Analysis of electricity theft influence propagation using multiplex and heterogeneous networks

Luiz C. Borro[1], Mayara C. Maioli[1], Tales F. B. Souza[2], Daniel C. Pinto[2]

[1]Fundação CPqD – Campinas – SP – Brazil

[2]CPFL Energia – Campinas – SP – Brazil

{lborro,mmaioli}@cpqd.com.br, {tfonteboa,dcarvalho}@cpfl.com.br

Abstract. *In developing countries, electricity theft is a common type of non-technical losses (NTL, i.e., losses associated with electricity that is consumed but not billed by some type of anomaly), financially affecting not only distribution system operators (DSO) but also customers. Similarly to frauds in other contexts, there is evidence that electricity theft is highly influenced by social interactions. Here we propose a multiplex and heterogeneous network model to evaluate how social and professional interactions influence on electricity theft. Particularly, by employing a variation of the random walk with restart algorithm we were able to derive a new exposure score for discriminating between fraudsters and regular customers.*

1. Introduction

Electricity theft is a major issue that mostly affects developing countries where resulting losses can range up to 40% of the distributed electricity [Glauner et al. 2016]. More specifically, electricity theft belongs to the class of non-technical losses (NTLs), which are related to the energy delivered to customers but not properly billed. NTLs could be due to any issue in the meter-to-cash process, also encompassing faulty meters and equipment.

The most evident effect of NTLs is financially harming distribution system operators (DSOs) by loss of revenue, which is often passed to regular customers through tariffs. For instance, in Brazil, according to its national power sector regulator (ANEEL), in 2016 NTLs accounted for 6.72% of all electrical energy injected into the distribution grids, generating a loss of R$ 4.5 billion (approximately US$ 1.2 billion)[1]. In addition to the financial effects, NTLs can also impact stability and reliability of electrical power grids, leading to electrical supply problems such as black-outs [Glauner et al. 2016, Messinis and Hatziargyriou 2018].

Tackling electricity theft is a challenge for DSOs due to difficulties in accurately detecting and characterizing irregular customers. DSOs usually deploy inspection campaigns to identify such users and minimize losses, even though those campaigns can be costly and not always effective. Moreover, the effectiveness of an inspection campaign is directly related to the selection of customers that must be inspected, which is a difficult task even for experts [Ramos et al. 2018].

In the last ten years, DSOs have been steadily investing in developing novel methods for NTL detection aiming to improve the identification and profiling of irregular users,

[1]Compiled by the authors with information obtained from http://www.aneel.gov.br

therefore minimizing costs related to ineffective inspections. Considering NTL detection methods recently reported in the literature, there is a trend of using supervised machine learning techniques to automatically recognize customers involved in electricity theft or any other kind of fraud [Messinis and Hatziargyriou 2018, Ramos et al. 2018].

Those machine learning-based prediction methods mostly rely on customers consumption information, almost neglecting the evidence that electricity theft is a social phenomena, which can be characterized by factors such as place of residence, level of education and income, among others. Moreover, according to Glauner and coworkers [Glauner et al. 2016], the perpetration of electricity theft can be influenced by the behavior of neighboring customers, since neighbors are likely to share their knowledge of electricity theft as well as the outcome of inspections. Therefore, properly characterizing those neighborhood interactions could help improving the accuracy of NTL detection models, especially electricity theft. This assumption can be extended for another types of interactions, such as family relationships or professional interactions among DSO's workers and customers.

Based on this premise, in this article we propose a multiplex and heterogeneous network model to evaluate how social (neighborhood and family) and professional interactions among DSO's workers and customers influence on electricity theft. Particularly, we employ a variation of the random walk with restart algorithm [Valdeolivas et al. 2019] to measure how the effect of electricity theft propagates through the network. Our aim is to derive network-based features that can be used in the development of models for detecting NTLs, more specifically electricity theft.

This article is organized as follows: Section 2 presents an overview of the literature, addressing the main techniques used in the detection of non-technical losses. Section 3 describes the datasets used in this work and explain how the exposure score can be obtained, based on the application of the random walk with restart algorithm on a multiplex and heterogeneous network. Section 4 presents the experiments performed and its respective results. Finally, Section 5 presents conclusions and discussions of future works.

2. Related work

NTL detection methods can be broadly classified into three main categories [Messinis and Hatziargyriou 2018]: data oriented, grid oriented and hybrids. Data oriented methods make use of customer-centric data (for example energy consumption), employing data mining or machine learning techniques to learn patterns from sample data. On the other hand, grid oriented methods rely on data obtained from distribution grid sensors taking advantage of the physical rules that govern the underlying electrical network. Finally, hybrid methods combine algorithms and techniques belonging to the two aforementioned classes.

For the last ten years, most of the research on NTL detection have been focusing on data oriented solutions employing supervised machine learning methods, such as Support Vector Machines (SVM) [Coma-Puig et al. 2016, Jindal et al. 2016, Nagi et al. 2010, Nagi et al. 2011], statistical models [Faria et al. 2016], artificial neural networks (ANNs) [Coma-Puig et al. 2016, Costa et al. 2013] and decision trees [Costa et al. 2013, León et al. 2011]. Other methods included rule induction [Leon et al. 2011, Nagi et al. 2011], Bayesian classifiers [Monedero et al. 2012] and Op-

timum Path Forests (OPF) [Ramos et al. 2011].

Usually, data oriented methods depend on large samples characterizing distinct NTL profiles in order to derive predictive models that are effective enough in different electricity theft scenarios. NTL profile characterization is commonly performed in terms of features related to historical consumption data using summary statistics such as mean, max/min, standard deviation [Glauner et al. 2017] or employing mathematical transformations (Fourier, Wavelet and cosine) or even adjusting a statistical model as time series or polynomial regression.

Although the main source of information used by NTL detection models is based on consumption data [Glauner et al. 2017, Messinis and Hatziargyriou 2018], adding features that capture other aspects related to the electricity theft phenomena can significantly improve NTL detection models. Such features can convey insights about customer behavior (or current state) that are knowingly related to electricity theft, thus helping to discriminate between regular and fraudulent customers. In that sense, some recent works have been proposing the use of features related to socioeconomic level [Coma-Puig et al. 2016, Faria et al. 2016], climate conditions [Coma-Puig et al. 2016, Jindal et al. 2016] and geographical location [Faria et al. 2016]. Also, credit worthiness rating [Nagi et al. 2010, Nagi et al. 2008b, Nagi et al. 2008a], which reflects the propensity of a customer to repeatedly delay or avoid payments of electricity bills, and history of query of debits [Costa et al. 2013] are examples of features describing financial aspects of electricity theft.

Electricity theft can be viewed as a fraudulent event influenced by the customer's social interactions. For instance, when a fraudster does not receive any kind of financial or legal penalties, related customers, either by family ties, neighborhood or some other type social interaction, would be aware of the situation, thus being susceptible to follow a similar behavior. Moreover, in a broader context, fraudsters tend to transfer knowledge on how to commit fraud without being caught [Van Vlasselaer et al. 2017]. Would be expected, therefore, that a proper characterization of customers' social interactions might help to improve the detection of electricity theft cases.

Recently, the use of social network analysis techniques has been successfully employed in fraud detection problems [Baesens et al. 2015]. In such approaches, frauds are characterized as a social phenomena, in which group of fraudsters collaborate and share knowledge with close allies. So by using social network analysis techniques, it is possible to make inferences about how fraudsters are organized [Baesens et al. 2015, Van Vlasselaer et al. 2017]. Next, we present and discuss works that use social network analysis in the fraud detection context.

Van Vlasselaer and collaborators [Van Vlasselaer et al. 2017] studied the impact of network information for social security fraud detection. They created a bipartite graph connecting companies to their past and present resources. In this case, resources include address, equipment, buyers, suppliers, employees, etc. By exploring the neighborhood of known fraudulent companies, they propagated fraudulent behavior through the network and inferred a fraud exposure score, which was used for the derivation of a set of network-based features. Intrinsic features, describing the current characteristics of a company, and network-based features were used to develop a novel fraud detection model, which

showed promising results.

In [Akoglu et al. 2013], the authors employ social network analysis techniques that exploits the network effects and automatically detect fraudulent users and fake reviews in online review networks. Finally, Van Vlasselaer and coworkers [Van Vlasselaer et al. 2015a] developed a novel approach to detect fraudulent credit card transactions conducted in online stores. The authors derived a time-dependent suspiciousness score by exploiting a network of credit card holders and merchants. The combination of intrinsic features (extracted from the characteristics of incoming transactions and the customer spending history) and network-based features (based on the suspiciousness score) led to performance gains in terms of accuracy.

So far, to the best of our knowledge, we have not found in the electric sector context, approaches that employ network structures to analyze the relationship of interdependence among individuals (customers and DSO's workers), evaluating how different kinds of social interaction would influence electricity theft.

3. Material and methods

3.1. Datasets

The data used in this study were supplied by a Brazilian DSO. More precisely, five datasets comprising information about more than 300,000 customers (active and nonactive as well) belonging to a selected city were provided for the period ranging from January 2014 to September 2018. The five datasets are the following:

Customers
: Customer information, including location (latitude/longitude), customer class, contract status (active/nonactive), connection voltage and type of meter.

Kinship
: Family relations between DSO's customers: father, mother, brother, sister, uncle, aunt, cousin, spouse, mother-in-law and offspring. Up to 10 relatives per customer are listed.

Energy consumption
: Customers' billed energy per month. It contains more than 15 million meter readings.

Inspections
: Historical data regarding inspections conducted by the DSO in order to detect NTLs. Each record contains the target customer, the inspection date and the respective result. More than 70,000 customers were inspected at least once during the considered period.

Services
: Historical data on services performed by the DSO (new connection requests, repairs, disconnections, etc.). For each service record, it contains the type of service, the DSO's workers involved, the date of execution and the consumer unit[2]. During the considered period, almost 900,000 services were performed by a staff of 640 workers.

[2]Residence or store that consumes electricity from the distribution system.

In Brazil, following ANEEL regulations, the electric power consumer units are classified into two groups: A and B. Group A (high voltage) comprises users that receive energy in voltage equal to or greater than 2.3 kilovolts (kV) while group B (low voltage) is characterized by consumer units serviced at a voltage lower than 2.3 kV. In this last group are present residential, commercial or small size industrial consumers. Due to the levels of relationships being considered in this article, only group B consumers are part of the study.

3.2. Electricity theft influence propagation

Firstly, based on the datasets described in Section 3.1, we defined two classes of nodes (customers and DSO's workers) and their respective intra- and inter-class interactions. As proposed in [Valdeolivas et al. 2019], we modeled those interactions as a multiplex and heterogeneous network.

A multiplex network is a collection of two or more networks sharing the same set of nodes, but with edges representing different types of connections. In this work, we have a multiplex network representing family and neighborhood interactions between customers. For the neighborhood network, two given customers are defined as neighbors if the Haversine distance[3] between their respective consumer units is less than 50 meters. The Ball Tree algorithm [Omohundro 1989] was used to detect neighbors based on the geographical coordinates from the Customers dataset.

On the other hand, a heterogeneous network is the composition of two networks, each one having different types of nodes and edges, which are linked through a bipartite graph. Here, interactions between customer and DSO's workers is defined according the services performed in consumer units (Services dataset). A customer x and worker y are linked if there was any kind of service performed by y in any consumer unit owned by x.

Finally, workers that performed any kind of service in a given consumer unit are also linked, shaping a DSO's workers' team network. The multiplex and heterogeneous network can be schematically displayed as presented in Figure 1.

Following the multiplex and heterogeneous network notation used in [Valdeolivas et al. 2019], let n the number of customer nodes, it follows that the $n \times n$ adjacency matrix for each customer layer is defined as $A^{[\alpha]} = (a_{i,j}^{[\alpha]})_{i,j=1,\cdots,n}$, with $\alpha = 1$ denoting the family network and $\alpha = 2$ denoting the neighborhood network, where $a_{i,j}^{[\alpha]} = 1$ if the customer node i is connected with node j on layer α, and 0 otherwise. As considered in [Valdeolivas et al. 2019], auto-interactions are not considered, meaning that $a_{i,i}^{[\alpha]} = 0 \quad \forall \quad i = 1, \cdots, n$.

Also, let m the number of DSO's workers and $A_w = (a_{i,j}^{[w]})_{i,j=1,\cdots,m}$ the adjacency matrix for the workers' team network, where $a_{i,j}^{[w]} = 1$ if the worker node i is connected with node j, and 0 otherwise. Interactions between customers and workers can be represented through the bipartite adjacency matrix $B_{n \times m}$ where $B_{i,j} = 1$ if the customer node i is connected with worker node j, and 0 otherwise. As noted by [Valdeolivas et al. 2019], in order to build an heterogeneous with multiplex layers graph, nodes from every layer from the multiplex graph need to be linked with their respective nodes from the worker's

[3]Great-circle distance between two points on a sphere given their longitudes and latitudes.

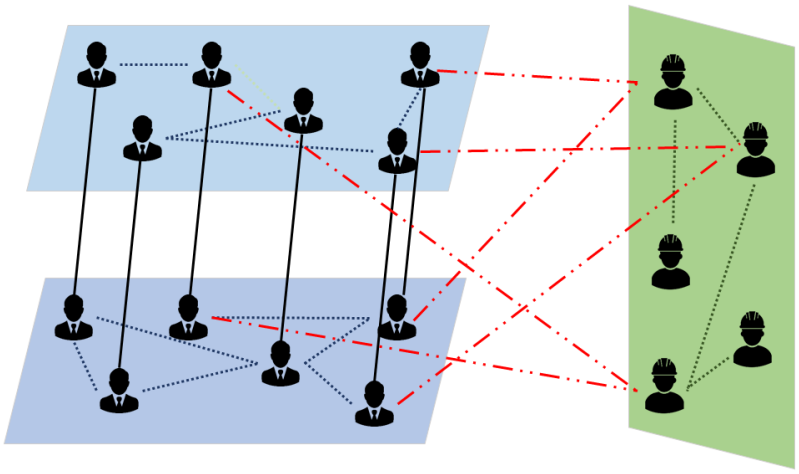

Figure 1. Multiplex and heterogeneous network representing interactions among customers and DSO's workers. Solid line represents the possibility of a customer node jumping to another layer, that is, jumping from the neighborhood network to the family network, and vice versa. Dotted line represents the edges within each layer (neighbors, family, teams' workers). Finally, dotted-dashed line represents the bipartite connections among DSO's workers and customers.

team network, according to the bipartite association. To do so, two identical bipartite graphs $B_{n\times m}$ have to be defined.

Then, the multiplex and heterogeneous adjacency matrix can be defined as

$$A = \begin{bmatrix} (1-\delta)A^{[1]} & \delta I_n & B \\ \delta I_n & (1-\delta)A^{[2]} & B \\ B^T & B^T & A_w \end{bmatrix} \tag{1}$$

where $\delta \in [0, 1]$ quantifies the probability of a customer node staying in a layer or jumping to another layer and I_n denotes the $n \times n$ identity matrix.

Given the matrix M obtained from column normalization of A, we want to propagate the effect of a limited number of fraudsters (customers that who have committed energy theft) through the network. By doing that, our aim is to derive an **exposure score**, that can be used as a metric for the propensity of a customer to commit energy theft. In order to derive the exposure score, we employed a variation of the random walk with restart (RWR) as proposed in [Valdeolivas et al. 2019].

In the RWR, at each interaction, a particle traveling through the heterogeneous network can also restart by jumping to a random node with a restart probability $r \in [0, 1]$. Considering the matrix M, by iteratively solving the RWR equation

$$p_{t+1}^T = (1-r)Mp_t^T + rp_0^T \tag{2}$$

we can derive the exposure score for each node of the heterogeneous network. More precisely, when the difference between vectors p_{t+1} and p_0 is negligible, the stationary probability is reached and the values of those vectors can be seen as a metric of importance for each node composing the network.

If we restrict the restart of the traveling particle to a limited number of specific nodes, we will restrict the exploration of the network to the neighborhood of those nodes, also known as seeds. By doing so, the solution of the RWR equation will represent a proximity measure between the seeds and all other nodes in the network. In this case, the vector p_{t+1} will store the fraud exposure score for both customers and DSO's workers. Since the vector p_0 is the initial probability distribution, its entries will be nonzero only for the nodes representing the seeds. Particularly, the seeds will be represented by nodes related to customers known to have committed energy theft.

4. Experiments

Here we describe the performance evaluation of the proposed exposure score in terms of discriminating fraudsters from regular customers.

For each year in the period 2014-2017, we divided the target city into a 50×50 grid. Then, we computed the electricity theft propensity for each grid cell i, j, which is given by the proportion of identified fraudsters among inspected customers belonging to cell i, j. For a given year, only customers belonging to the top 20 regions regarding electricity theft propensity and their respective interactions were used for building a multiplex and heterogeneous network as described in Section 3.2. Seed nodes were defined based on inspections results of the first eight months of the year. Customers that were inspected within the remaining four months were then used to define the **test set**.

The exposure score was validated by their Enrichment Factors (EF) and by Receiver Operating Characteristic (ROC) curve analysis over the test set. Enrichment Factors measure the quality of the exposure score, indicating how many more fraudsters we find within an "early recognition fraction" of the list of inspections relative to a random distribution. Enrichment factors are calculated as follows:

$$EF_{x\%} = \frac{Fraudsters_{x\%}/N_{x\%}}{Fraudsters/N} \tag{3}$$

where $Fraudsters_{x\%}$ is the number of identified fraudsters in the top $x\%$ of the rank-ordered inspection list, $Fraudsters$ is the total number of identified fraudsters, $N_{x\%}$ is the number of performed inspections in the $x\%$ of the inspection list, and N is the total number of inspections.

The ROC curve analysis is a well-recognized method used as an objective way of evaluating the ability of a given metric to discriminate between two distinct populations [Triballeau et al. 2005]. The ROC curve is represented by plotting the fraction of true positives (true positive rate, TPR) versus the fraction of false positives (false positive rate, FPR). The area under the curve (AUC) is a practical way of evaluating the metric's performance. If the AUC is close to 0.5 (random), the metric can be considered poor. The greater the AUC, the more effective the exposure score is in discriminating regular customers from fraudsters. Table 1 summarizes the inspection data for each each considered test set.

4.1. Preliminary results

In Table 2, we present the results for the test experiments. Figure 2 depicts the ROC curve analysis for each test set. More precisely, the ROC curve applied to the retrospective

Table 1. Summarized inspection data for each test set. For each considered year, the test set encompasses customers that were inspected in the last four months.

Year	#Customers	#Inspected customers	#Identified fraudsters
2014	10,913	782	332
2015	10,055	595	302
2016	11,376	850	453
2017	11,710	1,052	434

analysis of the exposure score is a plot of the true positive rates (TPR, y-axis) versus false positive rates (FPR, x-axis) for all for all ranked customers belonging to an inspection list. Each point of the ROC curve represents a unique TPR/FPR pair corresponding to a particular fraction of the inspection list.

Results from the ROC curve analysis (see AUC in Table 2 and ROC curves in Figure 2) suggest that the proposed exposure score is a good discriminator between fraudsters and regular customers. Results of different years were quite similar, indicating that the fraud patterns for the considered regions (top 20 regions regarding energy theft propensity) had little change over time.

Comparing our proposed approach with other NTL methods is challenging because they are trained in different data sets, many of them proprietary. For that reason, we focused on evaluating if our exposure score is useful for customer prioritization considering inspections lists, which are mostly based on experts' knowledge on electricity theft. Since EFs are more reliable towards the early recognition problem, our preliminary results (Table 2) indicate that the proposed exposure score could be helpful in prioritizing fraudsters over regular customers in inspections lists defined by DSO's experts.

Table 2. Obtained results for each test set using $\delta = 0.5$ and $r = 0.75$. Enrichment factors (EF) were calculated for 1%, 5% and 10%.

	Year	2014	2015	2016	2017
	1%	1.81	1.70	1.77	2.30
EF	**5%**	2.17	1.85	1.81	2.35
	10%	2.17	1.89	1.83	2.35
	AUC	0.865	0.845	0.847	0.853

5. Conclusions

In this work, we used a multiplex and heterogeneous network model to evaluate how social (neighborhood and family) and professional interactions among DSO's workers and customers influence on electricity theft. By employing a variation of the random walk with restart, we were able to derive a new exposure score for discriminating between fraudsters and regular customers. Our results indicate that our exposure score can be used for customer prioritization in inspections campaigns.

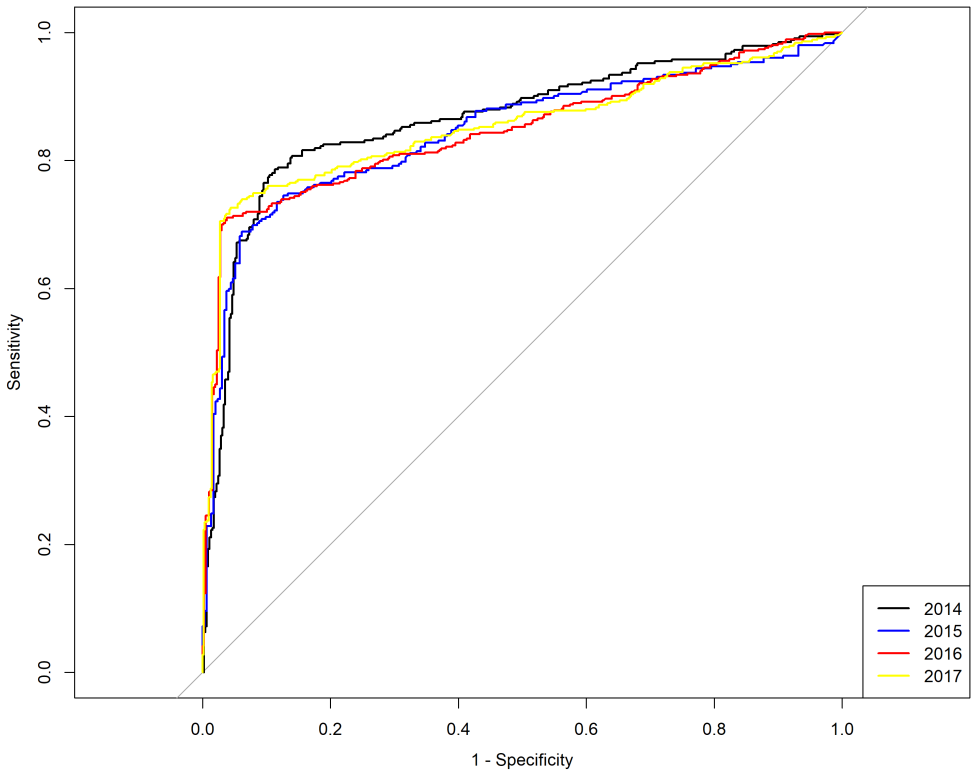

Figure 2. ROC curves analysis for each experimental setup. Sensitivity = TPR; FPR = 1 – Specificity.

As a future development, our intention is to also consider regions that are less susceptible to energy theft in order to improve the discriminating power of our exposure score. In addition, we plan to evaluate our proposed fraud score along with other commonly used features for NTL detection, such as consumption data.

Finally, since fraud schemes are highly adaptive and evolve over time [Van Vlasselaer et al. 2015b], we are also planning to develop a new time-weighted network model that are able to characterize fraudsters according to the recency of their frauds and social interactions.

6. Acknowledgements

This work was supported by ANEEL's research & development program (Project ID PD-0063-3039/2018). The authors would like to thank all members of CPqD's cognitive computing team for the insightful and helpful discussions.

References

Akoglu, L., Faloutsos, C., Chandy, R., and Faloutsos, C. (2013). Opinion Fraud Detection in Online Reviews by Network Effects. In *Proceeding of the 7th International AAAI Conference on Weblogs and Social Media*, pages 2–11.

Baesens, B., Van Vlasselaer, V., and Verbeke, W. (2015). Social network analysis for fraud detection. In *Fraud Analytics: Using Descriptive, Predictive, and Social Network Techniques*, pages 207–278. Wiley, Hoboken, NJ.

Coma-Puig, B., Carmona, J., Gavalda, R., Alcoverro, S., and Martin, V. (2016). Fraud Detection in Energy Consumption: A Supervised Approach. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 120–129. IEEE.

Costa, B. C., Alberto, B. L. A., Portela, A. M., Maduro, W., and Eler, E. O. (2013). Fraud Detection in Electric Power Distribution Networks using an Ann-Based Knowledge-Discovery Process. *International Journal of Artificial Intelligence & Applications (IJAIA)*, 4(6).

Faria, L. T., Melo, J. D., and Padilha-Feltrin, A. (2016). Spatial-Temporal Estimation for Nontechnical Losses. *IEEE Transactions on Power Delivery*, 31(1):362–369.

Glauner, P., Meira, J. A., Dolberg, L., State, R., Bettinger, F., and Rangoni, Y. (2016). Neighborhood features help detecting non-technical losses in big data sets. In *IEEE/ACM 3rd International Conference on Big Data Computing Applications and Technologies (BDCAT)*. IEEE.

Glauner, P., Meira, J. A., Valtchev, P., State, R., and Bettinger, F. (2017). The Challenge of Non-Technical Loss Detection Using Artificial Intelligence: A Survey. *International Journal of Computational Intelligence Systems*, 10(1):760.

Jindal, A., Dua, A., Kaur, K., Singh, M., Kumar, N., and Mishra, S. (2016). Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid. *IEEE Transactions on Industrial Informatics*, 12(3):1005–1016.

León, C., Biscarri, F., Monedero, I., Guerrero, J. I., Biscarri, J., and Millán, R. (2011). Integrated expert system applied to the analysis of non-technical losses in power utilities. *Expert Systems with Applications*, 38(8):10274–10285.

Leon, C., Biscarri, F., Monedero, I., Guerrero, J. I., Biscarri, J., and Millan, R. (2011). Variability and Trend-Based Generalized Rule Induction Model to NTL Detection in Power Companies. *IEEE Transactions on Power Systems*, 26(4):1798–1807.

Messinis, G. M. and Hatziargyriou, N. D. (2018). Review of non-technical loss detection methods. *Electric Power Systems Research*, 158:250–266.

Monedero, I., Biscarri, F., León, C., Guerrero, J. I., Biscarri, J., and Millán, R. (2012). Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees. *International Journal of Electrical Power & Energy Systems*, 34(1):90–98.

Nagi, J., Keem Siah Yap, Sieh Kiong Tiong, Ahmed, S. K., and Nagi, F. (2011). Improving SVM-Based Nontechnical Loss Detection in Power Utility Using the Fuzzy Inference System. *IEEE Transactions on Power Delivery*, 26(2):1284–1285.

Nagi, J., Mohammad, A. M., Yap, K. S., Tiong, S. K., and Ahmed, S. K. (2008a). Non-Technical Loss analysis for detection of electricity theft using support vector machines. In *2008 IEEE 2nd International Power and Energy Conference*, pages 907–912. IEEE.

Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., and Mohamad, M. (2010). Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines. *IEEE Transactions on Power Delivery*, 25(2):1162–1171.

Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., and Mohammad, A. M. (2008b). Detection of abnormalities and electricity theft using genetic Support Vector Machines. In *TENCON 2008 - 2008 IEEE Region 10 Conference*, pages 1–6. IEEE.

Omohundro, S. M. (1989). *Five balltree construction algorithms*. International Computer Science Institute Berkeley.

Ramos, C. C., Souza, A. N., Chiachia, G., Falcão, A. X., and Papa, J. P. (2011). A novel algorithm for feature selection using Harmony Search and its application for non-technical losses detection. *Computers & Electrical Engineering*, 37(6):886–894.

Ramos, C. C. O., Rodrigues, D., de Souza, A. N., and Papa, J. P. (2018). On the Study of Commercial Losses in Brazil: A Binary Black Hole Algorithm for Theft Characterization. *IEEE Transactions on Smart Grid*, 9(2):676–683.

Triballeau, N., Acher, F., Brabet, I., Pin, J.-P., and Bertrand, H.-O. (2005). Virtual Screening Workflow Development Guided by the "Receiver Operating Characteristic" Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *Journal of Medicinal Chemistry*, 48(7):2534–2547.

Valdeolivas, A., Tichit, L., Navarro, C., Perrin, S., Odelin, G., Levy, N., Cau, P., Remy, E., and Baudot, A. (2019). Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics*, 35(3):497–505.

Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., and Baesens, B. (2015a). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, 75:38–48.

Van Vlasselaer, V., Eliassi-Rad, T., Akoglu, L., Snoeck, M., and Baesens, B. (2015b). AFRAID. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 - ASONAM '15*, pages 659–666, New York, New York, USA. ACM Press.

Van Vlasselaer, V., Eliassi-Rad, T., Akoglu, L., Snoeck, M., and Baesens, B. (2017). GOTCHA! Network-Based Fraud Detection for Social Security Fraud. *Management Science*, 63(9):3090–3110.