DOI: 10.48011/sbse.v1i1.2318

Exploratory Factor Analysis of Distribution Networks Characterization by Metrics of Complex Networks Theory ⋆

Michel Bessani, Diego A. da Mata

Graduate Program in Electrical Engineering, Federal University of Minas Gerais, Belo Horizonte
e-mail:{mbessani;diegoadmata}@ufmg.br

Abstract: Distribution networks are responsible for supplying electricity to most consumers, and also are the power system part where the majority of electricity interruptions occurs. These infrastructures are in urban and rural regions and are organized to meet different geographical and operational restrictions and the energy demand. In this study, the use of metrics from complex networks theory to describe the organization of such important systems in topological and electrical perspectives was evaluated. A variety of metrics were extracted from different distribution networks. They were calculated considering topological and active power flow in nominal conditions information. The values obtained were investigated using exploratory factor analysis approach. Results indicated that the metrics can be grouped into three distinct factors, and there is a metric, unrelated to such factors, which describes how the power flow is distributed over the network structure. Considering the importance of such systems and the various possibilities of the operational and topological organization, the knowledge of metrics capable of characterizing, in a systemic perspective, is significant for the analysis of current and future challenges related to energy distribution. This topic and its applications will be further explored in future research.

Keywords: Distribution networks, Complex networks, Systems characterization, Exploratory factor analysis.

1. INTRODUCTION

Power Distribution Systems (DS) operate, in general, as a set of radial distribution networks (Ahmadi and Martí, 2015), because of different reasons (Kazmi et al., 2017), as reduced implementation and protection costs, and less complex control and automation requirements. As each distribution network is radial, the power flows from the distribution feeders to the loads, resulting on the high contribution of such part of the power grid to the electricity interruption experienced by consumers and its consequences (Zidan et al., 2016).

Besides these operational aspects, the load spatial organization together with geographical characteristics (Zvoleff et al., 2009) needs to be pondered in the planning, construction and expansion of such systems. A distribution network can be organized in different manners, because of the various requirements and resources available. Such an organization reflects both in the connectivity among the network elements and in the electrical characteristics of the elements composing such systems.

An approach capable of describing how systems are organized is the Complex Networks (CN) theory (Barabási et al., 2016). Examples are the analyzes of muscle networks organization at distinct conditions (Boonstra et al., 2015), and the Linux Kernel structure (Gao et al., 2014). The studies of engineered systems topological organization date back to the late 1990s, as Watts and Strogatz (1998) which investigated the small-world property of the US power grid.

CN applications on power systems studies investigated systems dynamic behaviors, as cascade failures (Pahwa et al., 2014), synchronization (Rohden et al., 2012), and vulnerability (Bessani et al., 2017; Zang et al., 2019). Others application can be seen in the modeling of such systems to aid in different problems solution or investigation, for example, the planning of power distribution networks (Cuadra et al., 2017), and the study of different distribution networks resilience loss during extreme weather events (Bessani et al., 2018).

In this manuscript, we analyze metrics from CN theory to describe the organization of power distribution networks. The traditional CN metrics (Costa et al., 2007) characterize the connectivity among the elements, while extended CN metrics, or hybrid metrics (Cuadra et al., 2015), can describe how the electricity flows through such networks structure in nominal operating conditions, named hybrid characterization.

This study is an exploratory investigation of the metrics used in (Bessani et al., 2018), where the resilience drop of distribution networks during extreme weather was related to topological and hybrid CN metrics for such networks. This metrics will be investigated using Exploratory Fac-

⋆ This work has been supported by the Brazilian agency CAPES.

tor Analysis (Fabrigar and Wegener, 2011) allowing an examination on how they are interrelated and a better understanding of how they characterize the networks.

We organized the remainder of this manuscript as follows. Section 2 presents the fundamental of the used metrics, including the approach to embed the power flow quantities in these. The distribution networks used, and the methodology applied to investigate the metrics are described in Section 3. Section 4 presents and discusses the results followed by Section 5 that concludes this manuscript.

2. BACKGROUND

Given a distribution network composed by N buses and M branches, it can be represented as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of vertices, v_i, describing the N buses, and $\mathcal{E}$ is a set of edges, e_j, that represents the M branches connecting the network buses.

2.1 Topological Metrics

The topological metrics (Costa et al., 2007; Barabási et al., 2016) used in this study represent, in distinct perspectives, how the vertices are connected on a graph. They are briefly described below.

- Order (N): the cardinality of $\mathcal{V}$;
- Average degree ($\overline{k}$): the average number of connections, or degree (k), the vertices have. It is a general descriptor of the number of connections present in $\mathcal{G}$;
- Density (d): the ratio between the size (M) of $\mathcal{G}$, i.e., the cardinality of $\mathcal{E}$, and the maximum number of edges a graph with an order equal to N can have, which is calculated as:
$$d = \frac{2M}{N(N-1)}; \tag{1}$$
- Diameter (D): The largest shortest path among all the vertices in $\mathcal{G}$. The diameter indicates the shortest distance between the two vertices more separated in $\mathcal{G}$;
- Average shortest path (L): the average length of the shortest paths for all the pairs of network vertices. This metric shows how the vertices are separated in average;
- Degree entropy (H): the uncertainty on the number of connections the $\mathcal{G}$ vertices have and is calculated as:
$$H = -\sum_k p(k) \log p(k), \tag{2}$$
where $p(k)$ is the probability of a vertex to have k connections. Higher values of entropy indicate a higher variability in the number of connections the vertices have;
- Assortativity coefficient (r): quantifies how similar are the degrees of connected vertices, and ranges from -1 to 1; Values near to 1 indicates that vertices are connected with other vertices with a similar degree, and near to -1 indicates that high degree vertices are connected to low degree ones and vice-versa.
- Degree probability distribution (α and x_{min}): knowing the network vertices k, we can describe the degree probability distribution - $p(k)$. One type of probability distribution that can be used is the power-law, which form is $p(x) \propto x^{-\alpha}$. In general, the power-law distribution describes only values greater than a minimum (x_{min}), and in this case, we are modelling the distribution tail (Clauset et al., 2009). Given the k of the network vertices, we can adjust a power-law distribution to model the degree probability distribution as $p(k) \propto k^{-\alpha}$ for values of $k \geq x_{min}$. The higher the value of α, the lower is the probability of finding a vertex with higher k and also the x_{min} value describes where the distribution tail begins, being directly related with the observed values of k.

2.2 Hybrid Metrics

A distribution network also has electrical characteristics related to the power, current and voltage over its elements. It is already known that the use of pure topological metrics leads to an incomplete description of such systems (Bompard et al., 2012), and there are an extensive amount of proposals to incorporate the electrical features in the topological metrics (Cuadra et al., 2015), which were named as extended or hybrid metrics.

The hybrid metrics evaluated here are based in the ones used in the analysis presented in Bessani et al. (2018) study, where the active power flowing through the distribution networks branches in nominal operating conditions were used as the weight of the graph edges. By knowing these weights, it is possible to calculate the weighted version of all the metrics previously presented in Subsection 2.1, except the Order and Density that will result in the same value.

The use of the power flow as edges weights will result in metrics describing how the power flow is arranged over the network structure. As an illustration, the weighted vertices degree k^w will be the sum of the power flowing through the edges connected to each vertex, and the average weighted degree ($\overline{k}^w$) will describe the average amount of power that flows through the branches connected to a network bus. The weighted assortativity (r^w) will reflect if buses with large amounts of power flow are connected with similar buses or to ones where a small amount of power flow is demanded.

2.3 Exploratory Factor Analysis

Exploratory Factor Analysis (EFA) is a multivariate statistical approach to model the relationships between observable variables through a smaller set of latent variables, or hidden variables, called factors (Salkind, 2010), which describe the underlying relationship between measured variables. This approach is similar to Principal Components Analysis (PCA), the difference is that EFA assumes that the total variance can be partitioned into common and unique variance, while PCA does not consider unique variance.

EFA results in a model that expresses each variable (x_i) as a linear combination of the factors ($f_1, \ldots, f_m$), which describe the common variance, plus an error term e_i related to the unique variance:

$$x_i = \mu + \lambda_{i1} f_1 + \ldots + \lambda_{im} f_m + e_i, \quad (3)$$

where λ_{ij} is the j-th factor loading of the i-th variable.

The factors loadings are obtained by solving the following equation (Fabrigar and Wegener, 2011):

$$R - D_\Psi - \Lambda\Lambda^T = 0, \quad (4)$$

where R is the correlation matrix calculated as the covariance matrix of the standardized variables x_i, D_Ψ is the covariance matrix among the unique factors assumed that all are uncorrelated with one other, and Λ is the factor loading matrix, where each λ_{ij} represents the loading of x_i to the j-th factor.

The matrix $R - D_\Psi$ is also known as the reduced correlation matrix and can be used to determine the number of factors by the Kaiser criterion (Fabrigar and Wegener, 2011), where the number of factors is equal to the number of eigenvalues of the reduced correlation matrix that are greater than one. The errors are related to the unique variance of the variables that cannot be described by the linear combination of the factors.

After the first Λ estimate, we can rotate it to obtain a better estimation, and the most used is the orthogonal rotation by *varimax* (Fabrigar and Wegener, 2011), which aim to remove the correlation among the factors (Salkind, 2010). Using the rotated Λ, we can calculate the communality of each variable, which is the squared sum of its factors loading and describes the proportion of the variance of the variable expressed by the factors. The communality can be used to evaluate how the observed variables are represented by the latent ones. Moreover, the observation of a variable with a low communality indicates that it describes an aspect that is discrepant from the other variables.

3. MATERIALS & METHODS

3.1 Distribution Networks Used

The distribution networks used in this study were obtained from seven distinct distribution systems. They were presented in the literature for diverse purposes and vary in size, number of feeders, and electrical characteristics. They all are available at the REDS: Repository of distribution systems (Kavasseri and Ababei, 2015). Since each feeder is a radial distribution network, 34 distribution networks were obtained assuming nominal operation conditions from this set of distribution systems. They are listed in the Table below.

3.2 Metrics Extraction and EFA

For each system presented in Table 1, the following steps were performed:

(1) Extract the radial distribution networks relative to each system feeder;
(2) Calculate the nine topological metrics presented in the Subsection 2.1 and store these values;
(3) Calculate the power flow for each distribution network considering nominal operating conditions;

Table 1. Distribution systems used in this study. Each feeder was modeled as an independent network where the metrics were calculated.

System	Feeders	Source
bus_13_3	3	(Civanlar et al., 1988)
bus_29_1	1	(Eminoglu and Hocaoglu, 2005)
bus_32_1	1	(Baran and Wu, 1989)
bus_83_11	11	(Su et al., 2005)
bus_135_8	8	(Guimaraes and Castro, 2005)
bus_201_3	3	(Guimaraes and Castro, 2005)
bus_873_7	7	Created by NSDU Power Group

(4) Calculate the seven hybrid metrics by using the power flow over the branches of the network as edges weights and store such values.

After calculating all the 16 metrics for the 34 distribution networks, EFA can be performed. First, the Kaiser-Meyer-Olkin (KMO) test (Cerny and Kaiser, 1977) was performed into the observed values to verify whether the factor analysis should be performed on the generated data set. Then, the factor loading matrix was estimated by minimizing the residual of (4) and, after that, the loading matrix was rotated using the *varimax* approach to obtain the final loading matrix. With the loading matrix and the residues related to that matrix, AFE can be performed to explore how these variables are related to each other and the underlying relationship between them highlighted on the factors loading.

4. RESULTS & DISCUSSION

The first result presented is the correlation matrix for all the metrics calculated for the 34 distribution networks, this is shown in Figure 1. The correlation values presented are Spearman's coefficient, which is a nonparametric (distribution-free) statistic to measure the relationship strength between two variables. Most variables resulted in positive or close to zero correlations. The only variable that showed negative correlations between the others is d, which is exactly the opposite of the $\overline{k}$ correlation with the other variables. Another aspect is the small correlation values observed between α^w and the other variables, except x^w_{min}, which is the other metric related to the weighted degree probability distribution.

The KMO test for the set of metrics resulted in a value of 0.73, indicating that the use of an EFA for the data set is applicable. The Λ was estimated and its eigenvalues were calculated to decide how many loading factors should be adopted in the model. The eigenvalues were presented in Figure 2, and we can see that only the first three factors have their eigenvalues greater than one, resulting in a model with three factors.

After defining the number of factors, the final matrix Λ was obtained, composed of three orthogonal factors that represent the latent, or hidden, variables describing the relationship between the calculated variables. Table 2 shows the loading factor for each variable, and the values in bold are the highest loading observed for each variable. The first observation that should be highlighted is about metrics $\overline{k}$ and d, which resulted in the same absolute values

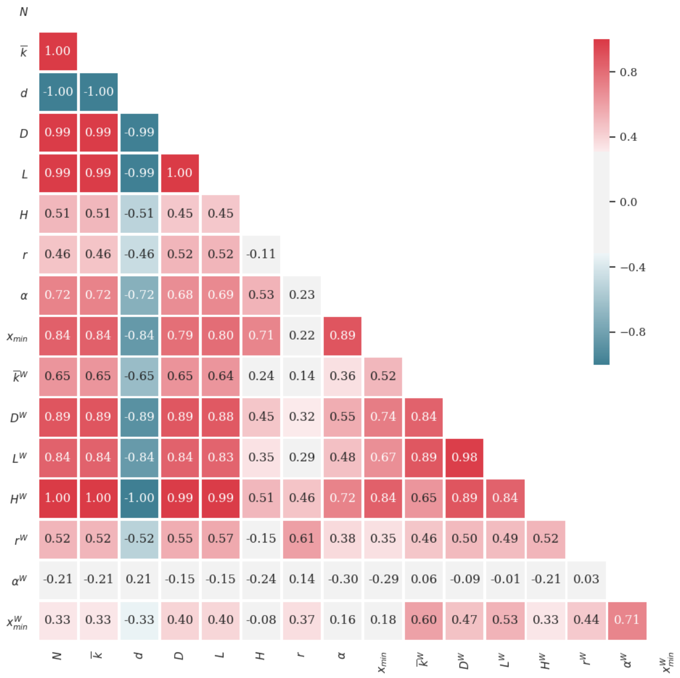

Figure 1. Spearman correlation between all the metrics calculated for the distribution networks.

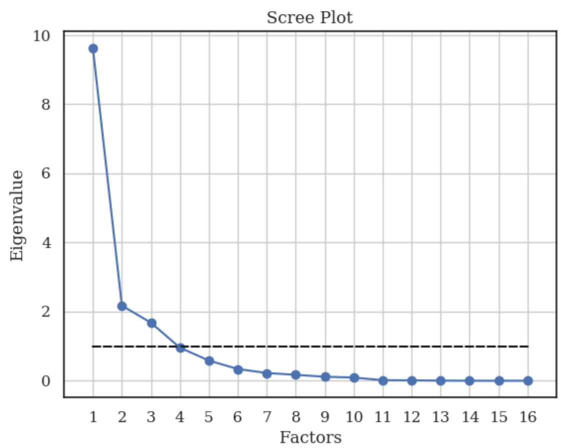

Figure 2. Scree Plot representing the eigenvalue of each factor. The doted line indicates the adopted Kaiser criterion of eigenvalues greater than one to define the number of factors to be used.

for the three factors, but with an inverted sign. This shows that they share, in an inverse way, the same information about the distribution networks, this relationship is similar to the information on the correlation matrix of Figure 1. In addition, the three factors loading have close values, reflecting that these two metrics are similarly related to the three factors.

Factor 1 loading is highest for the N, D, L, D^w, $\overline{k}^w$, L^w, H^w, and x^w_{min} variables, which describe topological and hybrid characteristics. These variables have already been shown to be highly correlated, as shown in Figure 1, and all of them are related to the scale of the system. As a distribution network increases in size and energy demand, these variables are also expected to increase in value due to

Table 2. Factor loadings of each feature for the three factors. The bold values indicate the higher loading observed for each variable.

Features	Factor 1	Factor 2	Factor 3
N	**0.87**	0.42	-0.01
$\overline{k}$	0.45	**0.63**	0.62
d	-0.45	**-0.63**	-0.62
D	**0.90**	0.26	0.22
L	**0.91**	0.24	0.24
H	0.16	**0.75**	-0.26
r	0.16	0.07	**0.83**
α	0.26	**0.76**	0.36
x_{min}	0.33	**0.92**	0.15
D^w	**0.88**	0.35	-0.09
$\overline{k}^w$	**0.86**	0.25	0.17
L^w	**0.96**	0.26	-0.03
H^w	**0.73**	0.57	0.34
r^w	0.27	0.24	**0.81**
α^w	-0.09	-0.10	**0.19**
x^w_{min}	**0.67**	0.01	0.21

the greater number of necessary elements. This is reflected in the topological metrics, and amount of energy that must flow through the network affecting the weighted (hybrid) metrics.

Factor 2 loading is highest for the metrics $\overline{k}$, d, H, α and x_{min}, and they are all related to the expected number of connections that each vertex has in the network. This set of variables can be seen as metrics that describe how the buses in the networks are connected in a pure topological manner. A network with more branches, i.e., higher $\overline{k}$ and d, will tend to have greater variability in the expected number of connections for each vertex, which is quantified by H and by the degree distribution described by the α and x_{min} variables.

Factor 3 loading is highest only for the r, r^w and α^w, being the loading for this last variable only 0.19, showing also a small relationship between the α^w and this factor. The r and r^w are both related with the similarity between connected vertices, the former to the similarity in the number of connections, and the second to the similarity between the amount of power flowing through the connected vertices. An explanation is that a vertex with more connections will be the one where more power is flowing and vice versa.

The communalities of the variables were calculated using the factors loading, these are presented in Figure 3. Most of the variables resulted in high communalities, indicating that they can be well represented by the three factors. Two variables resulted in lower communalities, the x^w_{min} and the α^w, which communalities are 0.49 and 0.05, respectively. These values indicate that the x^w_{min} is related to the other variables but also cannot be fully explained by the three factors, and the α^w communality suggests that this variable describes an aspect that differs from the information measured by the others.

Observing the networks in this data set, a hypothesis that can be defined is that these two variables, x^w_{min} and α^w, are directly related to the organization of the power flow through the structure of the network. This is evidenced by the higher values of x^w_{min} and α^w observed for

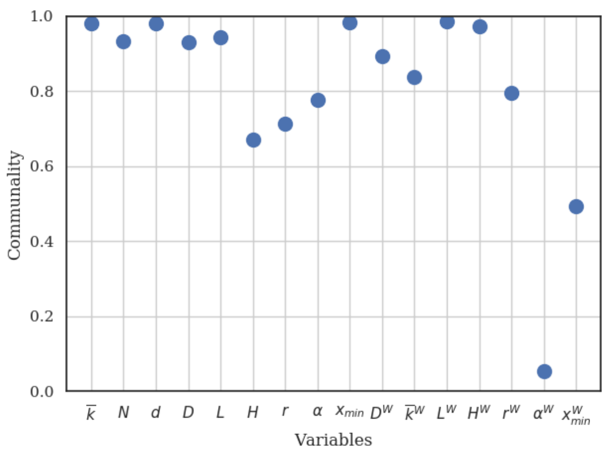

Figure 3. Communalities observed for each variable after obtaining the factors loading.

networks where the majority of power is flowing through the main trunk, while networks, where the power is flowing more homogeneously over their structure, resulted in lower values of x^w_{min} and α^w.

This is illustrated by two distribution networks in Figure 4. These networks resulted in $x^w_{min} = 16.63$ and $\alpha^w = 5.41$, and $x^w_{min} = 17.55$ and $\alpha^w = 15.36$ respectively for Figure 4(a) and Figure 4(b). The other features calculated for these networks are presented in Table 3. They share topological similarities as shown by the unweighted metrics, but the hybrid metrics reflect the difference between them when the power flow over their structure is considered. The main difference among them is how the structure is used to distribute the electricity to the buses, the network in Figure 4(b) have the main trunk where most of the power flow is concentrated, while the power flow in Figure 4(a) is more distributed throughout the structure. The values of x^w_{min} and α^w highlight this difference.

Table 3. Topological and hybrid features calculated for the networks presented in Figure 4.

	Figure 4(a)	Figure 4(b)
N	109	71
$\bar{k}$	1.98	1.97
d	0.02	0.03
D	56	40
L	23.96	15.54
H	0.93	1.55
r	-0.19	-0.21
α	4.04	4.59
x_{min}	2	3
$\bar{k}^W$	8.04	8.30
D^W	388.02	288.21
L^W	149.14	107.28
H^W	6.77	6.15
r^W	0.94	0.64
α^W	5.41	15.36
x^W_{min}	16.63	17.55

5. CONCLUSION

In this manuscript, different topological metrics describing how the elements in a distribution network were connected, and hybrid metrics describing how the power is flowing through the network structure in nominal conditions, were calculated. These metrics were analyzed, in an exploratory manner, to assess how they are characterizing these systems. The use of EFA allowed evaluating how these metrics can be related between each one and also to latent variables by knowing the number of factors and their respective loading.

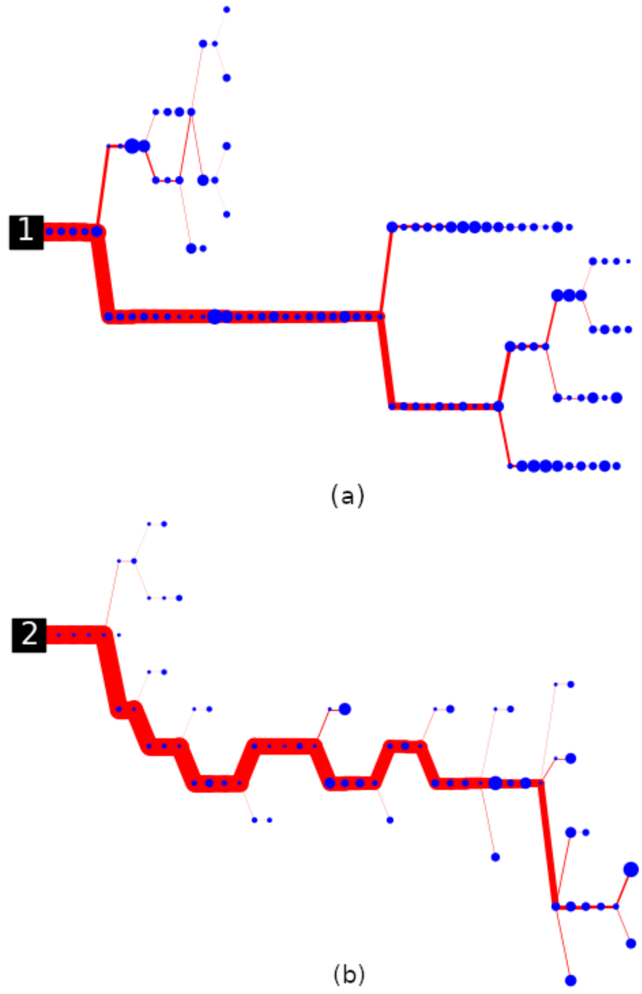

Figure 4. Illustration of networks with different power flow over their structure. Network (a) is the feeder 1 of the system bus_873_7, and network (b) is the feeder 2 of the systems bus_201_3. In the graphs representation, the edges' width is proportional to the power flowing through them, and the vertices size are proportional to the real power demand at each one.

The results indicated that three latent variables are related to the calculated metrics. The first latent variable is related to the scale of the distribution networks, the second describes the variability in the number of connections between the vertices, and the third is related to the similarity between the connected vertices. There are two metrics, related to the weighted degree probability distribution, that resulted in small communalities, and seem to describe how the power flow is organized over the network structure, as illustrated in Figure 4.

Considering the importance of such a system, together with the various current and future possibilities of operational and topological organization, the knowledge of metrics capable of a system-level characterization is significant to analyze the challenges related to power distribution. The results presented here should be further explored in future research. These metrics must be analyzed in a higher number of distribution networks, and also to evaluate how those metrics can be used to infer different behaviours of such systems. Future research is the use of such metrics to characterize resilience drop due to extreme weather events and to describe the effects of distributed energy resources in the loading of power distribution networks.

ACKNOWLEDGEMENTS

We would like to thank Julio Massignan for the constructive discussions during the development of this study.

REFERENCES

Ahmadi, H. and Martí, J.R. (2015). Mathematical representation of radiality constraint in distribution system reconfiguration problem. *International Journal of Electrical Power & Energy Systems*, 64, 293–299.

Barabási, A.L. et al. (2016). *Network science*. Cambridge university press.

Baran, M.E. and Wu, F.F. (1989). Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Power Engineering Review*, 9(4), 101–102.

Bessani, M., Massignan, J.A., Fanucchi, R.Z., Camillo, M.H., London, J.B., Delbem, A.C., and Maciel, C.D. (2018). Probabilistic assessment of power distribution systems resilience under extreme weather. *IEEE Systems Journal*, 13(2), 1747–1756.

Bessani, M., Massignan, J.A., London, J.B., Maciel, C.D., Fanucchi, R.Z., and Camillo, M.H. (2017). A hierarchical framework for complex networks robustness analysis to errors. In *2017 Annual IEEE International Systems Conference (SysCon)*, 1–8. IEEE.

Bompard, E., Pons, E., and Wu, D. (2012). Extended topological metrics for the analysis of power grid vulnerability. *IEEE Systems Journal*, 6(3), 481–487.

Boonstra, T.W., Danna-Dos-Santos, A., Xie, H.B., Roerdink, M., Stins, J.F., and Breakspear, M. (2015). Muscle networks: Connectivity analysis of emg activity during postural control. *Scientific reports*, 5, 17830.

Cerny, B.A. and Kaiser, H.F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate behavioral research*, 12(1), 43–47.

Civanlar, S., Grainger, J., Yin, H., and Lee, S. (1988). Distribution feeder reconfiguration for loss reduction. *IEEE Transactions on Power Delivery*, 3(3), 1217–1223.

Clauset, A., Shalizi, C.R., and Newman, M.E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4), 661–703.

Costa, L.d.F., Rodrigues, F.A., Travieso, G., and Villas Boas, P.R. (2007). Characterization of complex networks: A survey of measurements. *Advances in physics*, 56(1), 167–242.

Cuadra, L., Pino, M., Nieto-Borge, J., and Salcedo-Sanz, S. (2017). Optimizing the structure of distribution smart grids with renewable generation against abnormal conditions: A complex networks approach with evolutionary algorithms. *Energies*, 10(8), 1097.

Cuadra, L., Salcedo-Sanz, S., Del Ser, J., Jiménez-Fernández, S., and Geem, Z. (2015). A critical review of robustness in power grids using complex networks concepts. *Energies*, 8(9), 9211–9265.

Eminoglu, U. and Hocaoglu, M.H. (2005). A new power flow method for radial distribution systems including voltage dependent load models. *Electric power systems research*, 76(1-3), 106–114.

Fabrigar, L.R. and Wegener, D.T. (2011). *Exploratory factor analysis*. Oxford University Press.

Gao, Y., Zheng, Z., and Qin, F. (2014). Analysis of linux kernel as a complex network. *Chaos, Solitons & Fractals*, 69, 246–252.

Guimaraes, M.A. and Castro, C.A. (2005). Reconfiguration of distribution systems for loss reduction using tabu search. In *IEEE Power System Computation Conference (PSCC)*, volume 1, 1–6.

Kavasseri, R. and Ababei, C. (2015). REDS: Repository of distribution systems. *On line]. Available: http://venus.ece.ndsu.nodak.edu/kavasseri/reds.html.*

Kazmi, S.A.A., Shahzad, M.K., Khan, A.Z., and Shin, D.R. (2017). Smart distribution networks: A review of modern distribution concepts from a planning perspective. *Energies*, 10(4), 501.

Pahwa, S., Scoglio, C., and Scala, A. (2014). Abruptness of cascade failures in power grids. *Scientific reports*, 4, 3694.

Rohden, M., Sorge, A., Timme, M., and Witthaut, D. (2012). Self-organized synchronization in decentralized power grids. *Physical review letters*, 109(6), 064101.

Salkind, N.J. (2010). *Encyclopedia of Research Design*, volume 1. SAGE.

Su, C.T., Chang, C.F., and Chiou, J.P. (2005). Distribution network reconfiguration for loss reduction by ant colony search algorithm. *Electric Power Systems Research*, 75(2-3), 190–199.

Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684), 440.

Zang, T., Gao, S., Huang, T., Wei, X., and Wang, T. (2019). Complex network-based transmission network vulnerability assessment using adjacent graphs. *IEEE Systems Journal*.

Zidan, A., Khairalla, M., Abdrabou, A.M., Khalifa, T., Shaban, K., Abdrabou, A., El Shatshat, R., and Gaouda, A.M. (2016). Fault detection, isolation, and service restoration in distribution systems: State-of-the-art and future trends. *IEEE Transactions on Smart Grid*, 8(5), 2170–2185.

Zvoleff, A., Kocaman, A.S., Huh, W.T., and Modi, V. (2009). The impact of geography on energy infrastructure costs. *Energy Policy*, 37(10), 4066–4078.