster that are dissimilar to each other are purple and orange. They are not neighboring countries, and the nodes of these clusters are not directly connected to each other.		
Participants	1	2	7

Table 1: Amount of correct, partly correct and incorrect answers to the general questions of the visualization.

but it is the point at which it is determined whether two clusters are equal to each other. One participant was partially correct, saying "*...Green and orange inhabitants do like to eat meat.*". This person therefore assumes that neighboring countries share an opinion. This is somewhat true, because nodes that share an opinion are closer together and will have more connections. The other 8 participants did not have the correct answer. They interpreted it as saying that if clusters are of the same size (have the same number of inhabitants) then they resemble each other. For example one participants said "*Green and orange, light blue and yellow, purple and blue, because all sizes are quite similar and seem to be the same*". In this visualization it is a coincidence that, for example, the Orange and Green clusters are almost the same size and also happen to be equal to each other, but the number of inhabitants has nothing to do with this.

The last question of this part "Which clusters are (somewhat) different to each other? Why?" was correct answered by one, partly correct by two, and incorrect by eight participants. The participant who answered this question correct is the said "*Purple and orange, as they contain almost no connections*". This again sounds a bit short, but it is the point at which it is determined whether two clusters are different to each other. The two participants who are partly correct said "*Orange inhabitants did answer almost all questions. They do eat meat. Yellow inhabitants are more socialist persons that like to control what other people eat, while light blue inhabitants are more autonomous*" and "*Orange and yellow seem to have quite different opinions*". These participants stated that countries that are different also have a different opinion, which is true. But it misses the part that therefore the nodes in these clusters do not, or almost not, share connections. The other seven participants stated that countries which differ the most in size are the most different. This is not true. One of them said, for example, "*Orange and blue, because one is much bigger than the other.*".

Questions about the interpretation per cluster The answers to these questions are coded. Subsequently, these codes are compared with the interpretation of the researchers. The results of ten participants are shown in table 2. This table shows if the participants interpretation of each cluster is correct, partly correct, or incorrect compared to the interpretation of the researchers. The three questions that were asked per cluster in the survey had to ensure that the participants would arrive at the correct interpretation. The answers of the three questions are therefore looked at together, from which it is concluded whether the interpretation is correct, half correct or incorrect. An answer is half correct if not all parts (what represents the cluster, what is the most salient opinion, what are the most important claims) of the answer are named. For example, if the participant answers correctly what the cluster represents, but does not name the most important claims.

The Orange cluster was partly correctly interpreted by five of the ten participants. Four of these five noted well from the visualization that only meat from sustainable and organic production may be consumed. However, they missed the part that the other participant did notice, he said "*... the loss of many intense pleasure experiences" would be lost if we wouldn't eat meat or animal products anymore*". This participant was able to deduce from the visualization that this cluster represents the people who believe that eating meat is in principle deemed for permissible reasons of taste and cultural tradition. However, this participant again missed the part that meat only from sustainable and organic production may be consumed. So therefore none of them were totally correct. The other five participants were incorrect. Two of them said they had no clue what the cluster was about. The other three misinterpreted the visualization. One of these three said for example, "*vegetarians*".

The light blue cluster was answered correctly by two of the ten participants. For example, one of them indicated that the cluster was about this "*People who like eating meat, and do not wish to change. Eating meat should not be re-*

Researchers' opinion of each cluster:	**Correct**	**Partly correct**	**Incorrect**
Orange: In this country, eating meat is in principle deemed permissible for reasons of taste and cultural tradition. However, the problems of modern mass farming are also acknowledged, which explains the wide-shared belief that only meat from sustainable and organic production may be consumed.	0	5	5
Light blue: In this country, meat is considered to be way too delicious to be abstained from. Animals are not recognized as beings with a right to life. And modern mass-farming, it is believed, doesn't represent much of a problem anyway. So, in this country, you may basically eat what you want.	2	8	0
Yellow: In this country, killing and eating animals is seen as something very natural. The traditional human diet is essentially viewed as unproblematic. Any moral demands to change individual eating habits are rejected as paternalistic and out-of-place.	0	3	7
Green: In this country, there exist a profound awareness and recognition of the diverse problems of mass farming – ranging from violations of animal rights in the farm to global climate change. A (possibly drastic) reduction of meat consumption and a boycott of products from mass farming are considered to be appropriate answers to these problems. Many even demand that the consumption of all animal products whatsoever be reduced.	2	4	4
Purple: In this country, animal products of any kind are a no-go. In view of all the problems of farming animals and the consumer's irreducible uncertainty about what is really going on in a farm, a vegan diet is unanimously demanded. From a culinary point of view, the vegan cuisine is considered as no less delicious than the traditional one.	0	3	7
Blue: In this country, one must not eat meat, while other animal products may very well be consumed. Killing animals for meat production is considered to be unethical. Farming animals in order to produce, e.g., eggs or milk, it is believed however, is not necessarily problematic. Vegetarian dishes are judged to be (at least) as delicious and healthy as traditional ones.	0	10	0

Table 2: Amount of correct, partly correct and incorrect interpretation of each cluster.

duced and reducing meat consumption reduces overall happiness. Animal well being was not important. Freedom to consume meat is important." So these two participants indicated that this cluster represents people who like to eat meat, and that the welfare of the animals is not an issue, which is in line with the researchers' interpretation. The other eight participants were partially correct. These mainly indicated that it concerns people who like to eat meat, but missed the part that the welfare of the animals is not an issue here.

Three of the ten participants partly completed the yellow cluster correctly. One indicated that it is nature to hunt and eat animals, but said nothing about the individual eating habits should not be adjusted. The other two indicated that each person decides for himself what is good and what is not, but again missed the part about eating meat is something natural.The other seven participants were incorrect. These indicated that the cluster represents people who do eat meat, but are also open to changes, or that they were vegans. One of them wasn't quite sure what it was about.

The green cluster was interpreted well by two of the ten participants. These two indicated that it represents this cluster of people who try to reduce meat consumption as much as possible. They do this for the climate, but also for the welfare of the animals. This is in line with the researchers' interpretation. Four out of ten were partially correct, indicating well that the cluster represents people trying to reduce meat consumption. But according to these four participants, this was only for climate considerations. So they missed the part that it is also done for the welfare of the animals. And the other four participants were incorrect, indicating they either didn't know or eating less meat due to health consideration.

Three of the ten participants gave an almost correct interpretation of the cluster. These indicated that the cluster represents people who are vegan. That they eat no meat and no animal products. However, they did not indicate that an important claim is that vegan dishes are no less tasty, and this is something the researchers do explicitly mention. The other seven were incorrect, indicating that the cluster represents people who are vegetarian. This is not correct because the people in this cluster do not only eat no meat, but also no animal products.

In the last cluster, blue, all ten participants gave a partially correct answer. They all indicated that the cluster represents people who do not eat meat but do eat animal products. However, they did not indicate that the claim that the dishes are no less tasty is also important, and the researchers did explicitly state this.

(a) Word cloud of the negative points of the interface OpMAP

(b) Word cloud of the positive points of the interface OpMAP

Figure 12: Word clouds

General impression of the interface To analyse this part, the data has been broken down into parts and “codes” have been created and labeled. Each label consists of one word, so that a word cloud could be generated. To code and label the data and to make the word cloud Nvivo version 20 was used. This has been done per sub question. To get the general impression of the interface, we asked participants to give three positives and three points for improvement. Of these, two word clouds have been generated that show the most important points. The word cloud about the positives points can be found in Figure 12b.

From this word cloud it can be seen that the most given positive points are the colors in the visualization, the filter buttons, and the graphs. The participants found that the colors are easy to distinguish from each other, and that the clusters are also easy to distinguish. In addition, the filter buttons made it easier to find out how the claims were answered in the clusters. And the graphs that are used (the bar charts and the table) are very clear. In addition, some of the participants also indicated that it is nice that the visualization is so interactive, because you can click on the labels and sort the bar charts. And also that it is useful that you can see what the group size is and how the claims have been answered in a cluster (how many accepted and rejected). And they also think that the visualization works smoothly and that it looks nice.

The points to be improved can be seen in Figure 12a. The biggest problem that, according to the participants, needs to be improved is that you can't see how similar the clusters are. It would be nice if you could view two clusters at the same time. In addition, there is also the problem that the clusters have different claims that are in the bar charts and claims that are not. This makes it difficult, according to the participants, to compare the clusters with each other. In addition, the participants also indicated that it is not clear why the network is depicted as it is now. There is great uncertainty as to why the size of the clusters is this way. For example, one participant stated: "*1: What is the figure with the colors supposed to mean, they seem to be connected to each other, but mostly to themselves. Why is this figure made like this, maybe use another figure 2: The small groups like purple and blue gain significant amount of attention, while their numbers of participants are really low, this screws opinions*".

We also asked about the general impression of the visualization. This showed that the visualization looks nice, and that the functions that are there (clicking on the cluster labels, clicking on the filter buttons) work well. However, 7 out of 10 indicated that they were not quite sure what the clusters really meant. And 3 of these indicated that the network itself is also difficult to interpret; "*why do the clusters look like this?*", "*why this size?*", and "*why are certain clusters next to each other and others not?*"

The System Usability Scale In this part we also asked whether the participants enjoyed the exploration of opinions in OpMAP. We will analyze this by looking at how many participants agreed and how many disagreed. Because the participants answered this question on the basis of a number from 1 to 5, where 1 stands for strongly disagree and 5 for strongly agree.

Three participants indicated that they agreed with the statement: "I enjoyed the exploration of opinions in OpMAP". Three were neutral about this, and four disagreed.

In order to be able to analyze the results of the ten questions of the SUS, the score has been calculated. This has been done for each participant. The average score has also been calculated. The steps to calculate the SUS score are as follows [14]:

1. Divide the odd number and even number questions.
2. Subtract 1 from the score of each of the odd questions.
3. Subtract the score from each of the even questions from 5.
4. Add the scores of all ten questions and multiply this number by 2.5.

The next and final step is to classify the scores. A score of 80.3 or higher would be excellent. A score between 80.3 and 68 would be good. Between 68 and 51 will be okay. And a score of 51 or less is awful [28].

The scores for the System Usability Scale per participant and the main score can be found in Figure 13. Here you can see that 5 out of 10 give the usability of OpMAP a score of 51 or higher. The mean score is 53. Which means that the usability is okay, but not really great.

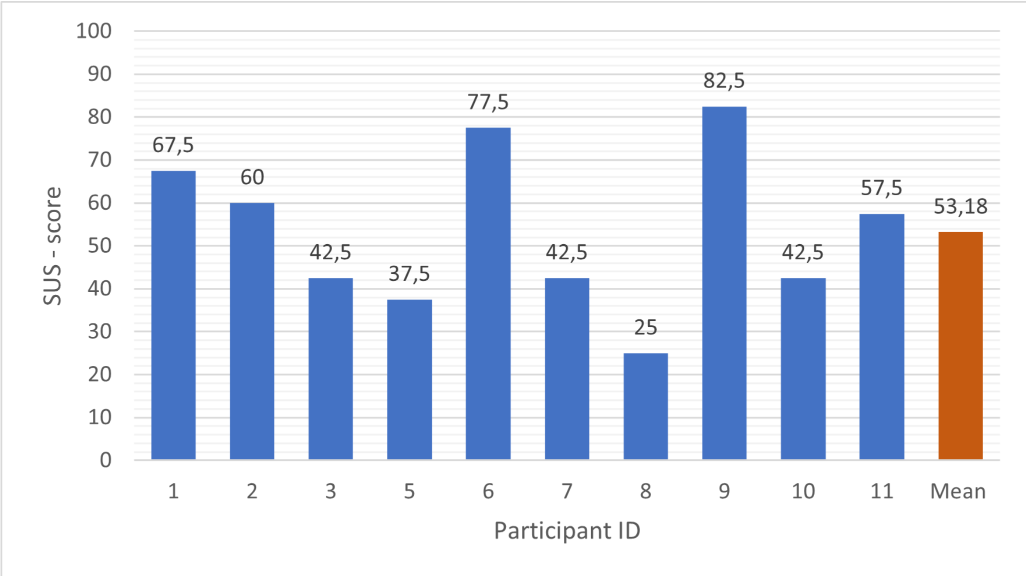


Figure 13: SUS score of each participant and the mean

4.5 Limitations

The current study has a number of limitations. The first limitation is that it was quite difficult for the participants to get OpMAP working on their own device at home. This could have influenced the study, because the participants may have become too frustrated by this. An on-site experiment with OpMAP already prepared on the device could prevent this. In addition, a survey was used, in which the questions were not randomized. As a result, questions about a cluster may sometimes have gotten better results, as the participants gradually got more experience with OpMAP and they came to understand the interface better and probably the data better. Likewise, the large amount of open questions could lead to answer reduction. By not using a randomized survey this could happen for the same question. Also, a survey does not always lead to very detailed answers. We would have preferred to hold a semi-structured interview, because this would probably lead to more insights. In addition, a semi-structured interview could also remedy that people are less motivated during the experiment, because you as a researcher can let the participant stay with it. However, due to the measures surrounding COVID-19, it was not possible to hold sufficient interviews. That's why we chose to stick to surveys. Another limitation relates to the functionality of OpMAP. It was decided to disable the feature that showed node-specific data when clicking on one, so that using OpMAP for the survey would not lead to confusions. However, this could mean that the usability of OpMAP was reduced. As a result, the participants could have missed information about the exact data of all opinions in the clicked cluster, which could make the interpretation of a cluster more difficult.

4.6 Conclusion

The goal of this Part B of the study was to gain more insights in potential improvements to better facilitate understanding of the clustering presented in OpMAP. First, a novel design for the OpMAP interface was developed to allow for in-depth analysis of complex opinion. Second, a qualitative survey was conducted to learn about users' experiences when using OpMAP to explore and analyse complex opinions.

In an attempt to answer the first sub-question of this part (RQ1), "*How can detailed statistics of the clusters be visualized to aid the interpretation of the clusters?*", an interactive stacked bar chart was implemented in the interface. The stacked bar chart provided a clear visualization of the composition of the complex opinions and allowed for filtering on characteristics of the opinions. Moreover, the stacked bar chart was linked with the opinion network and the data table was visible at all times to enhance cross-referencing between the interactive elements. In addition, the existing data table was reorganized and slightly adjusted to better enable sorting and analyzing individual claims. Lastly, the overall accessibility of the interface was improved by introducing color-blind friendly colors.

Returning to the second sub-question (RQ2), "*Are users able to interpret the clusters in of the opinion network?*", we can conclude that this is not quite the case yet. The participants indicated that they found it very difficult to interpret the clusters, and this is also apparent when we compare the answers with the researchers' interpretation. One of the functions that has been removed for this user study is clicking on nodes. This prevented participants from viewing individual opinions within a cluster. As a result, this may have made it much more difficult to properly interpret a cluster. With this feature enabled, the participants might have been able to see better why two clusters are equal, and why not. However, it appears that half find the system usable. The advantages of the visualization are that there are enough functions to filter and order the date (filter buttons) and it is very interactive. However, the participants would like to see more similarity between the cluster. That all claims have been answered, or are visible. And would they like more information about why the network looks the way it does, and why certain clusters are so much bigger.

5 Conclusion

The goal of the present study was to explore an extension of OpMAP to support the interpretation of complex opinion. This goal was shaped into the following research question: "*How can OpMAP be extended to support the interpretation of complex opinions?*" To answer this question, the study was divided in two main parts; part A focused on the implementation of complex and high-dimensional data, and part B focused on the improvement and verification of the usability of the OpMAP interface.

In short, the study showed that it is possible to visualize complex, high-dimensional data, but not in OpMAP. Providing an answer to the first sub-question (RQa): "*Can OpMAP be extended to feature complex, high-dimensional data?*". To answer this question the EVS data set was reduced to two dimensions with t-SNE and subsequently clustered with DBSCAN. This resulted in meaningful and pronounced clusters. However, integration with OpMAP appeared to be difficult as OpMAP's functionality could net yet successfully be extended to support variables of scalar and categorical types and the drawing technique (i.e. SVG) was not able to draw the required amount of nodes and edges in its current implementation.

In addition, the study showed that the OpMAP interface can be improved by introducing a novel visualization (stacked bar chart), reshaping the existing data table and enhancing linking between the interactive elements. Providing an answer to the second sub-question (RQb): "*RQb: How can OpMAP be improved to facilitate better understanding of clustering presented in OpMAP?*" The novel design was implemented and a qualitative survey was conducted to gain more insight in the usability of the interface. Participants indicated to enjoy working with the re-designed interface and appreciated the interactive stacked bar chart with filtering. However, the provided visualizations and information was not sufficient to support in-depth analysis and the interpretation of clusters. Participants were not able to interpret the clusters or to gain insights in the general opinion represented in the clusters.

Future work Albeit that integration with OpMAP was too time-consuming and difficult to realize for this research, future work should be aimed at achieving this. This work should be primarily focused on extending OpMAP to support all variable types (i.e. scalar, categorical, etc.) as this is what we did not succeed at in this research. Furthermore, for completeness, it is advised that future work is also directed at converting OpMAP's drawing technique from SVG to Canvas or WebGL. This way, it should be possible to visualize all required nodes and edges.

Regarding the usability of OpMAP and the interpretation of clustering, future research should be aimed at better investigating the needs of users. The stacked bar chart appeared to provided some information to the user, but not enough to really help him/her grasp the meaning of the clusters. Interesting would also be to see if a general explanation about clustering would be enough to guide the analysis without the need for pre-labeling and interpretation of the clusters by experts.

6 References

[1] Networkx: Network analysis in python. Available at https://networkx.org.

[2] sklearn.cluster.agglomerativeclustering. *Scikit Learn*. Available at https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.ht

[3] sklearn.cluster.dbscan. *Scikit Learn*. Available at https://scikit-learn.org/stable/modules/generated/dbscan-function.html?highlight=dbscansklearn.cluster.dbscan.

[4] sklearn.cluster.kmeans. *Scikit Learn*. Available at https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html.

[5] sklearn.cluster.optics. *Scikit Learn*. Available at https://scikit-learn.org/stable/modules/generated/sklearn.cluster.OPTICS.html.

[6] Sklearn.manifold.tsne. *Scikit Learn*. Available at https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html.

[7] sklearn.metrics.silhouette_score. *Scikit Learn*. Available at https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html.

[8] Veggie-debate: What should a good diet be like? *OpMAP*.

[9] Bangor, A., Kortum, P. T., and Miller, J. T. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction 24*, 6 (2008), 574–594.

[10] BELKIN, M., AND NIYOGI, P. Using manifold stucture for partially labeled classification. *Advances in neural information processing systems 15* (2002).

[11] BETZ, G., H. M. M. T. V. S. S. Applying argumentation to structure and visualize multi-dimensional opinion spaces. *Argument Computation 10*, 1 (2019).

[12] BJØRKELUND, E., B. T. H. . N. K. A study of opinion mining and visualization of hotel reviews. *In Proceedings of the 14th International Conference on Information Integration and Web-based Applications Services* (2012), 229–238.

[13] BOEING, G. Clustering to reduce spatial data set size. *arXiv preprint arXiv:1803.08101* (2018).

[14] BROOKE, J. Sus: a 'quick and dirty' usability scale. *Usability evaluation in industry 189* (1996).

[15] CHEN, Y. S., C. L. H. Y. T. . T. Y. Visualization system for analyzing user opinion. *In 2015 IEEE/SICE International Symposium on System Integration (SII)* (2012), 646–649.

[16] DEEB, S. The molecular basis of variation in human color vision. *Clinical Genetics 67*, 5 (2005), 369–377.

[17] DEMARTINES, P., AND HÉRAULT, J. Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. *IEEE Transactions on neural networks 8*, 1 (1997), 148–154.

[18] EBERHARDT, C. Rendering One Million Datapoints with D3 and WebGL.

[19] EVS/WVS. European Values Study and World Values Survey: Joint EVS/WVS 2017-2021 Dataset (Joint EVS/WVS). *GESIS Data Archive*. Dataset Version 2.0.0.

[20] GORODOV, E. Y., AND GUBAREV, V. V. Analytical review of data visualization methods in application to big data. *Journal of Electrical and Computer Engineering 2013* (2013), 1–7.

[21] HINTON, G. E., AND ROWEIS, S. Stochastic neighbor embedding. *Advances in neural information processing systems 15* (2002).

[22] HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology 24*, 6 (1933), 417.

[23] JIMENEZ, L. O., AND LANDGREBE, D. A. Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 28*, 1 (1998), 39–54.

[24] L. SONG, A.J. SMOLA, K. B., AND GRETTON, A. Colored maximum variance unfoldings. *Advances in Neural Information Processing Systems 21* (2007).

[25] LIU, B. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies 5*, 1 (2012), 1–167.

[26] ROWEIS, S. T., AND SAUL, L. K. Nonlinear dimensionality reduction by locally linear embedding. *science 290*, 5500 (2000), 2323–2326.

[27] SAMMON, J. W. A nonlinear mapping for data structure analysis. *IEEE Transactions on computers 100*, 5 (1969), 401–409.

[28] SASMITO, G. W., . N. M. Usability testing based on system usability scale and net promoter score. *International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (2019), 540–545.

[29] SHAMIM, A., B. V. . T. M. Evaluation of opinion visualization techniques. Information visualization. *Argument Computation 14*, 4 (2015), 339–358.

[30] TENENBAUM, J. B., SILVA, V. D., AND LANGFORD, J. C. A global geometric framework for nonlinear dimensionality reduction. *science 290*, 5500 (2000), 2319–2323.

[31] TIAN, Z., ZHAI, X., VAN DRIEL, D., VAN STEENPAAL, G., ESPADOTO, M., AND TELEA, A. Using multiple attribute-based explanations of multidimensional projections to explore high-dimensional data. *Computers amp; Graphics 98* (2021), 93–104.

[32] VAN DER MAATEN, L., AND HINTON, G. Visualizing data using t-sne. *Journal of machine learning research 9*, 11 (2008).

[33] VAN DER MAATEN, L., POSTMA, E., VAN DEN HERIK, J., ET AL. Dimensionality reduction: a comparative. *J Mach Learn Res 10*, 66-71 (2009), 13.

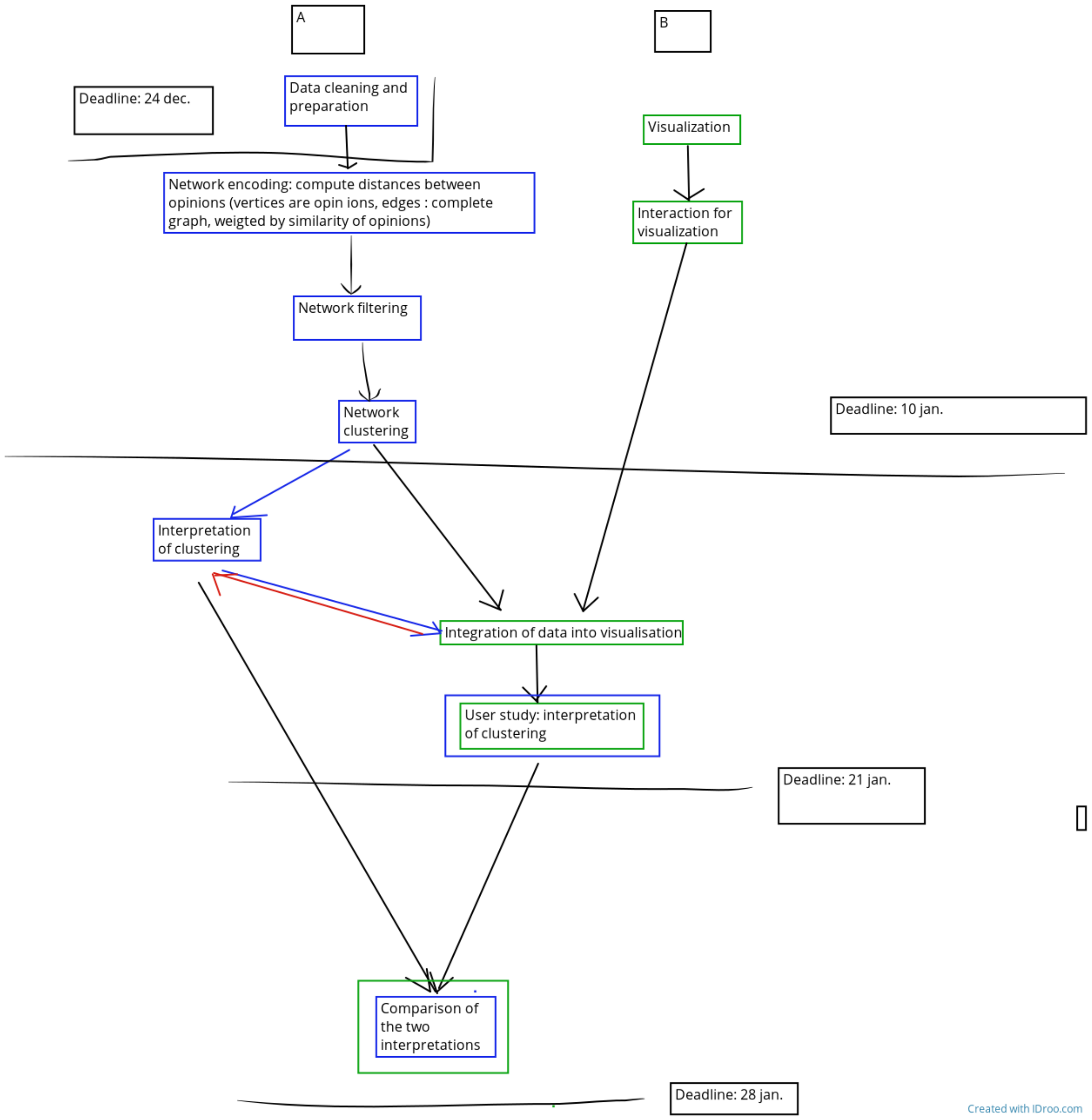
[34] WEINBERGER, K. Q., AND SAUL, L. K. An introduction to nonlinear dimensionality reduction by maximum variance unfolding. In *AAAI* (2006), vol. 6, pp. 1683–1686.

[35] WONG, B. Points of view: Color blindness. *Nature Methods 8*, 6 (2011), 441–441.

[36] WU, Y., W. F. L. S. A. N. C. W. Z. H. . Q. H. OpinionSeer: interactive visualization of hotel customer feedback. *IEEE transactions on visualization and computer graphics 16*, 6 (2010), 1109–1118.

A Project Overview

A schematic overview of the various steps within the project and the division of labor among the two groups. Blue boxes refer to Group A, green boxes refer to Group B. Horizontal lines show the planned deadlines for the steps above the line.



B Complete overview of survey

Welcome to the study Visualization in OpMap

On the next page you will be provided with information about this study. Please read this information carefully, as it contains information regarding your personal data. After reading the information, you can sign the form to give your consent and continue to the survey.

Note: this survey is ideally opened on a laptop or desktop computer. It is not possible to complete this survey on a mobile device.

New page

Information letter

Aim of the study The aim of this study is to get insight into the usability of a novel visualization method. By collecting data on this subject, we aim to answer questions about the interpretability, readability and understandability of a visualization called OpMap.

Data collection and storage Personal data will not be collected. Data will remain confidential and will be anonymised before being stored. Only the researchers wil have acces to the full dataset and only anonymized data will be used in scientific publication. Your data will be stored for at least 10 years. Anonymised versions of the data could be published in scientific literature. In addition, any data collected may be used for follow-up research or research with another purpose.

Content of the study This questionnaire will take approximately 15 minutes. In this study you will be asked to explore the visualization and to answer several questions about it afterwards. For taking part in our study you will receive nothing but gratitude. There are not potentially harmful side effects of taking part in this study.

Taking part in our research is voluntary and may be terminated at any moment. Termination will not have any consequences and may be done without providing a reason for doing so. Data that is collected up to the point of termination may be used for research.

Contact information If you have questions or remarks regarding this research, please contact AAL van Uden (a.a.l.vanuden@students.uu.nl). If you would rather address your remarks to a person independent of this research, please contact L. Hardman (lynda.hardman@cwi.nl). Formal complaints can be directed to klachtenfunctionaris-fetcsocwet@uu.nl.

This information letter was last reviewed on January 24th, 2022.

New page

Consent form

By giving your permission in the field below, you agree to participate in the study Visualization in OpMap on a voluntary basis, are over 18 years of age, and know you may withdraw from the study at any time for any reason. You also give permission for using your data: Data may be stored anonymously for at least 10 years and could be shared anonymously with other researchers or in open access databases.

o I agree, take me to the study.

o I don't agree, end the study

New page

Setting up the OpMap visualization

First, you are asked to explore the OpMap visualization yourself. Please download the file below if you do not yet have the software on your compute. Follow the steps below:

OpMAP application
1. Open the link
2. Download the zip-file by clicking on the green code button, and then click on "Download ZIP.
3. Unpack the zip-file on your computer.

MacOS
3. Open your Terminal
4. Navigate to the folder using cd (e.g. – cd downloads/opmap_vis

5. Run the following command: – python -m SimpleHTTPServer 8000
5. Open a browser of your preference
6. Go to the following url: http://localhost:8000/
7. The visualization should load and become visible.
8. Zoom the screen down, holding down the option, command, and minus (-) button at the same time until the screen reads 67%.

Windows
3. Right-click on the file.
4. Choose open in windows terminal.
5. Run the following command: python -m http.server 8000
5. Open a browser of your preference
6. Go to the following url: http://localhost:8000/
7. The visualization should load and become visible.
8. Zoom the screen down, holding down the ctrl button and the - button at the same time, until the display reads 67%.

Were you able to get the visualisation up and running?

- o Yes

New page

Exploring the OpMap visualization

Now that you have the visualization up and running, you have the opportunity to explore it yourself.

Short introduction to the visualization
The OpMap visualization is a reconstruction of the debate about eating habits. Participants were given multiple statements about food and vegetarianism. For each statement, they had to indicate if they agreed with (accepted), disagreed with (rejected), or did not care (empty) about the statement. All the answers of a participant taken together form his/her opinion within the debate. The distribution of opinions is visualised in an opinion network.

Instructions
Please go to the visualization and explore the features for as long as you wish. We recommend to take your time and explore for at least 5 minutes.

Once done with exploring, please answers the question below with 'yes' to continue the survey. Remember, you can always return to the visualization throughout the survey.

I have explored the visualization.

- o Yes

New page

Data Analysis with the OpMap visualization

The OpMap visualization is designed to analyze the distribution of complex opinions among participants.

We are curious what kind of information you were able to retrieve from the OpMap visualization.

Note: you can always return to the visualization throughout the survey.

The visualization shows clusters of data points. First, you will get some general questions about the clustering. Second, you will get questions about each of the clusters.

New page

1. General Questions about the Clusters

Note: you can always return to the visualization throughout the survey.

Which cluster represents the opinion of the majority of the participants?

- o Orange
- o Green
- o Light Blue
- o Purple
- o Blue
- o Yellow

Which clusters are (somewhat) similar to each other? Why?

Which clusters are (somewhat) different to each other? Why?

New page

2. Questions about the interpretation per cluster

Orange

Note: you can always return to the visualization throughout the survey.

What do you think does the ORANGE cluster represent? Please explain.

What is the most salient opinion in the ORANGE cluster? How would you describe it?

Which questions or claims were important or not important for the ORANGE cluster? Why?

New page

Green

Note: you can always return to the visualization throughout the survey.

What do you think does the GREEN cluster represent? Please explain.

What is the most salient opinion in the GREEN cluster? How would you describe it?

Which questions or claims were important or not important for the GREEN cluster? Why?

New page

Light Blue

Note: you can always return to the visualization throughout the survey.

What do you think does the LIGHT BLUE cluster represent? Please explain.

What is the most salient opinion in the LIGHT BLUE cluster? How would you describe it?

Which questions or claims were important or not important for the LIGHT BLUE cluster? Why?

New page

Purple
Note: you can always return to the visualization throughout the survey.

What do you think does the PURPLE cluster represent? Please explain.

What is the most salient opinion in the PURPLE cluster? How would you describe it?

Which questions or claims were important or not important for the PURPLE cluster? Why?

New page

Blue
Note: you can always return to the visualization throughout the survey.

What do you think does the BLUE cluster represent? Please explain.

What is the most salient opinion in the BLUE cluster? How would you describe it?

Which questions or claims were important or not important for the BLUE cluster? Why?

New page

Yellow
Note: you can always return to the visualization throughout the survey.

What do you think does the YELLOW cluster represent? Please explain.

What is the most salient opinion in the YELLOW cluster? How would you describe it?

Which questions or claims were important or not important for the YELLOW cluster? Why?

New page

3. General impression of the OpMap visualization
Note: you can always return to the visualization throughout the survey.

What is your overall opinion on the OpMap visualization? Please explain.

Please name 3 positive points / tops of the visualization

Please name 3 points of improvement / tips of the visualization

New page

4. SUS
Please rate your level of agreement with each of the following statements:

	1. Strongly Disagree	2	3	4	5. Strongly agree
I would like to use OpMap visualisation to analyze opinions of people collected in other surveys	o	o	o	o	o
I found the OpMap visualization unnecessarily complex.	o	o	o	o	o
I thought the OpMap visualization was easy to use.	o	o	o	o	o
I think that I would need the support of a technical person to be able to use this OpMap visualization.	o	o	o	o	o
I found the various functions in this OpMap visualization were well integrated.	o	o	o	o	o
I thought there was too much inconsistency in this OpMap visualization.	o	o	o	o	o
I would imagine that most people would learn to use this OpMap visualization very quickly.	o	o	o	o	o

I found the OpMap visualization very cumbersome / awkward to use.	o	o	o	o	o
I felt very confident using the OpMap visualization.	o	o	o	o	o
I needed to learn a lot of things before I could get going with this OpMap visualization.	o	o	o	o	o
I enjoyed the exploration of opinions in OpMap	o	o	o	o	o

New page

5. Demographics

What is your age?

What is your gender?

o Male

o Female

o Non-binary / third gender

o Prefer not to say

New page

This is the end of the survey.

If you have any remarks or message you would like to share with us, please fill in the text field below.

Press Next to submit your answers.

General remarks / messages:

C Code repository for project

https://github.com/kimvgen/MLDI-main