CLUSTERING OF LOW VOLTAGE FEEDERS FORM A NETWORK PLANNING PERSPECTIVE

Michiel Nijhuis
Eindhoven University of Technology
The Netherlands
m.nijhuis@tue.nl

Madeleine Gibescu
Eindhoven University of Technology
The Netherlands
m.gibescu@tue.nl

Sjef Cobben
Liander N.V. – The Netherlands
Eindhoven University of Technology
sjef.cobben@alliander.com

ABSTRACT

The utilisation of low voltage (LV) networks is changing due to electrification of heating and transportation and the increase in distributed generation. To ensure that existing grids are reinforced in the most timely and cost-efficient manner, LV network planning is performed. The LV grid of the Netherlands alone consists of over 300.000 feeders, which makes accurately assessing the future loading and the required alterations in the grid for all feeders individually enormously time consuming. The clustering of feeders to a set of generic types, which can be studied in detail can provide a suitable alternative. A fuzzy k-medians clustering approach is proposed based on the data of the LV networks (88.000 feeders) of Liander (largest Dutch DSO). The main network parameters: impedances, cable length, number of branches and branch depth and the number and type of connected customers are used in combination with the graph theory concepts of degree distribution, sequence and the centrality of the power, impedance and length. By using these parameters feeders can be rebuilt from the cluster centres which would represent the structure and loading of the original feeder. Based on these generic feeders an initial assessment of the capacity and voltage deviations allows for the classification of the current LV-network of Liander as 92,55% low risk, 5,91% medium risk and 1,54% high risk feeders.

INTRODUCTION

The energy transition, from a fossil fuel based energy supply to more sustainable sources is changing the loading of the low voltage (LV) grids. The electrification of transportation and heating loads increases the electricity demand of households, while rooftop PV systems generate electricity at the households and can introduce bidirectional power flows. The LV-grids has been built over the past decades and was not designed to meet these load changes. Therefore LV network planning is becoming more and more important to deal with the reinforcing of the LV-grid in the most cost-effective way.

One of the main difficulties with the planning and assessment of the LV-grid is the large number of LV-feeders and MV/LV substations. In the Netherlands alone there are over 120.000 MV/LV substations with over 300.000 LV-feeders connected to them. The analysis of each substation and/or each feeder individually becomes a computationally intensive task. The application of mitigation measures, for a LV-feeder which is found inadequate to deal with the future loading, needs to be standardised in order to ensure a more cost-effective solution. The creation of a limited amount of generic LV-feeders can increase the effectiveness of the LV network planning. As the generic LV-feeders must have a close resemblance to the actual feeders in the field, a clustering approach on the data of the whole LV-grid is the most suitable method for creating these generic feeders.

The characterisation and clustering of electrical networks has previously been studied. For the evaluation of reliability and susceptibility to threats, clustering based on graph theory is already being used especially for the transmission network [1][2][3]. In [4] a small number of networks are defined, based on the length of the feeder, the number of connected customers and the number of branches. Though some analysis can be performed on these representative networks, they are not classified based on enough detail to be usable for network planning. A more extensive approach is required to be able to create generic grids with a strong relation to the existing low voltage grids.

In this paper a new clustering approach based on a number of LV-feeder characteristics and parameters derived from these characteristics is presented. First the feeder characterisation is described followed by a description of the clustering approach. Hereafter the clustering is applied to the LV-grid of Liander and an initial risk assessment of the LV-grid is performed to show a possible application of the clustering method.

FEEDER CHARACTERISATION

The LV-feeders within The Netherlands are constructed from the 1900's onwards mostly by municipality owned electric utilities. Depending on the local situation the most adequate LV-feeder topology was used. This created a large variety in the topology of the LV-feeders, which are not all equally capable of dealing with the future trends in electricity consumption. To give an indication of the differences in the LV-feeders the occurrence of feeder length, number of customers and cable types are shown inFigure 2 Figure 1.

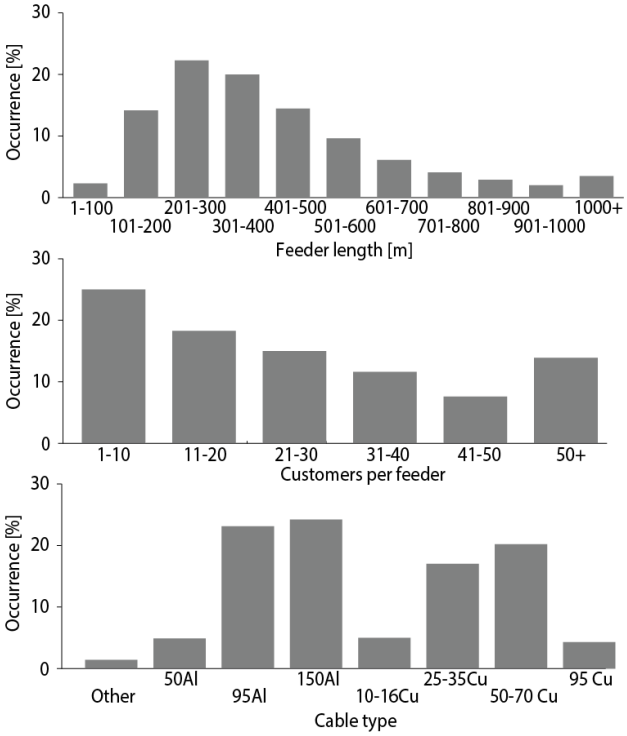

Figure 1: Distribution of the feeder length, customers per feeder and cable types in the LV-grid of a Dutch DSO.

These characteristics only give an indication of the LV-feeder type as it is not yet possible to assess the adequacy of a LV-feeder solely based on this information. Next to these parameters, the distribution of the loads over the cable, the branching within a feeder and the location and the lengths of the cables should all be known. To be able to gain enough insight to assess the LV-feeder adequacy the characterisation must be detailed enough to allow for load flow calculations. The parameterisation of a LV-feeder therefore should include more metrics on the LV-feeder then the ones depicted above, as explained below.

Clustering parameters

To characterise a LV-feeder a number of parameters are used, to illustrate these parameters they are described based on the example LV-feeder in Figure 2. To characterise the loading of a feeder, the number of household (houses in the figure) and non-household loads (arrows in the figure) connected to the feeder and the combined total and average yearly energy usage (indicated by the flows towards the transformer) of these loads are used. For the topology of the feeder the number of branches (4 from a-g, c-j, i-k and e-m) of the feeder and the maximum branch depth (3 between bus a and k from branch a-c, c-i and i-k) are used. For the characterisation of the cables the length (distance between the buses) of the average distance and the average impedance per meter (thickness of the lines between the buses) between buses and the total cable length and total impedance is utilised. To gain more insight in the distribution of the connections over the cables the average impedance at the point of connection is included as an additional parameter.

From a planning perspective the age of the cables in the LV-feeder is also an important parameter, as it gives an indication when the cable sections of a LV-feeder should be replaced, as an aging LV-cable is more prone to failure and the reliability becomes the main criteria for the cable replacement. The quality of the data on the age of the cable is however, not sufficient to be included in the clustering for the case of Liander.

Additionally, a number of parameters is derived from these characteristics, mainly to gain insight into the sequence and distribution of these characteristics, as it makes a large difference if for instance loads are mostly connected a short distance or a relatively long distance from the LV-busbar. To include this kind of parameters in the clustering, complex network analysis is applied to the LV-feeders [5]. The network is converted to a weighted graph by converting the buses to the vertices and the cables to the edges between the vertices. The edges are subsequently weighted according to one of the three following measures: The distance, the impedance and the power flow between two buses. The degree of vertex *n*, denoted as *d(n)*, is the sum of the weights of the edges $e_{i\text{-}j}$ which contain *n* as begin or end vertex. This is illustrated for vertex c in Figure 2 by the formula below:

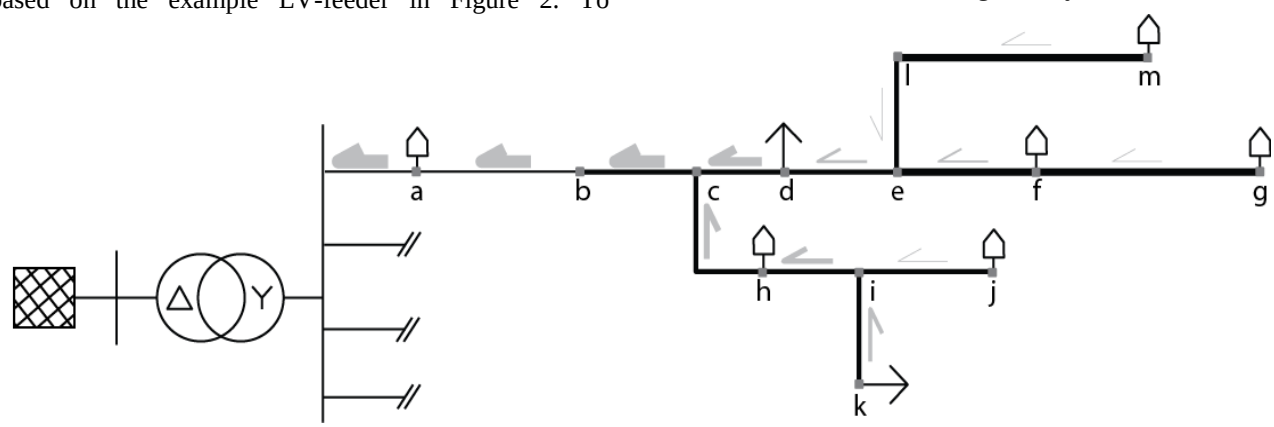

Figure 2: Overview of the parameterisation of the LV-feeders; thickness of line segments denote average impedance; gray arrows denote average yearly energy consumption.

$$d(c) = e_{b-c} + e_{c-d} + e_{c-h} \quad (1)$$

As the number of vertices vary between LV-feeders the degree of the LV-feeder is expressed in four different ways, the mean degree (μ_d), the standard deviation of the degree (σ_d), the maximum closeness centrality (c_{c_d}):

$$\max_{i \in C}\left(c_{c_d}(i) = \frac{1}{\sum_{n \in C} d_c(i,n)}\right) \quad (2)$$

With $d_c(i,n)$ defined as:

$$d_c(i,n) = e_{i-(i+1)} + \cdots + e_{(n-1)-n} \quad (3)$$

and the maximum eigenvector centrality of the degree (c_{e_d}):

$$\max_{i \in C}\left(c_{e_d}(i) = \lambda \sum_{u \in C_i} c(u)\right) \quad (4)$$

where $c(u)$ is the centrality of vertex u, C_i the set of all vertices connected by a single edge to vertex i and λ is the largest eigenvalue of the adjacency matrix (matrix of all vertices containing one if they are connected and zero if they are not connected) of i. By using these four measures for the impedance, distance and the power flow in combination with the characteristics described earlier, the feeder which is reconstructed from this data should have a close resemblance to the original feeder.

The majority of the resulting parameters have a distribution which resembles an exponential distribution with additional modes. Clustering methods generate the best results if the distribution of the data is not heavily skewed. To generate a more symmetric distribution while not affecting the modes in the data, a logarithmic transformation is applied to the non-discrete parameters (all but the number of (non)-household loads, branches and branch depth).

CLUSTERING

In order to generate the generic low voltage feeders, a clustering approach based on the data of the low voltage networks of Liander (largest Dutch DSO) is utilised. The two main approaches applicable to clustering of electrical network data are k-means and hierarchical clustering [6]. A hierarchical clustering approach would generate more accurate clustering results for the following three reasons: It can be expected that no clear independent clusters exist, outliers exist due to the power law distribution of certain network data [7] and the number of clusters is still unknown. The computational burden however makes hierarchical clustering infeasible if all the distribution feeders are included in the clustering. Therefore an adjusted k-mean approach is chosen; fuzzy k-medians clustering. This approach utilises the median instead of the mean as cluster centre and has a soft cluster assignment [8] as is explained in more detail below.

Fuzzy k-medians clustering

In a standard k-medians clustering procedure, the clusters are generated by first selecting random data points as cluster centres. All the observations are subsequently assigned to the nearest cluster centre based on the squared Euclidean distance. For each cluster a new centre is computed by calculating the median of the observations in the cluster. This procedure is repeated until the cluster centres no longer change between iterations. If no convergence is reached within 30 iterations the randomly chosen initial cluster centres are rejected. The clustering is performed with different randomly selected observations until 5 sequential initial clusters centres no longer generate a better result. This is judged by an increase of the average distance between the cluster centres and a decrease of the average distance between points in the same cluster and its centre.

As many of the parameters which are used for the classification procedure are continuous, a fuzzy assignment of the observations to the cluster centres is introduced. This fuzzy assignment assigns each observation to each cluster centre with a certain weight based on the relative distance to the cluster centre. These weights are used for the computation of the new location of the cluster centres. The fuzzy allocation of the clusters to the cluster centres is computed with the following weights:

$$w_{i,j} = \frac{d_{i,j}}{\sum_{j=1}^{n_c} d_{i,j}} \quad (5)$$

where $w_{i,j}$ is the weight associated with the distance between observation i and cluster centre j, $d_{i,j}$ is the distance between observation i and cluster centre j and n_c is the number of cluster centres.

Determination of the number of clusters

The number of clusters is an input variable for the clustering algorithm. To determine which number of clusters generates the best result, a number of metrics are computed to gain insight in the effects of increasing or decreasing the number of clusters. The silhouette of the clusters and cross validation are used to evaluate the adequacy of the number of clusters chosen [8]. With the silhouette, the difference between a certain observation and the other observations in the same cluster is compared to the difference between the observation and all of the observations in the nearest cluster. The silhouette is then averaged for each cluster and for all the

Table 1: Overview of the most common types of LV-feeders, after applying the clustering algorithm.

Cluster #	Length [m]	Branches [#]	Z_{POC} [mΩ]	Main cable type	Customers [#]	Occurrence [%]
1	184	1	24	150AL	17	6.4
2	270	2	35	70CU	24	4.5
3	266	1	43	95AL	39	4.5
4	218	1	26	50CU	19	4.4
5	362	3	56	150AL	32	4.1
6	290	2	74	50AL	26	3.4
7	386	2	57	95AL	49	3.4
8	633	5	107	150AL	70	3.3

clusters leading to the following formula:

$$S = \frac{1}{n_c}\sum_{c_a=1}^{n_c}\frac{1}{n_a}\sum_{i=1}^{n_a}\frac{\sum_{i=1}^{n_b} d_{i,c_b} - \sum_{i=1}^{n_a} d_{i,c_a}}{\max_{a,b\in n_c}\left(\sum_{i=1}^{n_b} d_{i,c_b} - \sum_{i=1}^{n_a} d_{i,c_a}\right)}$$

with S being the silhouette number, a being the cluster observation i belongs to with cluster centre c_a , and b being the nearest cluster to observation i. In addition to the silhouette, the mean square error of the cluster centres between two independently clustered subsets of the observations and the average distance between the observations and the clusters centres of these two subsets is computed:

$$MSE=\left(\frac{1}{n_c}\sum_{c_a=1}^{n_c}\frac{1}{n_a}\sum_{i=1}^{n_a}\hat{d}_{i,c_a}-\frac{1}{n_c}\sum_{c_a=1}^{n_c}\frac{1}{n_a}\sum_{i=1}^{n_a}d^*_{i,c_a}\right)^2 + \left(\frac{1}{n_c^2}\sum_{c_a=1}^{n_c}\sum_{c_b=1}^{n_c}\hat{d}_{c_a,c_b}-\frac{1}{n_c^2}\sum_{c_a=1}^{n_c}\sum_{c_b=1}^{n_c}d^*_{c_a,c_b}\right)^2 \quad (6)$$

with d^* denoting the distance for the first clustered subset of the data and $\hat{d}$ the distance for the second clustered subset of the data.

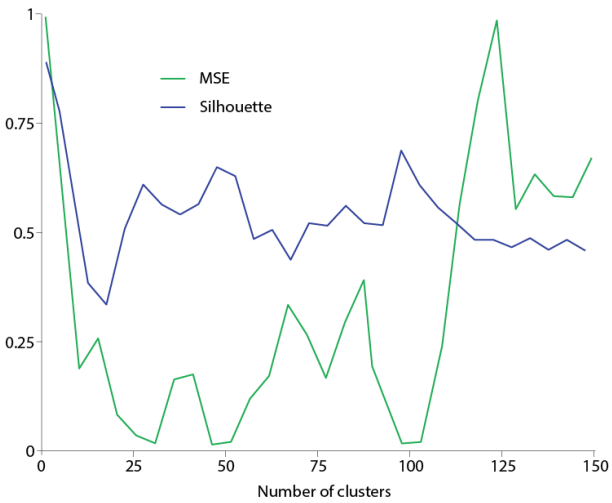

Figure 3: Cluster separation for various cluster sizes.

These measures are both computed for the number of clusters varying from 10 to 150 clusters. The resulting silhouette and normalised MSE values for the LV-feeder data set analysed are plotted in Figure 3. A silhouette close to one indicates an ideal number of clusters, while a MSE close to zero indicates the same. From the figure it can be seen that the MSE is lowest for a number 35 clusters, around 50 clusters and around 100 clusters. For the silhouette the ideal number of clusters lies around the 50 or the 100 value as can be seen from the spikes in the graph at these numbers of clusters. On closer inspection of both indicators around the 100 clusters mark, the number of clusters for the Liander network data is chosen to be 94 as it scores well on both indicators simultaneously.

RESULTS

The clustering approach described in the last section was applied to the network data of Liander. The 88.000 LV-feeders are parameterised and clustered, resulting in 94 classes. The main characteristics of the eight most common types of feeders, which account for about a third of the total LV-grid, are depicted in Table 1.

The table shows that the top seven LV-feeders have quite some similarities, while feeder 8 is significantly longer, with more connections and a higher average impedance. The most common feeder has a relatively low impedance, length and number of customers, indicating that no problems are expected for this feeder type.

Feeder assessment

To show a possible use of the clustered feeders an assessment of these feeders based on the voltage limits and overloading of the cables criteria is performed. A load flow is computed with generic 15-min load profiles [9] for the entire year 2014 for the resulting feeders. From this load flow, the minimum and maximum voltage and the maximum loading of the cables can be determined.. For all 94 generic LV-feeders these calculations have been performed. The results are converted to a high, medium or low risk based on the

risk levels given in Table 2.

Table 2: Risk levels used for feeder assessment

	High risk	Medium risk	Low risk
Voltage	<0.92 & >1.08	0.92-0.94 &1.06-1.08	0.94-1.06
Overloading	>1	0.9-1	<0.9

The risk level for the voltage deviations in the table are based on the voltage limits and do not account for a possible voltage drop at the MV side of the MV/LV transformer (the voltage at the LV side is assumed to be 1 p.u.). With these risk levels and the occurrence of the classes in the LV-grid of Liander an estimation of the currently present risks in the low voltage grid can be computed. The resulting risks for the entire LV-grid of Liander are shown in Table 3.

Table 3: Risks in the LV-network

	High risk	Medium risk	Low risk
Voltage	1.54%	5.91%	92.55%
Overloading	0.97%	2.88%	96.15%

The risks with respect to the voltage deviations and overloading is calculated for the current situation and therefore is low, as the current grid is well functioning. By using scenarios for future loads, the adequacy of the LV-grid can also be assessed with respect a certain future loading.

CONCLUSION

A clustering approach for the LV-feeder data has been proposed. This approach utilises a fuzzy k-medians clustering technique of parameterised LV-feeders. By parameterising the LV-feeders based on common characteristics such as cable length, loading, number of customers and impedance in combination with graph-theory concepts of degree and centrality, representative LV-feeders can be created from these parameters. With the use of the LV-grid of Liander, the clustering approach was tested and the LV-grid was clustered in 94 classes. An initial risk assessment on these clusters gives the indication that the current risks with respect to overloading and voltage deviations are not large

Future work

The classes of LV-feeders which have been determined from the clustering approach should be used in combination with a scenario assessment on the future loading of the grid in order to gain more insight on the risk the energy transition may be imposing on the grid.
The initial assessment can be expanded with the possible safety risk within the LV-grid of Liander. A more detailed analysis on the actual safety risks and the estimation of the 'hidden' earthing in the LV-grid should be performed to gain more insight into the acceptable risks levels for the earthing in the LV-grid.
The approach in this paper should be extended to include the distribution of feeders at a MV/LV substation as well as the upstream MV-grid to gain full insight into the characteristics and risks to the distribution grid.

REFERENCES

[1] A.J. Holmgren, 2006, "Using graph models to analyze the vulnerability of electric power networks", *Risk Analysis*, vol. 26, no. 4, pp 955-969.

[2] Y. Xu, A.J. Gurfinkel, P.A. Rikvold, "Architecture of the Florida power grid as a complex network", *Physica A: Statistical Mechanics and its Applications*, vol. 401, pp 130-140.

[3] P. Schulyz, 2014, "A random growth model for power grids and other spatially embedded infrastructure networks", *The European Physical Journal Special Topics*, pp 2593-2610.

[4] J. Dickert, , M. Domagk, , P. Schegner, , 2013, "Benchmark low voltage distribution networks based on cluster analysis of actual grid properties", *Proceedings PowerTech*.

[5] G.A. Pagani, M. Aiello, 2011, "Towards Decentralization: A Topological Investigation of the Medium and Low Voltage Grids", *IEEE Transactions on Smart Grid*, vol. 2, no. 3, pp 538-547.

[6] R.J. Sanchez-Garcia, , M. Fennelly, S. Norris, N. Wright, G. Niblo, J. Brodzki, J.W. Bialek, 2014, "Hierarchical Spectral Clustering of Power Grids", *IEEE Transactions on Power Systems*, vol. 29, pp 2229-2237.

[7] P. Crucitti,, V. Latora, M. Marchiori, 2004, "A topological analysis of the Italian electric power grid", *Physica A: Statistical Mechanics and its Applications*, vol. 338, no. 1, pp 92-97.

[8] C.K. Reddy, C.C. Aggarwal, 2013, *Data Clustering: Algorithms and Applications*, CRC Press, Boca Raton, US

[9] Energy Data Services Netherlands, (online) Nov 2014, *verbruiksprofielen*, http://www.edsn.nl/wp-content/uploads/2012/08/Profielen-Elektriciteit-2013-versie-1_00.zip